FROM PHYSICS TO ECONOPHYSICS AND BACK: METHODS AND INSIGHTS

EDITED BY: Siew Ann Cheong, Takayuki Mizuno, Wei-Xing Zhou, Gabjin Oh, Anirban Chakraborti and Damien Challet

PUBLISHED IN: Frontiers in Physics

frontiers Research Topics

frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88976-519-5
DOI 10.3389/978-2-88976-519-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

FROM PHYSICS TO ECONOPHYSICS AND BACK: METHODS AND INSIGHTS

Topic Editors:
Siew Ann Cheong, Nanyang Technological University, Singapore
Takayuki Mizuno, National Institute of Informatics, Japan
Wei-Xing Zhou, East China University of Science and Technology, China
Gabjin Oh, Chosun University, South Korea
Anirban Chakraborti, Jawaharlal Nehru University, India
Damien Challet, Université Paris-Saclay, France

Citation: Cheong, S. A., Mizuno, T., Zhou, W.-X., Oh, G., Chakraborti, A., Challet, D., eds. (2022). From Physics to Econophysics and Back: Methods and Insights. Lausanne: Frontiers Media. doi: 10.3389/978-2-88976-519-5

Table of Contents

05 ***Editorial: From Physics to Econophysics and Back: Methods and Insights***
Anirban Chakraborti, Damien Challet, Siew Ann Cheong, Takayuki Mizuno, Gabjin Oh and Wei-Xing Zhou

10 ***Analysis of Global Remittance Based on Complex Networks***
Shigang Wen, Yu Tan, Mengge Li, Yunke Deng and Chuangxia Huang

19 ***The Time-Dependent Lead-Lag Relationship Between WTI and Brent Crude Oil Spot Markets***
Yanhong Yang, Yinghui Shao, Haolin Shao and Xin Song

25 ***Tick Size and Market Quality Using an Agent-Based Multiple-Order-Book Model***
Ruwei Zhao, Yian Cui and Xuenu Liu

33 ***The Dynamic Correlations Between Mass Media News and New Media News in Stock Market***
Zuochao Zhang, Yongjie Zhang and Dehua Shen

44 ***Research on Dynamic Evolution of Systemic Risk in Chinese Banking System***
Hong Fan and Wanqin Sheng

59 ***Stock Network Stability After Crashes Based on Entropy Method***
Ming-Yuan Yang, Fei Ren and Sai-Ping Li

68 ***Effective Transfer Entropy Approach to Information Flow Among EPU, Investor Sentiment and Stock Market***
Can-Zhong Yao and Hong-Yu Li

82 ***Is Export a Probe for Domestic Production?***
Francesco Saltarelli, Valeria Cimini, Andrea Tacchella, Andrea Zaccaria and Matthieu Cristelli

93 ***Grandpa, Grandpa, Tell Me the One About Bitcoin Being a Safe Haven: New Evidence From the COVID-19 Pandemic***
Ladislav Kristoufek

99 ***Evaluation of the Heterogeneous Spatial Distribution of Population and Stores/Facilities by Multifractal Analysis***
Mariko I. Ito and Takaaki Ohnishi

111 ***Rational Belief Bubbles***
Hyun-U Sohn and Didier Sornette

125 ***Confucianism and Firm Performance***
Lisong Wang, Yongjie Zhang, LiNa Wang and Jingkai Fu

134 ***A Perspective on Correlation-Based Financial Networks and Entropy Measures***
Vishwas Kukreti, Hirdesh K. Pharasi, Priya Gupta and Sunil Kumar

141 ***Stock-Index Tracking Optimization Using Auto-Encoders***
Chi Zhang, Shuang Liang, Fei Lyu and Libing Fang

156 ***Does Crude Oil Market Efficiency Improve After the Lift of the U.S. Export Ban? Evidence From Time-Varying Hurst Exponent***
Yinghui Shao

164 ***Initial Value Dependence of Urban Population's Growth-Rate Distribution and the Long-Term Growth***
Atushi Ishikawa, Shouji Fujimoto, Arturo Ramos and Takayuki Mizuno

174 ***Measuring Complexity in Financial Data***
Gaurang Singh Yadav, Apratim Guha and Anindya S. Chakrabarti

183 ***Time Scales and Characteristics of Stock Markets in Different Investment Horizons***
Ajit Mahata and Md. Nurujjaman

190 ***Analyses of Daily Market Impact Using Execution and Order Book Information***
Kenta Yamada and Takayuki Mizuno

198 ***Model of Continuous Random Cascade Processes in Financial Markets***
Jun-ichi Maskawa and Koji Kuroda

210 ***Analysis of Stock Price Motion Asymmetry via Visibility-Graph Algorithm***
Ruiyun Liu and Yu Chen

223 ***Inequality Measures: The Kolkata Index in Comparison With Other Measures***
Suchismita Banerjee, Bikas K. Chakrabarti, Manipushpak Mitra and Suresh Mutuswami

238 ***Lending Diversification and Interconnectedness of the Syndicated Loan Market***
Gabjin Oh and A-young Park

249 ***Analysis of the Global Banking Network by Random Matrix Theory***
Ali Namaki, Reza Raei, Jamshid Ardalankia, Leila Hedayatifar, Ali Hosseiny, Emmanuel Haven and G. Reza Jafari

257 ***Community Detection of Dynamic Complex Networks in Stock Markets Using Hybrid Methods (RMT-CN-LPAm+ and RMT-BDM-SA)***
Acep Purqon and Jamaludin

270 ***Network-centric Indicators for Fragility in Global Financial Indices***
Areejit Samal, Sunil Kumar, Yasharth Yadav and Anirban Chakraborti

281 ***Using Topological Data Analysis (TDA) and Persistent Homology to Analyze the Stock Markets in Singapore and Taiwan***
Peter Tsung-Wen Yen and Siew Ann Cheong

300 ***Characteristics of Principal Components in Stock Price Correlation***
Wataru Souma

312 ***Regional Economic Convergence in China: A Comparative Study of Nighttime Light and GDP***
Qiao-Li Xiao, Yue Wang and Wei-Xing Zhou

326 ***Network Structures for Asset Return Co-Movement: Evidence From the Chinese Stock Market***
Huai-Long Shi and Huayi Chen

frontiers | Frontiers in Physics

EDITORIAL
published: 22 July 2022
doi: 10.3389/fphy.2022.969516

Editorial: From Physics to Econophysics and Back: Methods and Insights

Anirban Chakraborti[1,2], *Damien Challet*[3], *Siew Ann Cheong*[4]*, *Takayuki Mizuno*[5], *Gabjin Oh*[6] *and Wei-Xing Zhou*[7]

[1]School of Computational and Integrative Sciences, Jawaharlal Nehru University, Delhi, India, [2]School of Engineering and Technology, BML Munjal University, Kapriwas, India, [3]Université Paris-Saclay, CentraleSupélec, Laboratoire de Mathématiques et Informatique pour la Complexité et les Systémes, Gif-sur-Yvette, France, [4]School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore, [5]Information and Society Research Division, National Institute of Informatics, Tokyo, Japan, [6]Division of Business Administration, Chosun University, Gwangju, South Korea, [7]School of Business, East China University of Science and Technology, Shanghai, China

Keywords: econophysics, financial markets, market models, banking and macroeconomics, information filtering

Editorial on the Research Topic

From Physics to Econophysics and Back: Methods and Insights

The term "Econophysics" was coined by H. Eugene Stanley in 1995 during a statistical physics conference on the Dynamics of Complex Systems in Kolkata, India to refer to the then emerging interdisciplinary field of physicists working on problems in economics and finance [1]. An interdisciplinary area of research straddling computer science, economics, finance, mathematics, and physics, econophysics started out drawing heavily upon theories and methods developed in nuclear physics and statistical physics. From the use of Random Matrix Theory (RMT) to discriminate between signal and noise in financial time series data [2, 3], to the use of the Ising model and variants to explain stylized facts of stock markets in terms of the microscopic dynamics of traders [4–7], econophysicists have since gone on to develop methods and insights inspired by specific problems. These include the DebtRank measure of systemic risk in banking networks [8], and the discovery of unusual Brownian motion dynamics in order books [9–11], among many others.

Unfortunately, scientists and the public are generally unaware of these contributions. Even within the broader physics community, the fruits of econophysics remain relatively unknown. In fact, every now and then we will find physicists, mathematicians, electrical engineers, or computer scientists reinventing the wheel, and publish results that have been obtained by econophysicists 5–10 years ago. The main reason for this predicament, and also econophysics methods and insights not catching on in the broader physics community is that econophysicists tend to publish in a variety of journals with diverse audiences. As econophysics matures as a field—it is now more than 20 years old, we feel that it has progressed to a stage where we have derived new methods and results found nowhere else. We believe these have the potential to contribute towards deeper understanding in other areas of physics. This is our motivation for launching a Research Topic in Frontiers in Physics, read by serious physicists from different research fields, so that econophysics can give back to the broader physics community. Additionally, early econophysicists came from very different backgrounds, from those starting out in statistical physics, to those moving on from nuclear physics, to former string theorists and former condensed matter physicists. They have all benefitted greatly from cross fertilization amongst themselves, as well as with economists, mathematicians, and computer scientists. Having such a Research Topic appear in Frontiers in Physics will give the cross fertilization between physicists and scientists from other disciplines a renewed push.

OPEN ACCESS

Edited and reviewed by:
Matjaž Perc,
University of Maribor, Slovenia

***Correspondence:**
Siew Ann Cheong
cheongsa@ntu.edu.sg

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *15 June 2022*
Accepted: *20 June 2022*
Published: *22 July 2022*

Citation:
Chakraborti A, Challet D, Cheong SA, Mizuno T, Oh G and Zhou W-X (2022) Editorial: From Physics to Econophysics and Back: Methods and Insights. Front. Phys. 10:969516. doi: 10.3389/fphy.2022.969516

In this Research Topic, which we called *From Physics to Econophysics and Back: Methods and Insights*, we now have a Research Topic of 30 articles. We organized them into six groups:

1. Methods;
2. Models;
3. Financial times series;
4. Financial time series cross sections;
5. Banking and macroeconomics; and
6. Urban complexity.

While Methods and Models are clearly about new and existing econophysics methods, new methods are also introduced in the next four groups of papers focused on developing insights.

To begin, we find many claims in finance. Some are based on rigorous statistical analysis, while others are based on anecdotal evidence. For example, financial time series are expected to be more complex during crisis periods than in calm periods. In the first group of three papers on methods, we find first the paper by Yadav et al. who used the block decomposition method [12–14] to probe the algorithmic complexity in financial time series cross section data. Doing this for the daily returns of $N = 300$ stocks in NASDAQ between 1972 and 2018, over a sequence of 4-years overlapping sliding time windows, Samal et al. found that the 2007–2009 Global Financial Crisis did not register on the results of the first two measures, which were linear, but was clearly present on the results of the last two measures, which were nonlinear. Another claim tested in this Research Topic is a recent one suggesting that cryptocurrencies, such as Bitcoin, might also serve as safe havens. A safe haven is an instrument or investment sector that guarantees the value of one's capital during periods of financial turmoil. Of these, gold and other precious metals have been extensively tested [16–19], but the results for other safe havens have been largely inconclusive. Using the COVID-19 pandemic as the backdrop, Kristoufek used the method of quantile correlation to test the possibility of using Bitcoin as a safe haven for investors . In the method of quantile correlation, we can evaluate separately the correlations between small positive/negative deviations and those between large positive/negative deviations, and in so doing avoid diluting statistically significant signals coming from particular quantiles. Even with this care, Kristoufek found that gold continued to show promise as a safe haven during the COVID-19 pandemic, but neither Bitcoin nor the broader index of cryptocurrencies were able to do so. Finally, we find the paper by Wang et al., testing the impact of corporate culture (Confucianism) on firm performance. This is a claim that sounds plausible, but is very difficult to check. To perform this check, Wang et al. assume that Confucianist ideals and thinking become embedded into the fabric of local communities through the teachings of Confucianism schools after a long time, and the closer communities are to these schools, the stronger the influence. Therefore, as a proxy variable measuring the influence of Confucianism on firms, Wang et al. used the distance from the firms to schools known since the Qing Dynasty. They then measured the performance of a firm by its return on assets. For the 20,121 Chinese firms listed either on the Shanghai Stock Exchange or the Shenzhen Stock Exchange, they found that Confucianist influence does indeed improve firm performance, at the 1% level of statistical confidence.

In the second group on market models, we find four papers on models familiar to physicists, as well as those unfamiliar to the physics community. For the former, Maskawa and Kuroda wrote down a continuous random cascade model to investigate intermittency and multifractality in financial time series. Models on energy cascades are commonly used in the study of turbulent fluid flow. After estimating the parameters of the resulting Fokker-Planck equation for 111 component stocks of the FTSE 100 index on the London Stock Exchange between November 2007 and January 2009, Maskawa and Kuroda were able to reproduce from model simulations multifractal features seen in their earlier empirical study. The next paper by Sohn and Sornette extended rational expectation theory from economics, to explain why economic bubbles arise even when all agents have rational expectations . In economics, agents are treated as having independent beliefs. Sohn and Sornette showed that, when these beliefs are correlated, the time scale at which the market processes information can slow down dramatically, giving rise to a bubble. This is reminiscent of how the central limit theorem results in a normal distribution when a large number of statistically independent random variables with finite variance are combined [25, 26], but in distributions with fat tails when the random variables are correlated [27–31]. The remaining two papers in this group are on order book models. In the first, Yamada and Mizuno, reported an empirical study pointing to a linear correlation $r = \alpha\Delta V + \beta$, where r is the return, ΔV the executed order imbalance (number of bids–number of asks), α and β are stock-dependent constants. This empirical observation is important for building models of price impact of different trade volumes. In the second, Zhao et al. developed a multi-order book agent-based model, based on rules on asset design, investor design, price prediction by the agent, adaptive asset allocation, and order placing . Through simulations, Zhao et al. found that market liquidity decreases with increasing tick size.

Next, we find the group of seven papers on the analysis of single financial time series data. In the paper by Mahata and Nurujjaman, the authors first used empirical mode decomposition (EMD) to write the stock price time series as the sum of a set of intrinsic mode functions (IMF). The advantages of EMD over traditional spectral methods like Fourier transform or wavelet analysis are the far fewer basis functions (the IMFs) needed, these basis functions can be determined empirically from the data, and the basis functions represent the natural time scales found in the data. The Hilbert transform was then computed for each IMF, before R/S scaling was carried out to estimate its Hurst exponent H. Analyzing the stock indices of 12 different countries and the prices of six stocks from December 1995 to July 2018, Mahata and Nurujjaman found that indices and prices are mean-reverting at short time scales, but have long-range correlations at long time scales. In the paper by Liu and Chen, the authors applied the visibility graph method to discriminate between periods of concave price movements (accelerating price change) and periods of convex

price movements (decelerating price change) in noisy time series data. Analyzing eight stock market indices from 28 June 1999, to 28 June 2019, they rediscovered the well-known asymmetry between accelerating price change and decelerating price change. In the next paper by Zhang et al., the authors explored machine learning methods to predict the movement of a stock index. Zhang et al. developed a feed-forward auto-encoder neural network with seven hidden layers using the daily closing prices between 1 January 2010 and 31 December 2018 (2,187 trading days) of the CSI 300 index and its constituent stocks. The model would first be trained on a four-month-window of the data, and then its index tracking performance tested on the 6 months following this time window. This performance was found to be better than conventional methods for stock selection and index tracking. Following this, we then find two papers on information flow in stock markets. Like most complex systems, the stock market is an open system. Economists argue that market crashes are exogenous events, i.e., they are triggered by information flowing in from outside of the stock market [37, 38], whereas econophysicists are open to endogenous explanations in terms of the interactions between stocks in the market [39–42]. In the first of these two, Zhang et al. analyzed the cross correlations between mass media news from 726 sources and new media news from 1,488 sources with the returns of 3,026 stocks, and found the existence of strong cross correlations both at equal times, as well as at various lags. In the second paper, Yao and Li used the transfer entropy method to investigate impacts on the stock market from economic policy uncertainties as well as investor sentiments. As far as the Chinese stock market is concerned, they found no information flow directly from economic policy to the stock market, nor indirectly from economic policy to investor sentiments to the stock market. Finally, the stock market is only one of many markets available to financial investors. Another popular market is the commodities market. In this Research Topic, we have two papers on the crude oil market. In the first paper Yang et al. investigated the lead-lag relationship between two important global crude oil indices between 20 May 1987 and 10 October 2017, and found the Western Texas Intermediate (WTI) leading the Brent slightly over the entire period. In the second paper Shao computed the Hurst exponent H. in 1-year, 2-years, and 4-years rolling windows of the WTI and Brent closing spot prices from 14 October 2011 to 6 March 2020. They found the WTI approaching $H = 0.5$, while Brent deviating from $H = 0.5$ after US lifted their export ban on 18 December 2015.

Another defining feature of complex systems is the large number of interacting variables. In a typical stock market like the New York Stock Exchange, investors can choose from more than 20,000 financial instruments, the majority of which are stocks. As we have just mentioned, interactions between stocks create endogenous forces affecting the prices of individual stocks. Therefore, instead of studying the time series of a single stock, or that of a stock index, econophysicists have also developed methods to investigate cross sections of time series. Because the stocks in these cross sections are interacting, they are often represented as networks. In this Research Topic, we have seven papers looking into various aspects of stocks as a network. In the first of these Kukreti et al. reviewed recent work on correlations-based networks of the stock market, and proposed the use of structural entropy and eigen-entropy for monitoring how these networks change over time. Then in the second of these Shi and Chen investigated the co-movement of asset returns over 120-days rolling windows advancing 1 day at a time, by first decomposing the daily log returns of 28 sector indices between 5 January 2000 and 29 March 2019 on the Chinese stock market using the French-Fama Five Factor Model into the value-weighted market portfolio return (MKT), the portfolio size (SMB), the portfolio value (HML), the portfolio probability (RMW), and the investment factor (CMA). Then, they constructed in each rolling window the minimum spanning tree (MST) based on the Spearman rank correlations between the 29 sector indices and the five factors. They found that the MST having a star-like structure over the entire period, with MKT as the hub, and this star-like structure changing over different parts of the market cycle. One common application of financial networks is to understand the market's response to crises. Related to this we have three papers. In the third of these Samal et al. instead of focusing on a single stock market, Samal et al. computed the cross correlations between the daily closing prices of 69 global financial market indices between 2000 and 2014. They then compared the networks obtained by simple thresholding (keeping cross correlations above some threshold level) and the minimal spanning trees within growth periods as well as crisis periods and found that the discrete edge-based Ricci curvature can be used as an indicator of fragility in global financial markets. In the fourth of these Yang et al. probed whether the network of stocks became stable after a market crash. To do so, they constructed the planar maximally filtered graphs (PMFG) [51] of the constituent stocks of the Shanghai Stock Exchange 180 index within stable and crash periods and computed the entropies of their degree distributions. They found that the stock market did indeed stabilize after market crashes. In the fifth of these Yen and Cheong used the increasingly popular topological data analysis (TDA) method to investigate the persistent homology of the cross correlations between stocks in the Singapore and Taiwan stock exchanges, as well as how these evolve over time. Based on how the Betti numbers change from one time window to the next, they found hints of multiple stages in market crashes. Lastly, in this group of papers we find two on the identification of communities and principal components in stock markets. In the sixth paper Purqon and Jamaludin tested two hybrid methods for detecting communities in the threshold network of cross correlations. While the community structures discovered by the two methods are not the same, these communities were nevertheless meaningful to human experts. Finally, in the seventh paper Souma computed the cross correlations between 445 component stocks of the S&P 500 index over the period 2010 to 2019 and used two methods to extract the meaningful part of the cross correlations. In the first method, he assumed that the eigenvalues of a fully noisy correlation matrix would follow the Pastur-Marcenko distribution, and be bounded between $\lambda_{\min}$ and $\lambda_{\max}$. In the second method, Souma generated a null model through random rotational shuffling of the cross correlations, using which he extracted the meaningful part of the correlation matrix. He then complexified the meaningful part

of the correlation matrix using the Hilbert transform, before analyzing the leads and lags between stocks.

The next group of five papers in this Research Topic deals with the latest research problems in banking and macroeconomics. The paper by Wen et al. describes network structure properties for global remittance and found the key economics group using a community detection method. The impact that export has on domestic production is described by Saltarelli et al. using data from the World Input-Output Database. Recently, the interbank loan structure has been used to study the systemic risk in financial market. The paper by Xiao et al. focused on the connection between nighttime lights and GDP data, to probe regional economic convergence in China. Traditionally, properties of the banking system have been used to study a bank's profit and risk in global financial market. However, it can also be used to investigate the systemic risk of financial system using networks constructed from interbank loan information. Using random matrix theory, Namaki et al. describe the evolution of global bank network to examine the roles of individual countries. Constructing credit and interbank networks using real-world data, Fan and Sheng investigated the systemic risk that might result from credit risk and contagion effect in the banking system. Finally, the paper by Oh and Park provided a quantitative relationship between properties of the interbank network and bank performances, using syndicated loan data from the United States.

Finally, in this Research Topic, we also have two papers dealing with urban complexity, and one paper on a new measure of inequality. In the first paper on urban complexity, Ishikawa et al. analyzed municipal population data for the United States, Italy, and Spain over a period of 10 years, and found that small initial urban populations tend to decrease, but the probability for cities to expand does not depend on the initial population. Over 100 years, however, the populations of some cities increase exponentially while those of other cities decrease exponentially. In fact, large cities can also stop growing exponentially. Recognizing the heterogeneous spatial distribution of urban population in the second paper on urban complexity, Ito and Ohnishi used multifractal analysis to compare the spatial distributions of population, stores, and facilities, to find that stores and facilities are far more concentrated (within commercial districts) than human population. Finally, the paper Banerjee et al. surveyed the development of the Kolkata index for measuring social inequality, before comparing it against other measures of inequality like the Gini coefficient and the Pietra index.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Stanley HE, Afanasyev V, Amaral LA, Buldyrev SV, Goldberger AL, Havlin S, et al. Anomalous Fluctuations in the Dynamics of Complex Systems: from DNA and Physiology to Econophysics. *Physica A: Stat Mech its Appl* (1996) 224(1-2):302–21. doi:10.1016/0378-4371(95)00409-2
2. Laloux L, Cizeau P, Bouchaud JP, Potters M. Noise Dressing of Financial Correlation Matrices. *Phys Rev Lett* (1999) 83(7):1467–70. doi:10.1103/PhysRevLett.83.1467
3. Plerou V, Gopikrishnan P, Rosenow B, Nunes Amaral LA, Stanley HE. Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series. *Phys Rev Lett* (1999) 83(7):1471–4. doi:10.1103/PhysRevLett.83.1471
4. Cont R, Bouchaud J-P. Herd Behavior and Aggregate Fluctuations in Financial Markets. *Macroecon Dynam* (2000) 4(2):170–96. doi:10.1017/S1365100500015029
5. Stauffer D, Penna TJ. Crossover in the Cont–Bouchaud Percolation Model for Market Fluctuations. *Physica A: Stat Mech its Appl* (1998) 256(1-2):284–90. doi:10.1016/S0378-4371(98)00223-4
6. Stauffer D, Sornette D. Self-organized Percolation Model for Stock Market Fluctuations. *Physica A: Stat Mech its Appl* (1999) 271(3-4):496–506. doi:10.1016/S0378-4371(99)00290-3
7. Chowdhury D, Stauffer D. A Generalized Spin Model of Financial Markets. *Eur Phys J B* (1999) 8(3):477–82. doi:10.1007/s100510050714
8. Battiston S, Puliga M, Kaushik R, Tasca P, Caldarelli G. Debtrank: Too central to Fail? Financial Networks, the Fed and Systemic Risk. *Sci Rep* (2012) 2(1):1–6. doi:10.1038/srep00541
9. Yura Y, Takayasu H, Sornette D, Takayasu M. Financial Brownian Particle in the Layered Order-Book Fluid and Fluctuation-Dissipation Relations. *Phys Rev Lett* (2014) 112(9):098703. doi:10.1103/PhysRevLett.112.098703
10. Kanazawa K, Sueshige T, Takayasu H, Takayasu M. Kinetic Theory for Financial Brownian Motion from Microscopic Dynamics. *Phys Rev E* (2018) 98(5):052317. doi:10.1103/PhysRevE.98.052317
11. Kanazawa K, Sueshige T, Takayasu H, Takayasu M. Derivation of the Boltzmann Equation for Financial Brownian Motion: Direct Observation of the Collective Motion of High-Frequency Traders. *Phys Rev Lett* (2018) 120(13):138301. doi:10.1103/PhysRevLett.120.138301
12. Soler-Toscano F, Zenil H, Delahaye J-P, Gauvrit N. Calculating Kolmogorov Complexity from the Output Frequency Distributions of Small Turing Machines. *PLoS ONE* (2014) 9(5):e96223. doi:10.1371/journal.pone.0096223
13. Zenil H, Hernández-Orozco S, Kiani N, Soler-Toscano F, Rueda-Toicen A, Tegnér J. A Decomposition Method for Global Evaluation of Shannon Entropy and Local Estimations of Algorithmic Complexity. *Entropy* (2018) 20(8):605. doi:10.3390/e20080605
14. Zenil H, Kiani NA, Zea AA, Tegnér J. Causal Deconvolution by Algorithmic Generative Models. *Nat Mach Intell* (2019) 1(1):58–66. doi:10.1038/s42256-018-0005-0
15. Baur DG, McDermott TK. Is Gold a Safe haven? International Evidence. *J Banking Finance* (2010) 34(8):1886–98. doi:10.1016/j.jbankfin.2009.12.008
16. Baur DG, Lucey BM. Is Gold a Hedge or a Safe haven? an Analysis of Stocks, Bonds and Gold. *Financial Rev* (2010) 45(2):217–29. doi:10.1111/j.1540-6288.2010.00244.x
17. Hood M, Malik F. Is Gold the Best Hedge and a Safe haven under Changing Stock Market Volatility? *Rev Financial Econ* (2013) 22(2):47–52. doi:10.1016/j.rfe.2013.03.001
18. Lucey BM, Li S. What Precious Metals Act as Safe Havens, and when? Some US Evidence. *Appl Econ Lett* (2015) 22(1):35–45. doi:10.1080/13504851.2014.920471
19. Montgomery DC, Runger GC. *Applied Statistics and Probability for Engineers*. Hoboken, NJ, USA: John Wiley & Sons (2010).
20. Feller W. *An Introduction To Probability Theory And its Applications, Volumes 1 and 2*. Hoboken, NJ, USA: John Wiley & Sons (1957).
21. Tsallis C. Nonextensive Statistical Mechanics, Anomalous Diffusion and central Limit Theorems. *Milan J Math* (2005) 73(1):145–76. doi:10.1007/s00032-005-0041-1
22. Moyano LG, Tsallis C, Gell-Mann M. Numerical Indications of Aq-Generalised central Limit Theorem. *Europhys Lett* (2006) 73(6):813–9. doi:10.1209/epl/i2005-10487-1

23. Umarov S, Tsallis C, Abe S, Herrmann H, Quarati P, Rapisarda A, et al. On Multivariate Generalizations of the Q-central Limit Theorem Consistent with Nonextensive Statistical Mechanics. *AIP Conf Proc* (2007) 965:34. doi:10.1063/1.2828756
24. Tirnakli U, Beck C, Tsallis C. Central Limit Behavior of Deterministic Dynamical Systems. *Phys Rev E* (2007) 75(4):040106. doi:10.1103/PhysRevE.75.040106
25. Umarov S, Tsallis C, Steinberg S. On a Q-central Limit Theorem Consistent with Nonextensive Statistical Mechanics. *Milan J Math* (2008) 76(1):307–28. doi:10.1007/s00032-008-0087-y
26. Zhang C, Liang S, Lyu F, Fang L. Stock-index Tracking Optimization Using Auto-Encoders. *Front Phys* (2020) 8:388. doi:10.3389/fphy.2020.00388
27. Albuquerque R, Koskinen Y, Yang S, Zhang C. Resiliency of Environmental and Social Stocks: An Analysis of the Exogenous COVID-19 Market Crash. *Rev Corporate Finance Stud* (2020) 9(3):593–621. doi:10.1093/rcfs/cfaa011
28. Ludvigson SC, Ma S, Ng S. Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response? *Am Econ J Macroeconomics* (2021) 13(4):369–410. doi:10.1257/mac.20190171
29. Song R, Shu M, Zhu W. The 2020 Global Stock Market Crash: Endogenous or Exogenous? *Physica A: Stat Mech Its Appl* (2022) 585:126425. doi:10.1016/j.physa.2021.126425
30. Sornette D. Endogenous versus Exogenous Origins of Crises. In: *Extreme Events in Nature and Society*. Berlin, Heidelberg: Springer (2006). p. 95–119. doi:10.1007/3-540-28611-X_5
31. Guhathakurta K, Bhattacharya B, Chowdhury AR. Using Recurrence Plot Analysis to Distinguish between Endogenous and Exogenous Stock Market Crashes. *Physica A: Stat Mech its Appl* (2010) 389(9):1874–82. doi:10.1016/j.physa.2009.12.061
32. Fry JM. Exogenous and Endogenous Market Crashes as Phase Transitions in Complex Financial Systems. *Eur Phys J B* (2012) 85(12):1–6. doi:10.1140/epjb/e2012-30234-8
33. Tumminello M, Aste T, Di Matteo T, Mantegna RN. A Tool for Filtering Information in Complex Systems. *Proc Natl Acad Sci U.S.A* (2005) 102(30):10421–6. doi:10.1073/pnas.0500298102

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chakraborti, Challet, Cheong, Mizuno, Oh and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 02 April 2020
doi: 10.3389/fphy.2020.00085

Analysis of Global Remittance Based on Complex Networks

*Shigang Wen, Yu Tan, Mengge Li, Yunke Deng and Chuangxia Huang**

Hunan Provincial Key Laboratory of Mathematical Modeling and Analysis in Engineering, School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha, China

Remittance, one money flow between immigrants and their relatives, is a major source of foreign exchange revenue for economies. Consisting of economies linked by the money flows, global remittance constitutes a network. In this paper, we use bilateral remittances of 210 economies for the time period 2010–2016 to construct a global remittance network (GRN) and then investigate the network's structural properties. We study the degree distribution of the network and find that it is of heterogeneity. Analyses of centrality measures reveal that some key economies, such as the United States, France, India and China, are always ranked the highest. We also detect 6 communities in the network, where economies in the same regional economy cooperative organizations tend to be classified in the same community. Intra-community flows account for 66.07% of total remittances, indicating that economies present the characteristic of regionalization. In addition, the results of the topological stability test show that GRN is fragile to node removal, particular the selective removal based on betweenness centrality.

Keywords: remittance flow, complex network, degree distribution, key economies, community structure

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Chengyi Xia,
Tianjin University of Technology, China
Satyam Mukherjee,
Indian Institute of Management Udaipur, India
Yuming Chen,
Wilfrid Laurier University, Canada

****Correspondence:***
Chuangxia Huang
cxiahuang@amss.ac.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *23 November 2019*
Accepted: *10 March 2020*
Published: *02 April 2020*

Citation:
Wen S, Tan Y, Li M, Deng Y and Huang C (2020) Analysis of Global Remittance Based on Complex Networks. Front. Phys. 8:85. doi: 10.3389/fphy.2020.00085

1. INTRODUCTION

With the intensifying globalization, human migration is an increasing social phenomenon over the world. It leads to a huge transfer in the flow of capital, labor, and knowledge across international boundaries [1]. From a capital view, one special money flow associated with migration is well known as remittance. According to the World Bank, global remittances reach $613 billion in 2017 and involve more than 258 million migrants all over the world [2]. Since the remittance flows among numerous suppliers and recipients present complicated relationships, as far as global remittances are concerned, it is rather natural to count as a complex system.

Complex networks have proven to be a very powerful approach to characterize and analyze a broad array of different complex systems [3–7], such as human migration [8–10], biological mathematics [11, 12], technological [13–17] and finance systems [18–24]. These highly inter-coupled systems have been the focus of a great number of researches, which have investigated influential nodes of the systems. The analysis of key nodes for complex systems is of crucial importance, as it can help in ensuring the more efficient spread of information for technological system [13, 14], hindering crisis propagation for finance systems, and so on [19, 25–27]. In addition to the above discussed complex systems, recently, the global remittance has been incipiently investigated from a complex network perspective [28, 29].

Existing studies on remittance network mainly focused on node degree and strength, degree distribution and circular pattern [28, 29]. These complex network analysis tools can well identify influential remittance suppliers and recipients as well as collaboration pattern between two economies. However, some economies may also play important roles as bridge in the network

[30]. The ability of an economy to transfer the remittance from one economy to another has not been studied yet. On the other hand, in addition to the circular pattern between economy pair, community has been regarded as another important topology structure in networks [8, 30]. Yet little attention has been paid so far to uncover the community structure of global remittance and it still remains as a challenging open problem. Such research not only allows a deep and clear understanding of local properties of the remittance network itself, but also provides good guidance for the remittance management of policy makers and ensures greater regional cooperation.

Motivated by the above discussion, in this paper, by employing bilateral remittance of 210 economies over the period from 2010 to 2016, we construct a global remittance network (GRN). Moreover, we present a detailed study of the key economies for GRN, including degree, degree distribution, node strength, betweenness centrality, community structure, topological stability and so on. The main contributions of this paper can be summarized as follows: (1) Differing somewhat from previous studies that identify influential remittance suppliers and recipients, we also evaluate the importance of economies by taking fully into consideration of the intermediary roles of economies. (2) Literature dealing with the collaboration mechanism of economies in GRN mainly focuses on circular pattern between economy pairs; analysis of economies in community appears to be scarce. Our research intends to fill the existing gap in the literature. (3) This paper originally investigates the effect of key node removal on the topological stability of GRN, and finds that the network is fragile to node removal, particular the selective removal based on betweenness centrality.

The remainder of the paper is arranged as follows. In section 2, the employed methodology related to the empirical investigation is proposed. Section 3 provides details about the bilateral remittance data. Section 4 presents the empirical results and relates them to economic theory. Finally, conclusions are drawn in section 5.

2. METHODOLOGY

This section first proposes a network modeling approach to characterize the global remittance. Then, it introduces a variety of complex network analysis tools, such as degree, degree distribution, node strength, betweenness centrality, community structure and so on.

2.1. Construction of GRN

The global remittance network is represented by a set $G = (V, E)$, where economies $V = (v_1, v_2, \cdots, v_n)$ are denoted as the network nodes and remittance flows set E as the network edges [31]:

$$E = e_{ij}, \tag{1}$$

where e_{ij} is accumulated remittance from economy i to economy j ($i \neq j$) during a certain period, while self-loops ($i = j$) are not permitted. Therefore, an adjacency matrix $\mathbf{E}$ is constructed to characterize the directed weighted network.

2.2. Topological Properties of Networks

2.2.1. Degree, Degree Distribution and Node Strength

Degree k_i of a node is the number of edges that the node is connected to. In directed networks, the notion of degree can be further extended to the in-degree k_i^{in} and out-degree k_i^{out}, which are defined as follows [20]:

$$k_i^{in} = \sum_{j=1}^{N} a_{ji}, \quad k_i^{out} = \sum_{j=1}^{N} a_{ij}, \tag{2}$$

where a_{ij} denotes the indicator function that yields 1 if there is a directed edge from node i to j, and 0, otherwise. While N is the total number of nodes in the network. Moreover, the identity $k_i = k_i^{in} + k_i^{out}$ holds.

Since different nodes have various degrees, the degree distribution is used to describe the probability distribution of these degrees over the network. The degree distribution is given by:

$$P(k) = \frac{n_k}{N}, \tag{3}$$

where n_k is the number of nodes with degree k.

In weighted networks, the edges between nodes are no longer treated as binary interactions, but are measured by their magnitudes. Thus, node strength s_i is proposed to measure the sum of edge weights of i toward its neighbors. Likewise the degree, the notion of node strength can be decomposed into the in-strength s_i^{in} and out-strength s_i^{out}. The in-strength and out-strength in (4) below are extended from (2), respectively [20]:

$$s_i^{in} = \sum_{j=1}^{N} e_{ji}, \quad s_i^{out} = \sum_{j=1}^{N} e_{ij}, \tag{4}$$

where s_i^{in} represents the sum of edge weights inbound to node i, while s_i^{out} computes the sum of edge weights outbound from node i.

2.2.2. Betweenness Centrality

Intuitively, a node will be critical if a large number of shortest paths pass through the node. Therefore, betweenness B_i is proposed to measure the intermediality or bridge property of a node—the ability to transfer flows from one node to another [32]. The formulas in directed network are as follows:

$$B_i = \sum_{\substack{j,l \\ j \neq l \neq i}} \frac{N_{jl}(i)}{N_{jl}}, \quad C_B(v_i) = \frac{B_i}{(N-2)(N-1)}, \tag{5}$$

where N_{jl} is the number of shortest paths from node j to l; $N_{jl}(i)$ is the number of these shortest paths that pass through node i. $C_B(v_i)$ is normalized to a range from 0 to 1.

2.2.3. Community Structure

Numerous real-world networks take the form of community structures, within which the connections between nodes are stronger than connections between nodes of different

communities. In order to detect reasonable communities, Louvain algorithm is utilized [8, 33], which is based on modularity optimization. The notion of modularity evaluates the density of connections inside communities compared to connections between communities. In the case of weighted directed network, the modularity Q is defined as [8]:

$$Q = \frac{1}{M}\sum_{i,j}[e_{ij} - \frac{s_i^{in}s_j^{out}}{M}\delta(\sigma_i, \sigma_j)], \tag{6}$$

where $M = \sum_{ij} e_{ij}$. The function $\delta(\sigma_i, \sigma_j)$ only yields 1 in case $\sigma_i = \sigma_j$, and 0 otherwise.

The procedure of Louvain algorithm is described as follows [33]:

Step 1, all nodes are assigned a different community. For each node i, we calculate the gain of modularity ΔQ when i is placed in its neighbors' communities. By comparing the ΔQ, one can detect the optimal community for node i. If ΔQ is maximum and positive, then node i is placed in the new community. Otherwise, i stays in its original community. The formula of ΔQ is as follows:

$$\Delta Q = \left[\frac{\sum_{in}+2s_{i,in}}{2M} - (\frac{\sum_{tot}+s_i}{2M})^2\right] - \left[\frac{\sum_{in}}{2M} - (\frac{\sum_{tot}}{2M})^2 - (\frac{s_i}{2M})^2\right], \tag{7}$$

where $\sum_{in}$ stands for the sum of the weights of the links inside the new community, $\sum_{tot}$ denotes the sum of weights of the links incident to all nodes in the new community, $s_i = s_i^{in} + s_i^{out}$, $s_{i,in}$ is the sum of the weights of the links from i to nodes in the new community.

Step 2, a new network is constructed whose nodes are now the communities found during the Step 1. In the new network, the weights of the links between nodes are given by the sum of the weights of the links between nodes in the corresponding two communities. We reapply Step 1 to the resulting weighted network.

Step 3, the above steps are iterated until the maximum Q is obtained.

3. DATA

The global remittance matrices are available from the World Bank[1], containing 210 countries or regions of the world over the period from 2010 to 2016. The calculation of bilateral remittance is based on International Monetary Fund (IMF) Balance and Payments Statistics database as well as data released from central banks, national statistical agencies, and World Bank country desks [2].

[1]Bilateral remittance flows worldwide associated with this paper can be found at https://www.worldbank.org/en/topic/migrationremittancesdiasporaissues/brief/migration-remittances-data.

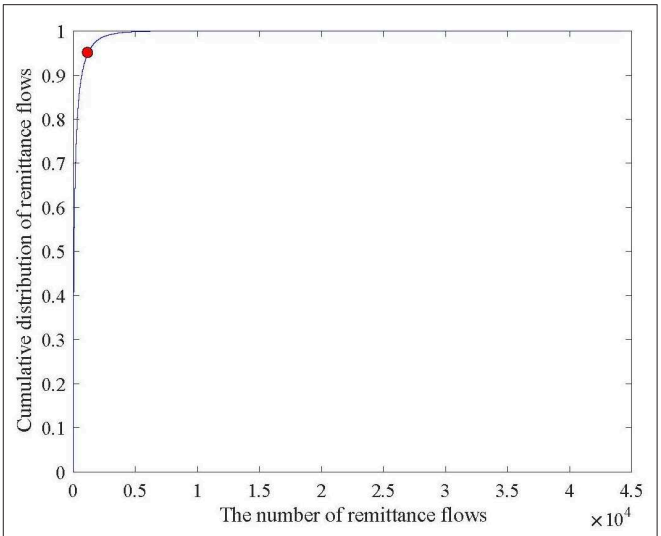


FIGURE 1 | The cumulative distribution of remittance.

4. EMPIRICAL RESULTS

4.1. Global Remittance Network

The global remittance consists of 210 economies and their bilateral remittance flows. As far as the economies with low remittances are concerned, it is natural to categorize them as either low population countries or isolated regions, because they take low levels of participation in remittance system. The zero-value remittances of these economies make few contributions to the total remittance flow [28]. Thus, specific thresholds are set to filter out zero-value flows and ensure enormous remittance flows in the network. In order to identify a reasonable threshold, we analyze the magnitudes of all accumulated remittance flows.

Figure 1 shows the cumulative distribution of all accumulated remittance flows. One should focus on the significant changes of slope below and above 95th percentile of cumulative distribution [30], implying that sum of the top 1,200 edges plays an important role in the global remittance system. In order to highlight the significant remittance relationships among economies, we filter the original network with 210 nodes and 12,687 edges into a core network with 172 nodes and 1,200 edges by setting the 95th percentile as a threshold. **Figure 2** shows the GRN before and after the filtering procedure. Note that some low population countries and isolated regions are removed from the GRN, including American Samoa, Aruba, Channel Islands and so on.

4.2. Complex Network Analysis

4.2.1. Degree Distribution

Figure 3 displays the cumulative distribution of degree on log-log axes in the GRN. Intuitively, there exists a large number of low-degree nodes and a small number of high-degree nodes. Thus, the degree distribution is of asymmetry. To further identify the degree distribution, the method of maximum-likelihood and Kolmogorov-Smirnov (KS) test are applied.

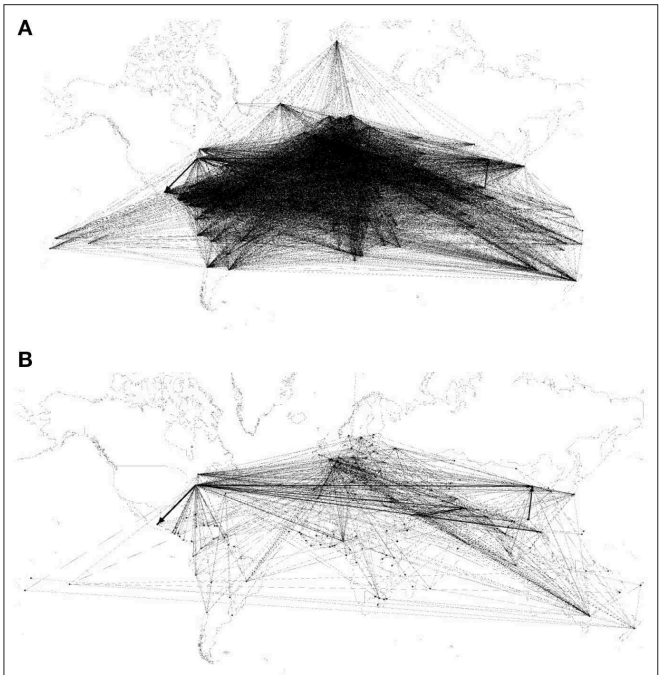


FIGURE 2 | Global remittance network. **(A)** Before filtering. **(B)** After filtering.

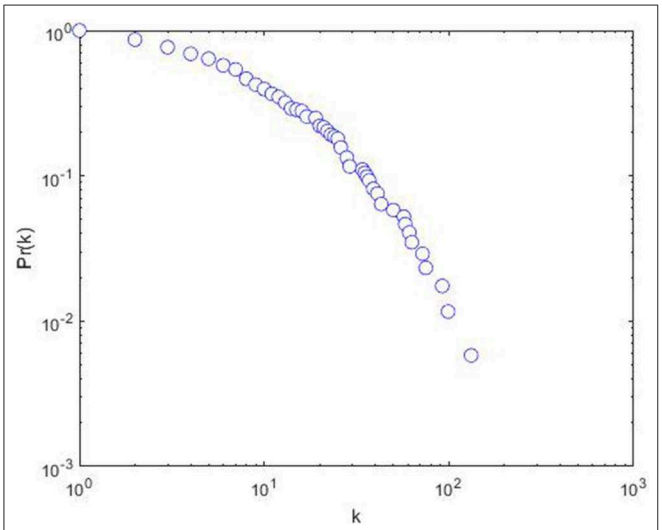


FIGURE 3 | The cumulative distribution of degree on log-log axes.

Table 1 presents the results of goodness-of-fit for three common statistical distributions [34], i.e., exponential, log-normal and power law distribution. Note that $P = 0.00$ in the case of exponential distribution. Thus, the exponential distribution is ruled out. However, the P-values of log-normal and power law are equal to 0.23 and 0.38, respectively, indicating that the log-normal and power law distribution are not ruled out. Both log-normal and power law distribution are likely fits since they can exhibit a large number of low-degree nodes and few high-degree nodes [35]. It implies that a small number of core nodes play important roles in the GRN. This also provides a theoretical basis for our subsequent analysis of key nodes in section 4.2.2.

TABLE 1 | The results of goodness-of-fit for degree distribution.

	Exponential	Log-normal	Power law
Distribution	$e^{-\lambda x}$	$\frac{1}{x\sigma\sqrt{2\pi}}e^{-(lnx-\mu)^2/2\sigma^2}$	$x^{-\alpha}$
Parameter	$\lambda = 13.95$	$\mu = 1.93, \sigma = 1.22$	$\alpha = 2.79$
P-value	0.00	0.23	0.38

The upper part of the table reports the exponential, log-normal and power law distribution, respectively. The parameters in each distribution is computed by the method of maximum-likelihood fitting. Furthermore, KS test is applied to examine the distribution. If P-value is reasonably large (say, $P > 0.1$) [34], then the corresponding distribution is not ruled out.

4.2.2. Centrality Measures

In order to identify key economies, centrality measures are employed to evaluate the importance of economies. **Figure 4** shows top 15 key economies in terms of degree, node strength and betweenness centrality measures, respectively. Note that the rankings based on different network centrality measures are distinct from one another, indicating that various economies tend to play distinct roles in GRN. However, a number of key economies, such as the United States, France, the United Kingdom, Germany, Spain, Italy, China, Russian Federation and India, are always at the forefront of the top 15 key economies, highlighting their importance in the remittance system.

For the out-degree and out-strength centrality case, the higher the economy's centrality is, the larger influence it possesses in the network. Note that the United States ranks first in both out-degree and out-strength centrality due to its economic power and immigration history. Further calculations show that US alone reimburses around 24.60% of total remittances to the rest of the world as payment, indicating that the country plays a significant role as a remittance supplier in GRN. In addition, affluent democracies and major oil and gas producing countries, such as Canada, the United Kingdom, Germany, France, Hong Kong (China), Japan, Saudi Arabia, United Arab Emirates and Kuwait, also become primary remittances suppliers.

In terms of in-degree and in-strength centrality, economies with high centrality are major remittance recipient countries, mainly including developing countries with considerable labor force like India, China, Philippines etc. A possible interpretation is that most of labor abundant countries reap up remittances by strategically converting abundant labor into capital [36]. In addition, developed countries like France, Germany, Belgium also obtain high rank of the centralities. The results do not come as surprises given that European Union (EU) citizens are allowed to move freely within the EU labor market. This leads to the fact that labor mobility in common market promotes the movement of remittances [37]. Thus, developed countries also become remittance recipient countries.

In addition to degree and node strength centrality, an economy's importance in the GRN can be evaluated by

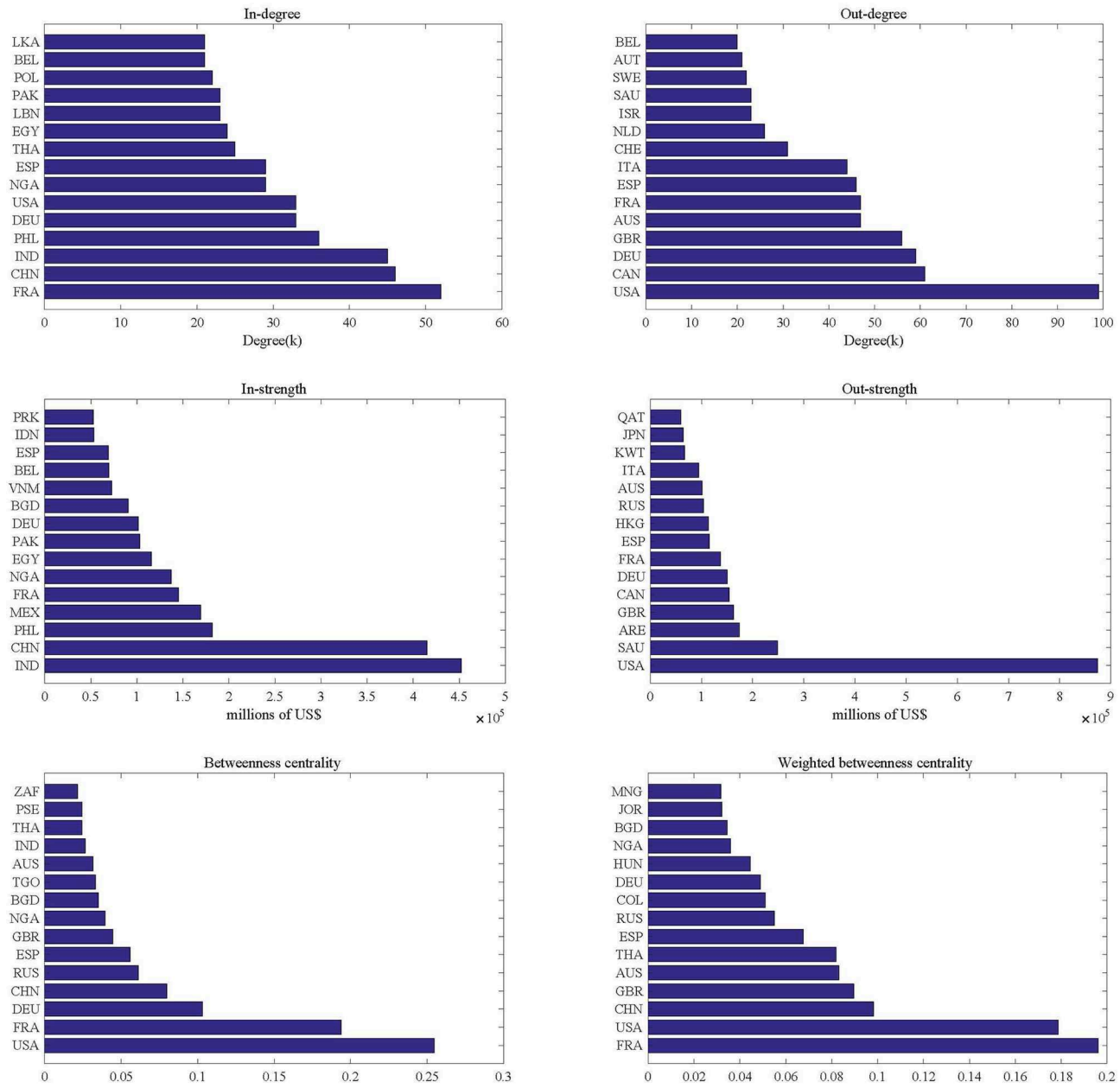


FIGURE 4 | Top 15 important economies based on centrality measures.

betweenness centrality, which measures the ability to transfer remittance flows from one economy to another. As shown in **Figure 4**, whether weighted or not, the betweenness centralities of the United States and France are significantly larger than those of other economies. Therefore, the United States and France play critical roles as bridge in the remittance system. There could be two possible explanations on the intermediary roles of the United States and France. On one hand, the United States is a nation of migrants, which has close cooperations with other economies in terms of remittance. On the other hand, there are many major corridors in remittance between France and its neighbor countries due to the geographical proximity[2].

4.2.3. Regional Community

The detection of communities is very important as community structures often correspond to functional organizations [14, 33, 38]. In terms of the remittance network, communities not

[2]It is supported by Eurostat statistics explained, which can be found at http://ec.europa.eu/eurostat/statistics-explained/index.php/Personal_remittances_statistics.

only uncover its regional structure features, but also provide insights for understanding the collaboration mechanism among economies in the same community. **Figure 5** visualizes the regional community structure in GRN.

As shown in **Figure 5**, the GRN is obviously divided into 6 communities where economies in the same community are labeled by a specific color. For each community, we present their major members and intra-remittances:

- C1 - The largest community, dominated by most of the economies in EU and South America, generates $ 6.35E+05 million intra-community remittances.
- C2 - The second-largest community, mainly consisting of Australia, the United Kingdom, the Nordic countries and Africa, contributes to $ 1.54E+05 million intra-community remittances.
- C3 - The third-largest community, led by the United States as well as Central America, transfers $ 4.46E+05 million intra-community remittances.
- C4 - The fourth-largest community, represented by the Middle East, possesses the largest intra-community remittances that is $ 6.47E+05 million.
- C5 - The fifth-largest community, locating in East Asia and Southeast Asia, holds $ 3.15E+05 million intra-community remittances.
- C6 - The smallest community, dominated by Eastern Europe, has $ 1.51E+05 million intra-community remittances.

It is interesting to note that economies in the same regional economy cooperative organizations tend to be included in the same community, such as EU and Latin American Integration Association (LAIA) in C1, Economic Community of West African States (ECOWAS) in C2, members of Central American Free-Trade Agreement(CAFTA) in C3, Gulf Cooperation Council Countries (GCC) in C4, Association of Southeast Asian Nations(ASEAN) in C5 and Eurasian Economic Union (EAEU) in C6. Thus, the community structure of the GRN appears to have significant correlations with geographical and economic factors. A possible interpretation is that labor mobility in regional cooperative organizations promotes the movement of remittances. Additionally, further calculations show that the intra-community remittances of C1–C6 account for 66.07% of total remittances in GRN, highlighting the regional integration of the GRN.

Figure 6 describes the intra-community and inter-community remittance flows. As **Figure 6** shows, the economies belonging to the same community form sub-units with close remittance relationships. However, economies not only participate in remitting within the same community, but also engage themselves into the inter-community transfers, which account for 33.93 % of total remittances.

4.2.4. Topological Stability Test of GRN

Remittance system may have some extreme risks such as migration or remittance restriction, to which corresponds node and edge removal in GRN. Such removal will affect the stability of GRN and the normal operation of the system. The approaches employed in robustness test of networks are usually node attack method, which removes some nodes and all edges connected to them from the network [39]. There are two kinds of node attack method: one is stochastic removal method and the other

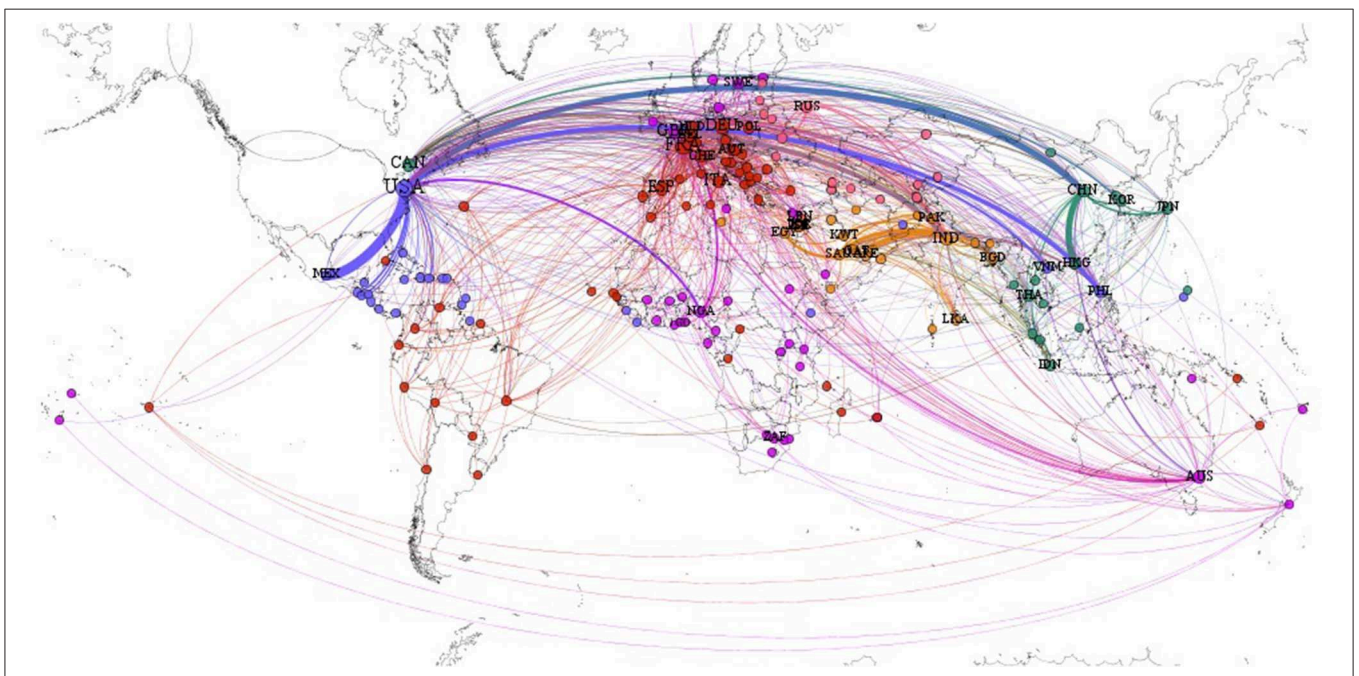


FIGURE 5 | The regional community structure in GRN. Notes: The coloring of nodes and their linkages are the same for each community, i.e., red for C1, purple for C2, blue for C3, orange for C4, green for C5 and pink for C6. The size of a linkage corresponds to the magnitude of remittance flow.

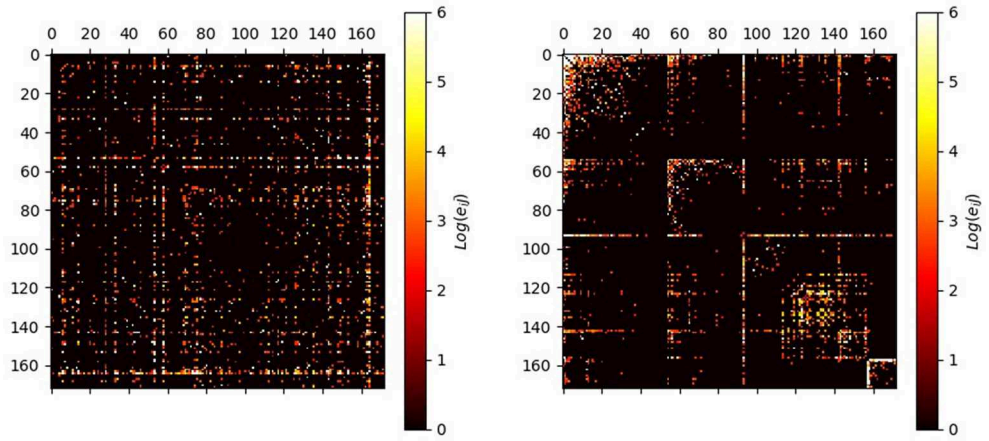


FIGURE 6 | Remittance matrix before and after permutation based on community structure.

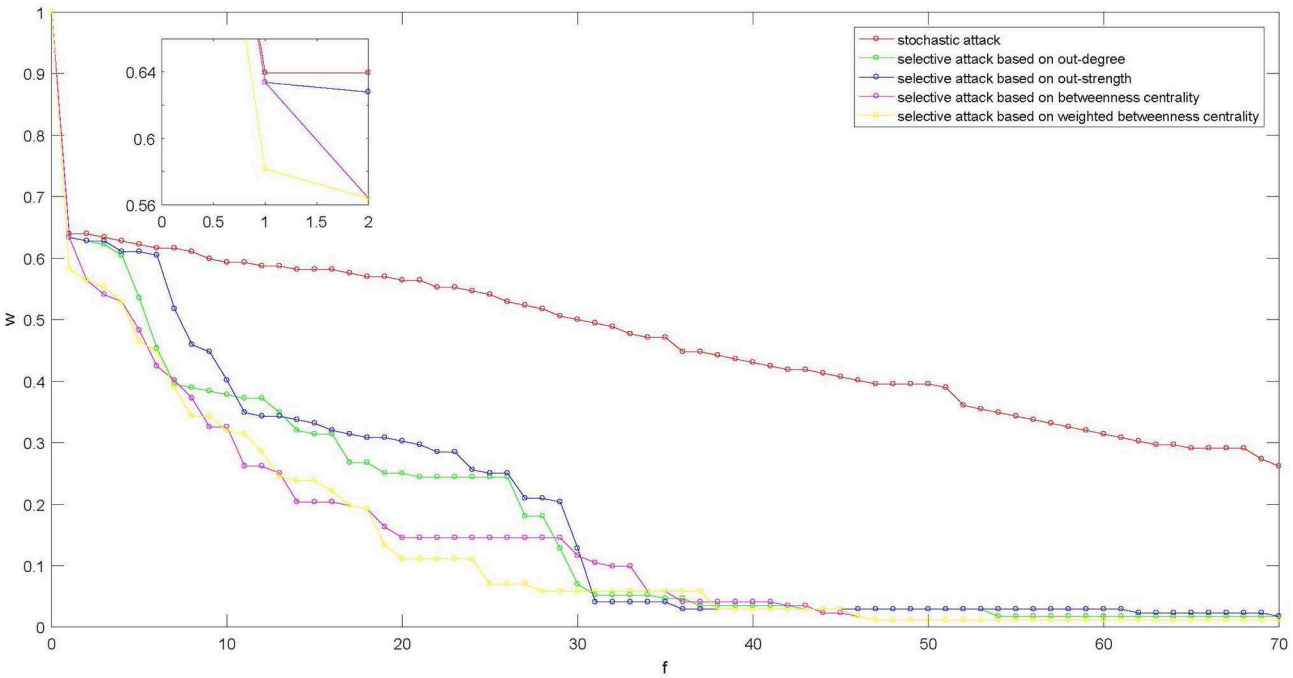


FIGURE 7 | The maximum SCC size of the GRN. Notes: f is the number of removal nodes. The selective removal removes nodes in decreasing order of their centrality.

is selective removal method. The former removes some nodes from the network in a random manner, while the latter removes the nodes in a special order. If the structural properties, such as connectivity, have no significant difference after the attack, it is natural to define that the network is robust. Since the maximum strongly connected component(maximum SCC) size measures the connectivity of a directed network, we evaluate topological stability by calculating the ratio w of the maximum SCC size of the new network to that of the original network.

Figure 7 displays the topological stability test of the GRN. As shown in **Figure 7**, no matter for the selective removal or stochastic removal, when the first nodes are removed, $w < 0.64$, implying that network connectivity drops rapidly under these two kinds of node attack methods. However, the impact of selective removal on connectivity is larger than that of stochastic removal. The same phenomenon can also be found at the aggregate level. Note that after a number of nodes removal, the impact of selective removal on the network is significantly larger than that of stochastic removal. Furthermore, in comparing the stochastic removal methods, the connectivity of the network under betweenness (weighted betweenness) centrality-based attack is lower than that under out-degree(out-strength)-based

attack. This indicates that selective attacks on the nodes with high betweenness (weighted betweenness) centrality are much more effective in the GRN. Therefore, in order to avoid the connectivity collapse of the remittance system, policy makers should pay more attention to major transshipment nodes.

5. CONCLUSION

To a large extent, the existing literature on the analysis of key economies for global remittance predominantly focuses on the influential remittance suppliers and recipients as well as circular patterns between economy pair. Literature dealing with the intermediary roles of the key economies, community structure and topological stability of remittance network seems to be scarce; such studies are however significant for us to understand the local properties as well as the global properties of the remittance network.

In this paper, by employing complex network approach, we present a detailed discussion of key economies for global remittance, and the findings can be summarized as follows: (i) the degree distribution is of heterogeneity, suggesting that GRN is a heterogeneous network; (ii) some key economies, such as the United States, France, India and China, are always at the forefront of centrality measures; (iii) 6 communities are detected, where economies in the same regional economy cooperative organizations tend to be classified in the same community; (iv) 66.07% of total remittances belongs to intra-community flows, indicating that economies present the characteristic of regionalization; (v) GRN is fragile to node removal, particular the selective removal based on betweenness centrality.

Although this paper builds a global remittance network and investigates its key economies, there are still some topics worth further studying. For example, economies play different roles in the network due to its heterogeneity. The positions of economies in the network might have an impact on remittance flows. Thus, it is worth further studying the relationship between remittance flows and economies' positions. This will be our future study.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.worldbank.org/en/topic/migrationremittancesdiasporaissues/brief/migration-remittances-data.

AUTHOR CONTRIBUTIONS

SW and CH: conceptualization and project administration. SW: methodology, formal analysis, resources, and data curation. YT: software, investigation, and visualization. ML and YD: validation. SW, CH, and YD: writing-original draft preparation. YD: writing-review and editing. CH: supervision. SW, CH, and ML: funding acquisition.

FUNDING

Partially supported by the National Natural Science Foundation of P. R. China (Nos. 11971076, 71471020), Hunan Provincial Natural Science Foundation (No. 2019JJ50650), Scientific Research Fund of Hunan Provincial Education Department (No. 18C0221), and Innovation Fund Designated for Graduate Students of Hunan Province (No. CX2018B571).

ACKNOWLEDGMENTS

The authors are extremely grateful to three reviewers and the editor for their valuable comments and suggestions, which have contributed a lot to the improved presentation of this paper.

REFERENCES

1. Bettin G, Lucchetti R, Pigini C. A dynamic double hurdle model for remittances: evidence from Germany. *Econ Model*. (2018) **73**:365–77. doi: 10.1016/j.econmod.2018.04.012
2. Group WB. *Migration and Remittances: Recent Developments and Outlook-Transit Migration*. Washington, DC: The World Bank (2018).
3. Huang C, Cao J, Wen F, Yang X. Stability analysis of SIR model with distributed delay on complex networks. *PLoS ONE*. (2016) **11**:e0158813. doi: 10.1371/journal.pone.0158813
4. Hu H, Yuan X, Huang L, Huang C. Global dynamics of an SIRS model with demographics and transfer from infectious to susceptible on heterogeneous networks. *Math Biosci Eng*. (2019) **16**:5729–49. doi: 10.3934/mbe.2019286
5. Gong X, Lin B. Time-varying effects of oil supply and demand shocks on China's macro-economy. *Energy*. (2018) **149**:424–37. doi: 10.1016/j.energy.2018.02.035
6. Song DM, Jiang ZQ, Zhou WX. Statistical properties of world investment networks. *Phys A Stat Mech Appl*. (2009) **388**:2450–60. doi: 10.1016/j.physa.2009.03.004
7. Dai YH, Xie WJ, Jiang ZQ, Jiang GJ, Zhou WX. Correlation structure and principal components in the global crude oil market. *Empiric Econ*. (2016) **51**:1501–19. doi: 10.1007/s00181-015-1057-1
8. Davis KF, D'Odorico P, Laio F, Ridolfi L. Global spatio-temporal patterns in human migration: a complex network perspective. *PLoS ONE*. (2013) **8**:e53723. doi: 10.1371/journal.pone.0053723
9. Fagiolo G, Mastrorillo M. Does human migration affect international trade? A complex-network perspective. *PLoS ONE*. (2014) **9**:e97331. doi: 10.1371/journal.pone.0097331
10. Li M, Xiao Q, Wang Y, Zhou W. Statistical properties of the mutual transfer network among global football clubs. *Int J Mod Phys B*. (2018) **32**:1850320. doi: 10.1142/S0217979218503204
11. Huang C, Zhang H, Cao J, Hu H. Stability and Hopf bifurcation of a delayed prey-predator model with disease in the predator. *Int J Bifurcat Chaos*. (2019) **29**:1950091. doi: 10.1142/S0218127419500913
12. Huang C, Su R, Cao J, Xiao S. Asymptotically stable high-order neutral cellular neural networks with proportional delays and D operators. *Math Comput Simulat*. (2020) **171**:127–35. doi: 10.1016/j.matcom.2019.06.001
13. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identification of influential spreaders in complex networks. *Nat Phys*. (2010) **6**:888–93. doi: 10.1038/nphys1746
14. Albert SR, Clara G, Sergio G, Alex A. Information transfer in community structured multiplex networks. *Front Phys*. (2015) **3**:61. doi: 10.3389/fphy.2015.00061
15. Iztok F, Iztok F, Matjavz P. Toward the discovery of citation cartels in citation networks. *Front Phys*. (2016) **4**:49. doi: 10.3389/fphy.2016.00049

16. Dai Z, Wen F. Some improved sparse and stable portfolio optimization problems. *Finan Res Lett*. (2018) **27**:46–52. doi: 10.1016/j.frl.2018.02.026
17. Dai Z, Wang F. Sparse and robust mean-variance portfolio optimization problems. *Phys A Stat Mech Appl*. (2019) **523**:1371–8. doi: 10.1016/j.physa.2019.04.151
18. Lu Q, Gu C, Qin X, Yang H, Wu G. State network approach to characteristics of financial crises. *Phys A Stat Mech Appl*. (2018) **492**:1120–8. doi: 10.1016/j.physa.2017.11.042
19. Majapa M, Gossel SJ. Topology of the South African stock market network across the 2008 financial crisis. *Phys A Stat Mech Appl*. (2016) **445**:35–47. doi: 10.1016/j.physa.2015.10.108
20. Han R, Li M, Chen W, Zhou W, Stanley HE. Structural properties of statistically validated empirical information networks. *Phys A Stat Mech Appl*. (2019) **523**:747–56. doi: 10.1016/j.physa.2019.03.010
21. Wen F, Ye Z, Yang H, Li K. Exploring the rebound effect from the perspective of household: an analysis of China's provincial level. *Energy Econ*. (2018) **75**:345–56. doi: 10.1016/j.eneco.2018.08.018
22. Wen F, Zhao C, Hu C. Time-varying effects of international copper price shocks on China's producer price index. *Resour Policy*. (2019) **62**:507–14. doi: 10.1016/j.resourpol.2018.10.006
23. Wen F, Yang X, Zhou W. Tail dependence networks of global stock markets. *Int J Finan Econ*. (2019) **24**:558–67. doi: 10.1002/ijfe.1679
24. Yang X, Wen S, Zhao X, Huang C. Systemic importance of financial institutions: a complex network perspective. *Phys A Stat Mech Appl*. (2019) **545**:123448. doi: 10.1016/j.physa.2019.123448
25. Xie W, Li M, Xu H, Chen W, Zhou W, Stanley HE. Quantifying immediate price impact of trades based on the k-shell decomposition of stock trading networks. *Europhys Lett*. (2016) **116**:28006. doi: 10.1209/0295-5075/116/28006
26. Gong X, Wen F, Xia XH, Huang J, Pan B. Investigating the risk-return trade-off for crude oil futures using high-frequency data. *Appl Energy*. (2017) **196**:152–61. doi: 10.1016/j.apenergy.2016.11.112
27. Wen F, Xu L, Ouyang G, Kou G. Retail investor attention and stock price crash risk: evidence from China. *Int Rev Finan Anal*. (2019) **65**:101376. doi: 10.1016/j.irfa.2019.101376
28. Lillo F, García L, Santander V. Dynamics of global remittances: a graph-based analysis. *Math Soc Sci*. (2017) **87**:64–71. doi: 10.1016/j.mathsocsci.2017.02.005
29. Lillo F, Garay JAM. The global remittance network: an inflow and outflow analysis. *J Math Sociol*. (2019) **43**:59–75. doi: 10.1080/0022250X.2018.1496917
30. Chen B, Li JS, Wu XF, Han MY, Zeng L, Li Z, et al. Global energy flows embodied in international trade: a combination of environmentally extended input-output analysis and complex network analysis. *Appl Energy*. (2018) **210**:98–107. doi: 10.1016/j.apenergy.2017.10.113
31. Yang X, Wen S, Liu Z, Li C, Huang C. Dynamic properties of foreign exchange complex network. *Mathematics*. (2019) **7**:832. doi: 10.3390/math7090832
32. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: generalizing degree and shortest paths. *Soc Netw*. (2010) **32**:245–51. doi: 10.1016/j.socnet.2010.03.006
33. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech*. (2008) **2008**:155–68. doi: 10.1088/1742-5468/2008/10/P10008
34. Clauset A, Shalizi CR, Newman MEJ. Power-law distributions in empirical data. *SIAM Rev*. (2009) **51**:661–703. doi: 10.1137/070710111
35. Broido AD, Clauset A. Scale-free networks are rare. *Nat Commun*. (2019) **10**:1017. doi: 10.1038/s41467-019-08746-5
36. Mamun MA, Sohag K, Uddin GS, Shahbaz M. Remittance and domestic labor productivity: evidence from remittance recipient countries. *Econ Model*. (2015) **47**:207–18. doi: 10.1016/j.econmod.2015.02.024
37. Javorcik BS, Özden Ç, Spatareanu M, Neagu C. Migrant networks and foreign direct investment. *J Dev Econ*. (2011) **94**:231–41. doi: 10.1016/j.jdeveco.2010.01.012
38. Guimera R, Amaral L. Functional cartography of complex metabolic networks. *Nature*. (2005) **433**:895–900. doi: 10.1038/nature03288
39. Wen L, Guan L, Shen J, Song L, Cui L. A complex network for studying the transmission mechanisms in stock market. *Phys A Stat Mech Appl*. (2017) **484**:345–57. doi: 10.1016/j.physa.2017.04.043

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wen, Tan, Li, Deng and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 29 April 2020
doi: 10.3389/fphy.2020.00132

The Time-Dependent Lead-Lag Relationship Between WTI and Brent Crude Oil Spot Markets

Yanhong Yang[1†], *Yinghui Shao*[2*†], *Haolin Shao*[3,4] *and Xin Song*[5*]

[1] Faculty of Education, East China Normal University, Shanghai, China, [2] School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, China, [3] Department of Statistics, Graduate School of Arts and Sciences, Columbia University, New York, NY, United States, [4] School of Entrepreneurship and Management, Shanghaitech University, Shanghai, China, [5] Business School, University of Shanghai for Science and Technology, Shanghai, China

Based on daily returns, we comprehensively characterize the lead-lag relationship between Brent and WTI crude oil spot markets from 1987 to 2017 with the non-parametric symmetric thermal optimal path (TOPS) method. The empirical results indicate that WTI spot price leads Brent spot price slightly, which provides support to the price leadership of WTI over Brent. However, the lead-lag relationship is volatile and sensitive to extreme events like geopolitical conflict and policy shift. Due to the concerns about future oil supply triggered by the two Gulf wars, both WTI and Brent experienced extreme uncertainty and co-moved closely during wartime. Notably, the TOPS method captures the structural break in the WTI-Brent price spread in 2011 which is influenced by the U.S. oil export ban and transportation bottleneck. After the lift of the ban, the two benchmark prices have reconnected. The lead-lag signals basically coincide with major influential changes in the oil markets, which suggests that the TOPS method provides a viable approach to reflecting the impact of extreme events on the crude oil prices motion.

Keywords: oil prices, spot market, lead-lag relationship, symmetric thermal optimal path, WTI, Brent, extreme events

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Can Zhong Yao,
South China University of Technology, China
Federico Battiston,
Central European University, Hungary

***Correspondence:**
Yinghui Shao
yinghuishao@126.com
Xin Song
silvia_sx@163.com

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *20 January 2020*
Accepted: *06 April 2020*
Published: *29 April 2020*

Citation:
Yang Y, Shao Y, Shao H and Song X (2020) The Time-Dependent Lead-Lag Relationship Between WTI and Brent Crude Oil Spot Markets. Front. Phys. 8:132. doi: 10.3389/fphy.2020.00132

1. INTRODUCTION

Brent and Western Texas Intermediate (WTI) are the two predominant benchmarks for crude oil in global markets. Their price relationship has been changeable all the time. Research on properties of their movements have been continually growing. Using copulas, Reboredo [1] reports evidence of symmetric upper and lower tail dependence structure between crude oil prices. Adopting a rolling estimation, Kao and Wang [2] trace the changing path of WTI information share from 1991 to 2009. Their results demonstrate that WTI has lost the status of leading price to Brent since 2004. While Liu et al. [3] utilize high-frequency data from 2008 to 2011 and show that WTI is more important than Brent for price discovery process when they are co-integrated. Elder et al. [4] also find WTI maintains a dominant role over Brent in price discovery from 2007 to 2012. Coronado et al. [5] apply a one-tailed non-parametric Granger causality test to study the co-movements for oil prices and alert a bi-directional feedback pattern between Brent and WTI from 2013 to 2015.

Moreover, the price spread between WTI and Brent is also essential to reflect the competitiveness among the two benchmarks. Chen et al. [6] verify the structural change of WTI-Brent spread process (December 2010). Scheitrum et al. [7] identify a breakpoint in January 2011. Caporin et al. [8] document two structural breaks in February 2011 and October 2014. Ye and Karali [9]

detect three structural breaks in February 2005, the second is in December 2010 and the third is in April 2013.

Additionally, crude oil market is strongly associated to global economy, energy policy and financial market [10–12]. Correspondingly, crude oil price is fragile and subject to sudden incidents and shocks, such as political events, economic events (the Asian financial crisis in 1997), energy policies (the lift of U.S. crude oil export ban) and military conflicts (the Gulf War in 1990, the Iraq War in 2003, etc.). Thus, extensive interest has been devoted to the evident impacts of extreme and irregular events on crude oil markets. Zhang et al. [13] apply an empirical mode decomposition (EMD) based event analysis to evaluate the effects of the Persian Gulf War and the Iraq War on WTI and Brent. They find that impact of extreme events between the two prices exhibits no significant difference. Charles and Darné [14] discern that particular events have a bearing on large shocks in the crude oil prices based on the generalized autoregressive conditional heteroskedasticity (GARCH) models. Similarly, Basistha and Kurov [15] examine the impact of monetary policy shocks on energy prices and find different response patterns. Luong et al. [16] document that WTI and Brent have reintegrated since the lift of the U.S. crude oil export ban.

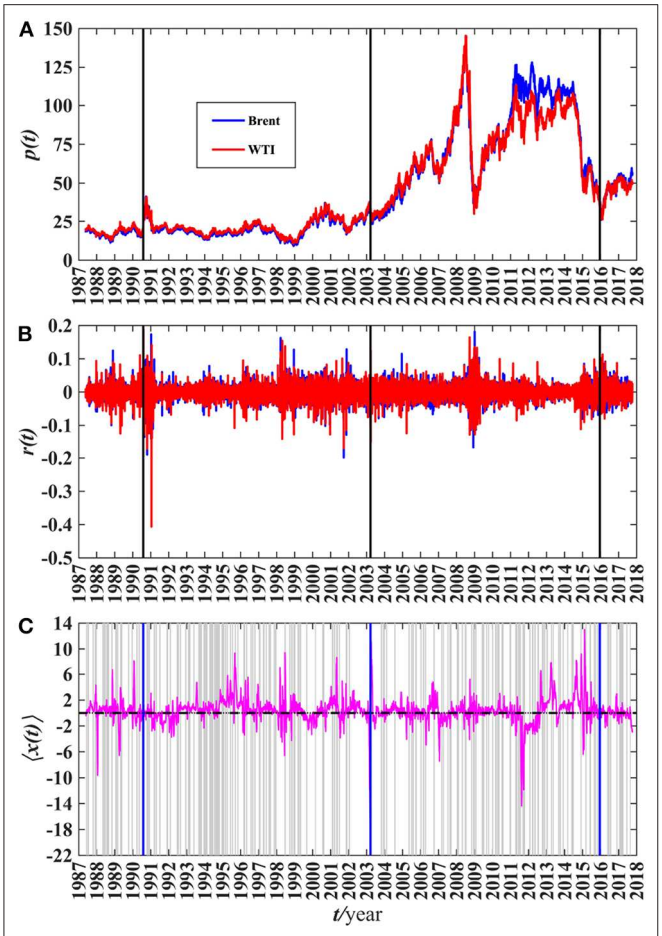


FIGURE 1 | Spot price, returns, and lead-lag path $\langle x(t)\rangle$ from 20 May 1987 to 10 October 2017, respectively labeled by **(A–C)**. The vertical lines mark the outbreaks of the Gulf War and the Iraq War as well as the lift of export ban. **(A)** Brent and WTI spot price. **(B)** Returns. **(C)** Lead-lag path $\langle x(t)\rangle$. The gray shades represent the periods when the consistency test is significant at the 5% level.

Though numerous researchers have investigated the crude oil markets and their interaction, the interlinkages and dynamics between WTI and Brent during extreme events is still a substantial point of discussion. In this paper, we investigate the time-dependent lead-lag relationship between the Brent and WTI crude oil spot price during 1987–2017 via the symmetric thermal optimal path (TOPS) method developed by Meng et al. [17]. The two non-parametric approaches can effectively identify the time-dependent lead-lag correlation between two time series, which have been applied to financial market [12, 18–23]. The TOPS method is an improved version of the novel thermal optimal path (TOP) method [24–26] with smaller biases and higher accuracy.

The TOP/TOPS method does not confirm the genuine causality but provides information on the time lag between two series [17, 24]. Moreover, crude oil price is highly sensitive to a variety of factors like policy changes and geopolitical events. Therefore, one should be careful when examining the price discovery of crude oil market via TOPS method. Further information is needed to determine the price leadership relationship. In this manuscript, we mainly focus on the time dependent lead-lag relationship between the two markets. Inspired by Jiang et al. [27], we separate the whole sample into four time periods by the outbreak of the Gulf War, the Iraq War and the lift of U.S. crude oil export ban to examine the lead-lag relationship during different periods. To test the significance of the lead-lag structure, we implement the self-consistency test introduced by Meng et al. [17].

The paper is organized as follows. In section 2, data and summary statistics are presented. Section 3 depicts the methodology. Section 4 provides the empirical results, and section 5 concludes the paper.

2. DATA DESCRIPTION

This study uses daily data collected from the US Energy Information Administration (EIA) website. The data is given in the U.S. dollar per barrel spanning from 20 May 1987 to 10 October 2017 and illustrated in **Figure 1**. Note that the three vertical solid lines in **Figure 1A** divide the whole period into four segments: before the Gulf War, before the Iraq War, after the Iraq War and after the lift of export ban.

We use logarithmic returns for the analysis. The logarithmic returns of the spot price are defined as follows:

$$r(t) = \ln p(t) - \ln p(t-1), \tag{1}$$

where $p(t)$ represents the crude oil spot price at time t. The logarithm returns are in illustrated in **Figure 1B**.

Table 1 summarizes the statistics for the two return series. The descriptive statistics are quite similar across the two markets. The ADF test shows the stationarity of daily Brent and WTI crude oil spot returns. In comparison with Brent spot returns, WTI spot returns offer slightly higher volatility levels. Both of the returns

TABLE 1 | Descriptive statistics for Brent and WTI spot returns.

	ADF	Mean $\times 10^2$	Max	Min	Std	Skew	Kurtosis	JB
	p-value							*p*-value
Brent	0.001	0.015	0.1813	−0.3612	0.0230	−0.5538	16.6006	0.001
WTI	0.001	0.013	0.1887	−0.4064	0.0247	−0.6879	17.4420	0.001

are left-skewed and leptokurtic. The high values for the kurtosis statistic suggest that the distribution of returns has larger, thicker tails than normal distribution. The Jarque–Bera test results also suggest that both returns are not normally distributed.

Specifically, the Pearson correlation coefficient between spot returns of Brent and WTI is 0.60047, which indicates that Brent spot returns are slightly positive correlated to WTI spot returns.

3. METHODOLOGY

Meng et al. [17] improve the TOP method [24–26] and propose the TOPS method. The TOPS method is briefly described as below.

Consider two standardized time series $X(t_1): t_1 = 1, 2, \cdots, N-1$ and $Y(t_2): t_1 = 1, 2, \cdots, N-1$, the distance matrix between $X(t_1)$ and $Y(t_2)$, denoted by $E_{X,Y}$, is defined as

$$\epsilon(t_1, t_2) = |X(t_1) - Y(t_2)|. \tag{2}$$

The lead–lag structure between two time series is detected by searching for the one-to-one mapping $t_1 \rightarrow t_2 = \phi(t_1)$ that satisfies

$$\min_{\{\phi(t_1), t_1=0,1,\ldots,N-1\}} E := \sum_{t_1=0}^{N-1} |X(t_1) - Y(\phi(t_1))|, \tag{3}$$

which is a global minimization with continuity constraint

$$0 \leq \phi(t_1 + 1) - \phi(t_1) \leq 1\,. \tag{4}$$

And rotate the coordinate system (t_1, t_2) to (t, x)

$$\begin{cases} t = t_2 + t_1 \\ x = t_2 - t_1. \end{cases} \tag{5}$$

Denote the probability for a path through (x, t) as $W(t, x)/W(t)$, where $W(t, x)$ is the partition function of x at fixed t and

$$W(t) = \sum_x W(t, x). \tag{6}$$

Notice that, the partition function $W(t, x)$ obeys the following recursion relation:

$$W(t, x) = [W(t-1, x-1) + W(t-1, x+1) + W(t-2, x)]e^{-\epsilon(t,x)/T}. \tag{7}$$

Then implement a time-reversed invariant node weight process by averaging the weights determined along the time-forward and time-backward directions:

$$\langle x(t) \rangle = \sum_x x \frac{\overrightarrow{W}(t, x)/\overrightarrow{W}(t) + \overleftarrow{W}(t, x)/\overleftarrow{W}(t)}{2}, \tag{8}$$

where the arrow $\rightarrow$ represents the recursive weight process in a time-forward direction, and the arrow $\leftarrow$ refers to the recursive weight process in a time-backward direction. Therefore, one can obtain the optimal thermal averaged path trajectory $\langle x(t) \rangle$.

4. RESULTS

This section presents analysis results for the lead-lag features between Brent and WTI from 20 May 1987 to 10 October 2017. Based on the empirical experiment [17], we implement the TOPS analysis with temperature $T = 2$, which is the optimal parameter.

To examine the lead-lag structure during extreme events, we divide the full-sample period into four sub-samples by the outbreaks of the Gulf War (2 August 1990), the Iraq War (20 March 2003), and the lift of export ban (18 December 2015). We further carry out self-consistency tests in a moving window with size of 10 days to assess whether the lead-lag path is significant, and illustrate the result in **Figure 1C**. Periods that pass the test are marked in gray shades. Basically the consistency test is significant. Explanatorily, the four different subsamples are denoted as Sub 1, Sub 2, Sub 3, and Sub 4 based on their chronological orders, respectively. **Table 2** shows the summary of $\langle x(t) \rangle$ during the sample period and **Figure 2** shows the boxplots and probability density curves of $\langle x(t) \rangle$.

As illustrated by the boxplots in **Figure 2A**, the 4 subsamples paths are highly similar except the outliers. See in **Figures 2B–F**, the probability density curve of the whole sample period TOPS path is almost symmetrical around 0. More accurately, positive $\langle x(t) \rangle$ takes more percentage of the whole paths (64.06%). The positive lead-lag paths reveal that WTI spot price is ahead of Brent spot price slightly with certain time lags, which means the price changes in WTI preceded the price changes in Brent. This is in accord with the fact that WTI holds the predominant position in the crude oil market except several episodes [3, 4, 7].

During the Sub 1 period from 20 May 1987 to 2 August 1990, the TOPS paths $\langle x(t) \rangle$ have 71.29% positive values, which indicates that Brent spot returns lag behind WTI spot returns over this period. The Iraq's invasion of Kuwait on 2 August, 1990 caused immediate disruption of the oil supply and triggered the oil price shock. The price of oil almost doubled in 2 months. WTI price peaked above $40 per barrel in October, while Brent price

TABLE 2 | Summary of the $\langle x(t)\rangle$ during whole period and the four sub-periods.

Period	Length	Mean	Median	Max	Min	Positive values %	Negative values %
Whole	7,520	0.33	0.35	13.02	−21.03	64.06	35.93
Sub 1	808	0.34	0.41	8.07	−9.64	71.29	28.59
Sub 2	3,100	0.44	0.44	9.44	−20.97	66.42	33.58
Sub 3	3,158	0.28	0.31	13.02	−21.03	62.16	37.84
Sub 4	454	−0.08	−0.03	2.94	−3.03	48.24	51.76

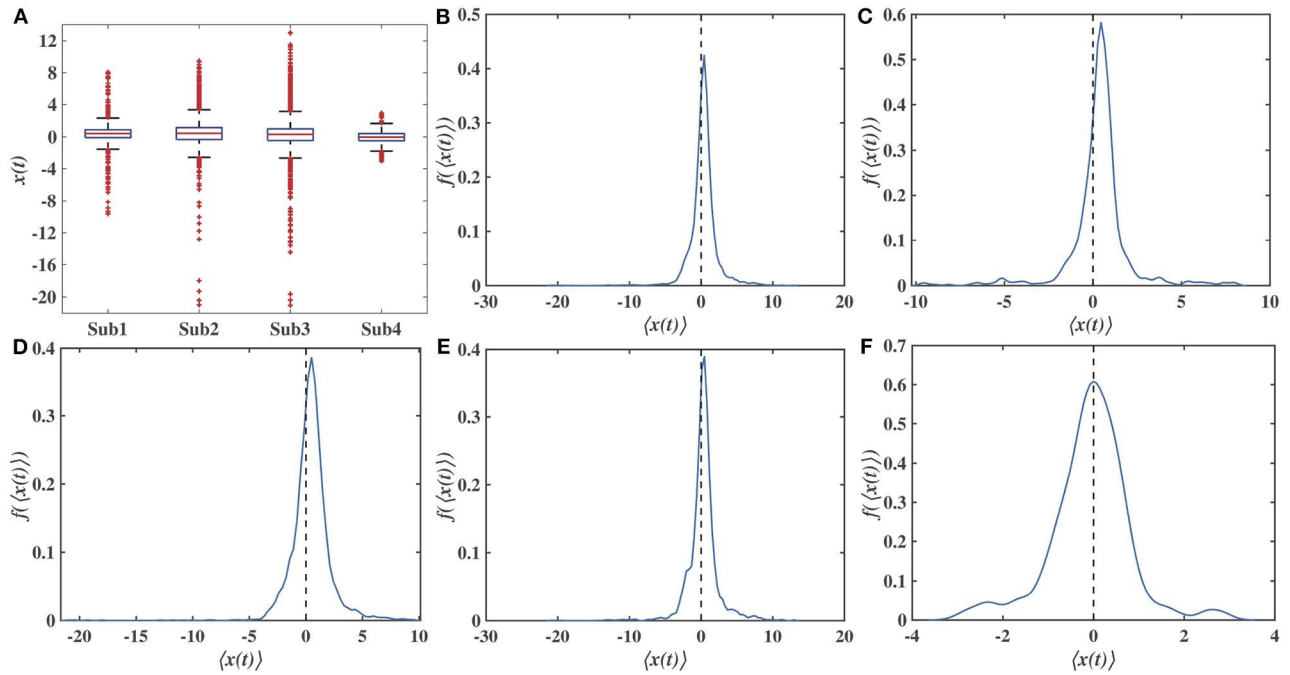


FIGURE 2 | Boxplots and probability density curve based on kernel density estimation of the $\langle x(t)\rangle$. **(A)** Boxplots of 4 subsamples. The box plots illustrate the 25 and 75% quartiles, the median and outliers. **(B–E)** Probability density curve of four subsamples. **(F)** Probability density curve of the whole sample period.

has risen to $41.45 per barrel on 27 September (see in **Figure 1A**). The alternate lead-lag dependence in September and October also suggests that the crude oil prices of WTI and Brent co-moved closely during periods of extreme market uncertainty [13].

As the situation gradually became clear, the uncertainty was eased and the market calmed down. From 1991 to the second half of 1992, Brent leads WTI (see in **Figure 1C**). Whereas, Brent is still lagged behind WTI in most cases during the Sub 2 period (with 66.42% positive $\langle x(t)\rangle$). Remarkably, the $\langle x(t)\rangle$ has two vibrations around the financial crises in 1995 and 1998, which coincides with the 1995 Russian crisis and the 1997 Asian financial crisis. Most intriguingly, we could observe a sharp vibration on 20 March 2003 in **Figure 1C**. The vibration is strong and temporal, which reflects the nervousness in the crude oil market triggered by the war. The weak alternate lead-lag relationship after the vibration implies that, the two markets react similarly under stress [13]. As concerns about long-term supply shortages eased and prices began to fall, a positive lead-lag path is observed in 2004.

In the Sub 3 period, the percentage that WTI spot leads Brent spot is 66.42%. The positive paths imply that WTI plays a leading role in most times while occasionally Brent takes the lead. For instance, the lead-lag relation reversion in 2007 might be associated with the financial crisis. Until 2011, WTI is more important than Brent for price discovery process [3, 4]. However, starting in 2011 the spread in the U.S. and international crude prices was enlarging, even reached -$29 per barrel (see in **Figure 1A**). WTI was trading at a large discount to Brent. Though several pipeline projects from Cushing to the Gulf Coast were built to expand transportation capacity after 2012 [16, 28], the U.S. crude oil could not be exported to the global marketplace to arbitrage away the spread due to the export ban [7] and transportation bottleneck [29]. Subsequently WTI was actually more a localized market than a globalized market. Then Energy Information Administration (EIA) replaced WTI with the North Sea Brent contract as the reference oil price [7]. The dramatic fluctuation of lead-lag path in 2011 (see in **Figure 1C**) supports the structural break that occurred in the WTI-Brent price spread

around 2000–2011 [6–9]. At this time, WTI is "viewed as a broken benchmark" [7]. Since 2013 the lead-lag paths switch to an upward oscillation, which implies Brent changes are lagging behind WTI changes with increasing time lag. However, WTI was disconnected from the global market due to the export ban and transportation bottlenecks, it might not be reasonable to claim that WTI serves as a price discovery center during this time.

However, things have changed in 2015. The significant Brent-WTI spread, the sharp drop in oil prices and a Congressional budget deal led to the U.S. policy shift on oil export [30]. On 18 December of this year, the forty-decades-old crude oil export ban was finally lifted. Since then US oil could be transported abroad. Also, trade and efficiency in the oil market increased [31]. Consequently, sustainable increases in pipeline capacity together with the lift helped WTI reconnect with international crude oil prices [16]. After the lift of export ban, WTI spot almost evolves simultaneously with Brent (see in **Figure 1A**). The WTI and Brent crude oil returns display a bi-directional lead-lag structure (see in **Figure 1C**) with 48.24% positive $\langle x(t) \rangle$ in the Sub 4 period, which is consistent with the fact that WTI and Brent have reconnected ever since the lift of the ban [16].

Basically, WTI leads Brent during the whole sample period except the last sub-period, when the two benchmark prices align closely and exhibit an alternate lead-lag structure. Percent of positive $\langle x(t) \rangle$ is highest during Sub 1 period, then decreases in the following periods, which might suggest that WTI and Brent are more and more connected.

5. CONCLUSION

This paper comprehensively investigates the lead-lag relationship between Brent and WTI crude oil spot prices from 20 May 1987 to 10 October 2017 based on the TOPS method. To examine the lead-lag structure at different periods, we divide the whole series into four sub-series by the outbreaks of the Gulf War, the Iraq War, and the lift of the U.S. export ban.

The results show that the WTI spot market is leading the Brent spot market slightly during the whole period, which is in line with the fact that WTI oil price plays the dominant role in the international oil market [3, 4, 7]. However, the lead-lag relationship is volatile and subject to economic events, geopolitical events and policy shifts. The percentage of positive $\langle x(t) \rangle$ during different sub-periods increases successively, which may indicate that WTI and Brent prices are reintegrating gradually.

Additionally, extreme events affected the path in varying degrees. Political events like the two Gulf wars triggered concerns about future oil supply. Both of the two benchmark prices experienced extreme uncertainty [13], which is reflected in the sharp but temporary increase of oil price. During wartime WTI and Brent co-move closely. A violent fluctuation of the $\langle x(t) \rangle$ is observed on the outbreak of the Iraq war, which might be related to the Iraq's oil output collapse at that time. The U.S. export ban also affected the WTI-Brent lead-lag relationship. The ban [7] together with transportation bottleneck [29] have long influenced the WTI-Brent spread. The turning point of the WTI-Brent spread structure [6–9] in 2011 accords with the upheaval of the $\langle x(t) \rangle$. The shift of energy policy, i.e., the lift of the export ban in 2015, has reconnected WTI to Brent [16]. The alternate lead-lag relationship lasts until the end of the sample.

These results are in line with other related studies, which reveal that The TOPS method has the potential to investigate the impact of extreme events on correlations between two time series. However, oil market is extremely fragile and affected by numerous factors. There still needs further exploration and stronger support to make the TOPS method a more active and convincing approach to event analysis. Given this, we will focus on these issues in future research.

DATA AVAILABILITY STATEMENT

The datasets for this study can be found in the https://www.eia.gov/ US Energy Information Administration (EIA) website.

AUTHOR CONTRIBUTIONS

All authors analyzed and discussed the results, contributed to the draft paper, and revisions. YS and XS conceived, designed the research, and wrote the manuscript. YY have implemented the empirical analysis. HS participated on the writing and editing of the paper.

FUNDING

This work was supported by the Peak Discipline Construction Project of Education at East China Normal University, and the National Natural Science Foundation of China (Grant No. 11805119).

ACKNOWLEDGMENTS

We would like to acknowledge the reviewers for their constructive comments.

REFERENCES

1. Reboredo JC. How do crude oil prices co-move? A copula approach. *Energy Econ*. (2011) **33**:948–55. doi: 10.1016/j.eneco.2011.04.006
2. Kao CW, Wan JY. Price discount, inventories and the distortion of WTI benchmark. *Energy Econ*. (2012) **34**:117–24. doi: 10.1016/j.eneco.2011.03.004
3. Liu WM, Schultz E, Swieringa J. Price dynamics in global crude oil markets. *J Futures Mark*. (2015) **35**:148–62. doi: 10.1002/fut.21658
4. Elder J, Miao H, Ramchander S. Price discovery in crude oil futures. *Energy Econ*. (2014) **46**:S18–27. doi: 10.1016/j.eneco.2014.09.012
5. Coronado S, Fullerton TM Jr, Rojas O. Causality patterns for brent, WTI, and argus oil prices. *Appl Econ Lett*. (2017) **24**:982–6. doi: 10.1080/13504851.2016.1245830
6. Chen W, Huang Z, Yi Y. Is there a structural change in the persistence of WTI-Brent oil price spreads in the post-2010 period? *Econ Model*. (2015) **50**:64–71. doi: 10.1016/j.econmod.2015.06.007

7. Scheitrum DP, Carter CA, Revoredo-Giha C. WTI and Brent futures pricing structure. *Energy Econ*. (2018) **72**:462–9. doi: 10.1016/j.eneco.2018.04.039
8. Caporin M, Fontini F, Talebbeydokhti E. Testing persistence of WTI and Brent long-run relationship after the shale oil supply shock. *Energy Econ*. (2019) **79**:21–31. doi: 10.1016/j.eneco.2018.08.022
9. Ye S, Karali B. *Estimating Relative Price Impact: The Case of Brent and WTI. No 235728, 2016 Annual Meeting, July 31–August 2*. Boston, MA: Agricultural and Applied Economics Association (2016).
10. Zhu H, Guo Y, You W. An empirical research of crude oil price changes and stock market in China: evidence from the structural breaks and quantile regression. *Appl Econ*. (2015) **47**:6055–74. doi: 10.1080/00036846.2015.1064076
11. Kang W, Ratti RA, Yoon KH. Time-varying effect of oil market shocks on the stock market. *J Bank Finance*. (2015) **61**:S150–63. doi: 10.1016/j.jbankfin.2015.08.027
12. Yao CZ, Kuang PC. A study of lead-lag structure between international crude oil price and several financial markets. *Phys A Stat Mech Appl*. (2019) **531**:121755. doi: 10.1016/j.physa.2019.121755
13. Zhang X, Yu L, Wang S, Lai KK. Estimating the impact of extreme events on crude oil price: an EMD-based event analysis method. *Energy Econ*. (2009) **31**:768–78. doi: 10.1016/j.eneco.2009.04.003
14. Charles A, Darné O. Volatility persistence in crude oil markets. *Energy Policy*. (2014) **65**:729–42. doi: 10.1016/j.enpol.2013.10.042
15. Basistha A, Kurov A. The impact of monetary policy surprises on energy prices. *J Futures Mark*. (2015) **35**:87–103. doi: 10.1002/fut.21639
16. Luong PV, Mizrach B, Otsubo Y. Location basis differentials in crude oil prices. *Energy J*. (2019) **40**:41–58. doi: 10.5547/01956574.40.SI2.pluo
17. Meng H, Xu HC, Zhou WX, Sornette D. Symmetric thermal optimal path and time-dependent lead-lag relationship: Novel statistical tests and application to UK and US real-estate and monetary policies. *Quant Financ*. (2017) **17**:959–77. doi: 10.1080/14697688.2016.1241424
18. Guo K, Zhou WX, Cheng SW, Sornette D. The US stock market leads the Federal funds rate and treasury bond yields. *PLoS ONE*. (2011) **6**:e22794. doi: 10.1371/journal.pone.0022794
19. Wang D, Tu J, Chang X, Li S. The lead-lag relationship between the spot and futures markets in China. *Quant Financ*. (2017) **17**:1447–56. doi: 10.1080/14697688.2016.1264616
20. Xu HC, Zhou WX, Sornette D. Time-dependent lead-lag relationship between the onshore and offshore Renminbi exchange rates. *J Int Financ Mark Inst Money*. (2017) **49**:173–83. doi: 10.1016/j.intfin.2017.05.001
21. Shao YH, Yang YH, Shao HL, Stanley HE. Time-varying lead-lag structure between the crude oil spot and futures markets. *Phys A*. (2019) **523**:723–33. doi: 10.1016/j.physa.2019.03.002
22. Jiang T, Bao S, Li L. The linear and nonlinear lead-lag relationship among three SSE 50 Index markets: the index futures, 50ETF spot and options markets. *Phys A*. (2019) **525**:878–93. doi: 10.1016/j.physa.2019.04.056
23. Yang YH, Shao YH. Time-dependent lead-lag relationships between the VIX and VIX futures markets. *The North American Journal of Economics and Finance*. (2019) 191013729. doi: 10.1016/j.najef.2020.101196
24. Sornette D, Zhou WX. Non-parametric determination of real-time lag structure between two time series: the "optimal thermal causal path" method. *Quant Financ*. (2005) **5**:577–91. doi: 10.1080/14697680500383763
25. Zhou WX, Sornette D. Non-parametric determination of real-time lag structure between two time series: the "optimal thermal causal path" method with application to economic data. *J Macroecon*. (2006) **28**:195–224. doi: 10.1016/j.jmacro.2005.10.015
26. Zhou WX, Sornette D. Lead-lag cross-sectional structure and detection of correlated-anticorrelated regime shifts: application to the volatilities of inflation and economic growth rates. *Phys A*. (2007) **380**:287–96. doi: 10.1016/j.physa.2007.02.114
27. Jiang ZQ, Xie WJ, Zhou WX. Testing the weak-form efficiency of the WTI crude oil futures market. *Phys A*. (2014) **405**:235–44. doi: 10.1016/j.physa.2014.02.042
28. Tian HZ, Lai WD. The causes of stage expansion of WTI/Brent spread. *Petrol Sci*. (2019) **16**:1493–505. doi: 10.1007/s12182-019-00379-z
29. Buyuksahin B, Lee TK, Moser JT, Robe MA. Physical markets, paper markets and the WTI-Brent spread. *Energy J*. (2013) **34**:129–51. doi: 10.5547/01956574.34.3.7
30. Colgan JD, de Graaf T. A crude reversal: The political economy of the United States crude oil export policy. *Energy Res Soc Sci*. (2017) **24**:30–5. doi: 10.1016/j.erss.2016.12.012
31. Melek NÇ, Ojeda E. Lifting the US crude oil export ban: prospects for increasing oil market efficiency. *Econ Rev*. (2017) **102**:51–74. doi: 10.18651/ER/2q17CakirMelekOjeda

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past supervisory role with the author YY.

Copyright © 2020 Yang, Shao, Shao and Song. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 08 May 2020
doi: 10.3389/fphy.2020.00135

Tick Size and Market Quality Using an Agent-Based Multiple-Order-Book Model

Ruwei Zhao[1], *Yian Cui*[2]* *and Xuenu Liu*[3]

[1] School of Business, Jiangnan University, Wuxi, China, [2] Research Institute, Shenzhen Stock Exchange, Shenzhen, China, [3] College of Intelligence and Computing, Tianjin University, Tianjin, China

This study investigated the dynamics between tick size and market quality using an agent-based multiple-order-book stock-market model. Given the multiple-order-book setting, we integrated the model with small-, medium-, and large-cap stocks and conducted the analysis from both a tick-size-series and cross-sectional perspective. The simulation results showed that small-cap stocks were of the lowest quality. Furthermore, quality was generally weakened as tick-size value increased, with expanded bid-ask spreads, elevated market volatility, and reduced market efficiency.

Keywords: tick size, market quality, agent-based modeling, multiple order books, market capitalization

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Yajun Xiao,
University College Dublin, Ireland
Wen Long,
University of Chinese Academy of Sciences, China

***Correspondence:**
Yian Cui
yacui@szse.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *15 January 2020*
Accepted: *06 April 2020*
Published: *08 May 2020*

Citation:
Zhao R, Cui Y and Liu X (2020) Tick Size and Market Quality Using an Agent-Based Multiple-Order-Book Model. Front. Phys. 8:135. doi: 10.3389/fphy.2020.00135

1. INTRODUCTION

As one of the fundamental transaction requirements, tick size (i.e., the minimum movement of stock price) has received considerable research attention with regard to its powerful effects on market quality [1–19]. In light of the newly introduced tick-size policy change by the Chicago Mercantile Exchange, Martinez and Tse [9] investigated the relationship between tick-size reduction and market quality in the foreign currency futures market. They found that reduced tick size powerfully promoted the market quality of the Mexican peso, dramatically dropping pricing errors, greatly improving pricing discovery capacity, and considerably reducing informed trading. Similarly, in consideration of the 2016 US SEC tick-size pilot program, Griffith and Roseman [15] studied the market-liquidity reaction to the rising movement of tick size. They found that market liquidity would be severely weakened by an increase in tick size. Lepone and Wong [18] examined market-quality evolution with tick-size changes in the Singapore stock market. They found significant drops in the bid-ask spread and order depth following the reduction of tick size, and lower-priced stocks were favored through disguised market markers with higher returns. However, most of these prior studies were based on policy changes made by regulating authorities, which are extremely rare.

To avoid the quantity deficiencies of tick-size changes, this study used an agent-based multiple-order-book model, adopting six gradually increasing tick sizes to investigate the dynamics between tick size and market quality. Many studies have investigated the feasibility of agent-based modeling as an effective vehicle for stock-market simulation [20–34]. However, most of those studies were oriented using one stock, which limited the credibility of the models. Thus, to be more in line with market reality, we followed Wei et al. [35] and constructed an agent-based stock-market model with three stocks as representatives of large, medium, and small stocks. We found that as tick size rises, the bid-ask spread generally increases, which is consistent with findings for liquidity indexes, showing a decreasing trend in market liquidity. We also found that small-cap stocks have the worst liquidity with the highest bid-ask spreads and lowest liquidity index values at cross-sectional levels.

We also took the return standard deviation as the measurement of market volatility. We found that market volatility increased dramatically with the rise in tick size, and the small-cap stocks were of the highest volatility with the largest values. Furthermore, we investigated market efficiency with three types of stocks. We found that the market efficiency demonstrates monotone increasing character with the expansion of tick size interval, and mid- and small-cap stocks had the highest and lowest market efficiency, respectively. Finally, we checked investor wealth evolution using tick-size changes and found that the wealth changes demonstrated a concave shape.

This study contributes to the literature in three ways. First, our agent-based stock-market model originates from a multiple-order-book perspective. Using the model, our study investigated the dynamics between tick size and market quality with small-, medium-, and large-cap stocks, which improves upon the limitations of prior one-stock models [27, 28]. Second, compared to previous tick-size studies [9, 18], our study covers six tick sizes, providing more detailed insights into the dynamics of tick size. Third, this study was based on the largest emerging Chinese stocks; thus, it can provide valuable risk-management suggestions for both international investors and regulatory authorities.

The rest of this paper is organized as follows. Section 2 illustrates the method. Section 3 describes the simulation results, and section 4 concludes the paper.

2. METHODS

2.1. Agent-Based Multiple-Order-Book Model

We introduced three types of stocks with variations in capitalization to construct an agent-based multiple-order-book model. In the model, investors adopt mixed heterogeneous beliefs and adaptive asset allocations in stock-price prediction and cash distribution, respectively.

2.1.1. Market Design

(1) Asset Design

In the Chinese stock market, there are three types of boards: the main board, the SME board (small and medium enterprise board), and the ChiNext board (growth enterprise market board). These three boards represent completely different types of stocks. The main board is constituted by large-cap stocks, while the SME and ChiNext boards consisted of medium- and small-sized stocks, respectively. Also, Xiong et al. [36] found that the average shares of the main board stocks, the Shenzhen 100 Index, SME Index, and ChiNext Index, are 1,975, 912, and 637 million shares, respectively. However, the average prices of the three boards are 20, 19, and 19 CNYs. In consideration of the actual market structure with three board indices, we set three stocks—one large-cap stock, one mid-cap stock, and one small-cap stock as well as a risk-free asset—in the agent-based multiple-order-book model. The shares of large-, mid-, and small-cap stocks are set as 2,000,000, 1,000,000, and 500,000, respectively, which are close to the one thousandths of real numbers, and the initial prices of the three stocks were 20 CNYs, as shown in **Table 1**. Also, we assumed there was no fundamental value correlation among the three stocks and that the interest rate of cash was zero.

TABLE 1 | Asset setting in the agent-based multiple-order-book model.

Asset	Shares	Initial price (CNY)	Average price fluctuation per minute (basis point)
Large-cap stock	2,000,000	20	3
Mid-cap stock	1,000,000	20	3
Small-cap stock	500,000	20	3

(2) Investor design

Considering the calculation capacity limits of the simulation platform, we set the number of total investors as $N = 5{,}000$. When investors engage in the market the first time, they are allocated a certain amount of initial stocks and cash. The initial position of investor i of stock j, $s_0^{i,j}$, is

$$s_0^{i,j} = s_{Mean}^{j}\varphi \tag{1}$$

where s_{Mean}^{j} is the average position of stock j allocated to each investor as the total shares of stock j over total investors. φ refers to the uniform distribution with the boundary limit, $[0.1, 1.9]$, complying with the heterogeneity of investors. Also, we assumed the initial cash of the investors was equal to the capitalization of stocks at the beginning of the simulation. In this view, the initial cash of investor i, α_0^i, is

$$\alpha_0^i = \sum_{j=L,M,S} p_0^j s_0^{i,j} \tag{2}$$

where p_0^j is the initial price of stock j. In addition, when the investor's wealth is too low, we recognize the investor as bankrupt. Meanwhile, a new investor will be enrolled in the simulation market to ensure the normal operation of the agent-based model.

2.1.2. Investor Price Prediction

Following Chiarella et al. [37], we adopted three types of investorsfundamental investor, technical investor, and a noise investor—which is in accordance with the market-participator reality in the Chinese stock market. In this view, in our model, the investor makes a price prediction based on a mixed-heterogeneity belief. The return prediction of investor i of stock j is

$$r_{t+\tau}^{i,j} = \left(x^i r_c^{i,j} + y^i r_f^{i,j} + z^i \varepsilon\right) / \left(x^i + y^i + z^i\right) \tag{3}$$

where ε is responsible for the investor's noise belief, and $r_f^{i,j}$ denotes the predicted return by investor i of stock j based on fundamental belief. Incorporated with investor trading horizon τ_i, $r_f^{i,j}$ is calculated as the natural log value of the fundamental value of stock j, denoted as f_t^j, over price p_t^j, as shown in Equation (4). Also, $r_c^{i,j}$ refers to the predicted return by investor i of stock

j based on technical belief, calculated as the natural log value of short-term average price $\bar{p}^j_{\tau_i/4}$ over long-term average price $\bar{p}^j_{\tau_i}$, as shown in Equation (5).

$$r_f^{i,j} = (\ln \frac{f_t^j}{p_t^j})/\tau_i \tag{4}$$

$$r_c^{i,j} = \ln \frac{\bar{p}^j_{\tau_i/4}}{\bar{p}^j_{\tau_i}} \tag{5}$$

In addition, x^i, y^i, and z^i represent the weight of technical, fundamental, and noise belief, following the constraints as $x^i = x^*\theta$, $y^i = (1 - x^* - z^*)\theta$, and $z^i = z^*\theta$, where θ is of uniform distribution with $[0, 1]$ as the boundary limit. x^*, y^*, and z^* are the given exogenous technical, fundamental, and noise parameters, respectively, in the agent-based model, and the sum of the three parameters is equal to one, as shown in Equation (6):

$$x^* + y^* + z^* = 1 \tag{6}$$

The trading horizon of investor i, τ^i, is determined by the relation between technical belief weight x^i and fundamental belief weight y^i. If the technical belief weight x^i grows, the trading horizon begins to be narrowed. Meanwhile, the larger the fundamental belief weight y^i, the longer the trading horizon, as shown in Equation (7). We set the initial trading horizon $\tau^* = 5days = 1200t$. The possibility of investor arrival at the market is inverse to trading horizon τ^i and follows Poisson distribution with the parameter of $\lambda^i = \omega/\tau^i$.

$$\tau^i = \tau^* \frac{1 + y^i}{1 + x^i} \tag{7}$$

In the end, the investor makes a stock-return prediction with regard to mixed fundamental, technical, and noise beliefs. The predicted price of stock j at time t, p_t^j, is derived from the price of stock j at time t, $p_{f,t}^{i,j}$, as shown in Equation (8):

$$p_{f,t}^{i,j} = p_t^j e^{r_{t+\tau}^{i,j}\tau^i} \tag{8}$$

2.1.3. Investor-Adaptive Asset-Allocation Model

Asset allocation is the most important procedure in the investor-trading decision-making process. In the 1950s, Markowitz [38] introduced the well-known "mean-variance" model as the first scientific calculation of optimal portfolio weights upon equities risks. However, based on the survey results of the *Shanghai Stock Exchange Market Quality Report 2016*, we found that, in the Chinese stock market, most individual investors show intensely irrational compulsions when making asset-allocation decisions. In fact, most Chinese investors favor stocks with better performances or that have been extensively covered by the media in recent weeks, indicating a hot-topic-chasing preference in investors' adaptive asset-allocation processes. Follow Brock and Hommes [39], we integrated the adaptive belief-transfer mechanism to investors' asset-allocation decision-making processes and constructed an investor-adaptive asset-allocation model.

In our model, when investors come to the stock market, they spread their wealth into the positions of four assets based on the predicted returns[1]. The distributed wealth ratio of asset j by investor i, $h_t^{i,j}$, is

$$h_t^{i,j} = \frac{e^{\beta\pi_{t+\tau}^{i,j}}}{\sum_{j=1}^{4} e^{\beta\pi_{t+\tau}^{i,j}}} \tag{9}$$

where $\pi_{t+\tau}^{i,j} = r_{t+\tau}^{i,j}$ is the predicted return of asset j by investor i, and β is the adaptive transfer degree.

With the wealth ratio allocated to each asset, we can get the cash balance of investor i in asset j at time t, $\alpha_t^{i,j}$, as

$$\alpha_t^{i,j} = h_i^j \alpha_t^i \tag{10}$$

where α_t^i is the total balance of investor i at time t.

2.1.4. Investor Order-Placing Rule

In the model, the investors decide the order direction, order type, and order size based on the analysis of the predicted return and temporal status of the order book. Following Gil-Bazo et al. [40], we assume that investors reserve part of their return, μ, as compensation for the transaction cost[2]. The reserved return is

$$\mu = 1\% p_t^{m,j} \tag{11}$$

where $p_t^{m,j}$ is the midpoint of the optimal bid and ask spread.

Table 2 shows the detailed order-placing rules of the investors. We follow this mechanism design proposed by Gil-Bazo et al. [40]. As shown in the table, investor i decides the order type by comparing the distance between asset predicted price, $p_{f,t}^{i,j}$, and reserved return, μ. For example, when both the bid and ask orders exist in the order book (scenario 1), if the difference between the predicted price, $p_{f,t}^{i,j}$, and reserved return, μ, is higher than the optimal ask price, the investor places a market bid order with a price of $p_{mb}^{i,j} < p_{f,t}^{i,j} - \mu$. If the predicted price, $p_{f,t}^{i,j}$, is not larger than the optimal ask price, a_t^j, plus the reserved return, μ; not smaller than the optimal bid price, b_t^j, minus the reserved return, μ; and larger than the midpoint of the optimal bid-ask spread, $p_t^{m,j}$, the investor would place a limit bid order with a price of $p_{lb}^{i,j} = p_{f,t}^{i,j} - \mu$.

In the end, the model determines the order size of the investor. When the investor decides to buy a stock, he or she will place all

[1] The predicted return of cash is zero.

[2] We do not implement a specific check of the size of reserved return. However, with the simulation tests, we find that a limited variation of reserved return presents no impact to the final results.

TABLE 2 | Order-placing rules of investors.

Scenarios	Order direction and type	Order price
Scenario 1: Both bid and ask orders exist in the order book		
$p_{f,t}^{i,j} - u > a_t^j$	Market buy	$p_{mb}^{i,j} < p_{f,t}^{i,j} - u$
$[p_{f,t}^{i,j} \geq p_t^{mj}]\&[a_t^j + u \geq p_{f,t}^{i,j} \geq b_t^j - u]$	Limit buy	$p_{lb}^{i,j} = p_{f,t}^{i,j} - u$
$[p_{f,t}^{i,j} < p_t^{mj}]\&[a_t^j + u \geq p_{f,t}^{i,j} \geq b_t^j - u]$	Limit sell	$p_{ls}^{i,j} = p_{f,t}^{i,j} + u$
$p_{f,t}^{i,j} + u < b_t^j$	Market sell	$p_{ms}^{i,j} > p_{f,t}^{i,j} + u$
Scenario 2: No bid orders in the order book		
$p_{f,t}^{i,j} - u > a_t^j$	Market buy	$p_{mb}^{i,j} < p_{f,t}^{i,j} - u$
$p_{f,t}^{i,j} - u \leq a_t^j$	Limit buy	$p_{lb}^{i,j} = p_{f,t}^{i,j} - u$
Scenario 3: No ask orders in the order book		
$p_{f,t}^{i,j} + u < b_t^j$	Market sell	$p_{ms}^{i,j} > p_{f,t}^{i,j} + u$
$p_{f,t}^{i,j} + u \geq b_t^j$	Limit sell	$p_{ls}^{i,j} = p_{f,t}^{i,j} + u$
Scenario 4: No bid and ask orders in the order book		
50% probability	Limit buy	$p_{lb}^{i,j} = p_{f,t}^{i,j} - u$
50% probability	Limit sell	$p_{ls}^{i,j} = p_{f,t}^{i,j} + u$

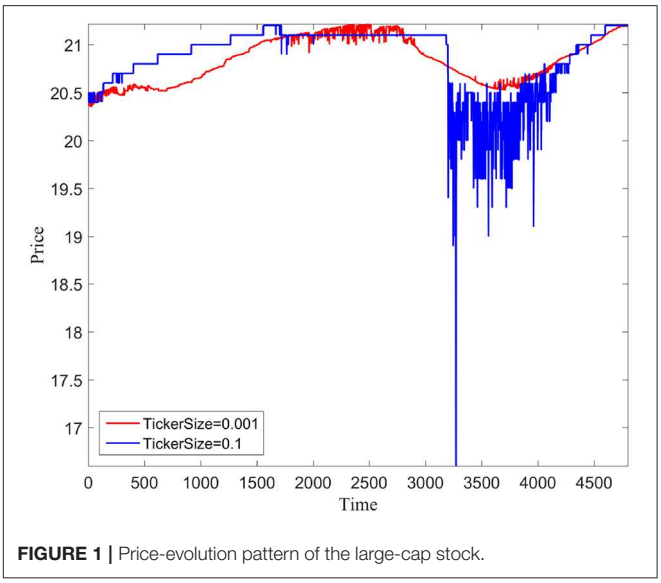


FIGURE 1 | Price-evolution pattern of the large-cap stock.

cash allocated to the stock into the bid order, and the size of the bid order is $b_t^{i,j} = \alpha_t^{i,j}/(p_{f,t}^{i,j} - \mu)$. Meanwhile, when the investor decides to sell a stock, he or she will leave no positions left.

3. SIMULATION RESULTS

3.1. Price Pattern

Figure 1 shows the dynamics of the price series of the large-cap stock with small and large ticker sizes (0.001 and 0.1, respectively). We can found that under a ticker size of 0.001, the price pattern shows powerful evidence for stock-liquidity sufficiency with a smooth and continuous line. Also, the price pattern of the small ticker size, 0.001, presents mild fluctuation with a narrow range, gradually increasing and descending from 20.5 to 21. Meanwhile, the price line with regard to the large ticker size, 0.1, is of adequate liquidity with a jagged shape, indicating a low degree of participation in the placement of bid-ask orders. In addition, we found that the price range of the large ticker size varies from 0 to 21, which is much larger than that of the small ticker size, indicating a higher degree of volatility.

Figure 2 shows the price-evolution series of the mid-cap stock with small and large ticker sizes. We can see that, similarly to **Figure 1**, the price pattern of the small ticker size shows more liquidity than the large ticker size with a flatter line. Also consistent with **Figure 1**, the price pattern of the small ticker size shows higher stability with a limited price range of 20.5–19.5 compared to the large ticker's wild range of 18–23. However, the price trends of the mid-cap stock show few differences from those of the large-cap stock. We can see in **Figure 2** that, unlike the rising trends at the beginning of the large-cap stock simulation, the price evolutions of the small and large ticker sizes of the mid-cap stock are of a declining character at the beginning of the simulation and gradually bounce back later in the simulation.

Figure 3 shows the price pattern of the small-cap stock under the circumstances of small and large ticker sizes. We can found that, different from **Figures 1**, **2**, the price evolution of the

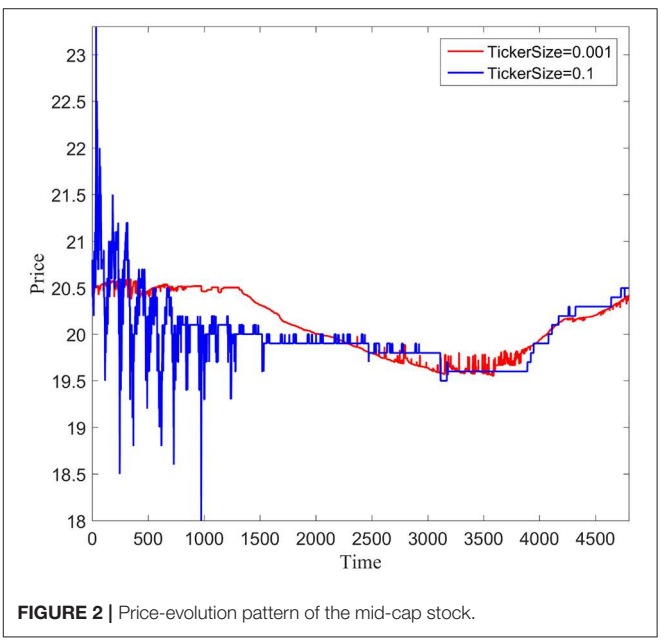


FIGURE 2 | Price-evolution pattern of the mid-cap stock.

small ticker size with the small-cap stock shows no noticeable curves, indicating a higher degree of stability. Consistent with **Figures 1**, **2**, the price pattern of the large ticker size shows more volatility with vast price changes from 0 to 80 in contrast to the near-zero price change of the small ticker. Also, we can see that the maximum price of the small-cap stock with a large ticker size is roughly four times that of the large- and mid-cap stocks, and the price pattern exhibits more peaks with those in **Figures 1**, **2**, indicating a further lower level of stability.

3.2. Market Liquidity

For market liquidity, we adopted three primary measurements. The first one lies in the bid-ask spread, which is widely used

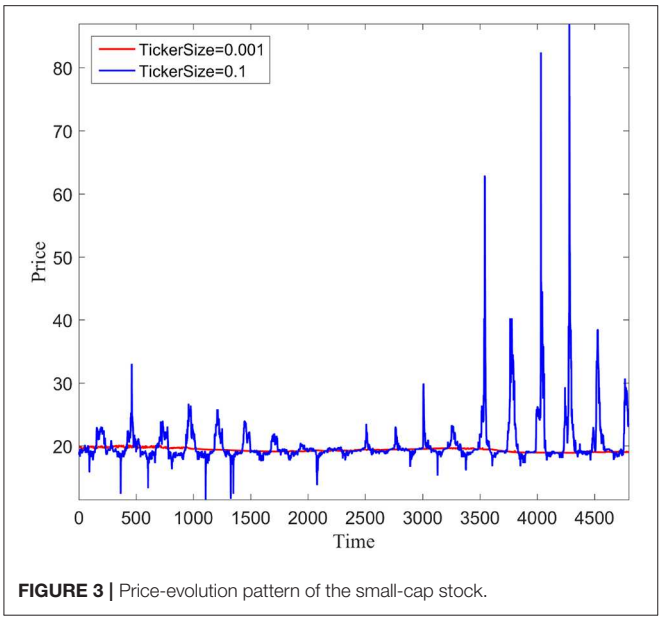


FIGURE 3 | Price-evolution pattern of the small-cap stock.

in prior research. Also, to obtain more comprehensive insight, we introduced a second measurement, order depth, which is calculated as the average of the sum of the bid-ask order volumes. The construction process is as follows

$$\begin{aligned} Depth_{Ask} &= \sum_{j=1}^{k} Volume_ask_{j,t} \\ Depth_{Bid} &= \sum_{j=1}^{k} Volume_bid_{j,t} \\ Depth_t &= \tfrac{1}{2}(Depth_{Ask} + Depth_{Bid}) \end{aligned} \tag{12}$$

where $Volume_ask_{j,t}$ is the j-th ask-order volume at time t, $Volume_bid_{j,t}$ is the j-th bid-order volume at time t, and $Depth_t$ is the order depth at time t, k=5.

The third is the liquidity index, L_{SSE}, proposed by the Shanghai Stock Exchange, one of the two leading exchanges in China. The index measures the elasticity of the trading CNY change upon the variation in stock price, Δ[3]. The more CNY initiated in the stock-price change, the greater the index, indicating a higher degree of liquidity. The specific calculation process is as follows

$$\begin{aligned} L_{Ask} &= \sum_{j=1}^{k-1} D_j A_j + A_k, \text{where } k = \{\min(m) | \tfrac{|A_m - A_1|}{A_1} \geq \Delta\} \\ L_{Bid} &= \sum_{j=1}^{k-1} D_j B_j + B_k, \text{where } k = \{\min(m) | \tfrac{|B_m - B_1|}{B_1} \geq \Delta\} \\ L_{SSE} &= \tfrac{1}{2}(L_{Ask} + L_{Bid}) \end{aligned} \tag{13}$$

where B_j and A_j are the j-th bid and ask quotations. D_j is the order depth with the j-th quotation.

[3]We set the price change magnitude, Δ, as 1%.

TABLE 3 | Bid-ask spreads of large-, mid-, and small-cap stocks.

Tick size	Large-cap stock	Mid-cap stock	Small-cap stock	Mean
0.001	0.0270	0.0261	0.0401	0.0311
0.005	0.0300	0.0264	0.0598	0.0387
0.01	0.0292	0.0267	0.0637	0.0399
0.02	0.0332	0.0352	0.0535	0.0406
0.05	0.0542	0.0641	0.0987	0.0723
0.1	0.1095	0.1102	0.2529	0.1575

TABLE 4 | Order depths of large-, mid-, and small-cap stocks.

Tick size	Large-cap stock	Mid-cap stock	Small-cap stock	Mean
0.001	577	443	360	460
0.005	1,899	1,485	970	1,451
0.01	4,098	3,034	2,113	3,082
0.02	10,077	6,791	4,334	7,067
0.05	22,155	1,1397	6,551	13,368
0.1	14,641	8,814	1,921	8,459

Table 3 reports the bid-ask spreads of the large-, mid-, and small-cap stocks with escalating levels of tick sizes from 0.001 to 0.1. We found clear variations in the spread traces at the cross-sectional and tick-size-series levels. From the cross-sectional perspective, we find that within each tick size, the small-cap stock has the lowest liquidity and the largest bid-ask spread value. Meanwhile, in view of the tick-size series, we found that market liquidity shows the greatest favor with a tick size of 0.001 by the minimum average bid-ask spread value. We could also see that the average bid-ask spread value shows a noticeable growing trend with increasing bid-ask spread values, indicating a lower level of market liquidity.

Table 4 shows the results of the second market-liquidity measuring instrument, the order depth, with the integration of trading volume, across the three types of stocks. Consistent with **Table 3**, we conducted the analysis from both tick-size-series and cross-sectional perspectives. We found that, within the tick-size series, the order depths generally show continuous growth, with the exception of tick size 0.05, which is in line with findings of Lepone and Wong [18] by the large. From the cross-sectional perspective, we found that at each tick-size level, the large-cap stock has the largest order depth, which is roughly 1.67 and 3.29 times those of the mid- and small-cap stocks, respectively.

Table 5 shows the results of the third market liquidity indicator, the liquidity indices, through large-, mid-, and small-cap stocks. Consistent with **Tables 3**, **4**, we initiated tick-size-series and cross-sectional checks of the dynamics of the liquidity indices. The average and large cap market-liquidity indices show apparent falls with increasing tick-size values from 0.005 to 0.1, indicating an impairing trend in market liquidity. We can also found that the large-cap stocks have the largest liquidity index values, showing a considerable edge in market liquidity at each tick-size level.

TABLE 5 | Liquidity indices of large-, mid-, and small-cap stocks.

Tick size	Large-cap stock	Mid-cap stock	Small-cap stock	Mean
0.001	571,711	416,259	318,596	435,522
0.005	3,102,311	1,820,927	770,116	1,897,785
0.01	2,622,049	1,945,937	933,257	1,833,748
0.02	2,832,469	1,736,470	698,103	1,755,681
0.05	2,252,589	1,123,120	886,983	1,420,897
0.1	752,019	407,532	52,052	403,868

TABLE 6 | Volatilities of large-, mid-, and small-cap stocks.

Tick size	Large-cap stock	Mid-cap stock	Small-cap stock	Mean
0.001	23	29	29	27
0.005	23	25	61	36
0.01	22	26	74	41
0.02	22	26	39	29
0.05	26	31	76	44
0.1	37	36	376	150

3.3. Volatility

For market volatility, we took the traditional standard deviation in stock price as the indicator. **Table 6** shows that the average market volatilities grew following the expansion of tick sizes, with the exception of tick size 0.02, which is generally consistent with findings of Griffith and Roseman [15]. Also, within each tick size, we can find that large-cap stocks usually have the least volatility and the smallest numbers. Meanwhile, mid- and small-cap stocks had the second- and first-highest volatilities with the second-largest and largest deviations, respectively.

3.4. Market-Pricing Efficiency

For market-pricing efficiency, we took two indicators, MAE and MRE, as the measuring instruments. MAE refers to the absolute distance between stock price, p_t, and stock fundamental value, f_t. Similarly, MRE is the relative distance between stock price, p_t, and stock fundamental value, f_t. The smaller the indicator number, the higher the pricing efficiency of the market. The detailed constructions of both indicators are as follows

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|p_t - f_t| \tag{14}$$

$$MRE = \frac{1}{T}\sum_{t=1}^{T}\frac{|p_t - f_t|}{f_t} \tag{15}$$

Tables 7, **8** show the results for the market-pricing-efficiency indicators, MAE and MRE, with large-, mid-, and small-cap stocks. We found that, if expanded intervals are taken, such as 10 times, the average MAEs and MREs of 0.1 (0.01) are much larger than those of 0.01 (0.001), respectively. In this view, we believe that the market efficiency generally falls with the increase of tick sizes, especially when the size intervals are of huge differences[4]. We can also find that in most cases, mid- and small-cap stocks have the best and worst market-pricing efficiencies, respectively, with smallest and largest MAE and MRE values for each tick size.

TABLE 7 | MAEs of large-, mid-, and small-cap stocks.

Tick size	Large-cap stock	Mid-cap stock	Small-cap stock	Mean
0.001	0.1981	0.1937	0.2204	0.2041
0.005	0.1766	0.1597	0.4808	0.2724
0.01	0.2035	0.1850	0.5038	0.2974
0.02	0.1821	0.1711	0.3382	0.2305
0.05	0.1880	0.2058	0.2973	0.2304
0.1	0.2890	0.2741	1.52014	0.6944

TABLE 8 | MREs of large-, mid-, and small-cap stocks.

Tick size	Large-cap stock (%)	Mid-cap stock (%)	Small-cap stock (%)	Mean (%)
0.001	0.95	0.97	1.14	1.02
0.005	0.85	0.80	2.49	1.38
0.01	0.98	0.92	2.60	1.50
0.02	0.87	0.85	1.75	1.16
0.05	0.90	1.03	1.54	1.16
0.1	1.39	1.36	7.89	3.55

TABLE 9 | Investor wealth and ticker sizes.

Ticker size	Investor wealth
0.001	28,519
0.005	28,323
0.01	28,267
0.02	28,288
0.05	28,378
0.1	28,439

3.5. Investor Wealth

Table 9 shows the results for the investor wealth dynamics with each distinct tick size. We found that the dynamic pattern of investor wealth has a concave character, and the maximum value of investor wealth corresponds to a tick size of 0.001. Also, with increasing tick-size value, investor wealth shows a falling trend with reduced values, reaching the lowest point at tick size 0.01. After that, investor wealth starts to rebound with growing values from tick sizes 0.02 to 0.1.

[4]The original evolutions of average MAEs and MREs in **Tables 7**, **8** demonstrate faulty monotonicity, with certain breaks at 0.02 and 0.05. We attribute this phenomenon as the result of narrowed interval setting from 0.005 to 0.05. Each interval from 0.005 to 0.05 is 2 or 2.5 times, much smaller than that between 0.1 (0.01) and 0.01 (0.001). Also, if we designate the 0.001 and 0.005, 0.01 and 0.02, and 0.05 and 0.1, as hypothetical low, medium and high tick size subgroups, it is evident to see that the averages of MAE and MRE grow with the escalation of subgroup sizes. In this view, the monotone increasing characters of MAEs and MREs are perfectly presented with the expansion of size intervals.

4. CONCLUSION

This study used an agent-based multiple-order-book stock-market model to investigate the relationship between tick size and market quality. We set six tick sizes from 0.001 to 0.1. With a gradual increase in tick size, market liquidity showed a significant drop with amplified bid-ask spreads and shrunken liquidity index values. Also, for market volatility, we found noticeable volatility decreases with declining tick sizes. In addition, we investigated market efficiency, which is a major element of market quality. We found that market efficiency would be weakened upon the increase of tick size, especially when the interval is of large distance. Furthermore, from a cross-sectional perspective, we found that small-cap stocks had the worst market quality with the least satisfactory indicator values.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

RZ and XL conceived the method and conducted the model. YC analyzed the results and detailed the method. All authors reviewed the manuscript.

FUNDING

We affirm that this work was supported by the National Natural Science Foundation of China (No. 71901107) and the Fundamental Research Funds for the Central Universities (Nos. JUSRP11996 and 2019JDZD16).

REFERENCES

1. Zhao X, Chung KH. Decimal pricing and information-based trading: tick size and informational efficiency of asset price. *J Bus Finance Account*. (2006) **33**:753–66. doi: 10.1111/j.1468-5957.2006.00622.x
2. Chakravarty S, Wood RA, Van Ness RA. Decimals and liquidity: a study of the NYSE. *J Financ Res*. (2004) **27**:75–94. doi: 10.1111/j.1475-6803.2004.00078.x
3. Aitken M, Comerton-Forde C. Do reductions in tick sizes influence liquidity? *Account Finance*. (2005) **45**:171–84. doi: 10.1111/j.1467-629x.2004.00128.x
4. Pan W, Song FM, Tao L. The effects of a tick-size reduction on the liquidity in a pure limit order market: evidence from Hong Kong. *Appl Econ Lett*. (2012) **19**:1639–42. doi: 10.1080/13504851.2011.650327
5. Goldstein MA, Kavajecz KA. Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE. *J Financ Econ*. (2000) **56**:125–49. doi: 10.1016/S0304-405X(99)00061-6
6. Alampieski K, Lepone A. Impact of a tick size reduction on liquidity: evidence from the Sydney Futures Exchange. *Account Finance*. (2009) **49**:1–20. doi: 10.1111/j.1467-629X.2008.00279.x
7. Kuo SW, Huang CS, Chen CC. Impact of the change in tick size on transaction costs and liquidity: an empirical investigation of the Taiwan Stock Exchange. *Asia Pac J Financ Stud*. (2010) **39**:524–51. doi: 10.1111/j.2041-6156.2010.01020.x
8. Hsieh TY, Chuang SS, Lin CC. Impact of tick-size reduction on the market liquidity—evidence from the emerging order-driven market. *Rev Pac Basin Financ Markets Policies*. (2008) **11**:591–616. doi: 10.1142/S0219091508001490
9. Martinez V, Tse Y. The impact of tick-size reductions in foreign currency futures markets. *Finance Res Lett*. (2019) **28**:32–8. doi: 10.1016/j.frl.2018.03.023
10. Comerton-Forde C, Grégoire V, Zhong Z. Inverted fee structures, tick size, and market quality. *J Financ Econ*. (2019) **134**:141–64. doi: 10.1016/j.jfineco.2019.03.005
11. Yamamoto R. Limit order submission risks, order choice, and tick size. *Pac Basin Finance J*. (2020) **59**:101261. doi: 10.1016/j.pacfin.2019.101261
12. Chung D, Hrazdil K. Liquidity and market efficiency: a large sample study. *J Bank Finance*. (2010) **34**:2346–57. doi: 10.1016/j.jbankfin.2010.02.021
13. Chung DY, Hrazdil K. Liquidity and market efficiency: analysis of NASDAQ firms. *Glob Finance J*. (2010) **21**:262–74. doi: 10.1016/j.gfj.2010.09.004
14. Chung KH, Kim KA, Kitsabunnarat P. Liquidity and quote clustering in a market with multiple tick sizes. *J Financ Res*. (2005) **28**:177–95. doi: 10.1111/j.1475-6803.2005.00120.x
15. Griffith TG, Roseman BS. Making cents of tick sizes: the effect of the 2016 US SEC tick size pilot on limit order book liquidity. *J Bank Finance*. (2019) **101**:104–21. doi: 10.1016/j.jbankfin.2019.01.017
16. Chung KH, Van Ness RA. Order handling rules, tick size, and the intraday pattern of bid-ask spreads for Nasdaq stocks. *J Financ Markets*. (2001) **4**:143–61. doi: 10.1016/S1386-4181(00)00021-5
17. Bacidore J, Battalio RH, Jennings RH. Order submission strategies, liquidity supply, and trading in pennies on the New York Stock Exchange. *J Financ Markets*. (2003) **6**:337–62. doi: 10.1016/S1386-4181(03)00003-X
18. Lepone A, Wong JB. Pseudo market-makers, market quality and the minimum tick size. *Int Rev Econ Finance*. (2017) **47**:88–100. doi: 10.1016/j.iref.2016.10.002
19. Chung KH, Chuwonganant C. Tick size and quote revisions on the NYSE. *J Financ Markets*. (2002) **5**:391–410. doi: 10.1016/S1386-4181(01)00029-5
20. Xiong X, Liang J, Cui Y, Zhang W, Zhang Y. Analysis of the spot market's T+1 trading system effects on the stock index futures market. *Eurasia J Math Sci Technol Educ*. (2017) **13**:7679–93. doi: 10.12973/ejmste/77941
21. Zhang Y, Zhang W. Can irrational investors survive? A social-computing perspective. *IEEE Intell Syst*. (2007) **22**:58–64. doi: 10.1109/MIS.2007.4338495
22. Wei L, Xiong X, Zhang W, He XZ, Zhang Y. The effect of genetic algorithm learning with a classifier system in limit order markets. *Eng Appl Artif Intell*. (2017) **65**:436–48. doi: 10.1016/j.engappai.2017.07.006
23. Wei L, Zhang W, Xiong X, Zhao Y. A multi-agent system for policy design of tick size in stock index futures markets. *Syst Res Behav Sci*. (2014) **31**:512–26. doi: 10.1002/sres.2292
24. Yeh CH, Yang CY. Examining the effectiveness of price limits in an artificial stock market. *J Econ Dyn Control*. (2010) **34**:2089–108. doi: 10.1016/j.jedc.2010.05.015
25. Dieci R, He XZ. Heterogeneous agent models in finance. In: Hommes C, LeBaron B, editors. *Handbook of Computational Economics*. Vol. 4. Amsterdam: Elsevier (2018). p. 257–328.
26. Hoffmann AOI, Jager W, Von Eije JH. Social simulation of stock markets: taking it to the next level. *J Artif Soc Soc Simul*. (2007) **10**:7. Available online at: http://jasss.soc.surrey.ac.uk/10/2/7.html
27. Zhang W, Bi Z, Shen D. Investor structure and the price-volume relationship in a continuous double auction market: an agent-based modeling perspective. *Phys A Stat Mech Appl*. (2017) **467**:345–55. doi: 10.1016/j.physa.2016.10.044
28. Zhao R, Xiong X, Shen D, Zhang W. Investor structure and stock price crash risk in a continuous double auction market: an agent-based perspective. *Int J Inform Technol Decis Making*. (2019) **18**:695–715. doi: 10.1142/S0219622019500081
29. Sornette D. Physics and financial economics (1776–2014): puzzles, Ising and agent-based models. *Rep Prog Phys*. (2014) **77**:062001. doi: 10.1088/0034-4885/77/6/062001
30. Zhou WX, Sornette D. Self-organizing Ising model of financial markets. *Eur Phys J B*. (2007) **55**:175–81. doi: 10.1140/epjb/e2006-00391-6

31. Sornette D, Zhou WX. Importance of positive feedbacks and overconfidence in a self-fulfilling Ising model of financial markets. *Phys A Stat Mech Appl*. (2006) **370**:704–26. doi: 10.1016/j.physa.2006.02.022
32. Jiang ZQ, Xie WJ, Zhou WX, Sornette D. Multifractal analysis of financial markets: a review. *Rep Prog Phys*. (2019) **82**:125901. doi: 10.1088/1361-6633/ab42fb
33. Meng H, Ren F, Gu GF, Xiong X, Zhang YJ, Zhou WX, et al. Effects of long memory in the order submission process on the properties of recurrence intervals of large price fluctuations. *Europhys Lett*. (2012) **98**:38003. doi: 10.1209/0295-5075/98/38003
34. Zhou J, Gu GF, Jiang ZQ, Xiong X, Chen W, Zhang W, et al. Computational experiments successfully predict the emergence of autocorrelations in ultra-high-frequency stock returns. *Comput Econ*. (2017) **50**:579–94. doi: 10.1007/s10614-016-9612-1
35. Wei L, Zhang W, Xiong X. The mechanism and solution for the liquidity stampede crisis in stock markets. *J Manag Sci China*. (2017) **20**:1–23. Available online at: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFDLAST2017&filename=JCYJ201703001&v=MjA2MjJDVVI3cWZZK1JuRnlublVyclBMeTdTWkxHNEg5Yk1ySTlGWllSOGVYMUx1eFlTN0RoMVQzcVRyV00xRnI=
36. Xiong X, Yian C, Wei L, Chen S, Zhang W. Adaptive asset allocation and commonality in multi order books. In: *The 23rd International Conference on Computing in Economics and Finance, June 28–30, 2017*, New York NY.
37. Chiarella C, Iori G, Perelló J. The impact of heterogeneous trading rules on the limit order book and order flows. *J Econ Dyn Control*. (2009) **33**:525–37. doi: 10.1016/j.jedc.2008.08.001
38. Markowitz H. Portfolio selection. *J Finance*. (1952) **7**:77–91. doi: 10.1111/j.1540-6261.1952.tb01525.x
39. Brock WA, Hommes CH. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *J Econ Dyn Control*. (1998) **22**:1235–74. doi: 10.1016/S0165-1889(98)00011-6
40. Gil-Bazo J, Moreno D, Tapia M. Price dynamics, informational efficiency, and wealth distribution in continuous double-auction markets. *Comput Intell*. (2007) **23**:176–96. doi: 10.1111/j.1467-8640.2007.00301.x

Disclaimer: The views expressed in this paper are those of the author(s) and do not necessarily represent the views of the Shenzhen Stock Exchange.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhao, Cui and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 25 May 2020
doi: 10.3389/fphy.2020.00179

The Dynamic Correlations Between Mass Media News and New Media News in Stock Market

*Zuochao Zhang, Yongjie Zhang and Dehua Shen**

College of Management and Economics, Tianjin University, Tianjin, China

In this paper, we examine the dynamic correlations between mass-media news and new-media news as well as their reprints in the Chinese stock market. We mainly find that: (1) there is a significant positive correlation between the four types of news and the correlation between reprints and their congener media news is stronger; (2) both the mass-media news and new-media news have significant positive autocorrelation. The new-media news and the reprints rely more on mass-media news. (3) both mass-media news and new-media news will make a positive response to the corporate events, but the reaction of new-media news including the new-media reprints is stronger than mass-media; (4) the results of dynamic correlation analysis indicate that there is a significant correlation between the four types of news over time. The correlation between mass-media news and new-media news is stronger when the stock market performs well. We attribute these results to the symbiotic relationship between mass media and new media in the Chinese stock market, where mass media devotes to provide original news with high quality and new media transfers these numbers of news to more readers to attract more attention and discussion.

Keywords: mass-media news, new-media news, contemporaneous correlation, mutual information, PVAR, dynamic correlation

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Yang Hu,
University of Waikato, New Zealand
Bin Liu,
University of Wollongong, Australia

****Correspondence:***
Dehua Shen
dhs@tju.edu.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *23 February 2020*
Accepted: *27 April 2020*
Published: *25 May 2020*

Citation:
Zhang Z, Zhang Y and Shen D (2020) The Dynamic Correlations Between Mass Media News and New Media News in Stock Market. Front. Phys. 8:179. doi: 10.3389/fphy.2020.00179

INTRODUCTION

To figure out how the investors obtain the information will help us to understand the financial market [1]. Media, as the information intermediary, plays an important role in the process of information transmission in the stock market. With the development of the Internet, new media becomes one of the main platforms for investors to get financial information. Unlike traditional media, the large portals collect and provide a large number of, even all, news from the various news source for the investors. Investors need not subscribe to the newspaper or watch the TV to get information like before and they can get the information from the internet easily and quickly. On the other hand, the reprints on the internet could disseminate financial information to broad investors. So the internet brings a great challenge for mass media. Although the internet provides these advantages for the new media, the mass media still co-exist with media and provides information for investors. The existing literature indicates that how investors obtain information is important to understand the financial market [1]. Facing the status quo of the co-exist of mass media and new media in such an internet age, to study the correlation between mass media and new media is of importance to figure out the role of media played in the process of the information transmission and will help us to furtherly understand the financial market.

A strand of literature has examined the role of the media in the financial market [1–3]. But they are all concentrated on one specific media, either mass media or new media [4–6]. However, the correlation between mass media and new media as well as the different roles they played in the finical market is not studied. How they co-exist and what impact on the process of information transmission that they have in the financial market are not clear, especially considering the reprints. In this paper, we examine the correlation between mass-media news and new-media news considering the reprints simultaneously. We find that mass media and new media exhibit a symbiotic relationship in the stock market, where mass media devotes to provide original news with high quality and new media transfers these numbers of news to more readers to attract more attention and discussion. And we also find that mass-media news and new-media news maintain a positive correlation over time and the correlation becomes stronger when the stock market performs well. These results will help us to comprehend the role of the information intermediary of media. Therefore, we add to the literature on the role of the media in the financial market. On the other hand, there are some studies about the correlation between mass media and new media [7–12]. But fewer of them are about the financial market and take reprint into account. Hence, we also add to the literature on the correlation between mass media and new media for the financial market. The closest research is our previous work [11, 12], which we used the sample data of stocks of CSI 300 and SSE index, respectively, to explore the correlation between mass-media news and new-media news. Comparing their work, the contributions of this paper are 3-fold. Firstly, we consider the reprint when we studying the correlation between mass-media news and new-media news. Secondly, we improve the classification criteria for mass-media news and new-media news and provide a more objective and reasonable classification criterion. Thirdly, The larger samples, the stocks of the full Chinese stock market, and the new methodology used in this paper could provide more comprehensive and effective results.

The remainder of this paper is organized as follows. Section Literature review gives the literature review of the role of the media in the financial market and the correlation between mass media and new media. Section Data and sample describes the sample data using in this paper and its' statistical properties. Section Result and discussions reports the empirical results and section conclusions provides the conclusions.

LITERATURE REVIEW

The Role of Media in the Financial Market

Our paper related to the literature about the role of media played in the stock market and this research problem has attracted a lot of attention from the researchers. In the earlier literature, researchers have concentrated the role of either mass media [2, 13–16] or new media [4, 6, 17–22] play in the financial market.

For the mass media, Klibanoff et al. [23] investigate the effect of the country-specific news appearing on the front page of The New York Times on the reaction of closed-end country fund prices to the asset value. They found that the prices react more when there is news on the front page of the New York Times. Chan [15] used the data of news headlines to study the reaction of stock price to news and no-news and they found different patterns of reversal and momentum for stocks accompanied by or unaccompanied by the news. Tetlock [2] explored the interaction between the media and the stock market by analyzing the content of the column of the Wall Street Journal. By constructing the pessimistic factor as the measure of media content, they found the predictability of the media content for market prices and market trading volume. Tetlock et al. [24] quantified the language used in all Wall Street Journal and Dow Jones News Service stories of finance and found that the quantifying language provides the firms' fundamental information. Fang and Peress [3] used the number of the newspaper as the measure of media coverage to explore the effect of media coverage on stock returns and found that there will be higher returns for the stocks with no media coverage than the stocks with high media coverage. Peress [1] investigate the national newspaper strikes to study the influence of media and diffusion of information on trading and price formation. They found that the media could improve the dissemination of information in the stock market.

For the new media, Tumarkin and Whitelaw [17] explored the effect of postings on the stock prices and indicated that the number of postings could not predict the return and trading volume of the stocks. Antweiler and Frank [6] studied the effect of the postings from Yahoo! Finance and Raging Bull on the stocks and found that the stock messages could help to predict the market volatility. Das and Chen [18] build the sentiment index by exacting the text of the postings on Yahoo's message board and found it is related to the stock index levels, volume and volatility. Zhang et al. [25] found that the number of Baidu News could explain the volatility persistence of SME PRICE INDEX in China. Shen et al. [26] used the number of Baidu News as the proxy for internet information flow to provide the evidence for the Mixture of Distribution Hypothesis.

The Correlation Between Mass Media and New Media

Our study is also related to the literature about the relationship between mass media and new media. The argument of the relationship between mass media and new media has existed since the birth of new media. There are some studies support the competitive relationship between the two types of media [7, 27–29]. Dimmick et al. [7] used the survey data of telephone to explore the relationship between mass media and new media and found that the new media has a competitive displacement effect on the mass media. Lee and Leung [27] study the "medium-centric" and "user-centric" approaches by using a random sample and the result indicated that the internet could displace the use of traditional media. Ha and Fang [28] explores the displacement effect of new media on mass media using the survey data of consumer time spent on media. They found that the new media indeed has a replacement effect on the mass media. On the other hand, some researchers argue that there is a complementary relationship between mass media and new media [8, 10]. Chyi and Lasorsa [10] used a random-sample telephone survey of the public's response to the print and online newspaper

to study the substantial overlap between online and print readerships and found a complementary relationship between them. Nguyen and Western [8] also found that mass media exist to complement new media to meet the needs of people to get information. Furthermore, Jang and Park [9] indicated there the substitutability and complementary relationship between mass media and new media exist simultaneously and the magnitudes of these relationships exhibit different for the different purposes of users.

DATA AND SAMPLE

News Classification Criteria

We classified the news into two types of news according to the data field of news sources provided by Wisers. And we called the two types of news as the mass-media news and new-media news. In prior researches, the news has been classified into mass-media news and new-media news according to the news' sources of the communication media [7–9, 11, 12]. They referred the news transferred on the internet as the new-media news and the news transferred through traditional channels, such as the newspaper, is regarded as the mass-media news. But the traditional media is also seeking the new developmental pattern in the internet era. They provided their news on the internet for the readers. And they can provide high-quality news relied on their professional interviewing and editing team. The prior literature found that the news category, reader gender and interest in a particular topic have greater influences on the readers' news consumption than whether the news appears in print or online after studying the news consumption of readers of the news presented in online and print versions of newspapers [30].

On the other hand, the relationship between print and Internet media is symbiotic in China [31]. Hassid and Repnikova indicated that Internet portal sites in China are enormously powerful to attract the attention or discussion of readers, but they couldn't provide their own news for the barrier of legal rights to interview and edit. And the portal sites have to rely on reprints from newspapers and wire services [31]. So Hassid and Repnikova argued that "portals need newspapers to provide content, and the newspapers, in turn, need portals to publicize their articles, attract readership, and pay subscription fees in China."

Considering these facts, we argue that the online news provided by the traditional media plays a similar role to the news in print. So we improve the methodology introduced by Zhang et al. [11] to categorize the news coming from the online traditional media into the mass-media news. There are two reasons for our classification criteria. Firstly, the online news provided by the traditional media shares the same interviewing and editing team with the news in print. So the online news and the news in print of traditional news are homogeneous. Secondly, for the legal rights to interview and edit, the traditional media and the new media play different positions during the transfer of information, even for the online news provided by the traditional media and the new media. According to this classification criteria, we classified our news data into two types of news, i.e., the mass-media news and new-media news. And the detailed definitions of mass-media news and new-media news are as follows:

(1) We regard the news provided by mass-media including the online news of them as the mass-media news. Such as the news provided by the newspapers, the TV, the radio, and the magazine.
(2) The new-media news is the news that is only provided by the Internet. Such as the news provided by the portal, the blog, and social media.

Furthermore, there is a large number of reprints with the development of the internet. Generally, there is a statement of the copyright for the original news on the website. So the website is not allowed to modify the content of news even for the headlines when it reprints one news. Based on this, we defined the directed reprint news as the news with the same headlines. For one news, we chose the first one appeared in the Wisers as the original news and the number of directed reprint news equal to the number of the news with the same headline minus one. Then we classified the reprint news into mass-media news and new-media news based on the classification criteria introduced in this paper.

Sample Select

The news data used in this paper comes from the database of Wisers which is provided by the Wisers Information Limited. Wisers is one of the largest databases of Chinese media including all the Chinese news of the Chinese stock market. We focus on the news sample of all A-share in the Chinese stock market dropping the stocks in special treatment. According to the illustration of the Wisers database, the news is crawled from the Internet based on keywords such as the names of the company, the names of the companies in the financial sector are often mentioned in the financial reports. These reports are not the news of these companies, but the Wisers may crawl them from the Internet. Therefore, the news of financial stocks may contain noise. And we drop the financial stocks in our sample. Our sample period spans from 1 January 2015 to 31 December 2018 with daily observations, and we get a sample including 3,067 stocks. Based on the classification criteria introduced in this paper, we get 726 types of mass-media sources and 1,488 types of new-media sources in our sample period. In this paper, we also used the trading data and financial data, such as the close price, the market value, the announcement and so on. And we get these data from the database of CSMAR and Wind, which are both the professional financial database in China.

Figure 1 illustrates the weekly number of four types of news. From top to bottom in **Figure 1**, the line represents the number of new-media news, new-media reprints, mass-media news, and mass-media reprints. As we can see from **Figure 1**, the number of new-media news, no matter for the total news or the reprints, is significantly larger than the mass-media news and the trend of the four types of news is similar. This implies that the new-media news provides a richer information environment for us and there is a strong correlation between the mass-media news and new-media news. **Table 1** reports the summary statistics for the mean daily number of the four types of news for our sample. We calculate the statistical property of the news' number for every

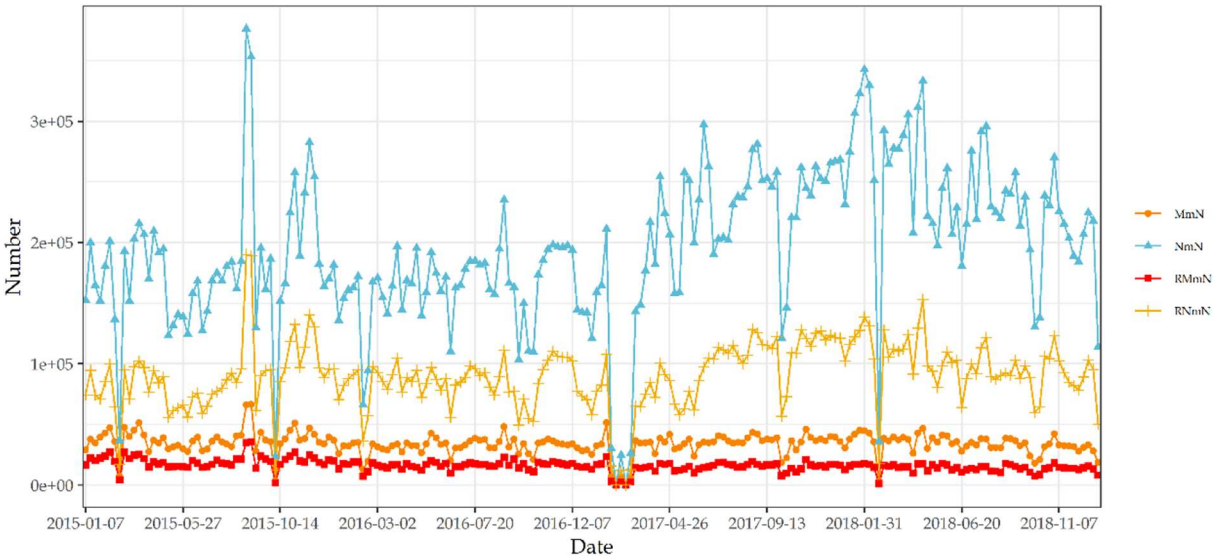


FIGURE 1 | The mean weekly number of four types of news.

TABLE 1 | Summary statistics for four types of news.

	MmN	*NmN*	*RMmN*	*RNmN*
Observations	3,067	3,067	3,067	3,067
Mean	1.54***	8.86***	0.72***	4.07***
	(19.24)	(24.31)	(17.70)	(20.86)
SD	3.80***	18.34***	2.38***	10.51***
	(36.99)	(37.76)	(38.17)	(36.01)
Q1	0.10***	1.07***	0.03***	0.27***
	(5.12)	(7.99)	(3.55)	(4.63)
Median	0.42***	3.73***	0.12***	1.12***
	(7.04)	(13.48)	(4.52)	(8.22)
Q3	1.49***	9.24***	0.44***	3.50***
	(13.26)	(19.26)	(8.06)	(13.64)
Max	60.18***	278.85***	40.87***	164.32***
	(37.82)	(31.40)	(40.69)	(29.98)
Min	0	0	0	0

This table presents summary statistics of the number of the four types of news at the firm-day level. MmN, NmN, RMmN, and RMmN denote the mass-media, new-media, the reproduction of mass-media, and the reproduction of new-media, respectively. Q1 and Q3 represent the 25th percentile and 75th percentile, respectively.
****Denotes t-test statistically significant at the 1% level.*

stock, and then, report the mean of these statistical properties in **Table 1**. From **Table 1**, we can find that the standard deviation of the new-media news is larger than the mass-media news, which implies that the new-media news is more likely to be influenced by the shocks.

RESULT AND DISCUSSIONS

In this section, we investigate the relationship between the four types of news defined in this paper. Firstly, we investigate the contemporaneous correlation between the four types of news. Secondly, we compare the reaction of the four types of news to different corporate events. Third, we conduct the Panel Var Model to explore the lead-lag relationship between the four types of news.

The Contemporaneous Correlation Between the Four Types of News

In this section, we calculate the Pearson correlation coefficient, Kendall correlation coefficient, and the Spearman correlation coefficient to investigate the contemporaneous correlation among the four types of news for every stock in our sample. And then we get the mean of these correlations for all stocks. **Table 2** reports the results of the correlation matrix for the four types of news. Panel A of **Table 2** shows the mean Pearson correlation coefficient among the four types of news. We can find that the correlation coefficients are all larger than 0.6 with a significant level of 0.01. These results indicate that there is a significant correlation among the four types of news defined in this paper and imply that the mass media and new media may pay attention to the same event every day although there is a big difference in the quantity of news that their provide. For these mean correlation coefficients, the correlation coefficient between the new-media reprints and new-media news is the largest which is larger than 0.9 with a significant level of 0.01 and the correlation coefficient between the mass-media news and the mass-media news reprints is also larger than 0.9 at a significant level of 0.01. On the other hand, the correlation coefficient between mass-media news and new-media reprints as well as the one between new-media news and mass-media reprints are around 0.7. The difference between these results shows that the two media both pay more attention to their congener media when they reprint news from other media. The larger correlation coefficient between mass-media news and new-media reprints than the one between new-media news and mass-media reprints implies that the mass media is paid more attention to when

TABLE 2 | The correlation matrix for the four types of news.

	MmN_t	NmN_t	$RMmN_t$	$RNmN_t$
PANEL A: PEARSON CORRELATION				
MmN_t	1			
NmN_t	0.745***	1		
	(397.1)			
$RMmN_t$	0.923***	0.678***	1	
	(996.7)	(325.8)		
$RNmN_t$	0.751***	0.932***	0.714***	1
	(411)	(1051.2)	(354.7)	
PANEL B: KENDALL CORRELATION				
MmN_t	1			
NmN_t	0.486***	1		
	(325.4)			
$RMmN_t$	0.701***	0.418***	1	
	(524.4)	(285.6)		
$RNmN_t$	0.544***	0.705***	0.501***	1
	(374.6)	(577.5)	(353.3)	
PANEL C: SPEARMAN CORRELATION				
MmN_t	1			
NmN_t	0.560***	1		
	(300.8)			
$RMmN_t$	0.728**	0.478***	1	
	(491.2)	(261.9)		
$RNmN_t$	0.596***	0.779***	0.542***	1
	(356.0)	(510.5)	(331.3)	

This table reports the correlation matrix for the daily number of the four types of news at the firm-day level. MmN_t, NmN_t, $RMmN_t$, and $RNmN_t$ denote the mass-media, new-media, the reproduction of mass-media, and the reproduction of new-media, respectively.
***Denotes t-test statistically significant at the 5% level.*
****Denotes t-test statistically significant at the 1% level.*

one media reprint news from other media. And we get similar conclusions from the results of Kendall correlation and Spearman correlation in Panel B and C in **Table 2**. We attribute these results that the mass media provides a large number of original news because of their advantage of interviewing and editing team. So the new media depend more on the mass media to some extent.

Then we calculate the Mutual Information to further study the relationship between the four types of news. In probability and information theory, the Mutual Information is a useful measure of interdependence between two variables. **Table 3** reports the result of Mutual Information among four news and we report the mean of Mutual Information for all stock. As we can see from **Table 3**, the Mutual Information is all larger than 0.3 with a significant level of 0.01 implying a significant positive correlation among the four types of news. Similarly, the Mutual Information between mass-media news and mass-media reprints (0.431) is larger than the one between new-media news and mass-media reprints (0.368) as well as the Mutual Information between new-media news and new-media reprints (0.915) is larger than the one between mass-media news and new-media reprints (0.427), which indicate that the two media both pay more attention to their congener media when they reprint news from other media.

TABLE 3 | The mutual information matrix for the four types of news.

	MmN_t	NmN_t	$RMmN_t$	$RNmN_t$
MmN_t	1			
NmN_t	0.492***	1		
	(60.59)			
$RMmN_t$	0.431***	0.368***	1	
	(73.42)	(55.53)		
$RNmN_t$	0.427***	0.915***	0.334***	1
	(60.08)	(90.39)	(56.90)	

This table reports the mutual information matrix for the daily number of the four types of news at the firm-day level. MmN_t, NmN_t, $RMmN_t$, and $RNmN_t$ denote the mass-media, new-media, the reproduction of mass-media, and the reproduction of new-media, respectively.
****Denotes t-test statistically significant at the 1% level.*

Also, the Mutual Information between new-media reprints and mass-media news (0.427) is larger than the one between mass-media reprints and new-media news (0.368) implying that mass-media news is paid more attention when the media choose which news to reprint.

The Lead-Lag Relation for the Four Types of News

In this section, we investigate the lead-lag relation among the mass-media news, new-media news, mass-media reprints, and new-media reprints with the following Panel Vector Autoregression (VAR) Model with one-period delay.

$$A_0 x_t = f_i + \sum_{j=1}^{2} A_j x_{i,t-1} + d_{i,t} + \mu_{i,t} \quad (1)$$

$$A_0 z_t = f_i + \sum_{j=1}^{4} A_j z_{i,t-1} + d_{i,t} + \mu_{i,t} \quad (2)$$

where ${x_{i,t}}' = (MmN_{i,t}, NmN_{i,t})$ and ${z_{i,t}}' = (MmN_{i,t}, NmN_{i,t}, RMmN_{i,t}, RNmN_{i,t})$ are the vector of endogenous variables for the model (1) and model (2), respectively.

Table 4 reports the coefficients of the PVAR model. Panel A shows the result of the model (1). We can see that the coefficient of the lagged number of mass-media news is positive and statistically significant, while the coefficient of the lagged number of new-media news is negative and statistically significant in column 1 of panel A. This indicates that the daily number of mass-media news has a significant auto-correlation and relies on less on the lagged number of new-media news. In column 2 of panel A, the coefficients of the lagged number of mass-media news and new-media news are both positive and statistically significant, implying that there is an auto-correlation for the daily number of new-media news and the number of mass-media news could lead the number of new-media news. Considering the reprints in panel B of **Table 4**, the coefficients of the lagged number of mass-media news are all positive and statistically significant when we use the current daily number of mass-media

TABLE 4 | Results of panel vector autoregression (PVAR) model for four types of news.

	MmNt	***NmNt***		
PANEL A RESULTS OF MODEL (1)				
MmN_{t-1}	0.056***	0.114***		
	(0.001)	(0.004)		
NmN_{t-1}	−0.004***	0.003***		
	(0.000)	(0.001)		
	MmN_t	***NmN_t***	***$RMmN_t$***	***$RNmN_t$***
PANEL B RESULTS OF MODEL (2)				
MmN_{t-1}	0.182***	0.438***	0.072***	0.231***
	(0.002)	(0.001)	(0.001)	(0.005)
NmN_{t-1}	0.002***	0.028***	0.002***	0.010***
	(0.000)	(0.002)	(0.000)	(0.001)
$RMmN_{t-1}$	−0.210***	−0.528***	−0.076***	−0.277***
	(0.003)	(0.01)	(0.002)	(0.007)
$RNmN_{t-1}$	−0.0136***	−0.055***	−0.007***	−0.021***
	(0.001)	(0.003)	(0.000)	(0.002)

This table reports the results of the Panel Vector Autoregression (PVAR) Model for the four types of news. $A_0x_t = f_i + \sum_{j=1}^{2} A_j x_{i,t-1} + d_{i,t} + \mu_{i,t}$ *(model 1) and* $A_0z_t = f_i + \sum_{j=1}^{4} A_j z_{i,t-1} + d_{i,t} + \mu_{i,t}$ *(model 2). The* MmN_t, NmN_t, $RMmN_t$, *and* $RNmN_t$ *denote the mass-media, new-media, the reproduction of mass-media, and the reproduction of new-media, respectively.* MmN_{t-1}, NmN_{t-1}, $RMmN_{t-1}$, *and* $RNmN_{t-1}$ *denote the news number of mass-media, new-media, the reproduction of mass-media, and the reproduction of new-media lagged 1 day, respectively. Panel A and Panel report the results of model (1) and model (2) respectively.*

****Denotes t-test statistically significant at the 1% level.*

news, new-media news, mass-media reprints, and new-media reprints as the dependent variables. This implies that the number of mass-media news keeps the significant auto-correlation and could lead the number of new-media news, mass-reprints as well as new-media reprints when considering the reprints. At the same time, the coefficients of the lagged number of new-media news are also positive and statistically significant, which indicating that the lagged number of new-media news could lead the all the current number of all four types of media news including itself. Panel B of **Table 4** shows that the coefficients of the lagged number of mass-media news are all larger than the corresponding coefficient for the lagged number of new-media news. We can conclude that the number of four types of news relies more on the mass-media news including the mass media itself. We attribute this result to the different role that the mass media and new media play when they transfer the information. As we discussed above, the mass media is engaged to provide original news with high quality while the new media devotes to transfer to the readers. So the number of new-media news will increase when there is more mass-media news that could be chosen and transferred by the new media. At the same time, the journalists will seek "inspiration of creation" on the internet. So the number of new-media news may lead the mass-media news to some extent.

The last two columns of Panel B show that there is no autocorrelation for the mass-media reprints and new-media reprints. And the significant positive coefficients of lagged number mass-media news and new-media indicate that they rely more on the mass-media news and new-media news. The coefficients of the number of lagged mass-media reprints and new-media news are all negative and statistically significant. When the number of reprints is large, there may be a hot event in the stock market and the media will pay more attention to this event even overreact to it. So the negative coefficients are more likely due to mean-reversion.

Factors Drive the Coverage of News

In this section, we investigate the different reactions for the four types of news defined in this paper to the different shocks. We analyze the influence of corporate events on the coverage and reprints of the mass media and new media.

Following the work introduced by Drake et al. [32], we defined 5 corporate event indicator variables to investigate the reactions of the mass media and new media to these events. The first indicator, EARN is the indicator of the earnings announcement. When there is earning announcement for stock in one day, the value of the day, as well as the following day, is set to one for this stock and to zero otherwise. The second indicator, DIVEQ, is about the dividend- and equity-related event. When one firm issue a dividend- or equity-related announcement, the value of the DIVIEQ of the day, as well as the following day, is set to one and to zero otherwise. The third indicator, M&A, is the mergers and acquisitions-related events. The value of M&A on the day, as well as the following day, of the mergers or acquisition-related announcement, is set to one, and to zero otherwise. The fourth indicator, Emerg, represents the emergencies defined by the database of CSMAR. Similarly, on the days 0 and +1 of this events, the value of Emerg equal to one, and to zero otherwise. Our last indicator, Other, denotes the other corporate events included in the database of CSMAR In addition to the above four types of events.

Then we conducted the panel regression with a fixed effect to explore the reaction of the mass media and new media to these corporate events. Our model is as follows:

$$MCOV_{it} = \beta_1 EARN_{it} + \beta_2 DIVEQ_{it} + \beta_3 M\&A_{it} + \beta_4 Emerg_{it} + \beta_5 Other_{it} \quad (3)$$

where $MCOV_{it}$ is the news number of firm i in day t which include mass-media news, new-media news, mass-media reprints, and new-media reprints. And the other variables are defined above.

Table 5 reports the results of Model (3). From **Table 4**, we can see that the coefficients of all the events for the four types of news are significantly positive. This indicates that both the mass media and new media will increase the number of coverage to the firm when there appears a corporate event. And the coefficient of the Emerg is larger than the other coefficients of the event indicator implied that the emergences are more likely to attract the media to cover. The coefficients for new-media news are larger than the ones for mass-media news and the coefficients for new-media reprints are also larger than mass-media reprints, even larger than the one for mass-media news, which indicates

TABLE 5 | Events that drive the coverage of news.

	MmN_t	NmN_t	$RMmN_t$	$RNmN_t$
$EARN_t$	0.45***	6.22***	0.05***	2.40***
	(26.85)	(70.26)	(4.64)	(47.87)
$DIVEQ_t$	0.79***	4.69***	0.44***	2.29***
	(59.69)	(66.80)	(47.14)	(57.64)
$M\&A_t$	1.21***	6.23***	0.64***	2.94***
	(72.04)	(70.21)	(54.82)	(58.49)
$EMerg_t$	1.74***	10.28***	1.04***	6.11***
	(9.03)	(10.09)	(7.70)	(10.59)
$Other_t$	1.23***	6.27***	0.69***	3.01***
	(201.37)	(194.05)	(160.47)	(164.5)
Adj. R-Squared	0.013	0.015	0.008	0.010

This table reports the results of the panel regressions of news numbers on the different events about the firms. $MCOV_{it} = \beta_1 EARN_{it} + \beta_2 DIVEQ_{it} + \beta_3 M\&A_{it} + \beta_4 Emerg_{it} + \beta_5 Other_{it}$. *The* $MmN_t, NmN_t, RMmN_t$, *and* $RNmN_t$ *denote the mass-media, new-media, mass-media reprints, and new-media reprints, respectively.* $EARN_t$ *denotes the earnings announcement.* $DIVEQ_t$ *denotes the dividend- or equity-related event.* $M\&A_t$ *represents the mergers and acquisitions-related events.* $Emerg_t$ *are the emergencies that are defined in the CSMAR.* $Other_t$ *denotes the other events which are announced by the firms.*
***Denotes t-test statistically significant at the 1% level.*

that the reactions to corporate events of new media are stronger than the mass media. When there appears an event, there will be a large number of new-media news to cover it. It could make more investors aware of this event. So this result could provide a piece of evidence that the new media could improve the information environment for investors.

Dynamic Correlation for the Four Types of News

To investigate the dynamic correlation between the four types of news, the rolling window analysis is conducted in this section. We choose 240 business days (∼1 year) as the length of the rolling windows selected in this paper. Figures (1) and (4) in **Figure 2** shows the dynamic correlation between mass-media news and new-media news, mass-media news and mass-media reprints, new-media news and new-media reprints, as well as mass-media reprints and new-media reprints, respectively. Figure (5) in **Figure 2** shows the varieties of the Chinese stock market in the corresponding time. And we choose the index price of the CSI A-share index as the indicator of the performance of the Chinese stock market. The CSI A-share index is compiled by the CSI company and could reflect the performance of Chinese performance well. As we can see in **Figure 2**, there are significant positive correlations between the four types of news over time and there is a decreasing trend for all four types of dynamic correlation in this paper. There is also a decreasing trend for the Chinese stock market in the corresponding time, which can be seen in (5) of **Figure 2**. We attribute the decreasing trend of the dynamic correlation to the decreasing media attention when there is a poor performance of the market, especially for the new-media news. And we provide further evidence of this by conducting the quantile regression of the dynamic correlation on the index price of the CSI A-share index as well as the number of news on the CSI A-share index:

(1) The dynamic correlation between mass-media news and new-media news

(2) The dynamic correlation between mass-media news and mass-media reprints

(3) The dynamic correlation between new-media news and new-media reprints

(4) The dynamic correlation between mass-media reprints and new-media reprints

(5) The index price of CSI A-share index

FIGURE 2 | The dynamic correlation for four types of news and stock market.

$$correlation_{(\tau)t} = \alpha_{(\tau)} + \beta_{(\tau)} Index_t \quad (4)$$

TABLE 6 | The results of the quantile regression of the dynamic correlation on the stock market.

τ	$CorMN_t$	$CorMrM_t$	$CorNrN_t$	$CorMrN_t$
0.1	0.025***	0.023***	0.016***	0.036***
	(23.56)	(43.29)	(27.46)	(41.21)
0.2	0.024***	0.025***	0.020***	0.036***
	(36.54)	(22.31)	(17.86)	(23.65)
0.3	0.024***	0.027***	0.031***	0.039***
	(30.93)	(25.08)	(27.74)	(21.50)
0.4	0.024***	0.041***	0.038***	0.054***
	(30.03)	(16.19)	(21.94)	(19.47)
0.5	0.026***	0.041***	0.042***	0.056***
	(18.14)	(23.15)	(28.57)	(31.61)
0.6	0.029***	0.041***	0.042***	0.057***
	(23.62)	(20.81)	(17.76)	(37.22)
0.7	0.031***	0.036***	0.046***	0.052***
	(30.43)	(11.89)	(17.76)	(19.54)
0.8	0.033***	−0.021***	−0.011***	0.044***
	(25.12)	(−11.42)	(−11.03)	(12.11)
0.9	0.041***	−0.015***	−0.008***	0.010***
	(18.84)	(−3.75)	(−2.78)	(8.40)

This table reports the results of the quantile regressions of the dynamic correlation for four types of news on the stock market. $correlation_{(\tau)t} = \alpha_{(\tau)} + \beta_{(\tau)}Index_t$. The symbols $CorMN_t$, $CorMrM_t$, $CorNrN_t$, and $CorMrN_t$ denote the dynamic correlation between mass-media news and new-media news, mass-media news and mass-media reprints, new-media news and new-media reprints, as well as mass-media reprints and new-reprints, respectively. Quantile is the quantiles we used in this paper including 10, 20, 30, 40, 50, 60, 70, 80, and 90%. $\alpha_{(\tau)}$ is the constant of the quantile regression. $\beta_{(\tau)}$ denotes the coefficient of four types of dynamic correlations. And τ represents the different quantile.

****Denotes t-test statistically significant at the 1% level.*

where $correlation_{(\tau)t}$ is the correlation of the four types of news in time t at τ, $Index_t$ is the index price of the CSI A-share index in time t, $\alpha_{(\tau)}$ is the constant of the quantile regression. $\beta_{(\tau)}$ denotes the coefficient of four types of dynamic correlations. And τ represents the different quantile.

Table 6 reports the results of the quantile regression of the dynamic correlation for the four types of news on the index price of the CSI A-share index and we report the coefficients only. To ensure the readability of the table, all coefficients in **Table 6** are multiplied 1000. For the dynamic correlation between mass-media news and new-media news, the coefficients in the first column in **Table 6** are all significant positive for all quantile levels. And there is an increasing trend with the increase of the quantile level. This result indicates that there is a stronger correlation between mass-media news and new-media news when the stock market performs well. From the other three columns in **Table 6**, we can also see the similar increasing trend of the coefficients of different quantile levels except for the 0.8 and 0.9 percentile. For the model of dynamic correlation between mass-media news and mass-media reprints as well as the model of dynamic correlation between new-media news and new-media reprints, the coefficients of 0.8 and 0.9 percentile are negative. Besides, the coefficients of the corresponding quantile levels are smaller than the other quantile levels. These results imply that the correlations between reprints and their corresponding media news become weak when the stock market performs well. We attribute these results to that the bullish market attracts more attention of the media, especially for new media. According to the results above, we can see that the number of new-media news has larger volatility, which has a larger standard deviation, than the mass-media news and the new media makes larger reactions to the shocks. Considering the purpose of mass media, which provides objective and balanced information for the investors, and the ability to interview and editing, the mass media are more likely to provide a more stable quantity of news for the investors than new media every day. So when the market performs well, the new media will provide more news to the investors and shows a stronger correlation with mass-media news. On the other hand, both the mass media and new media are attracted by the extremely good performance of the stock market, they will both provide a larger quantity of news for the investors. So when the stock market performs extremely well, the reprinted media will have more news to choose from and reprint. In this condition, the reprinted media will choose the essence new to reprint. Therefore, the correlation between the reprints and the media-news becomes smaller when the stock market performs extremely well.

Robust Test

In this section, we conduct robust tests of the results above. Firstly, we add financial stocks into our sample to repeat the analysis discussed above. Secondly, we choose the main media source of mass-media news and new-media news instead of all media sources to conduct the robust test. We choose the media sources as well as sub-sources of them that provide eighty percent of the number of news for mass media and new media, respectively, as the main media. We get similar results for the two robust tests, so we report the results of the main media only in this section.

Table 7 reports the correlation matrix for the four types of news of the main media. Although the value of the correlation coefficients is not the same, we can also find a significantly positive contemporaneous correlation between the four types of news. The correlation between mass-media news and mass-reprints as well as the one between new-media news and new-media reprints are larger than others. And the correlation between mass-media news and new-media reprints is larger than the one between new-media news and mass-media reprints. These results are consistent with the analysis for the sample of all news sources. **Table 8** reports the Mutual Information between the four types of news for main media. Also, the significant positive Mutual Information indicates the positive correlation between the four types of news. The Mutual Information about the reprints also indicates that the two media both pay more attention to their congener media when they reprint news from other media and mass-media news is paid more attention when the media choose which news to reprint. These results are all consistent with the ones of all media samples.

Table 9 shows the results of lead-lag correlation analysis for the main-media news. We also could find that the number of mass-media news could lead to the number of new-media news,

TABLE 7 | The correlation matrix of main-media news.

	MmN_t	NmN_t	$RMmN_t$	$RNmN_t$
PANEL A: PEARSON CORRELATION				
MmN_t	1			
NmN_t	0.704***	1		
	(359.9)			
$RMmN_t$	0.918***	0.633***	1	
	(916.8)	(299.7)		
$RNmN_t$	0.719***	0.908***	0.674***	1
	(391.6)	(835.4)	(334.3)	
PANEL B: KENDALL CORRELATION				
MmN_t	1			
NmN_t	0.469***	1		
	(328.9)			
$RMmN_t$	0.737***	0.404***	1	
	(539.7)	(295.1)		
$RNmN_t$	0.550***	0.682***	0.501***	1
	(389.1)	(539.6)	(370.4)	
PANEL C: SPEARMAN CORRELATION				
MmN_t	1			
NmN_t	0.532***	1		
	(310.1)			
$RMmN_t$	0.756**	0.457***	1	
	(520.4)	(275.0)		
$RNmN_t$	0.592***	0.749***	0.536***	1
	(386.3)	(475.3)	(358.7)	

This table reports the correlation matrix for the daily number of the main-media news at the firm-day level. MmN_t, NmN_t, $RMmN_t$, and $RNmN_t$ denote the mass-media, new-media, the reproduction of mass-media, and the reproduction of new-media, respectively.
***Denotes t-test statistically significant at the 5% level.*
****Denotes t-test statistically significant at the 1%.*

TABLE 8 | The mutual information matrix for the four types of main-media news.

	MmN_t	NmN_t	$RMmN_t$	$RNmN_t$
MmN_t	1			
NmN_t	0.387***	1		
	(68.53)			
$RMmN_t$	0.378***	0.295***	1	
	(81.14)	(60.50)		
$RNmN_t$	0.340***	0.763***	0.269***	1
	(69.99)	(90.03)	(64.34)	

This table reports the mutual information matrix for the daily number of the four types of main-media news at the firm-day level. MmN_t, NmN_t, $RMmN_t$, and $RNmN_t$ denote the mass-media, new-media, the reproduction of mass-media, and the reproduction of new-media, respectively.
****Denotes t-test statistically significant at the 1% level.*

TABLE 9 | Results of panel vector autoregression (VAR) model for main-media news.

	MmN_t	NmN_t		
PANEL A RESULTS OF MODEL (1)				
MmN_{t-1}	0.026***	0.0549***		
	(0.001)	(0.004)		
NmN_{t-1}	0.000***	0.010***		
	(0.000)	(0.001)		
	MmN_t	NmN_t	$RMmN_t$	$RNmN_t$
PANEL B RESULTS OF MODEL (2)				
MmN_{t-1}	0.055***	0.209**	0.027***	0.099***
	(0.002)	(0.01)	(0.001)	(0.006)
NmN_{t-1}	0.004***	0.030***	0.002***	0.010***
	(0.000)	(0.001)	(0.000)	(0.001)
$RMmN_{t-1}$	−0.038***	−0.197**	−0.010***	−0.089***
	(0.003)	(0.015)	(0.002)	(0.008)
$RNmN_{t-1}$	−0.007**	−0.045***	−0.005***	−0.015***
	(0.000)	(0.002)	(0.000)	(0.001)

This table reports the results of the Panel Vector Autoregression (VAR) Model for the main-media news. $A_0x_t = f_i + \sum_{j=1}^{2} A_j x_{i,t-1} + d_{i,t} + \mu_{i,t}$ (model 1) and $A_0z_t = f_i + \sum_{j=1}^{4} A_j z_{i,t-1} + d_{i,t} + \mu_{i,t}$ (model 2). The MmN_t, NmN_t, $RMmN_t$, and $RNmN_t$ denote the mass-media, new-media, the reproduction of mass-media, and the reproduction of new-media, respectively. MmN_{t-1}, NmN_{t-1}, $RMmN_{t-1}$, and $RNmN_{t-1}$ denote the news number of mass-media, new-media, the reproduction of mass-media, and the reproduction of new-media lagged 1 day, respectively. Panel A and Panel report the results of model (1) and model (2) respectively.
***Denotes t-test statistically significant at the 5% level.*
****Denotes t-test statistically significant at the 1% level.*

new-media reprints, as well as mass-media reprints. The number of new-media news could lead the mass-media news, mass-media reprints, and new-media reprints, but the coefficients of the lagged number of new-media news are all smaller than the corresponding coefficients of the lagged number of mass-media news. Also, the coefficients of the number of lagged mass-media reprints and new-media reprints indicate a mean-reversion of the news coverage. **Table 10** reports the results of the panel regressions of main media news numbers on the different events about the firms for main-media news. From **Table 10**, we can see that there are significantly positive reactions to the corporate events and the new media has a more significantly positive response to these shocks, which are consistent with the results of the sample of all news sources.

For the analysis of the dynamic correlation between the four types of news for the main media, there is also a positive correlation over time. **Table 11** reports the results of the quantile regression of the dynamic correlation on the index price for the CSI A-share Index for main-media news. Also, the results indicate that the correlation between mass-media news and new-media news becomes stronger and the coefficients related to the reprints are also weak when the market performs well. And these are consistent with the above.

CONCLUSIONS

In this paper, we investigated the roles of mass media and new media play in the process of information transmission in the Chinese stock market. Firstly, we conducted the contemporaneous correlation analysis among the mass-media

TABLE 10 | Events that drive the coverage of main-media news.

	MmN	***NmN***	***RMmN***	***RNmN***
$EARN_t$	0.45***	6.22***	0.05***	2.40***
	(26.85)	(70.26)	(4.64)	(47.87)
$DIVEQ_t$	0.79***	4.69***	0.44***	2.29***
	(59.69)	(66.80)	(47.14)	(57.64)
$M\&A_t$	1.21***	6.23***	0.64***	2.94***
	(72.04)	(70.21)	(54.82)	(58.49)
$EMerg_t$	1.74***	10.28***	1.04***	6.11***
	(9.03)	(10.09)	(7.70)	(10.59)
$Other_t$	1.23***	6.27***	0.69***	3.01***
	(201.37)	(194.05)	(160.47)	(164.5)
Adj. R-Squared	0.013	0.015	0.008	0.010

This table reports the results of the panel regressions of the main media news number on the different events about the firms for main-media news. $MCOV_{it} = \beta_1 EARN_{it} + \beta_2 DIVEQ_{it} + \beta_3 M\&A_{it} + \beta_4 Emerg_{it} + \beta_5 Other_{it}$. *The* $MmN_t, NmN_t, RMmN_t$, *and* $RNmN_t$ *denote the mass-media, new-media, mass-media reprints, and new-media reprints, respectively.* $EARN_t$ *denotes the earnings announcement.* $DIVEQ_t$ *denotes the dividend- or equity-related event.* $M\&A_t$ *represents the mergers and acquisitions-related events.* $Emerg_t$ *are the emergencies that are defined in the CSMAR.* $Other_t$ *denotes the other events which are announced by the firms.*

****Denotes t-test statistically significant at the 1% level.*

TABLE 11 | The results of the quantile regression of the dynamic correlation on the stock market for main-media news.

τ	$CorMN_t$	$CorMrM_t$	$CorNrN_t$	$CorMrN_t$
0.1	−0.014***	0.025***	0.025***	0.015***
	(−14.05)	(13.13)	(31.13)	(12.00)
0.2	−0.11***	0.029***	0.027***	0.019***
	(−8.49)	(23.92)	(25.55)	(11.13)
0.3	0.003*	0.030***	0.033***	0.026***
	(1.75)	(25.44)	(15.06)	(16.48)
0.4	0.015***	0.039***	0.038***	0.032***
	(6.87)	(13.63)	(7.92)	(13.58)
0.5	0.021***	0.050***	0.075***	0.039***
	(15.61)	(23.70)	(13.47)	(21.80)
0.6	0.023***	0.053***	0.078***	0.042***
	(12.30)	(25.51)	(16.37)	(15.88)
0.7	0.026***	0.053***	0.077***	0.046***
	(12.68)	(11.10)	(23.18)	(25.62)
0.8	0.030***	−0.030***	−0.011***	0.047***
	(23.82)	(−11.60)	(−14.20)	(14.78)
0.9	0.039***	−0.006**	−0.015***	0.040***
	(8.82)	(−2.20)	(−2.75)	(11.34)

This table reports the results of the quantile regressions of the dynamic correlation for four types of main-media news on the stock market. $correlation_{(\tau)t} = \alpha_{(\tau)} + \beta_{(\tau)} Index_t$. *The symbols* $CorMN_t$, $CorMrM_t$, $CorNrN_t$, *and* $CorMrN_t$ *denote the dynamic correlation between mass-media news and new-media news, mass-media news and mass-media reprints, new-media news and new-media reprints, as well as mass-media reprints and new-reprints, respectively. Quantile is the quantiles we used in this paper including 10, 20, 30, 40, 50, 60, 70, 80, and 90%.* $\alpha_{(\tau)}$ *is the constant of the quantile regression.* $\beta_{(\tau)}$ *denotes the coefficient of four types of dynamic correlations. And* τ *represents the different quantile.*

**Denotes t-test statistically significant at the 10% level.*

***Denotes t-test statistically significant at the 5% level.*

****Denotes t-test statistically significant at the 1% level.*

news, new-media news, mass-media reprints, and new-media reprints. We found that there are significant positive correlations among these four types of news and the reprints have a stronger correlation with the same type of media news. The results also show that the correlation between mass-media news and new-media reprints is larger than the one between new-media news and mass-media reprints. Secondly, we studied the lead-lag relationship between mass-media news and new-media news. We found that mass-media news and new-media news exhibit high positive autocorrelation, but the results of mass-media reprints and new-media reprints are reverses. The mass-media reprints and new-media reprints rely more on the mass-media news and new-media news. The lagged number of mass-media news could lead the number of new-media news, mass-media reprints, as well as new-media reprints and the lagged number of new-media news, could also lead the number of mass-media news, mass-media reprints, as well as new-media reprints. But the coefficients of the lagged number of new-media news are smaller than the corresponding ones of the number of lagged mass-media news, which implies that new media relies more on mass media and mass media relying more on themselves may seek inspiration on the internet to some extent. Thirdly, we explored the reactions of mass-media news and new-media news to corporate events. We found that all four types of news positively react to these shocks, but the new media exhibits a stronger positive reaction to these shocks. Finally, we studied the dynamic correlation between the four types of news. We found that there are significant positive correlations between the four types of news over time. And the correlation between mass-media and new-media news becomes stronger when the stock market performs well. The correlation between the reprints and the corresponding media news becomes smaller when the stock market performs extremely well.

We attribute these results to the different roles that the mass media and new media play in the process of information transmission in the Chinese stock market. The mass media has the advantages of the interviewing and editing ability and the legal rights of interviewing and editing. While the new media, especially for the larger portals could attract more attention and discussion of readers. So mass media and new media exhibit a symbiotic relationship, where mass media devotes to provide original news with high quality and new media transfers these numbers of news to more readers to attract more attention and discussion. So the number of new-media news relies on more mass-media news. On the other hand, the higher autocorrelation for these four types of news as well as the higher contemporaneous correlation between reprints and the same type of news indicates that the competition relation also exists between mass media and new media.

Admittedly, we only focus on the number of news in this paper. And this paper didn't combine with the stock market completely. It is promising to further investigate the roles of mass media and new media in the content of news and to compare the different effects on the stock market of the two types of media based on their different roles. We leave these for future research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

ZZ: data curation, investigation, methodology, software, validation, visualization, roles, and writing—original draft. YZ: conceptualization, funding acquisition, project administration, supervision, writing—review, and editing. DS: conceptualization, formal analysis, funding acquisition, project administration, resources, roles/writing—original draft, writing—review, and editing.

FUNDING

This work is supported by the National Natural Science Foundation of China (71771170, 71790594, and 71701150) and the Young Elite Scientists Sponsorship Program by Tianjin (TJSQNTJ-2017-09).

REFERENCES

1. Peress J. The media the diffusion of information in financial markets: evidence from newspaper strikes. *J Financ.* (2014) **69**:2007–43. doi: 10.1111/jofi.12179
2. Tetlock PC. Giving content to investor sentiment: the role of media in the stock market. *J Financ.* (2007) **62**:1139–68. doi: 10.1111/j.1540-6261.2007.01232.x
3. Fang L, Peress J. Media coverage the cross-section of stock returns. *J Financ.* (2009) **64**:2023–52. doi: 10.1111/j.1540-6261.2009.01493.x
4. Zhang YJ, Song WX, Shen DH, Zhang W. Market reaction to internet news: information diffusion and price pressure. *Econ Model.* (2016) **56**:43–9. doi: 10.1016/j.econmod.2016.03.020
5. Dougal C, Engelberg J, García D, Parsons CA. Journalists and the stock market. *Rev Finan Stud.* (2012) **25**:639–79. doi: 10.1093/rfs/hhr133
6. Antweiler W, Frank MZ. Is all that talk just noise? The information content of internet stock message boards. *J Financ.* (2004) **59**:1259–94. doi: 10.1111/j.1540-6261.2004.00662.x
7. Dimmick J, Chen Y, Li Z. Competition between the internet traditional news media: the gratification-opportunities niche dimension. *J Media Econ.* (2004) **17**:19–33. doi: 10.1207/s15327736me1701_2
8. Nguyen A, Western M. The complementary relationship between the internet and traditional mass media: the case of online news and information. *Inform Res.* (2006) **11**:151–83. Available online at: http://informationr.net/ir/11-3/paper259.html
9. Jang S, Park M. Do new media substitute for old media? A panel analysis of daily media use. *J Media Econ.* (2016) **29**:73–91. doi: 10.1080/08997764.2016.1170021
10. Chyi HI, Lasorsa DL. An explorative study on the market relation between online print newspapers. *J Media Econ.* (2002) **15**:91–106. doi: 10.1207/S15327736ME1502_2
11. Zhang Y, Zhang Z, Liu L, Shen D. The interaction of financial news between mass media and new media: evidence from news on Chinese stock market. *Phys A.* (2017) **486**:535–41. doi: 10.1016/j.physa.2017.05.051
12. Zhang Z, Zhang Y, Shen D, Zhang W. The dynamic cross-correlations between mass media news, new media news, and stock returns. *Complexity.* (2018) 2018:7619494 . doi: 10.1155/2018/7619494
13. Mitchell ML, Mulherin JH. The impact of public information on the stock market. *J Financ.* (1994) **49**:923–50. doi: 10.1111/j.1540-6261.1994.tb00083.x
14. Liang B. Price pressure: evidence from the dartboard column. *J Bus.* (1999) **72**:119–34. doi: 10.1086/209604
15. Chan WS. Stock price reaction to news no-news: drift reversal after headlines. *J Finan Econ.* (2003) **70**:223–60. doi: 10.1016/S0304-405X(03)00146-6
16. Zhang Y, Zhang Y, Shen D, Zhang W. Investor sentiment and stock returns: evidence from provincial TV audience rating in China. *Phys A.* (2017) **466**:288–94. doi: 10.1016/j.physa.2016.09.043
17. Tumarkin R, Whitelaw RF. News or noise? Internet postings and stock prices. *Financ Anal J.* (2001) **57**:41–51. doi: 10.2469/faj.v57.n3.2449
18. Das SR, Chen MY. Yahoo! for Amazon: sentiment extraction from small talk on the web. *Manage Sci.* (2007) **53**:1375–88. doi: 10.1287/mnsc.1070.0704
19. Zhang W, Li X, Shen D, Teglio A. Daily happiness and stock returns: some international evidence. *Phys A.* (2016) **460**:201–9. doi: 10.1016/j.physa.2016.05.026
20. Jin X, Shen DH, Zhang W. Has microblogging changed stock market behavior? Evidence from China. *Phys A.* (2016) **452**:151–6. doi: 10.1016/j.physa.2016.02.052
21. Zhang W, Shen D, Zhang Y, Xiong X. Open source information, investor attention, and asset pricing. *Econ Model.* (2013) **33**:613–19. doi: 10.1016/j.econmod.2013.03.018
22. Shen DH, Li X, Zhang W. Baidu news coverage and its impacts on order imbalance and large-size trade of Chinese stocks. *Financ Res Lett.* (2017) **23**:210–16. doi: 10.1016/j.frl.2017.06.008
23. Klibanoff P, Lamont O, Wizman TA. Investor reaction to salient news in closed-end country funds. *J Financ.* (1998) **53**:673–99. doi: 10.1111/0022-1082.265570
24. Tetlock PC, Saar-Tsechansky M, Macskassy S. More than words: quantifying language to measure firms' fundamentals. *J Financ.* (2008) **63**:1437–67. doi: 10.1111/j.1540-6261.2008.01362.x
25. Zhang Y, Feng L, Jin X, Shen D, Xiong X, Zhang W, et al. Internet information arrival and volatility of SME PRICE INDEX. *Phys A.* (2014) **399**:70–74. doi: 10.1016/j.physa.2013.12.034
26. Shen DH, Zhang W, Xiong X, Li X, Zhang YJ. Trading and non-trading period internet information flow and intraday return volatility. *Phys A.* (2016) **451**:519–24. doi: 10.1016/j.physa.2016.01.086
27. Lee PSN, Leung L. Assessing the displacement effects of the internet. *Telemat Inform.* (2008) **25**:145–55. doi: 10.1016/j.tele.2006.08.002
28. Ha L, Fang L. Internet experience and time displacement of traditional news media use: an application of the theory of the niche. *Telemat Inform.* (2012) **29**:177–86. doi: 10.1016/j.tele.2011.06.001
29. Gentzkow M. Valuing new goods in a model with complementarity: online newspapers. *Amer Econ Rev.* (2007) **97**:713–44. doi: 10.1257/aer.97.3.713
30. D'Haenens L, Jankowski N, Heuvelman A. News in online and print newspapers: differences in reader consumption and recall. *New Media Soc.* (2004) **6**:363–82. doi: 10.1177/1461444804042520
31. Hassid J, Repnikova M. Why Chinese print journalists embrace the internet. *Journalism.* (2016) **17:**882–98. doi: 10.1177/1464884915592405
32. Drake MS, Thornock JR, Twedt BJ. The internet as an information intermediary. *Rev Acc Stud.* (2017) **22**:543–76. doi: 10.1007/s11142-017-9395-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhang, Zhang and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 03 June 2020
doi: 10.3389/fphy.2020.00199

Research on Dynamic Evolution of Systemic Risk in Chinese Banking System

Hong Fan and Wanqin Sheng*

Glorious Sun School of Business and Management, Donghua University, Shanghai, China

Systemic risk is not only related to the contagious risk of interbank market risk exposure, but also to the credit risk exposure. At the same time, the dynamic characteristics of banking system also affect the systemic risk. A large number of studies have revealed the contagious risk of interbank market risk exposure, but the systematic risk research that considers these factors at the same time is still rare. Therefore, the present paper constructs dynamically evolving banking systemic risk assessment model under double risk exposures in the same framework that includes two parts: one is the assessment of credit risk; the other is the contagious risk. On the assessment of credit risk, we use Geometric Brownian Motion in physics and Monte Carlo simulation calculation method based on the big data of the stock market. On the assessment of contagious risk, we use the method of Geometric Brownian Motion in physics and minimum density method. Then we use data of 205 Banks and 3,017 listed companies in 18 industries in Chinese to study the dynamic evolution law of the Chinese banking systemic risk. The results show that the evolution characteristics of credit risk of 18 real industries can be divided into 4 types which are stability after decline, increase in fluctuation, decrease in fluctuation, and fluctuation. The systemic risk of Chinese banking system with the double risk exposures gradually increased with the evolution of time, and then stabilized at a certain degree. Among them, the initial increase of contagious risk is fast, and then stable; the credit risk starts small, slowly increases, and eventually stabilizes. The evolution degree of contagious risk is higher than that of credit risk. Large banks have the strongest ability to withstand the impact of double risk exposures, while foreign banks have the weakest.

Keywords: systemic risk, geometric Brownian motion, credit risk exposure, interbank market risk exposure, dynamic evolution

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Dehua Shen,
Tianjin University, China
Takaaki Ohnishi,
The University of Tokyo, Japan

****Correspondence:***
Hong Fan
hongfan@dhu.edu.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *14 March 2020*
Accepted: *04 May 2020*
Published: *03 June 2020*

Citation:
Fan H and Sheng W (2020) Research on Dynamic Evolution of Systemic Risk in Chinese Banking System. Front. Phys. 8:199. doi: 10.3389/fphy.2020.00199

INTRODUCTION

After the international financial crisis in 2008 and the European debt crisis in 2010, countries have paid enough attention to strengthening financial supervision, maintaining financial stability, and preventing and controlling systemic risks. In October 2017, the report of the 19th National Congress of the Communist Party of China clearly pointed out that it is necessary to "improve the financial supervision system, keep the bottom line of no systemic financial risks," and raised the prevention of systemic risk to the height of national security. However, due to the complexity of systemic risk itself, the definition of systemic risk has not been unified in academia. But, some studies have clearly pointed out that systemic risk is a global and systematic risk, which is not

limited to the financial system. It is worth noting that the real industry plays a more and more important role in the process of economic development, and the relationship between the real industry and the financial industry is increasingly close. These correlations not only constitute the channel of risk transmission, but also further expand the negative impact on the entire economic system. When the important real industries encounter difficulties, the contagion risk brought by the recession makes our economy face great risks.

In financial system, the banking industry plays an important role. To prevent the systemic risk of the banking industry has become the core of maintaining financial stability. At present, the systemic risk is mainly studied on the contagious risk of interbank market risk exposure, while there are few studies on the double risk exposures of basic failure caused by credit risk exposure and contagious risk caused by interbank market risk exposure.

Since Allen and Gale [1] used the network model to study the impact of the interbank network structure on the banking systemic risk, the research on the banking systemic risk has been emerging. The mutual loan relationship between banks can solve the short-term liquidity shortage of banks and increase investment opportunities, but it can also improve the contagion risk between banks. Most researches show that the risk exposure of interbank market is the main channel of systemic risk. Allen and Gale [1] found that the fully connected network structure is more robust than the incomplete network structure, and the contagion risk is significantly smaller. Freixas et al. [2] found the probability of contagion risk depends on the setting of network model parameters by comparing the fully connected network structure with the ring network structure. Using the data of 27 Mexican banks in 2007, Martínez-Jaramillo [3] found that the vulnerability of the banking system is determined by the probability of banks default, the connectivity between banks, and the banks that are over exposed to the network. Mistrulli [4] found that the fully connected interbank market is more susceptible to contagion than the incompletely connected interbank market. Ladley [5] found that there is no interbank market structure that can maximize the stability under any conditions. Sui et al. [6] studied the problem of contagion risk in scale-free networks with different scale parameters and found that the higher the concentration, the more the number of banks in the network with contagion failure. Bao and Sun [7] studied financial structure and contagion risk by establishing a multi-agent simulation model and found that the hierarchical structure of center edge will increase the degree and scope of contagion risk compared with the fully connected structure.

Although the interbank market risk exposure is the core of the study of banking systemic risk, the impact of credit risk exposure on the systemic risk of banks cannot be ignored. At present, there are few studies on credit risk exposure based on VaR-based method, while those researches did not consider the effect of the interbank market risk exposure on systemic risk. In the existing literature, Adrian and Brunnermeier [8] proposed a CoVaR to measure the contribution of financial institution to the systemic risk. Huang et al. [9] measured the systemic risk as the marginal contribution of a financial firm to the distress insurance premium of the financial sector. A large number of contributions on VaR-based systemic risk measurement associate the systemic risk with the degree of interdependence among financial firms as in Billio et al. [10], Ang and Longstaff [11], and Diebold and Yilmaz [12]. In China, Li et al. [13] used 11 industry VaR indexes to build the systemic risk spillover network among industries in China from 2002 to 2017 to find that the financial industry is not the only systemic risk source and the systemic importance of information technology, materials, industry, and other real economic industries is increasing day by day. Zhai [14] established GARCH-Copula-CoVaR model based on industry indexes to empirically analyze the two-way risk spillover effect between each real industry and banking industry. Zhu and Ma [15] used the method of eigenvector centrality to measure the systemic risk of China's industry and found that the systemic risk of the financial industry is generally on the decline, while the systemic risk of the non-financial industry is at a relatively high level. Sun et al. [16] used the real estate industry index to measure the systemic risk spillover intensity of China's real estate industry to the financial industry based on the CoVaR model and quantile regression method.

Based on the perspective of complex network, few studies consider credit risk exposure and interbank market risk exposure at the same time; most of them are theoretical studies and lack of empirical data support. In addition, the systemic risk of those studies is mainly based on the static model without considering the evolution of the systemic risk over time. Lu and Wang [17] used the maximum entropy method to estimate the complete interbank network and set up a common loan network of banks based on the covariance of industry indexes, but they did not consider the credit risk of industry default and the evolution of the systemic risk over time. In their study, they found that the impact of different loan industry shocks on the degree of contagion risk is different. The crisis in the industry with more loan funds is more serious for the financial system. Wu et al. [18] studied theoretically the impact of the price fluctuation of common assets on the systemic risk and found that the indirect linked contagion of the sale of common assets is far greater than the direct linked contagion. Although this study involves credit risk exposure, the systematic risk in this study is based on the assumption of the model, not supported by the actual data, and the model of the credit risk exposure is static. Zhai and Bian [19] used the maximum entropy estimation method to construct the interbank network structure and the bipartite network structure between banks and real industries to identify the systemic important bank and the systemic important industry, but they did not study the effect of credit risk on the systemic risk. Caccioli et al. [20] found theoretically that interbank market risk exposure magnified the chain reaction of bank portfolio risk exposure. Glasserman and Young [21] also found theoretically that the probability of bank default caused by the direct impact of external asset loss is higher than that of bank contagion default, and the interconnected financial network can increase the expected loss and default degree. Chiu et al. [22] found that there is a relationship between creditor's rights and debts between the real industry and the banking industry and the risk will circulate among different departments. In addition,

the systemic risk should be discussed from the perspective of the whole industry. The banking industry and the real industry are closely linked through credit. Wang and Li [23] established an extended matrix model based on the credit market between the banking and real estate industry. They found that the financial system composed of the real estate industry and the banking industry is more vulnerable than the separate banking system and the risk transmission speed is significantly faster. Li et al. [24] studied the joint impact of Chinese Banks' internal lending market and real estate industry's credit risk exposure on the systemic risk of the banking system and found that the two had a great joint resonance effect on contagion risk. The study only considered real estate credit risk exposure, instead of all the industries.

To sum up, at present, most of the researches on credit risk exposure use VaR-based method to study the impact of credit risk spillover on systemic risk without considering the interbank market risk exposure, and these studies only qualitatively point out the importance of credit risk exposure of the real industry for the assessment of systemic risk and do not quantitatively analyze the default probability of the real industry and the quantitative effect of credit risk on systemic risk. Few studies consider both credit risk exposure and the interbank market risk exposure; however, most of them are theoretical studies and lack of empirical data support; in addition, the systematic risk of those studies is mainly based on the static model without considering the evolution of systemic risk over time. Therefore, the present paper studies the dynamical evolution of systemic risk in the same framework of both the credit risk exposure of the real industry and the contagion risk exposure of the interbank market. In order to construct the dynamical evolution of the systemic risk, the present paper uses the Geometric Brownian Motion to simulate the movement of assets of the real industry and banks, which is often used to estimate the dynamic motion of assets in economics [25, 26]. In Lehar [25], Lehar first used the Geometric Brownian Motion to simulate the movement of assets of banks and studied the systemic risk based on the correlation of assets of banks, but the study did not consider the interbank market and the real industry that banks loan to. After constructing the dynamic evolution of assets and liabilities, the present paper uses the simulation methods, complex network methods, and empirical data to build a dynamic evolution model of banking systemic risk assessment to analyze the impact of the direct shock of credit risk exposure of the real industry and the indirect shock of interbank market exposure on the systemic risk of China's banking system.

The innovation of this paper is mainly in three aspects. First, inspired by the research of Caccioli et al. [20] and Lehar [25], the present paper is the first to combine the credit risk exposure of the real industry and the risk exposure of the interbank lending market in the same framework to conduct the dynamic evolution study of systemic risk by using Geometric Brownian Motion in physics. Second, consider that most of studies used the maximum entropy method to estimate the interbank network and lacked real data support, the present paper uses real data of listed companies in the real industry to estimate the dynamic credit risk exposure of real industry, then uses the minimum density method and real banking data to estimate the more realistic dynamic interbank market risk exposure. Third, in the existing literature, the method of stress test for banks is to let one or one type of bank assets lose, while in this paper, the listed companies in the real industry are shocked, which causes those companies bankruptcy, thus the credit risk of the real industry is formed.

This paper is organized as follows. In section Model, we describe the dynamic evolution model of banking systemic risk assessment under double risk exposures. In section The Calculation of Banking Systemic Risk Under Double Risk Exposures, we propose how to calculate the systemic risk of the banking system. In section Data, we explain the source of data. In section Results, we discuss the assessment of dynamically evolving banking systemic risk under double risk exposures which include two parts: one is the assessment of credit risk of the real industry in section The Assessment of Credit Risk of the Real Industry; the other is the contagious risk section The Assessment of Systemic Risk under Double Risk Exposures. In section Discussion and Conclusions, we give conclusions.

MODEL

In this section, we first build a dynamic evolution model of banking systemic risk assessment under double risk exposures, as shown in **Figure 1**. We assume that there are a set $N = \{1, \cdots N\}$ of banks in the system, each bank has a dynamically evolved bank balance sheet structure as shown in **Figure 1A**. Each bank's assets include interbank lending assets $A_i(t)$ and external assets $V_i(t)$. Each bank's liabilities include interbank borrowing liabilities $L_i(t)$ and external liabilities $D_i(t)$. Each bank's assets minus liabilities is owner's equity $e_i(t)$. The double risk exposures faced by banks refer to the credit risk exposure of the real industry shown in **Figure 1B** and the interbank market risk exposure shown in **Figure 1C** (The payment matrix of the interbank market risk exposure can be described as Equation (16), which shows the interbank loan assets $A_i(t)$and the interbank loan liabilities$L_i(t)$.). At time step T, the dynamic evolution of the credit risk exposure of the real industry (see **Figure 1B**) results in dynamic external asset losses for the banks (see the dashed boxes $Loss_i(t)$ and $Loss_j(t)$ in **Figure 1A**). The interbank loan assets $A_i(t)$ and interbank loan liabilities $L_i(t)$ are dynamic, so the interbank market risk exposure is also dynamically evolving. As shown in **Figure 1D**, the dynamic interbank market network structure evolved from time step t_1 to time step t_{150}. At each time step, we use the clearing payment vector to conduct asset and liability liquidation on the banking system under double risk exposures and to calculate the systemic risk of the banking system at each time step.

Next, this paper constructs a dynamic evolution of credit risk exposure of the real industry corresponding to **Figure 1B** in section Dynamic Evolution of Credit Risk Exposure of the Real Industry and Estimation of the Interbank Market Risk Exposure constructs an interbank market risk exposure estimation model corresponding to **Figure 1C**. Section Default Algorithms Based on the Dynamic Evolution of Bank Balance Sheet Structure explains the default algorithms based on dynamic evolution

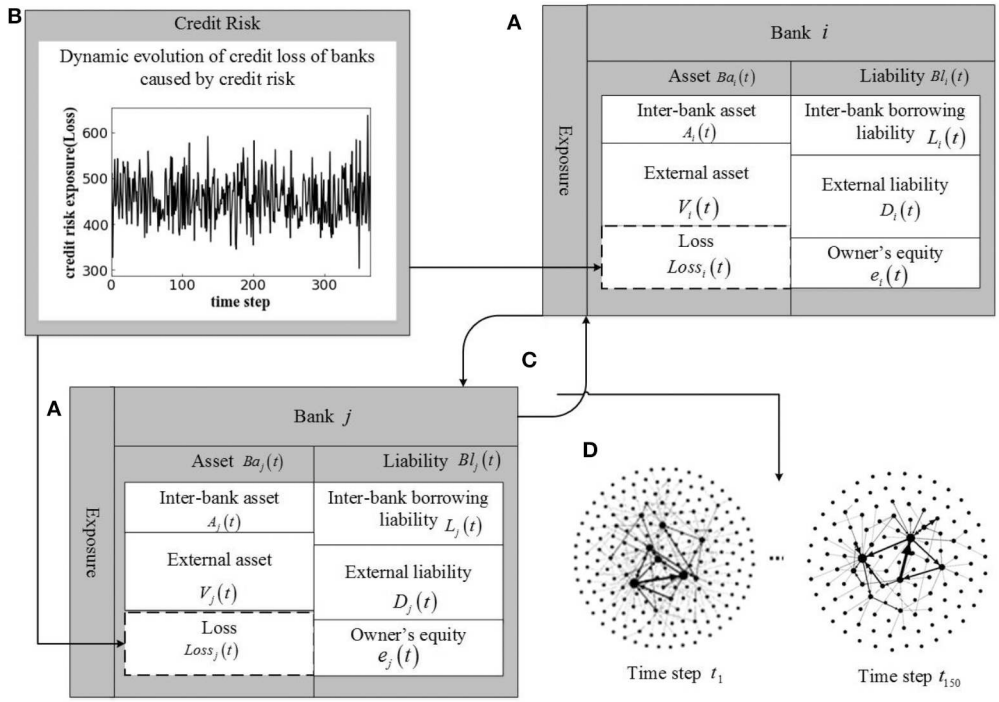


FIGURE 1 | A dynamic evolution model of bank systemic risk assessment under double risk exposures. **(A)** Represents dynamic evolution of bank balance sheet structure of bank *i* and *j*. **(B)** Plots the dynamic evolution curve of credit loss of banks caused by credit risk of the real industry. **(C)** Represents interbank market risk exposure. **(D)** Shows dynamic evolution of interbank market network structure from time step t_1 to time step t_{150}.

of bank balance sheet structure corresponding to **Figure 1A.** Section The Dynamic Evolution of Interbank Market Network discusses the dynamic evolution of interbank market network corresponding to **Figure 1D**.

Dynamic Evolution of Credit Risk Exposure of the Real Industry

The credit relationship between banks and the real industry constitutes a bank-real industry credit network. When credit default occurs in the real industry, it will cause loss of external assets of the related banks. Therefore, credit risk exposure of the real industry faced by banks includes three parts that are credit network between banks and the real industry, the distribution of credit default probability in the real industry, and loss of banks' external assets caused by credit risk exposure.

Credit Network Between Banks and Real Industry

Assuming that there is m real industries in China, the loan that bank i provide to the real industry are expressed as L_{ik}, the assets of bank i is V_i, the proportion of loans L_{ik} provided by bank i to the real industry k in the assets V_i of bank i is called loan proportion, expressed as R_{ik}.

$$R_{ik} = \frac{L_{ik}}{V_i} \quad i = 1, 2, \cdots, N, \quad k = 1, 2, \cdots, m \tag{1}$$

Considering that the distribution data of loan that a bank provides to the real industry is not disclosed in the annual report, it is impossible to calculate the loan proportion of all banks. Therefore, R_{ik}indicates the loan proportion provided by the banks that can obtain the data to the real industry.

In order to determine the loan proportion of all banks, we have divided N banks into a types, with Z banks under each type. Assuming that banks of the same type have the same loan proportion, Nbanks all provide loans to m real industries. R_{bik} is used to represent the proportion of loans provided by bank i under type b to real industry k. The calculation formula is as follows.

$$R_{bik} = \frac{\sum R_{ik}}{Z_b} \quad b = 1, 2, \cdots, a \tag{2}$$

$\sum R_{ik}$ is the sum of Equation (1) and Z_b indicates the number of banks under type b. In the end, we calculate the proportion of loans R provided by N banks to m industries is a matrix of $N \times m$.

$$R = \begin{bmatrix} R_{111} & \cdots & R_{11k} & \cdots & R_{11m} \\ \vdots & \ddots & \vdots & & \vdots \\ R_{bi1} & \cdots & R_{bik} & \cdots & R_{bim} \\ \vdots & & \vdots & \ddots & \vdots \\ R_{aN1} & \cdots & R_{aNk} & \cdots & R_{aNm} \end{bmatrix} \tag{3}$$

This paper assumes that every bank provides loans to every industry, then credit network Z between banks and real industry

can be expressed as a matrix of $N \times m$.

$$Z = \begin{bmatrix} V_1R_{111} & \cdots & V_1R_{11k} & \cdots & V_1R_{11m} \\ \vdots & \ddots & \vdots & & \vdots \\ V_iR_{bi1} & \cdots & V_iR_{bik} & \cdots & V_iR_{bim} \\ \vdots & & \vdots & \ddots & \vdots \\ V_NR_{aN1} & \cdots & V_NR_{aNk} & \cdots & V_NR_{aNm} \end{bmatrix} \quad (4)$$

Probability Distribution of Credit Default in Real Industry

Estimation of dynamic evolution of assets and liabilities of listed company in the real industry

Supposing the asset value V_g of the Listed Company in the real industry$g$ obeys Geometric Brownian Motion in physics [25], its drift rate is $\mu_g$and volatility is σ_g.

$$dV_g = \mu_g V_g dt + \sigma_g V_g dz \quad (5)$$

The owner's equity $E_g(t)$ of Listed Company g is regarded as call option, whose value is equal to the nominal value of liabilities $D_g(t)$ of Listed Company g in mature period T. Based on Black Scholes Model [26], we can obtain the relationship between owner's equity and asset value of Listed Company as Equation (6).

$$E_g(t) = V_g(t)N(d_t) - D_g(t)N(d_t - \sigma_g\sqrt{T}) \quad (6)$$

$$d_t = \frac{\ln\left(V_g(t)/D_g(t)\right) + \left(\sigma_g^2/2\right)T}{\sigma_g\sqrt{T}} \quad (7)$$

$$D_g(t) = D_g(0)e^{rt} \quad (8)$$

Where T is the maturity and$T$=365; t is the day before maturity. $E_g(t)$ is the daily equity value of Listed Company g, which can be obtained from the stock market. $D_g(0)$ is the initial liability value of Listed Company g, which can be obtained from the balance sheet. This paper assumes that the liabilities of all listed companies increase according to the risk-free rate of return r.

In this paper, we first set the initial drift rate $\mu_g(0)$ and volatility $\sigma_g(0)$ as any value, then get the initial asset value $V_g(0)$ from the balance sheet of the Listed Company and the owner's equity value of each day can be expressed as $E_g(0), E_g(1), \cdots, E_g(T)$. According to Equation (8), the liability can be expressed as $D_g(0), D_g(1), \cdots, D_g(T)$, then we estimate the asset value of each day as $\hat{V}_g(0), \hat{V}_g(1), \cdots, \hat{V}_g(T)$ according to Equations (6–8). Using the maximum likelihood function proposed by Duan J-C [27] to calculate the parameter of drift rate μ_g and volatility σ_g.

$$L\left(\mu_g, \sigma_g; \hat{V}_g(0), \hat{V}_g(1), \cdots, \hat{V}_g(T)\right) = -\frac{T}{2}\ln(2\pi\sigma_g^2 h) - \frac{T}{2}\sum_{k=1}^{T}\frac{\left(R^*{}_g(k^*) - (\mu_g - \frac{\sigma_g^2}{2})h\right)^2}{\sigma_g^2 h} - \sum_{k=1}^{T}\ln\hat{V}_t \quad (9)$$

In Equation (9), $R^*{}_g(k^*) = \ln\left(\hat{V}_g(t)/\hat{V}_g(t-1)\right), h = \frac{1}{365}$.

Finally, the drift rate μ_g and volatility σ_g estimated by Equation (9) are substituted into Equation (10) to estimate the assets of listed companies $V_g(t)$.

$$V_g(t) = V_g(0)e^{\mu_g - \left(\sigma_g^2/2\right)th + \sigma_g\sqrt{th}*z_g(t)} \quad (10)$$

In Equation (10), $z_g(t)$obeys the standard distribution of $N(0,1)$, $h = 1/T$.

Calculation of basic bankruptcy status of listed companies in real industry

Basic bankruptcy occurs when assets of listed companies in various industries are less than liabilities, which can be expressed by Equation (11).

$$V_g(t) - D_g(t) < 0 \quad (11)$$

In Equation (11), the assets $V_g(t)$ and liabilities $D_g(t)$ of listed companies have estimated in section Estimation of Dynamic Evolution of Assets and Liabilities of Listed Company in the Real Industry.

In this paper, the basic failure state variable is defined as integer of 0 or 1, which can be expressed by Equation (12):

$$S_g(t) = \begin{cases} 1 \ if \ V_g(t) - D_g(t) < 0 \\ 0 \qquad otherwise \end{cases} \quad (12)$$

Calculation of probability distribution of default in real industry

The number of listed companies in the industry k can be expressed as n_k.By using Monte Carlo simulation method, we calculate $S_g(t)$ according to Equation (12), then we calculate the default probability of industry k at time step t. This paper defines the default probability of industry as Equation (13).

$$P_k(t) = \frac{\sum S_g(t)}{n_k}, k = 1, 2, \ldots, m \quad (13)$$

In Equation (13), $P_{\mathrm{k}}(t)$ is the default probability of industry k at time step t, $\sum S_g(t)$ is the sum of the number of all the failed listed companies in the industry k at time step t. In this paper, $P_{\mathrm{k}}(t)$ that is the default probability curve with time changes is used to measure the evolution of credit default probability of the real industry.

Loss of Bank's External Assets Caused by Credit Risk Exposure

When listed companies in the real industry are unable to repay banks' loans in case of insolvency, the banks will suffer external asset losses due to the credit default of real industry. The loss of external assets of a bank can be calculated by combining Equation (1) and (13), which is expressed by Equation (14).

$$Loss_{ik}(t) = R_{bik} \times V_i(t) \times P_{\mathrm{k}}(t) \quad (14)$$

In Equation (14), $Loss_{ik}(t)$ is the loss of external asset loans from bank i provides to the real industry k at time step t, $V_i(t)$ is the asset value of bank i at time step t.

Finally, we combine the credit network between banks and real industry, the evolution sequence of bank's assets V_i, and the distribution of bank's loss of external assets $Loss_{ik}(t)$ to calculate the dynamic evolution of credit risk exposure between banks and the real industry, which can be expressed by Equation (15).

$$Loss(t) = \begin{bmatrix} V_1(t)R_{111} & \cdots & V_1(t)R_{11k} & \cdots & V_1(t)R_{11m} \\ \vdots & \ddots & \vdots & & \vdots \\ V_i(t)R_{bi1} & \cdots & V_i(t)R_{bik} & \cdots & V_i(t)R_{bim} \\ \vdots & & \vdots & \ddots & \vdots \\ V_N(t)R_{aN1} & \cdots & V_N(t)R_{aNk} & \cdots & V_N(t)R_{aNm} \end{bmatrix} \times \begin{bmatrix} P_1(t) \\ \vdots \\ P_k(t) \\ \vdots \\ P_N(t) \end{bmatrix} \tag{15}$$

Estimation of the Interbank Market Risk Exposure

Considering an interbank lending network formed by N banks, this network can be expressed in $N \times N$ matrix form by Equation (16).

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{iN} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nj} & \cdots & x_{NN} \end{bmatrix} \begin{matrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_N \end{matrix}$$
$$L_1 \quad \cdots \quad L_j \quad \cdots \quad L_N \tag{16}$$

In the interbank lending market, x_{ij} refers to the loans borrowed by bank$j$ from bank i; the sum of each row in the matrix A_i represents the total lending assets of bank i; the sum of each column in the matrix L_j represents the total lending liabilities of bank i. Since a bank itself has no lending relationship, the matrix in Equation (16) is modified to a new interbank lending network matrix with diagonal 0, which is expressed as X^* by Equation (17).

$$X^* = \begin{bmatrix} 0 & \cdots & x_{1j} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{i1} & \cdots & 0 & \cdots & x_{iN} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nj} & \cdots & 0 \end{bmatrix} \begin{matrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_N \end{matrix}$$
$$L_1 \quad \cdots \quad L_j \quad \cdots \quad L_N \tag{17}$$

In this paper, we use the minimum density method proposed by Anand et al. [28] to estimate the interbank market risk exposure matrix X^* .

In the minimum density method, parameter c is introduced, which represents the fixed cost of establishing connection between banks, that is, if there is a connection between banks, there will be a connection cost of c, then the minimum density method can be expressed as a constraint optimization problem for matrix X such as Equation (18):

$$\begin{aligned} &\min_X c\sum_{i=1}^{N}\sum_{j=1}^{N} 1_{[x_{ij}>0]} \quad st. \\ &\sum_{j=1}^{N} x_{ij} = A_i \forall i = 1, 2, \ldots, N \\ &\sum_{i=1}^{N} x_{ij} = L_j \forall j = 1, 2, \ldots, N \\ &x_{ij} \geq 0 \forall i, j \end{aligned} \tag{18}$$

The integer 1 is 1 only when there is a loan relationship between bank i and bank j, otherwise it is 0. Because in the real interbank system, there is a service charge for lending between two banks, so the significance of the objective equation is to reduce the cost as much as possible, so as to make the interbank network "less dense." The constraint equation is to meet the constraints of the total amount of interbank lending and interbank borrowing of each bank. After estimating the interbank network, we add this network to the time step evolution process to obtain the dynamically evolving interbank market risk exposure, see section The Dynamic Evolution of Interbank Market Network.

Default Algorithms Based on the Dynamic Evolution of Bank Balance Sheet Structure

Combining with the dynamically evolving balance sheet in the evolution process of time step t in **Figure 1A**, the owner's equity $e_i(t)$ of bank i at t time step can be calculated by Equation (19).

$$e_i(t) = \left(V_i(t) - Loss_i(t)\right) + A_i(t) - D_i(t) - L_i(t) \tag{19}$$

In Equation (19), we use the method proposed by Fan et al. [29] to estimate the asset sequence of bank i $V_i(t)$ and liability sequence $D_i(t)$; the borrowing asset sequence $A_i(t)$ and liability sequence $L_i(t)$ will be discussed in section The Dynamic Evolution of Interbank Market Network; $Loss_i(t)$ indicates the external asset loss of bank i caused by the impact of the credit risk exposure, which can be found in Equation (15).

If Equation (20) is satisfied, bank i will suffer the basic failure under the impact of the credit risk exposure of the real industry. We mark the basic failure state variable as $TagFR_i(t)$=1.

$$e_i(t) = \left(V_i(t) - Loss_i(t)\right) + A_i(t) - D_i(t) - L_i(t) < 0 \tag{20}$$

If Equation (20) does not hold, we first standardize the interbank market risk exposure $X^*(t)$ to obtain a new matrix $\prod(t)$, as shown in Equation (21). $L_i(t) = \sum_{j=1}^{N} x_{ji}(t)$ represents the borrowing liabilities of bank i. Then, the clearing vector mechanism proposed by Eisenberg and Noe [30] is cited to clear all the banks through the clearing payment vector $C^*(t)$, as shown in Equation (22).

$$\prod_{ij}(t) = \begin{cases} \frac{x_{ij}(t)}{L_i(t)} & L_i(t) > 0 \\ 0 & L_i(t) = 0 \end{cases} \tag{21}$$

$$C^*(t) = \begin{cases} L_i(t) & if \sum_{j=1}^{N}\prod_{ji} C_j^*(t) + e_i^*(t) \geq L_i(t) \\ \sum_{j=1}^{N}\prod_{ji} C_j^*(t) + e_i^*(t) & if 0 \leq \sum_{j=1}^{N}\prod_{ji} C_j^*(t) + e_i^*(t) < L_i(t) \\ 0 & if \sum_{j=1}^{N}\prod_{ji} C_j^*(t) + e_i^*(t) < 0 \end{cases} \tag{22}$$

The following Equation (23) is used to calculate the banks with contagious failure. Among them, $\sum_{j=1}^{N}\prod_{ji} C_j^*(t)$ refers to the interbank lending assets of bank i; $L_i(t)$ refers to the interbank borrowing liabilities of bank i; $\sum_{j=1}^{N}\prod_{ji} C_j^*(t) - L_i(t)$ refers to the owner's equity of interbank lending; $e_i^*(t) = \left(V_i(t) - Loss_i(t)\right) - D_i(t)$ refers to the owner's equity outside the interbank lending market.

$$\sum_{j=1}^{N}\prod_{ji} C_j^*(t) + e_i^*(t) - L_i(t) < 0 \tag{23}$$

Equation (23) indicates that the total owner's equity of bank i is <0 due to the impact of the interbank market risk exposure and Equation (20) does not hold at this time. Because of the failure of other banks, bank i can not recover all the interbank loans. Finally, bank i suffers contagious failure. We mark the contagious failure state variable as $TagCR_i(t)$=1.

The next step is to clean up all the failed banks in the banking system, then the banking network changes accordingly. We update the lending matrix $X^*(t+1)$, interbank lending assets $A_i(t+1)$, and interbank lending liabilities $L_i(t+1)$ at the next time step in section The Dynamic Evolution of Interbank Market Network. We cycle through the above operations at each time step until the end of the time step T.

The Dynamic Evolution of Interbank Market Network

After calculating the clearing payment vector $C^*(t)$ under the time step t, next we update the interbank borrowing assets $A_i(t+1)$ and borrowing liabilities $L_j(t+1)$. Combined with the minimum density method in section Default Algorithms Based on the Dynamic Evolution of Bank Balance Sheet Structure, we update the total interbank borrowing assets and total liabilities constraints to get a new borrowing matrix $X^*(t+1)$. We can see **Figure 2** for the specific evolution flow chart.

THE CALCULATION OF BANKING SYSTEMIC RISK UNDER DOUBLE RISK EXPOSURES

This paper defines a basic risk when the total assets of bank i are less than the total liabilities due to the loss of external assets under the impact of credit risk exposure of the real industry. When the total assets of bank j are less than the total liabilities due to the failure of bank i and the inability to recover all the interbank loans (the impact of the interbank market risk exposure), a contagious risk occurs. In the system, under the impact of double risk exposures, a bank's total assets are less than its total liabilities (including basic risk and contagious risk), the bank defaults. In this paper, the basic risk is represented by the basic failure probability $FR(t)$, the contagious risk is represented by the contagious failure probability $CR(t)$, and the systemic risk is represented by the system failure probability $SR(t)$, $SR(t) = FR(t) + CR(t)$.

$$FR(t) = \frac{\sum_{i=1}^{N} TagFR_i(t)}{N} \tag{24}$$

$$CR(t) = \frac{\sum_{j=1}^{N} TagCR_j(t)}{N} \tag{25}$$

$$SR(t) = \frac{\sum_{i=1}^{N} TagFR_i(t) + \sum_{j=1}^{N} TagCR_j(t)}{N} \tag{26}$$

DATA

This paper uses the balance sheet of Chinese banks in 2016 and the China Securities Regulatory Commission's 2012 version of industry classification Listed Company information in the CSMAR economic and financial database, as well as the data of the 2016 annual report publicly disclosed on the official website. China has a large number of banks, which are mainly divided into six types according to the definition of the CBRC annual report: policy banks, large commercial banks, joint-stock banks, urban commercial banks, rural commercial banks, and foreign banks. By analyzing the collected data, we have found that the same type of banks has the same order of magnitude of assets and liabilities. Moreover, assets and liabilities of policy banks and large commercial banks belong to the same order of magnitude. Consider that the nature and function of banks are irrelevant factors in this study, so policy banks and large commercial banks can be simplified into one type and named large banks. Finally, this paper has divided Chinese banks into five types: large banks, joint-stock banks, urban commercial banks, rural commercial banks, and foreign banks. Due to the limitation of data availability, this paper finally selects 205 banks, which are large banks (8 Banks), joint-stock banks (13 banks), urban commercial banks (102 banks), rural commercial banks (63 banks), and foreign banks (19 banks). It selects individual stock transaction amount in 244 days and balance sheet of 3,017 listed companies in 18 industries that are classified according to the China securities regulatory commission 2012 edition of industry classification guidelines, and then selects loan industry concentration data of 104 banks. For convenience of presentation, the chart in this article indicates 205 banks with numbers 1–205 (see **Appendix Table 1** for bank names) and 18 industries with numbers 1–18 (see **Appendix Table 2** for industry names). In addition, the descriptive statistics of data can be seen in **Appendix Tables 3–6**.

RESULTS

The Assessment of Credit Risk of the Real Industry

The Evolution Curve of Credit Risk of the Real Industry

In this paper, the stress test is carried out for the listed companies in 18 real industries to make each Listed Company lose 30% of its assets in the same real industry, so that some listed companies have credit default and the dynamic evolution chart of credit risk in each industry is obtained (see **Figure 1B**). According to the evolution characteristics of default probabilities in various industries, credit risks in 18 industries can be divided into 4 types that are stability after decline (see **Figure 3A**), increase in fluctuation (see **Figure 3B**), decrease in fluctuation (see **Figure 3C**), fluctuation (see **Figure 3D**).

The evolution of credit risk is characterized by the first decline and then stability in eight industries: finance industry, education industry, water conservancy, environment and public facilities management industry, leasing and business services industry, agriculture, forestry, animal husbandry and fishery industry, electricity, heat, gas and water production and supply industry, comprehensive industry, and accommodation and catering industry. The credit risk of the eight industries decreased within 100 time steps and then stabilized, among which the credit risk of the financial industry stabilized at 0.2, the credit risk of the accommodation and catering industry stabilized at 0.1, and the credit risk of the other six industries stabilized at 0.

The evolution of credit risk is characterized by increase in fluctuation in five industries: scientific research and technology services industry, mining industry, information transmission software and information technology services industry, manufacturing industry, culture, and sports and entertainment industry. Among them, the risk fluctuation range of scientific research and technical services industry is the largest, while the risk fluctuation range of culture, sports and entertainment industry is the smallest. The credit risk

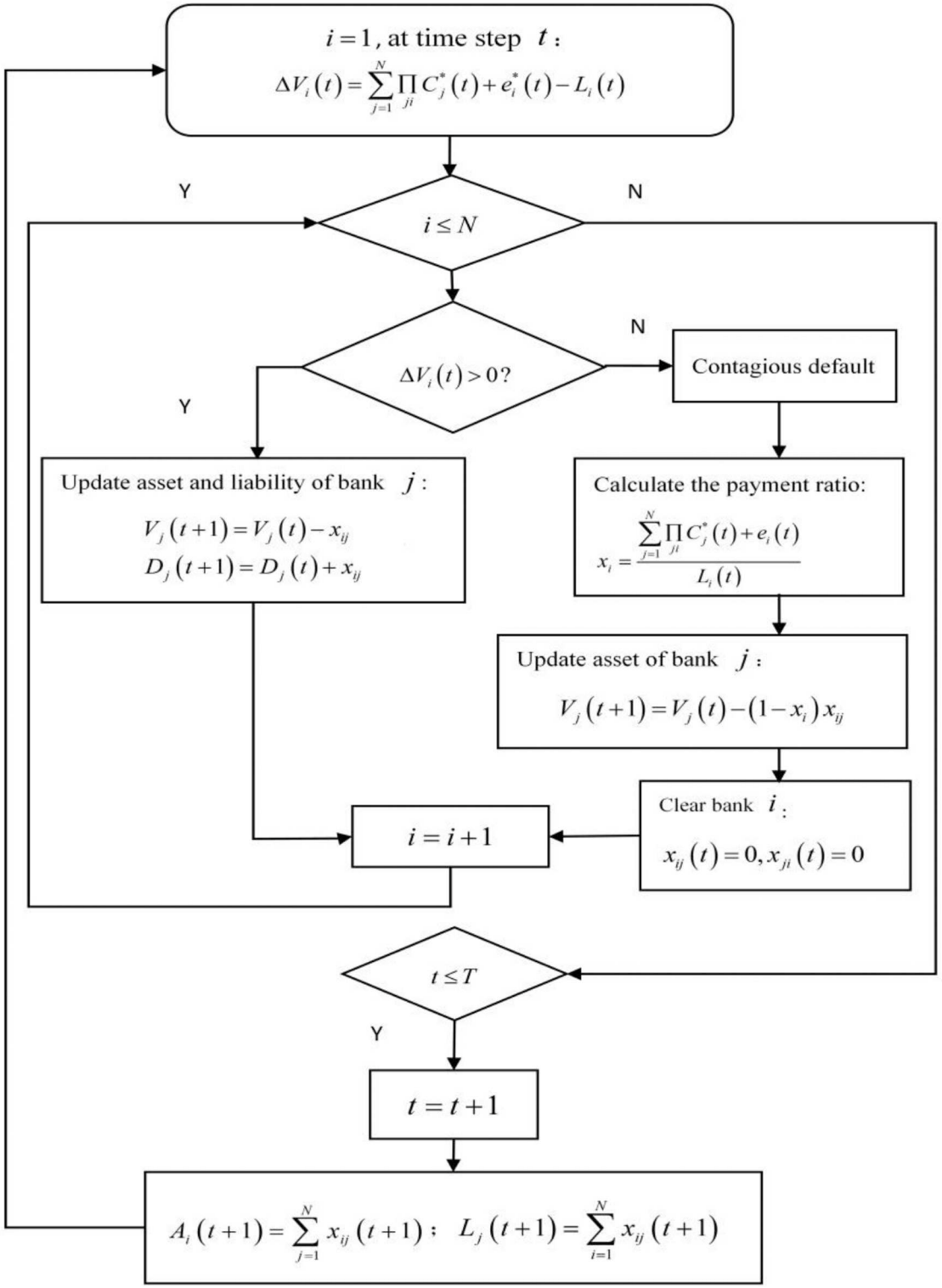


FIGURE 2 | Flow chart of dynamic evolution of interbank market network.

of information transmission software and information technology service industry and manufacturing industry showed an evolutionary trend of increasing continuously, while the credit risk of scientific research and technology service industry, mining industry, and culture, sports and entertainment industry showed an evolutionary trend of stabilizing.

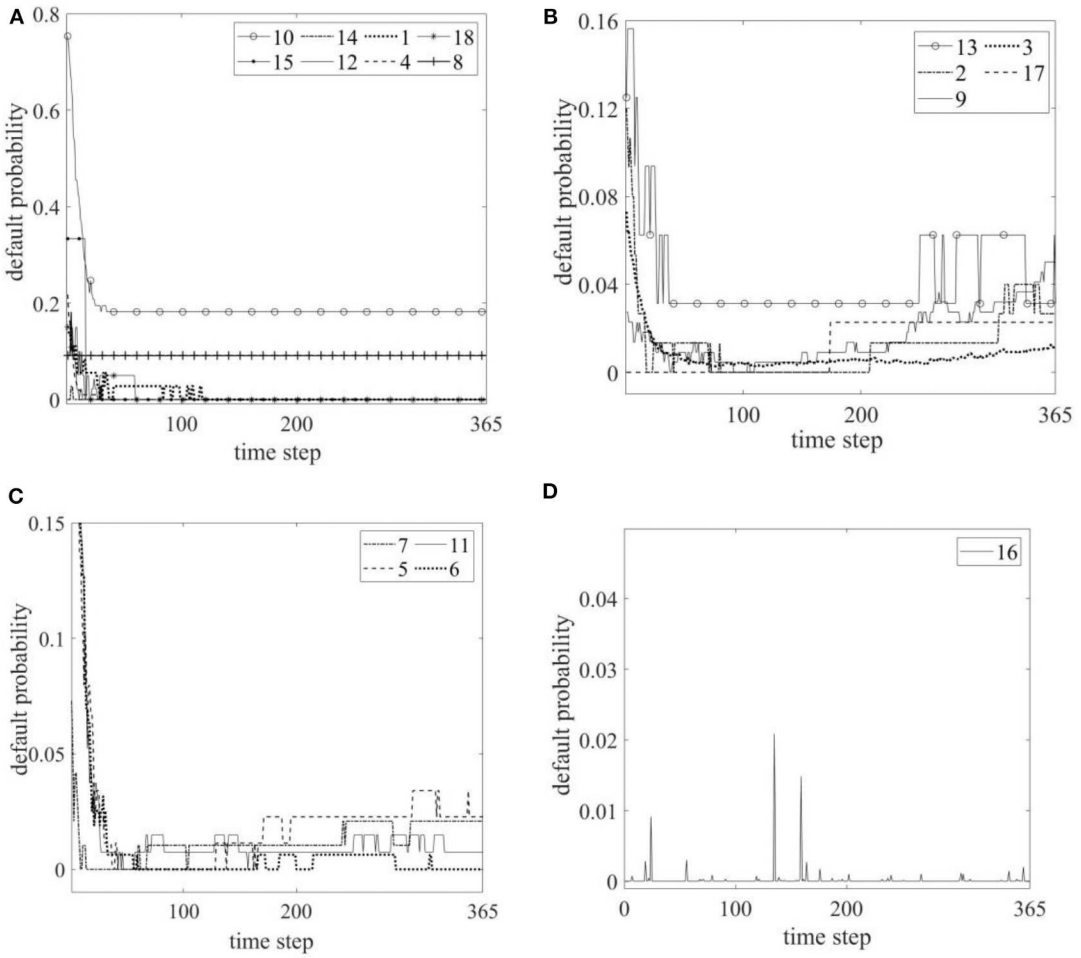


FIGURE 3 | Credit default probability curve of the real industry. **(A)** Curve of stability after decline. **(B)** Curve of increasing in fluctuation. **(C)** Curve of decreasing in fluctuation. **(D)** Curve of fluctuation.

The evolution of credit risk is characterized by decrease in fluctuation in four industries: transportation, warehousing and postal industry, construction industry, real estate industry, and wholesale and retail industry. The risk characteristics of these four industries are similar. The failure probability of the industry falls to a certain value within the first 100 time steps, and then fluctuates slightly within the probability range of 0 to 0.05.

The evolution of credit risk in the health and social work industry is characterized by fluctuation, which fluctuates around 0.

Stability of Credit Risk of the Real Industry

From **Figure 3** in The Evolution Curve of Credit Risk of the Real Industry, we can see that the credit risk of 18 real industries has different volatility, indicating that the credit risk of these industries is unstable. Therefore, this section further evaluates the stability of credit risk of the real industry by calculating the variance of credit risk. **Table 1** is the result of ranking the variance of credit risk of 18 real industries in descending order. It can be seen that the top three real industries with the largest variance are financial industry, education industry, and real estate industry, that is to say, the credit risk of these three real industries is the most unstable. The real industry with the least variance is health and social work industry, which has the most stable credit risk.

Mean Credit Risk of the Real Industry

From **Figure 3** in section Default Algorithms Based on the Dynamic Evolution of Bank Balance Sheet Structure and **Table 1** in section Stability of Credit Risk of the Real Industry, we can see that the dynamic evolution credit risk of real industries presents a fluctuating state, most industries are unstable. It is impossible to accurately assess the value of credit risk of different real industries. Therefore, the next step of this paper is to carry out Monte Carlo simulation in 1000 time steps on the basis of pressure test. In order to eliminate the simulation error as much as possible, this paper adopts the mean value for comparative analysis. We use $S_g(t)$ in Equation (12) and $AVG\left(S_g(t)\right)$ is taken as the mean value of the number of enterprise failures in the real industry during 1000 simulations. If the industry credit risk is expressed as P_K in 1,000 simulations, the calculation formula is: $P_K = \frac{\sum AVG(S_g(t))}{T \times n_k}$, $k = 1, 2, \ldots, 18$ indicates 18 industries, T is the dynamic evolution time step 1,000, n_k is number of listed companies in the k industry.

Figure 4 is the result of Monte Carlo simulation of mean credit risk of various industries under 1,000 times. From **Figure 4**, it

TABLE 1 | Statistical table of variance of credit risk of each real industry.

Industry number	Variance of credit risk
10	0.043942
15	0.041845
11	0.028669
13	0.028387
9	0.025056
5	0.023641
2	0.021085
6	0.020388
18	0.016346
1	0.014896
4	0.011940
3	0.011907
12	0.009350
17	0.008601
7	0.006394
8	0.002875
16	0.001452
14	0.001242

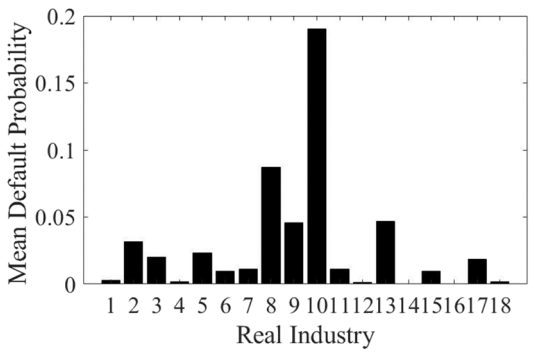


FIGURE 4 | The mean credit risk of the credit industry under the simulation of 1,000 times.

can be analyzed that the top four real industries with the highest mean credit risk are: finance, accommodation and catering, scientific research and technical services, and information transmission, and software and information technology services. It is suggested to strengthen the credit risk management of enterprises in the fields of finance, science and technology, and national life. Two real industries with the lowest mean credit risk are: water conservancy environment and public facilities management, and health and social work, while the other 12 real industries have medium mean credit risk.

Combined with the industry stability analysis results in section Stability of Credit Risk of the Real Industry, it can be found that the mean credit risk of the financial industry is the highest and the most unstable. The mean credit risk of accommodation and catering industry is high but stable. The mean credit risk and stability of scientific research and technical services, information transmission software and information technology services are relatively high. However, two real industries: water conservancy environment and public facilities management, health and social work, have the lowest mean credit risk and the highest stability.

Default State Analysis of Listed Companies in the Real Industry

The credit risk of the real industry is directly proportional to the number of default listed companies in the industry. In the process of dynamic evolution, the listed companies with higher probability of bankruptcy are more unstable. It can be considered that the enterprises with higher probability of bankruptcy contribute more to the credit risk of the industry.

Table 2 shows the top five listed companies in the default probability of 18 industries after stress test. These companies have the largest contribution to the credit risk of their industries, and the default status of listed companies in different industries is significantly different. A total of 17 listed companies in 18 real industries have a default probability of 1 during 1,000 times of dynamic evolution, which belongs to the state of extremely easy default. Among them, 14 listed companies belong to the financial industry. Combined with the mean credit risk results of the real industry in section Mean Credit Risk of the Real Industry, it can be seen that 20% of listed companies in the financial industry are very vulnerable to bankruptcy, and these companies contribute the most to credit risk of financial industry.

There are no more than three listed companies in five real industries: accommodation and catering industry, water conservancy environment and public facilities management industry, education industry, culture, sports and entertainment industry, and comprehensive industry. In the dynamic evolution process, only two listed companies have the default probability of more than 0.8, and other listed companies have the default probability of <0.1. There are 11 listed companies in the accommodation and catering industry, among which one is very vulnerable to bankruptcy and contributes the most to the credit risk of the accommodation and catering industry.

The top five listed companies in manufacturing, information transmission software and information technology services have a high default probability, which are all over 0.7. However, due to the large number of listed companies in the manufacturing industry, the mean credit risk is not high. The number of listed companies in the information transmission software and information technology services industry is close to one ninth of the manufacturing industry, so the mean credit risk is higher.

The Assessment of Systemic Risk Under Double Risk Exposures

Analysis of the Network Structure With Double Risk Exposures

Credit network structure between banks and the real industry

According to the model of credit network between banks and real industry in section Credit Network Between Banks and Real Industry and the data in section Data, the credit network structure diagrams between 5 types of banks and 18 real industries are constructed, as shown in **Figure 5**. Among them, 18 white circles marked with numbers represent the 18 real industries in which banks provide loans. The circle size reflects the number of listed companies in the real industry. The larger the circle, the more listed companies in the real industry. For example, the number of listed companies in the manufacturing industry is the largest and the circle is the largest. Five gray circles marked with letters indicate five types of banks, which are A (large bank), B (joint-stock bank), C (urban commercial bank), D (rural commercial bank), and E (foreign bank). The circle size indicates the number of banks under this type. The larger the circle, the more

TABLE 2 | Listed companies with the top five bankruptcy frequency in the real industry.

1	Bankrupt enterprise	A26	A29	A6	A33	A8
	Default probability	0.094	0.03	0.01	0.007	0.003
2	Bankrupt enterprise	B18	B9	B26	B51	B52
	Default probability	0.793	0.679	0.625	0.56	0.017
3	Bankrupt enterprise	C1572	C138	C1799	C769	C783
	Default probability	1	0.995	0.885	0.881	0.826
4	Bankrupt enterprise	D22	D25	D68	D64, D48	D20
	Default probability	0.026	0.012	0.009	0.007	0.006
5	Bankrupt enterprise	E64	E53	E44	E17	E72
	Default probability	0.861	0.83	0.274	0.038	0.027
6	Bankrupt enterprise	F63	F74	F75	F118	F60
	Default probability	0.097	0.056	0.031	0.03	0.026
7	Bankrupt enterprise	G7	G27	G84, G39, G45	G61	G48
	Default probability	0.993	0.332	0.006	0.004	0.003
8	Bankrupt enterprise	H6	H1			
	Default probability	1	0.001			
9	Bankrupt enterprise	I182	I17	I108	I3	I84
	Default probability	0.977	0.762	0.756	0.754	0.746
10	Bankrupt enterprise	J1, J24, J26, J28, J54, J59, J60, J61, J64, J67, J70, J72, J73, J75	J3	J5, J22	J45	J25
	Default probability	1	0.352	0.022	0.021	0.018
11	Bankrupt enterprise	K3	K110	K38	K33	K53
	Default probability	0.514	0.465	0.215	0.102	0.029
12	Bankrupt enterprise	L15	L17	L22, L25	L4	L36
	Default probability	0.011	0.01	0.008	0.007	0.002
13	Bankrupt enterprise	M1	M12	M27	M2	M20
	Default probability	1	0.382	0.343	0.032	0.03
14	Bankrupt enterprise	N12				
	Default probability	0.002				
15	Bankrupt enterprise	P1				
	Default probability	0.016				
16	Bankrupt enterprise	O14				
	Default probability	0.002				
17	Bankrupt enterprise	R44				
	Default probability	0.827				
18	Bankrupt enterprise	S2	S6	S12		
	Default probability	0.048	0.009	0.008		

banks under this type. The connection between the five types of banks and the real industry means that banks provide loans to the real industry. This paper assumes that each type of banks provides loans to 18 real industries. The line thickness indicates the proportion of loans. Industries with large proportion of loans are concentrated in the center of the map, while industries with small proportion of loans are scattered around.

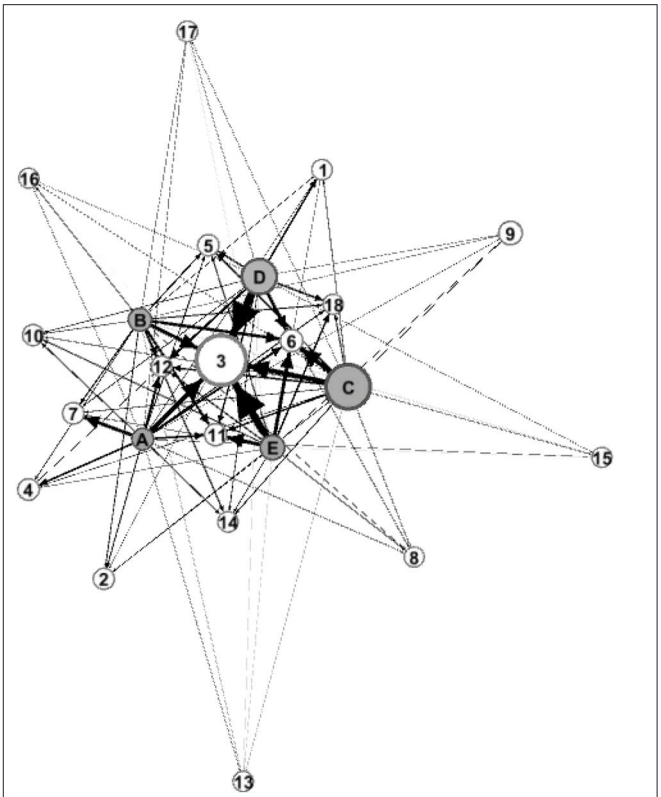


FIGURE 5 | Credit network between banks and real industry.

TABLE 3 | Top three real industries with the most loans from banks (the figure in the table is the industry number).

The type of bank	Top 1 industry	Top 2 industry	Top 3 industry
Large bank	3	7	12
Joint stock bank	3	6	11
Urban commercial bank	3	6	11
Rural commercial bank	3	6	12
Foreign bank	3	11	6

As can be seen from the thickness of the connection line in **Figure 5**, the top four real industries with the highest proportion of loans provided by the five types of banks to the real industry are 3 (manufacturing), 6 (wholesale and retail), 11 (real estate), and 12 (leasing and business services), while the lowest industry is 13 (scientific research and technical services). There are differences in the concentration of loans by industries among different types of banks. **Table 3** lists the industries with the highest concentration of loans among the five types of banks.

Dynamic evolution of interbank market network structure

We use the minimum density method in section Estimation of the Interbank Market Risk Exposure, interbank market network in section The Dynamic Evolution of Interbank Market Network, and data in section Data to obtain the dynamic evolution of interbank market network structure diagram, as shown in **Figures 6A,B**, which

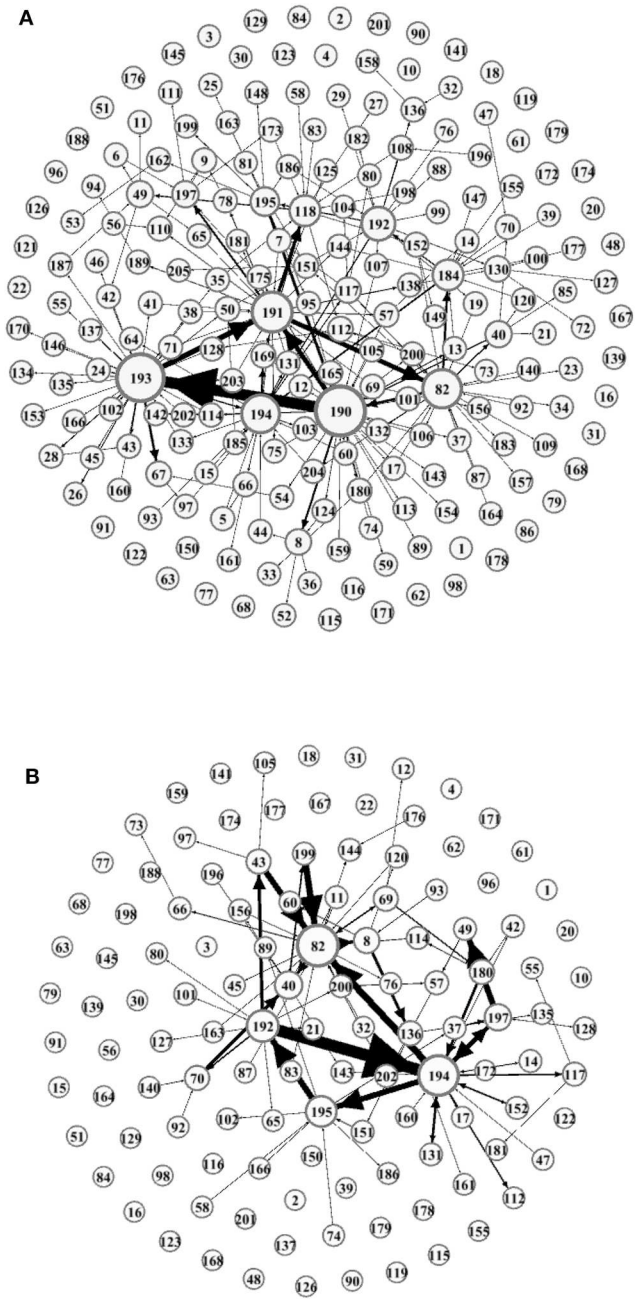


FIGURE 6 | Network structure of interbank market. **(A)** Represents Chinese interbank market network structure at time step t_1. **(B)** Represents Chinese interbank market network structure at time step t_{150}.

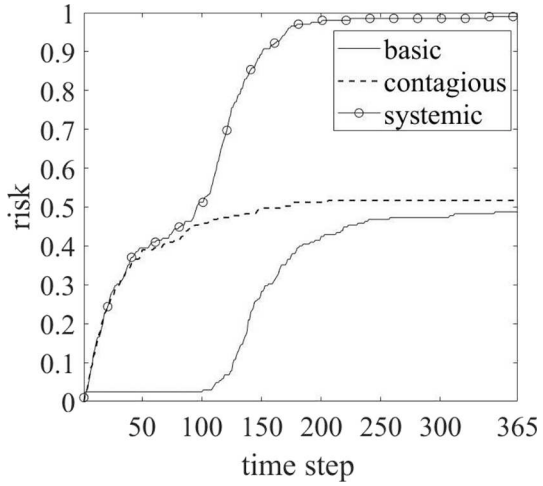


FIGURE 7 | Dynamic evolution curve of basic default, contagious, and systemic risk of Chinese banking system.

corresponds to **Figure 1D** in section Model. **Figure 6A** shows the network structure diagram of the interbank market at time step t_1. In the **Figure 6A**, network nodes marked with numbers represent 205 banks, the size of the node indicates the connection degree between banks. The larger the connection degree is, the larger the node is. The directed line between nodes indicates the interbank lending relationship, and the line thickness indicates the lending amount. From **Figure 6A**, it can be found that there are 8 nodes with relatively large connection degree at time step t_1, which are 190 (industrial and commercial bank of China), 193 (agricultural bank of China), 191 (China construction bank), 194 (bank of China), 82 (bank of communications), 118 (Shanghai Pudong Development Bank), 192 (China Minsheng bank), and 184 (China merchants bank). The connection between these eight nodes is relatively thick, of which the connection from 190 to 193 is the thickest, which indicates that Industrial and Commercial Bank of China lends the largest amount of funds to Agricultural Bank of China. In **Figure 6B**, the network structure of the interbank market at time step t_{150} has changed compared with time step t_1. The number of network nodes has been reduced to 120 (due to the removal of some default banks from the network), the node size and the connections between nodes have also changed (the interbank lending relationship has changed).

Dynamic Evolution of Systemic Risk in Chinese Banking System

Basic Default, Contagious, and Systemic Risk of Chinese Banks

Figure 7 represents the dynamic evolution law of basic default, contagious, and systemic risk under double risk exposures. It can be seen from the **Figure 7** that during the evolution of 365 time steps, the systemic risk has the characteristic of stabilizing after increasing and reaching a stable state at 250 steps. The degree of evolution of contagious risk is higher than that of basic default and the dynamic evolution characteristics of basic default and contagious risk are different. Contagious risk tends to stabilize after increasing and reaching a stable state at 200 steps. The probability of basic default is low and stable in the first 100 time steps, then jumping growth in 101 to 250 time steps, and finally slowing down and stabilizing in 250 steps.

Basic Risk, Contagious Risk, and Systemic Risk Characteristics of Different Types of Banks

On the basis of section Basic Default, Contagious, and Systemic Risk of Chinese Banks, this paper further analyzes whether there are differences in basic risk, contagious risk, and systemic risk characteristics of

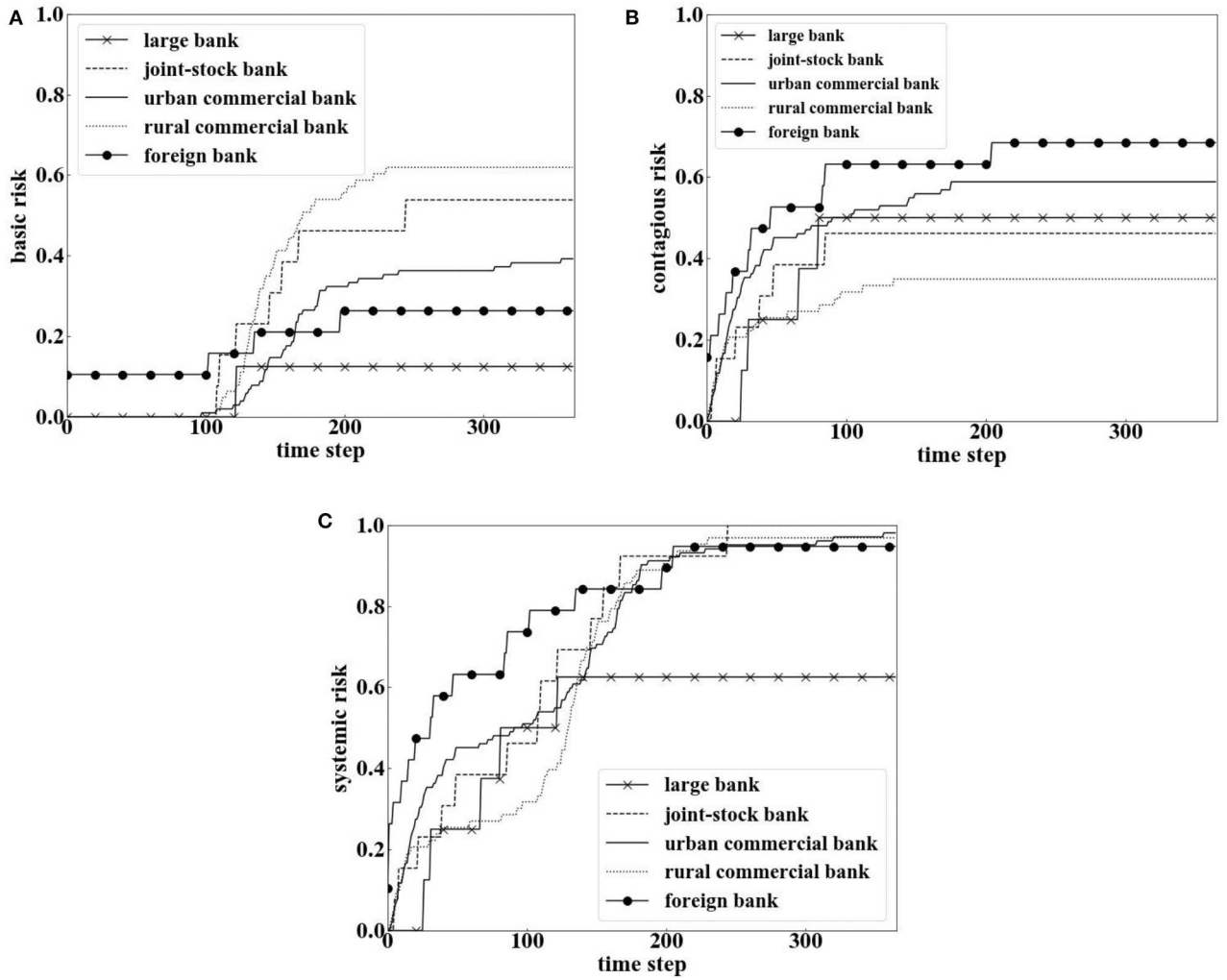


FIGURE 8 | Basic risk, contagious risk, and systemic risk evolution curve of different types of Chinese banks.

different types of banks under the impact of double risk exposures. **Figure 8A** represents the basic risk curve of five types of banks, **Figure 8B** represents the contagious risk curve of five types of banks, and **Figure 8C** represents the systemic risk curve of five types of banks. On the whole, the dynamic evolution trend of basic risk, contagious risk, and systemic risk in the five types of banks are stable after growth. However, the dynamic characteristics of basic risk, contagious risk, and systemic risk in the five types of banks are different.

In **Figure 8A**, we can see five types of banks reach a stable state with different time and basic risk finally. The order of basic risk from high to low is: rural commercial bank, joint-stock bank, urban commercial bank, foreign bank, and large bank. The order of time to reach stability from fast to slow is: large bank, foreign bank, rural commercial bank, joint-stock bank, and urban commercial bank.

In **Figure 8B**, we also can see five types of banks reach a stable state with different time and contagious risk finally. The order of contagious risk from high to low is: foreign bank, urban commercial bank, large bank, joint-stock bank, and rural commercial bank. The order of time to reach stability from fast to slow is: large bank, joint-stock bank, rural commercial bank, urban commercial bank, and foreign bank.

In **Figure 8C**, the systemic risk of large bank reaches a stable state at the earliest with the lowest risk, while the other four types of banks reach a stable state later with high systemic risk, of which the mean systemic risk of foreign bank is the highest. This reflects that large banks are stronger than the other four types of banks in resisting the impact of double risk exposures, while foreign bank are the weakest.

DISCUSSION AND CONCLUSIONS

This paper first used Geometric Brownian Motion in physics, Monte Carlo simulation, and real data of the real industry to estimate the dynamic evolution credit risk exposure of the real industry. Then, it used the minimum density method and Chinese real banking data to estimate the more realistic dynamic interbank market risk exposure. So far, we have successfully built up the double risk exposure. Finally, based on the double risk exposure, we use the data of assets, liabilities, and industry loans of 205 Chinese banks, as well as the data of 3017 listed companies' stock market to study the dynamic evolution law of the systemic risk of Chinese banking system. The important conclusions are as follows:

First, our dynamic evolution model highlighted the impact of credit risk exposure and enables to assess which real industries are more/less risky. In addition, we also assessed listed companies with top five bankruptcy frequency of the real industry. Therefore, this information can assist financial regulators to weigh credit loans by credit risk of the real industry/Listed Company. Financial industry, accommodation and catering industry, scientific research and technology service industry, and information transmission software, and technology service industry have higher credit risk, while financial industry, scientific research and technology service industry, and information transmission software and technology service industry have lower stability. It is suggested that financial regulators should pay special attention to the credit risk management of these four real industries when serving the development of the real economy. It is expected that this methodology would engender a reduction in credit loans to risky real industry/Listed Company, leading to a decrease in the systemic risk of Chinese banking system.

Second, the uniqueness of the dataset of Chinese banks and real industries not only permit us to construct credit network and interbank network, but also enables us to broaden other contagion channels apart from classical interbank market. Under the impact of double risk exposures, the systemic risk gradually rises and then reaches a stable state in the evolution process; the contagious risk initially grows faster and then stabilizes; however, the basic default is very small at first, then slowly rises, and finally stabilizes. Further results showed that the dynamic characteristics of the systemic risk in different types of banks are shown as growth first and then stabilization, among which the ability of large banks to resist the impact of double risk exposures is stronger than the other four types of banks, while the ability of foreign bank to resist is the weakest.

The present paper focused on the Chinese system market, however, in the future, the methods proposed can be extended to other global markets or a mature market in other countries. In addition, due to the use of different system risk measurement methods, the results will be different. In the future, we will use other methods to compare with the proposed method, so as to compare the robustness of the proposed method in the present paper.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

HF and WS performed analyses, discussed the results, and wrote the manuscript.

FUNDING

We are grateful for the support from national natural science foundation of China (71971054) and natural science foundation of Shanghai (19ZR1402100).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00199/full#supplementary-material

REFERENCES

1. Allen F, Gale D. Financial contagion. *J Polit Econ*. (2000) **108:**1–33. doi: 10.1086/262109
2. Freixas X, Parigi B, Rochet JC. Systemic risk, interbank relations and liquidity provision by the central bank. *J Money Credit Bank*. (2000) **32:**611–38. doi: 10.2307/2601198
3. Martínez-Jaramillo S, Pérez OP, Embriz FA, Dey FLG. Systemic risk, financial contagion and financial fragility. *J Econ Dyn Control*. (2010) **34:**2358–74. doi: 10.1016/j.jedc.2010.06.004
4. Mistrulli PE. Assessing financial contagion in the interbank market: maximum entropy versus observed interbank lending patterns. *J Bank Finan*. (2011) **35:**1114–27. doi: 10.1016/j.jbankfin.2010.09.018
5. Ladley D. Contagion and risk-sharing on the interbank market. *J Econ Dyn Control*. (2013) **37:**1384–400. doi: 10.1016/j.jedc.2013.03.009
6. Sui C, Chi GT, Wang ZY. Network structure and banking systemic risk. *J Manage Sci*. (2014) **17:**57–70. doi: 10.3969/j.issn.1007-9807.2015.12.002
7. Bao Q, Sun YX. Financial structure and financial risk contagion from the perspective of network. *Syst Eng Theor Prac*. (2014) **34:**2202–11. doi: 10.12011/1000-6788(2014)9-2202
8. Adrian T, Brunnermeier MK. *CoVaR*. Technical Report. Department of Economics Princeton University (2011). doi: 10.3386/w17454
9. Huang X, Zhou H, Zhu, H. Systemic risk contributions. *J Finan Serv Res*. (2011) **42:**55–83. doi: 10.1007/s10693-011-0117-8
10. Billio M, Getmansky M, Lo A, Pellizzon, L. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J Finan Econ*. (2012) **104:**535–59. doi: 10.1016/j.jfineco.2011.12.010
11. Ang A, Longstaff FA. Systemic sovereign credit risk: lessons from the U.S. and Europe. *J Monet Econ*. (2013) **60:**493–510. doi: 10.1016/j.jmoneco.2013.04.009
12. Diebold F, Yilmaz K. On the network topology of variance decompositions: measuring the connectedness of financial firms. *J Econometr*. (2014) **182:**119–34. doi: 10.1016/j.jeconom.2014.04.012
13. Li Z, Liu Q, Liang Q. Research on China's systemic risk prevention based on economic and financial correlation network. *Stat Res*. (2019) **36:**25–39. doi: 10.19343/j.cnki.11-1302/c.2019.02.003
14. Zhai YH. A study on risk spillovers between the real industry and the banking sector from the perspective of systemic risk management. *Int Finan Res*. (2019) **12:**74–84. doi: 10.13891/j.cnki.mer.2019.12.009
15. Zhu B, Ma YT. Industry characteristics, monetary policy and systemic risk-analysis based on the correlation network of "economy and finance". *Int Finan Res*. (2018) **4:**22–32. doi: 10.16475/j.cnki.1006-1029.2019.12.008
16. Sun L, Zhang Y L, Li JY. The real estate industry spillover effects on the systemic risk of financial institutions research - an empirical analysis of the point of view based on industry and enterprise. *Southern Econ*. (2020) **2:**1–16. doi: 10.16475/j.cnki.1006-1029.2018.04.003
17. Lu JG, Wang C. Study on bank risk contagion based on interbank borrowing and joint lending. *Mod Econ*. (2019) **12:**41–8. doi: 10.19592/j.cnki.scje.370208
18. Wu YY, Hu RD, Yuan ZZ. Network risk contagion and government rescue in common asset holding. *Finan Forum*. (2017) **8:**37–47. doi: 10.16529/j.cnki.11-4613/f.2017.08.006
19. Zhai YH, Bian QM. Bank lending and systemic risk prevention and control—research based on complex network model. *Wuhan Finan*. (2019) **11:**12–22. doi: 10.3969/j.issn.1009-3540.2019.11.002
20. Caccioli F, Farmer JD, Foti N, Rockmore D. Overlapping portfolios, contagion, and financial stability. *J Econ Dyn Control*. (2015) **51:**50–63. doi: 10.1016/j.jedc.2014.09.041
21. Glasserman P, Young HP. How likely is contagion in financial networks? *J Bank Finan*. (2015) **50:**383–99. doi: 10.1016/j.jbankfin.2014.02.006

22. Chiu W-C, Peña JI, Wang C-W. Industry characteristics and financial risk contagion. *J Bank Finan*. (2015) **50:**411–27. doi: 10.1016/j.jbankfin.2014.04.003
23. Wang H, Li S. A study on the measurement of systemic risk contagion in China's real estate industry and banking industry from an internal perspective. *Int Finan Res*. (2015) **9:**76–85. doi: 10.16475/j.cnki.1006-1029.2015.09.008
24. Li SZ, Ma JL, Zhu SS. The impact of real estate credit on the systemic risk of China's banking system – a simulated empirical study based on the internal lending network of the banking system. *J Zhongsan Univ*. (2019) **59:**186–96. doi: 10.13471/j.cnki.jsysusse.2019.03.020
25. Lehar A. Measuring systemic risk: a risk management approach. *J Bank Finan*. (2005) **29:**2577–603. doi: 10.1016/j.jbankfin.2004.09.007
26. Black F, Scholes MS. The pricing of options and corporate liabilities. *J Polit Econ*. (1973) **81:**637–54. doi: 10.1086/260062
27. Duan J-C. Correction: maximum likelihood estimation using price data of the derivative contract. *Math Finan*. (2000) **10:**461–2. doi: 10.1111/1467-9965.00105
28. Anand K, Craig B, Peter GV. Filling in the blanks: network structure and interbank contagion. *Quant Finan*. (2015) 15:625–36. doi: 10.1080/14697688.2014.968195
29. Fan H, Amalia AA, Gao QQ. The assessment of systemic risk in the Kenyan banking sector. *Complexity*. (2018) **2018**:1–15. doi: 10.1155/2018/8767836
30. Eisenberg L, Noe TH. Systemic risk in financial systems. *Manage Sci*. (2001) **47:**236–49. doi: 10.1287/mnsc.47.2.236.9835

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Fan and Sheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 12 June 2020
doi: 10.3389/fphy.2020.00163

Stock Network Stability After Crashes Based on Entropy Method

Ming-Yuan Yang[1,2], *Fei Ren*[1,3,4]* and *Sai-Ping Li*[5]*

[1] School of Business, East China University of Science and Technology, Shanghai, China, [2] China Academy of Financial Research, Zhejiang University of Finance and Economics, Hangzhou, China, [3] School of Science, East China University of Science and Technology, Shanghai, China, [4] Research Center for Econophysics, East China University of Science and Technology, Shanghai, China, [5] Institute of Physics, Academia Sinica, Taipei, Taiwan

This paper studies the stability of stock networks after crashes based on the entropy method. By measuring network stability using the entropy calculated with the degree distribution, we find that the entropy of a stock network is close to that of the Erdös-Rényi and Watts-Strogatz networks. We further introduce government rescue policies as a natural experiment and use the entropy measurement to study the influence of rescue policies after crashes on the network stability, finding that rescue policies only have short-term effects. Analysis of the relation between stock degrees and government purchasing behavior further confirms the effects of rescue policies on stock network stability.

Keywords: stock network, crashes, rescue policies, network stability, entropy, econophysics

OPEN ACCESS

Edited by:
Gabjin Oh,
Chosun University, South Korea

Reviewed by:
Michele Bellingeri,
University of Parma, Italy
Abolfazl Ramezanpour,
Shiraz University, Iran

***Correspondence:**
Fei Ren
fren@ecust.edu.cn
Sai-Ping Li
spli@phys.sinica.edu.tw

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *11 March 2020*
Accepted: *20 April 2020*
Published: *12 June 2020*

Citation:
Yang M-Y, Ren F and Li S-P (2020) Stock Network Stability After Crashes Based on Entropy Method. Front. Phys. 8:163. doi: 10.3389/fphy.2020.00163

1. INTRODUCTION

The stability of networks after random failures or attacks is a central issue in the study of complex networks [1–7]. A large amount of research has been devoted to this issue for various complex networks, such as biological, physical, social and financial networks, and has successfully revealed the relationship between network structure and stability after external attacks [8–11].

Stock markets are complex systems, which can be well-abstracted and described by complex networks. In research on stock networks, the network stability after stock market crashes has attracted much attention [11, 12]. Studies on network stability elucidate the topological reconfigurations of stock networks after crashes and help to improve our understanding of financial crises. However, these studies cannot be directly used to guide the government in stabilizing the market after crashes.

In recent years, an increasing body of work has concentrated on the question of how rescue policies, which are government policies aimed at stabilizing the stock market at times of crises, influence market stability after crashes [13, 14]. This question has been studied by measuring the market stability in terms of volatility, and it has been found that rescue policies can improve the stability of the stock market after crashes. The complex network method can reduce the immense complexity of financial markets to facilitate investigation while retaining the market's core information, and some researchers have studied market stability after crashes by using the network method [11, 12, 15]. Motivated by the studies mentioned above, we investigate the influence of rescue policies on the stability of stock markets after crashes from the perspective of complex networks.

In network research, most researchers measure the network stability based on network robustness with respect to random node removal or targeted attacks [11, 16, 17]. In this paper, we introduce the network entropy to measure the stability of complex networks. The network entropy

calculated with the degree distribution can provide an average measure of network heterogeneity, which is a simple but essential characteristic of complex networks and has a direct relationship with the network's resilience to both random failures and attacks [18–21]. To the best of our knowledge, this work is the first attempt to examine the influence of government rescue policies on stock market stability after crashes from the perspective of network stability.

2. DATASETS

The first dataset records the closing prices of all constituent stocks of the SSE 180 index, a benchmark index for the Shanghai market, at the end of every minute between 9:30 and 11:30 a.m. and between 1:00 and 3:00 p.m. on 717 consecutive trading days from 16 December 2013 to 22 November 2016. The dataset was obtained from Wind Information (http://www.wind.com.cn/) and contains ∼43,440 records per day on average, with a total size of 31,146,480.

The second dataset records closing prices of the Hang Seng Index (HSI) and all its constituent stocks trading in the Hong Kong Exchange at the end of every minute between 9:30 a.m. and 12:00 p.m. and between 1:00 and 4:00 p.m. on 422 consecutive trading days from 20 October 2014 to 5 July 2016, also from Wind Information. This dataset contains approximately 23,240 records per day on average, and the total size is 9,807,280.

The third dataset, from SinaFinance (http://finance.sina.com.cn/), records 45 news items on rescue policies for the stock market released by the Chinese government for the period from 16 December 2013 to 22 November 2016, with 34 of these news items released in the 2 months after 19 June 2016 during which more than 1,000 A-shares in the Chinese stock market hit the daily downward price fluctuation limit. To take into account the specific government rescue measures for stock market crashes, this dataset also includes information from Eastmoney (http://eastmoney.com/) on the list of stocks purchased by the government between 30 June 2015 and 30 September 2015.

3. METHODOLOGY

3.1. Crashes and Identification of Sub-periods

In this paper, crashes are identified by looking for large price changes within different time windows [22–24]. On 19 June 2015, more than 1,000 A-shares in the Chinese stock market hit the daily downward price fluctuation limit, and the SSE 180 index fell by 4.61%. On 7 January 2016, the circuit breakers mechanism was suspended, and the SSE 180 index dropped by more than 5%. By calculating the price changes in the SSE 180 index, we find that during the sampling period the price changes on these 2 days are the largest. Therefore, we identify these two large price changes in the Shanghai stock market as crashes. For the universality test, we also identify one crash occurring on 10 March 2015 in the Hong Kong stock market in a similar way.

Next, for the purpose of analyzing network stability during periods with and without rescue policies, we divide the sampling periods for the Shanghai and Hong Kong stock markets into different sub-periods. For the Shanghai stock market, sub-periods are determined by using the news of rescue policies recorded by SinaFinance. First, after the crash on 19 June 2015, most rescue policies were enacted by the government within the subsequent 2 months. These include: on 27 June, China's central bank, People's Bank of China, cut interest rates by 0.25%; and on 6 July, China Financial Futures Exchange (CFFEX) restricted index futures trading. In contrast, no rescue policies were put in place for months after the crash on 7 January 2016. To compare the stability of the stock network in the period after a crash and in a period when the market is stable, we also identify sub-periods of stable states. We find that the volatility of the SSE 180 index was low and stable from mid-June to early October of 2014 and over the second half of the year 2016. Therefore, for the Shanghai stock market, we divide the sampling period into four sub-periods: a sub-period with government rescue policies after a crash, a sub-period without government rescue policies after a crash, and two stable-state sub-periods. Considering that the impact of a crash would not be obvious when the sub-period is too short, and there may be external noise when the sub-period is too long, we choose the sub-period length to be 60 trading days. Note that sub-period lengths of 50 or 70 trading days give similar results.

For the Hong Kong stock market, for which there had been no rescue policies, we separate the sampling period into three sub-periods. The 60-day period following the crash on 10 March 2015 is the sub-period without government rescue policies. In analogy to the stable-state periods in the Shanghai stock market, we identify the 60-day periods following 22 December 2014 and 5 March 2016 as stable-state sub-periods in the Hong Kong stock market.

3.2. Construction of the Stock Network

The most common method of constructing a stock network is based on correlations of the stock price return. This method calculates the correlation coefficient of the stock price return and converts the coefficient matrix into a distance matrix [25].

Let $P_i(t)$ be the closing price of stock i at time t, and let $R_i(t)$ be the return of stock i at time t, given by

$$R_i(t) = \ln P_i(t) - \ln P_i(t-1). \tag{1}$$

Then the Pearson correlation coefficient $\rho_{i,j}$ between stocks i and j can be calculated as

$$\rho_{i,j} = \frac{\langle R_i R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sqrt{\langle R_i^2 - \langle R_i \rangle^2 \rangle \langle R_j^2 - \langle R_j \rangle^2 \rangle}}, \tag{2}$$

where $\langle \cdot \rangle$ refers to the time average over the period analyzed. Following the idea behind the construction of a complex network, we next transform the correlation matrix into a distance matrix D with elements $d_{i,j}$, where the distance between the two stocks i and j is defined as Mantegna [26]

$$d_{i,j} = \sqrt{2(1 - \rho_{i,j})}. \tag{3}$$

Treating stocks as network nodes, we now construct the stock network with the planar maximally filtered graph (PMFG) method [27, 28], in which network links are added according to the distance between nodes. The PMFG method is used to construct planar graphs and has the algorithmic advantage that planarity tests are relatively simple to perform; planar graphs also provide more information than minimum spanning tree graphs on the internal structure of a stock market. The PMFG procedure is carried out as follows. First, rank the distance elements between all pairs of nodes in ascending order. Second, add a link between the two nearest nodes, i.e., the pair with the smallest distance, if and only if the resulting graph after such a link insertion can still be embedded in a plane without crossing any links. Third, repeat the previous step until no more links can be added. The resulting PMFG contains $3N-6$ links, where N is the number of nodes. The average degree is $\langle k\rangle = 2(3N-6)/N$, which is roughly equal to 6 when N is large enough.

3.3. Network Entropy

Based on the theory of large deviations, Demetrius et al. [29] derived a fluctuation theorem, which states that network entropy and stability, as measured by the fluctuation decay rate after random perturbations, are positively correlated. By invoking this theorem, Demetrius and Manke [30] showed that the network entropy can quantitatively describe the homeostatic network properties under perturbations, a generic term for robustness. Meanwhile, Wang et al. [20] showed that the network heterogeneity, a simple but essential characteristic of a complex network, is in direct relationship with the network's resilience to both random failures and attacks, and that the heterogeneity can be measured by network entropy. This implies that the greater the entropy, the more stable and heterogeneous the network is. We therefore introduce the network entropy to study the stability of stock networks.

Following the definition of entropy introduced in Wang et al. [20], we define the stock network entropy as the entropy of the degree distribution, which is given by

$$H = -\sum_{k=1}^{N-1} P(k)\log P(k), \tag{4}$$

where N is the total number of nodes in the network and $P(k)$ is the degree distribution, which gives the probability of having a node with k links.

4. RESULTS AND DISCUSSION

4.1. Entropy of the Stock Network

To find out which kind of network structure the stock network is close to, we compare the entropy of the stock network to the entropies of several classes of networks, including the Erdös-Rényi (ER) random, Barabási-Albert (BA) scale-free and Watts-Strogatz (WS) small-world networks. Note that the average degree of the network of 180 stocks constructed by the PMFG method is approximately 6. Therefore, to make the entropies of the stock network and of the ER, BA and WS networks comparable, we calculate the entropy of the latter three networks with the same number of nodes, $N = 180$, and present the statistical characteristics of these networks in **Table 1**.

TABLE 1 | Statistical description of the ER random, BA scale-free, and WS small-world networks and the stock network.

	ER random	BA scale-free	WS small-world	Stock network
N	180	180	180	180
E	537	531	540	534
C	0.03	0.10	0.17	0.73
$\langle l\rangle$	3	2.80	3.50	3.58
$\langle k\rangle$	6	6	6	5.93
$\langle s\rangle$	6	6	6	1.63
$\langle E_w\rangle$	1	1	1	0.28

N represents the number of nodes, E the number of links, C the clustering coefficient, $\langle l\rangle$ the average shortest path length, $\langle k\rangle$ the average node degree, $\langle s\rangle$ the average node strength, and $\langle E_w\rangle$ the average link weight. The weights for the links in the ER random, BA scale-free and WS small-world networks are equal to 1. In the stock network, each link weight is defined as the correlation coefficient $\rho_{i,j}$ for stocks i and j [31].

First, we calculate the entropy of the ER random network as follows. For the ER network with N nodes and link probability p between each pair of nodes, the degree distribution $P(k)$ is given by the Poisson distribution [32],

$$P(k) = \frac{e^{-Np}(Np)^k}{k!}, \tag{5}$$

and the average degree $\langle k\rangle$ is Np. The entropy of the ER network with average degree $\langle k\rangle = 6$ for various N from 20 to 180 is plotted in **Figure 1A**.

Next, for the BA scale-free network with N nodes, minimal connectivity m, and scaling exponent α of the degree distribution, the entropy of the network can be expressed as Wang et al. [20]

$$H = \left(\log(\alpha-1) + \frac{\alpha}{1-\alpha}\right)\frac{1-N}{N} + \frac{\alpha}{1-\alpha}\frac{\log N}{N}, \tag{6}$$

and the average degree of the network is given by $\langle k\rangle = \frac{(\alpha-1)}{(2-\alpha)}m(N^{(2-\alpha)/(\alpha-1)}-1)$. The entropy of the BA network with $N = 180$ and $m = 1$ for various α from 1.5 to 2.5 is shown in **Figure 1B**.

For the WS small-world network first proposed in Watts and Strogatz [33], which starts with a ring lattice of N nodes where every node is connected to its first K neighbors and then has each edge of the lattice randomly rewired with probability p such that self-connections and duplicate edges are excluded, the degree distribution is given by Albert and Barabási [34]

$$P(k) = \sum C^n_{K/2}(1-p)^n p^{\frac{K}{2}-n}\frac{(\frac{pK}{2})^{k-\frac{K}{2}-n}}{(k-\frac{K}{2}-n)!}e^{-pK/2}, \tag{7}$$

and the average degree of the network is $\langle k\rangle = K$. We plot the entropy of the WS small-world network with average degree $\langle k\rangle = K = 6$ and $N = 180$ for various p from 0 to 1 in **Figure 1C**.

Using the network construction method and the definition of network entropy proposed in this paper, we present in **Figure 1D**

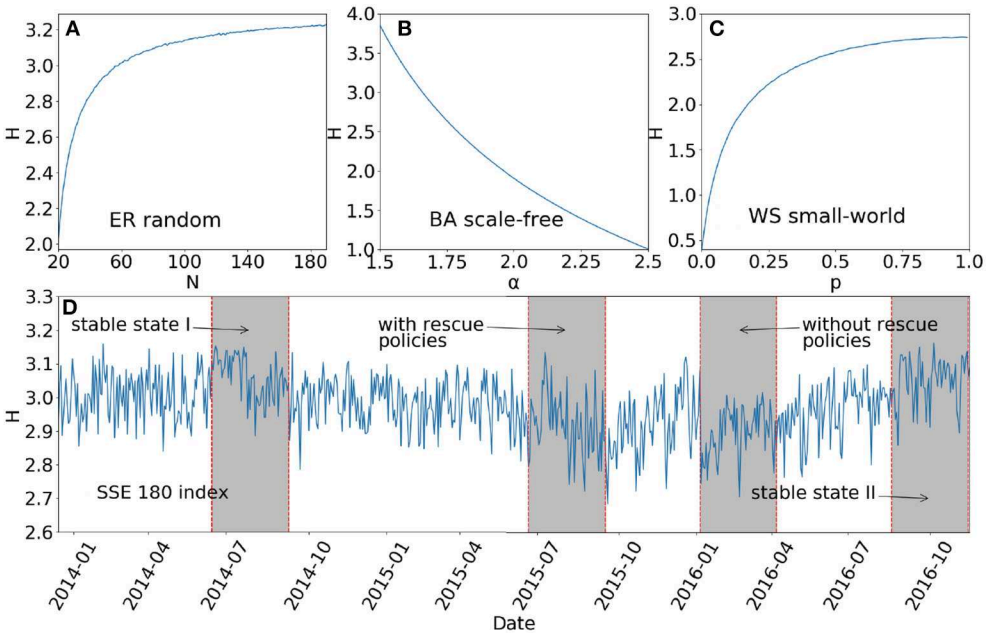


FIGURE 1 | Entropy of **(A)** the ER random network with $Np = 6$ for various N; **(B)** the BA scale-free network for $N = 180$, $m = 1$ and various α; **(C)** the WS small-world network for $K = 6$, $N = 180$ and various p; and **(D)** the stock network for the sampling period from 16 December 2013 to 22 November 2016 in the Shanghai stock market.

the entropy of the stock network during the sampling period from 16 December 2013 to 22 November 2016. Note that the stock network is constructed with the intraday stock returns for every trading day. We next compare the entropy of the stock network to that of the ER random, BA scale-free and WS small-world networks with the same average degree and number of nodes.

For the ER network with average degree $\langle k \rangle = 6$, the entropy is approximately 3.2 when $N = 180$, as shown in **Figure 1A**. For the BA network with $\langle k \rangle = 6$ and $N = 180$, the entropy is approximately 2 when $\alpha \approx 1.9484$, as shown in **Figure 1B**. As can be observed in **Figure 1C**, the entropy of the WS network with $\langle k \rangle = 6$ and $N = 180$ is always smaller than 2.7 for different values of p. From **Figure 1D**, one can see that the entropy H of the stock network satisfies $2.7 < H < 3.2$ during the sampling period, which is much larger than the entropy of the BA network but smaller than that of the ER network and a little bigger than that of the WS small-world network. This suggests that the topological structure of the stock network is closer to that of the ER and WS networks.

To reveal the influence of crashes on the stability of the stock network, we next analyze the network entropy in the four sub-periods, i.e., the two stable-state sub-periods and the sub-periods with and without rescue policies after a crash. First, the mean values of the entropy for the sub-periods with and without rescue policies after a crash are 2.9203 and 2.8992, respectively, and the mean values of the entropy for the sub-periods of stable states I and II are 3.0602 and 3.0487, respectively. This result suggests that in the periods after a crash the stock network is less stable and heterogeneous than during the stable-state periods, since the entropy measures the network stability and heterogeneity. We also conduct a t-test for the significance of the difference between the entropies of the stock network for different sub-periods, as shown in **Table 2**; the test finds no significant difference between the mean entropies of the sub-periods with and without rescue policies.

TABLE 2 | Results of t-test for testing the significance of the difference between the entropies of the stock network for different sub-periods in the Shanghai stock market.

	Stable state I	Stable state II	With rescue policies	Without rescue policies
Stable state I	0.0000			
Stable state II	−0.9159	0.0000		
With rescue policies	9.4892***	8.4878***	0.0000	
Without rescue policies	12.3273***	11.0788***	−1.3552	0.0000

*This table reports t-statistics from the t-test. The symbols *** denote significance at the 1, 5, and 10% levels, respectively.*

4.2. Entropy Evolution of the Stock Network

To examine how the effects of rescue policies on stock network stability change over time, we now analyze the evolution of the entropy of the stock network after a shock during sub-periods with and without rescue policies. One might imagine that these shocks are similar to after shocks following an earthquake.

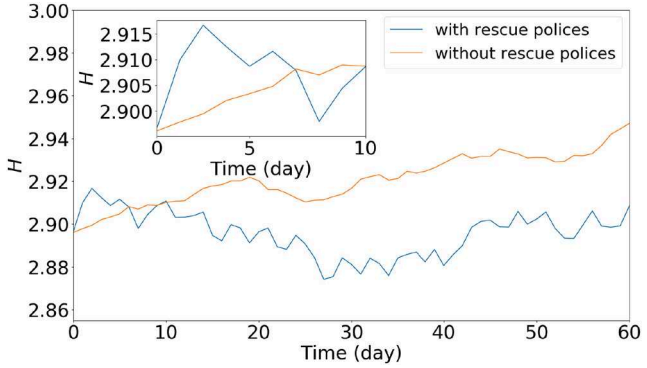


FIGURE 2 | Entropy evolution of the stock network during sub-periods with and without rescue policies after crashes in the Shanghai stock market.

A shock is defined here as a time at which the volatility $V(t) = |R(t)|$ exceeds a given threshold value $R\text{th} = S\delta$, where $R(t)$ is the logarithmic price return, δ is the mean volatility over the stable-state period, and the threshold S is positive [35]. Using the same method of shock identification, 43 and 30 shocks are identified with the threshold $S = 2$ during the sub-periods in the Shanghai stock market with and without rescue policies, respectively. Then, the entropy evolution is defined as the change in the entropy averaged over the shock after its occurrence in the sub-period. We present the entropy evolution of the stock network for these two sub-periods in **Figure 2**.

Figure 2 shows that the entropy evolution of the stock network for the sub-period with rescue policies after a crash increases rapidly at the early stage, then decreases and rises again slowly at later times. For the sub-period without rescue policies, the entropy increases slowly throughout the whole period. These results suggest that rescue policies can improve the heterogeneity and stability of the network only for a short time at the initial stage, whereas the heterogeneity and stability of the network will increase slowly over the period after a crash even without rescue policies.

4.3. Further Analysis of Stock Degree

In this section, we study the degrees of all stocks in the stock network over different sub-periods, with the aim of identifying the cause of the effect of rescue policies on the stock network entropy.

In **Table 3** we present the average degree of all stocks of conventional industries in the SSE 180 index during the four sub-periods, i.e., the two stable-state sub-periods and the sub-periods with and without rescue policies. **Table 4** presents industry information on constituent stocks in the SSE 180 index, including the industry codes, industry names, and number of stocks belonging to each industry.

As seen in **Table 2**, the average degree of stocks in the Finance & Insurance industry is large during the two stable-state sub-periods and the sub-period without rescue policies, but relatively small during the sub-period with rescue policies. On the other hand, the average degree of stocks in the Construction industry is large during the sub-period with rescue policies, and relatively small during the stable-state sub-periods and the sub-period without rescue policies. This suggests that the degrees of both the Finance & Insurance and the Construction industries are greatly influenced by government rescue policies, which ultimately lead to the change in network entropy during the sub-period with rescue policies.

TABLE 3 | Average degrees of constituent stocks in the SSE 180 index for each industry during two stable-state sub-periods and sub-periods with and without rescue policies.

Industry code	Stable state I	Stable state II	With rescue policies	Without rescue policies
A	5.5000	4.6167	4.6083	4.6917
B	5.7904	4.8030	5.0698	3.8398
C0	6.0500	5.7750	4.6000	3.3000
C4	5.4833	5.1167	5.2833	5.4567
C5	5.6278	5.3000	4.5667	5.0111
C6	6.9296	4.9479	5.8924	4.5700
C7	5.9214	5.4549	6.9706	6.8395
C8	5.8955	5.0963	4.6000	4.9537
D	4.6738	3.9033	4.3938	4.4050
E	3.7333	6.5021	8.3833	5.5519
F	4.6833	4.8375	5.2800	4.1530
G	6.2574	7.4639	6.3600	6.9667
H	5.0133	5.2521	3.1583	4.9528
I	7.1528	7.1635	5.1672	7.0186
J	5.0265	5.3250	4.7605	4.8358
K	5.2250	5.4600	3.7750	5.7583
L	3.7905	5.0500	4.9643	5.7167
M	4.7222	4.8875	7.3542	10.596

TABLE 4 | Industry information on constituent stocks in the SSE 180 index.

Industry code	Industry	Number of stocks	Industry code	Industry	Number of stocks
A	Agriculture	5	E	Construction	9
B	Mining	14	F	Transportation	9
C0	Food & Beverage	6	G	Information Technology	8
C4	Pharmaceuticals	8	H	Wholesale & Retail Trade	8
C5	Electronics	4	I	Finance & Insurance	31
C6	Metals & Non-metals	11	J	Real Estate	19
C7	Machinery	24	K	Social Services	3
C8	Pharmaceuticals	10	L	Communication & Cultural	4
D	Utilities	8	M	Comprehensive	3

The conventional industries are grouped based on the China Securities Regulatory Commission (CSRC) industry code. The basic information includes the industry code, the full name of the industry, and the number of chosen stocks belonging to each industry.

TABLE 5 | Top 10 stocks purchased by the government from 30 June 2015 to 30 September 2015 in the Shanghai stock market.

Code	Industry	Code	Industry
601669	Construction	600765	Machinery
600958	Finance & Insurance	600111	Metals & Non-metals
600023	Utilities	600118	Information Technology
600418	Machinery	600068	Construction
600518	Pharmaceuticals	600415	Comprehensive

To elucidate the relation between rescue policies and the degrees of stocks, we compare stocks purchased by the government and stocks with large degrees in the stock network.

Table 5 lists the top 10 stocks purchased by the government during the period from 30 June 2015 to 30 September 2015, which closely coincides with the period in which rescue policies were enacted. **Table 5** shows that the top stocks purchased by the government belong to the Construction industry during the sub-period with rescue policies, which is consistent with the results in **Table 3**. In **Table 6** we present the top 10 stocks with the highest mean value of stock degrees in the stock network for the two stable-state sub-periods (I and II) and the sub-periods with and without rescue policies.

From **Table 6** one observes that during the stable-state sub-periods, more than half of the top 10 stocks with the highest mean degree belong to the Finance & Insurance industry, which is consistent with the results in **Table 3**. Furthermore, no stock belongs to the Finance & Insurance industry in the sub-period with rescue policies, whereas more than half of the top 10 stocks belong to the Finance & Insurance industry in the sub-period without rescue policies; this is also consistent with the results in **Table 3**.

Finally, from **Table 6**, more than half of the top 10 stocks purchased by the government have large degrees in the sub-period with rescue policies, while only a few stocks purchased by the government have large degrees in the sub-period without rescue policies. All these observations imply that government purchasing behavior greatly affects the degrees of individual stocks and thus changes the entropy of the stock network during the sub-period with rescue policies.

TABLE 6 | Top 10 stocks with the highest mean value of degrees in the stock network for the two stable-state sub-periods and the sub-periods with and without rescue policies in the Shanghai stock market.

Stable state I		Stable state II	
Code	**Industry**	**Code**	**Industry**
600109	Finance & Insurance	601198	Finance & Insurance
601318	Finance & Insurance	600485	Information Technology
600030	Finance & Insurance	601555	Finance & Insurance
600111	Metals & Non-metals	600570	Information Technology
600999	Finance & Insurance	600061	Finance & Insurance
600837	Finance & Insurance	600150	Machinery
601088	Mining	600446	Information Technology
601601	Finance & Insurance	600030	Finance & Insurance
600036	Finance & Insurance	601688	Finance & Insurance
600739	Wholesale & Retail Trade	600109	Finance & Insurance
With rescue policies		**Without rescue policies**	
Code	**Industry**	**Code**	**Industry**
601727	Machinery	**600118**	Information Technology
600118	Information Technology	600895	Comprehensive
600031	Machinery	601211	Finance & Insurance
600150	Machinery	600150	Machinery
600111	Metals & non-metals	600739	Wholesale & Retail Trade
601669	Construction	601198	Finance & Insurance
600068	Construction	**600958**	Finance & Insurance
601390	Construction	601788	Finance & Insurance
600415	Comprehensive	600109	Finance & Insurance
600804	Information Technology	601099	Finance & Insurance

The stocks in this table are arranged from top to bottom in order of decreasing value of stock degree. The boldfaced stock codes in this table are also in the list of top 10 stocks purchased by the government during the sub-period with rescue policies.

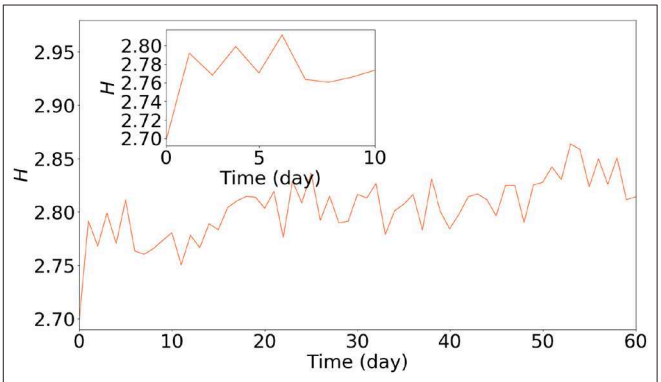


FIGURE 3 | Entropy evolution of the stock network during the 60-day period following the crash on 10 March 2015 in the Hong Kong stock market.

4.4. Universality and Robustness Tests

4.4.1. Universality Test

To test the universality of the empirical results for the Shanghai stock market, we calculate the entropy evolution for the stock network constructed with constituent stocks of the HSI index during the period following the crash on 10 March 2015 in the Hong Kong stock market, for which no government rescue policies existed.

In **Figure 3**, the entropy evolution of the stock network for the Hong Kong stock market increases slowly throughout the whole period, with a trend similar to the entropy evolution during the period without government rescue policies post-crash in the Shanghai stock market. This implies that the network stability also increases slowly in the Hong Kong stock market, where there were no government rescue policies.

In analogy to the analysis of stock degrees in the Shanghai stock market, we present in **Table 7** the top 10 stocks with the

TABLE 7 | Top 10 stocks with the highest mean value of degrees in the stock network for the two stable-state sub-periods and the sub-period without rescue policies in the Hong Kong stock market.

Stable state I		Stable state II		Without rescue policies	
Code	**Industry**	**Code**	**Industry**	**Code**	**Industry**
2318	Financials	0883	Energy	0883	Energy
0857	Energy	0941	Telecommunications	0941	Telecommunications
3968	Financials	2318	Financials	2628	Financials
0005	Financials	0005	Financials	2318	Financials
0941	Telecommunications	0700	Information Technology	0857	Energy
2628	Financials	2328	Financials	0388	Financials
0700	Information Technology	6030	Financials	3968	Financials
1336	Financials	0388	Financials	0005	Financials
6837	Financials	2601	Financials	1088	Energy
2601	Financials	1299	Financials	0386	Energy

The stocks in this table are arranged from top to bottom in order of decreasing value of stock degree. The conventional industries are based on the Hang Seng Industry Classification system (HSICS).

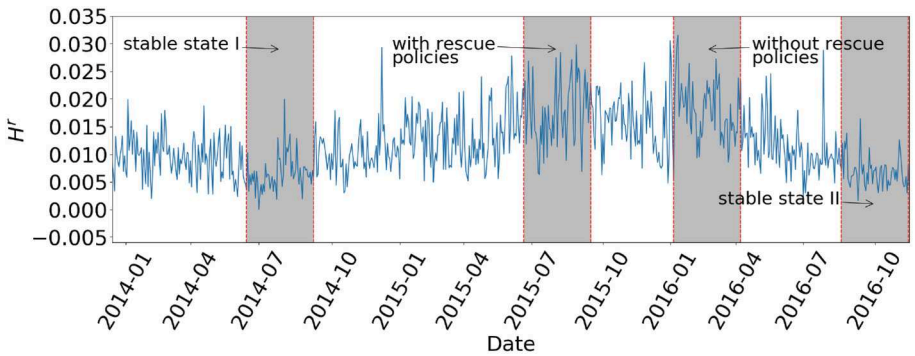


FIGURE 4 | Relative entropy of the stock network for the sampling period from 16 December 2013 to 22 November 2016 in the Shanghai stock market.

highest mean value of degrees in the stock network for the two stable-state sub-periods and the sub-period without rescue polices in the Hong Kong stock market.

In **Table 7**, more than half of the stocks belong to the Financials industry during the stable-state sub-periods and the sub-period without rescue policies following the crash that occurred on 10 March 2015 in the Hong Kong stock market, which is consistent with the results on stock degrees during the stable-state sub-periods and the sub-period without rescue policies in the Shanghai stock market.

4.4.2. Robustness Test

To ensure that the previous findings are robust for the network stability measured by entropy calculated with the node degree distribution, here we use another quantity, the relative entropy, to examine the relation between government rescue policies and network stability.

The relative entropy, which is also called the Kullback-Leibler (KL) divergence, is a measure of how a probability distribution differs from another probability distribution [36]. This is useful when we want to compare the degree distributions of, for example, the stable periods and the periods with and without rescue policies after crashes. For complex networks, the relative entropy between two node degree distributions can be defined as

$$H^r = -\sum_{k=1}^{N-1} P(k) \log \frac{P(k)}{Q(k)}, \tag{8}$$

where $Q(k)$ is the node degree distribution during the stable-state period.

Figure 4 presents the relative entropy of the stock network during the sampling period from 16 December 2013 to 22 November 2016 in the Shanghai stock market.

As in the analysis of the entropy of the stock network, we find that the mean values of the relative entropy for the sub-periods with and without rescue policies after a crash are 0.0176 and 0.0178, respectively, larger than the mean values of the relative entropy for the stable states I and II (0.0068 and 0.0065, respectively). We also give, in **Table 8**, the results of a t-test for the significance of the differences between the relative entropies of the stock network during the four sub-periods.

TABLE 8 | Results of *t*-test for testing the significance of the difference between the entropies of the stock network for different sub-periods in the Shanghai stock market.

	Stable state I	Stable state II	With rescue policies	Without rescue policies
Stable state I	0.0000			
Stable state II	0.5183	0.0000		
With rescue policies	−12.8112***	−12.9207***	0.0000	
Without rescue policies	−14.0500***	−14.2572***	0.2393	0.0000

*This table reports t-statistics from the t-test. The symbols *** denote significance at the 1, 5, and 10% levels, respectively.*

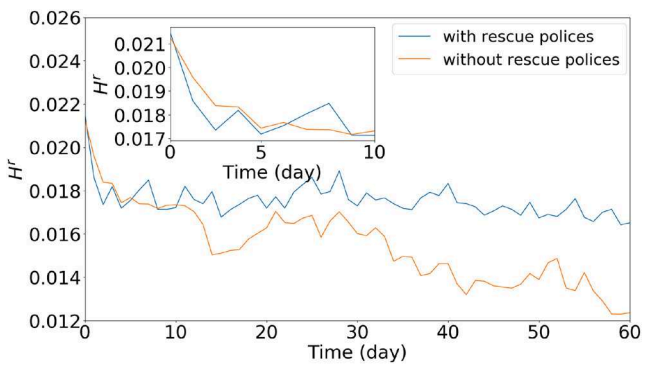


FIGURE 5 | Relative entropy evolution of the stock network during sub-periods with and without rescue policies after crashes in the Shanghai stock market.

Table 8 shows no significant difference between the relative entropies of the stock network for the stable-state sub-periods and the sub-periods with and without rescue policies. These results indicate that the stock network is less heterogeneous and stable during the sub-periods with and without rescue policies, which is consistent with the results obtained from the entropy of the stock network presented above.

Finally, we show in **Figure 5** the relative entropy evolution of the stock network for the sub-periods with and without rescue policies. Consistent with the entropy evolution results in **Figure 2**, the relative entropy evolution of the stock network for the sub-period with rescue policies decreases rapidly at first and much more slowly at later times. For the sub-period without rescue policies, the relative entropy evolution decreases slowly over the whole period. This implies that the rescue policies have only short-term influences on the stock market, and can only increase the heterogeneity and stability of the stock network for a short time. For the sub-period without rescue policies, the heterogeneity and stability of the stock network can also increase slowly.

5. CONCLUSION

In this paper, we have studied the influence of government rescue policies on the stability of stock networks after crashes in the Shanghai and Hong Kong stock markets based on the entropy method.

By analyzing the entropy of the stock network in different sub-periods, i.e., stable-state sub-periods and sub-periods with and without rescue policies after a crash, we find that rescue policies have only a short-term influence on the stability of the stock network after a crash, and can improve the network stability for only a short time. Over a longer time, the network stability during a sub-period with rescue policies could increase even more slowly than a sub-period without rescue policies. Further analysis of the relation between stock degrees and government purchasing behavior indicates that government purchasing can lead to significant changes in the degrees of specific stocks during the sub-period with rescue policies, and thus ultimately improve the stability of the stock network.

Our study focuses on the influence of government interventions on network stability measured using entropy. It is of theoretical interest for understanding the relation between external interventions and network topological structure, and further has practical significance for regulators and policymakers who are attempting to stabilize stock markets after crashes.

One can extend the present study to a microscopic analysis by looking at the effects of individual node strengths and link weights on the stability of the stock network after a crash. This can be done by using methods similar to that of Bellingeri et al. [37], for example, and is a topic for future research.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

M-YY worked on the calculations and drafting. FR and S-PL contributed to the concept design, critical revision, and finalization of the manuscript.

FUNDING

This work was partially supported by the National Natural Science Foundation of China (grant nos. 71532009, 71790594, and 71871094), the Humanities and Social Sciences Fund sponsored by Ministry of Education of the People's Republic of China (grant no. 17YJAZH067), and the Fundamental Research Funds for the Central Universities (2015).

REFERENCES

1. Holme P, Kim BJ. Growing scale-free networks with tunable clustering. *Phys Rev E*. (2002) **65**:026107. doi: 10.1103/PhysRevE.65.026107
2. Gallos LK, Argyrakis P, Bunde A, Cohen R, Havlin S. Tolerance of scale-free networks: from friendly to intentional attack strategies. *Phys A Stat Mech Appl*. (2004) **344**:504–9. doi: 10.1016/j.physa.2004.06.021
3. Lombardi A, Höfrnquist M. Controllability analysis of networks. *Phys Rev E*. (2007) **75**:056110. doi: 10.1103/PhysRevE.75.056110
4. Li J, Wu J, Li Y, Deng HZ, Tan YJ. Attack robustness of scale-free networks based on grey information. *Chinese Phys Lett*. (2011) **28**:058904. doi: 10.1088/0256-307X/28/5/058904
5. Wu FX, W L, Wang JX, Liu J, Chen LA. Transittability of complex networks and its applications to regulatory biomolecular networks. *Sci Rep*. (2014) **4**:4819. doi: 10.1038/srep04819
6. Dong GG, Du RJ, Hao HF, Tian LX. Modified localized attack on complex network. *Europhys Lett*. (2016) **113**:28002. doi: 10.1209/0295-5075/113/28002
7. Peng J, Lu W, Kurths J. Stochastic basin stability in complex networks. *Europhys Lett*. (2018) **122**:40003. doi: 10.1209/0295-5075/122/40003
8. Zdeborová L, Mézard M. The number of matchings in random graph. *J Stat Mech Theor Exp*. (2006) **5**:05003. doi: 10.1088/1742-5468/2006/05/P05003
9. Liu YY, Slotine JJ, Barabási AL. Controllability of complex networks. *Nature*. (2011) **473**:167–73. doi: 10.1038/nature10011
10. McCowan B, Beisner BA, Capitanio JP, Jackson ME, Cameron AN, Seil S, et al. Network stability is a balancing act of personality, power, and conflict dynamics in Rhesus Macaque societies. *PLoS ONE*. (2011) **6**:e22350. doi: 10.1371/journal.pone.0022350
11. Zhang W, Zhuang X. The stability of Chinese stock network and its mechanism. *Phys A Stat Mech Appl*. (2019) **515**:748–61. doi: 10.1016/j.physa.2018.09.140
12. Heiberger RH. Stock network stability in times of crisis. *Phys A Stat Mech Appl*. (2014) **393**:376–81. doi: 10.1016/j.physa.2013.08.053
13. Arouri M, Estay C, Rault C, Roubaud D. Economic policy uncertainty and stock markets: long-run evidence from the US. *Finance Res Lett*. (2016) **18**:136–41. doi: 10.1016/j.frl.2016.04.011
14. Yang MY, Li SP, Wu Y, Tang J, Ren F. Do government rescue policies reduce the market volatility after crash? Evidence from the Shanghai stock market. *Finance Res Lett*. (2019) **29**:117–24. doi: 10.1016/j.frl.2019.03.020
15. Yan XG, Xie C, Wang GJ. Stock market network's topological stability: evidence from planar maximally filtered graph and minimal spanning tree. *Int J Mod Phys B*. (2015) **29**:1550161. doi: 10.1142/S0217979215501611
16. Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex network. *Nature*. (2000) **406**:378–81. doi: 10.1038/35019019
17. Wang L, Sun YX. Robustness of pinning a general complex dynamical network. *Phys Lett A*. (2010) **374**:15–16. doi: 10.1016/j.physleta.2010.02.024
18. Cancho RFi, Solé RV. Optimization in complex networks. *Lect Notes Phys*. (2003) **625**:114–26. doi: 10.1007/978-3-540-44943-0_7
19. Solé RV, Valverde S. Information theory of complex networks: On evolution and architectural constraints. *Lect Notes Phys*. (2004) **650**:189–207. doi: 10.1007/978-3-540-44485-5_9
20. Wang B, Tang H, Guo C, Xiu Z. Entropy optimization of scale-free networks' robustness to random failures. *Phys A Stat Mech Appl*. (2006) **363**:591–6. doi: 10.1016/j.physa.2005.08.025
21. Billio M, Casarin R, Costola M, Pasqualini A. An entropy-based early warning indicator for systemic risk. *J Int Finan Markets Inst Money*. (2016) **45**:42–59. doi: 10.1016/j.intfin.2016.05.008
22. Mishkin FS, White EN. U.S. stock market crashes and their aftermath: Implications for monetary policy; 2002. *NBER working paper No. W8992*. Available online at: http://ssrn.com/abstract=315989.
23. Sornette D, Zhou WX. Predictability of large future changes in major financial indices. *Int J Forecast*. (2006) **22**:153–68. doi: 10.1016/j.ijforecast.2005.02.004
24. Choudhry T. World War II events and the Dow Jones industrial index. *J Bank Finan*. (2010) **34**:1022–31. doi: 10.1016/j.jbankfin.2009.11.004
25. Huang WQ, Zhuang XT, Yao S. A network analysis of the Chinese stock market. *Phys A Stat Mech Appl*. (2009) **388**:2956–64. doi: 10.1016/j.physa.2009.03.028
26. Mantegna RN. Hierarchical structure in financial markets. *Eur Phys J B*. (1999) **11**:193–7. doi: 10.1007/s100510050929
27. Tumminello M, Aste T, Di Matteo T, Mantegna RN. A tool for filtering information in complex systems. *Proc Natl Acad Sci USA*. (2005) **102**:10421–6. doi: 10.1073/pnas.0500298102
28. Aste T, Shaw W, Di Matteo T. Correlation structure and dynamics in volatile markets. *New J Phys*. (2010) **12**:085009. doi: 10.1088/1367-2630/12/8/085009
29. Demetrius L, Gundlach V, Ochs G. Complexity and demographic stability in population models. *Theor Popul Biol*. (2004) **65**:211–25. doi: 10.1016/j.tpb.2003.12.002
30. Demetrius L, Manke T. Robustness and network evolution–an entropic principle. *Phys A Stat Mech Appl*. (2005) **346**:682–96. doi: 10.1016/j.physa.2004.07.011
31. Bhattacharjee B, Shafi M, Acharjee A. Investigating the influence relationship models for stocks in Indian equity market: a weighted network modelling study. *PLoS ONE*. (2016) **11**:e0166087. doi: 10.1371/journal.pone.0166087
32. Erdös P, Rényi A. On random graphs. *Publ Math*. (1959) **6**:290–7. doi: 10.1109/ICSMC.2006.384625
33. Watts DJ, Strogatz SH. Collective dynamics in 'small-world' networks. *Nature*. (1998) **393**:440–2. doi: 10.1038/30918
34. Albert R, Barabási AL. Statistical mechanics of complex networks. *Rev Mod Phys*. (2002) **74**:47–97. doi: 10.1103/RevModPhys.74.47
35. Lillo F, Mantegna RN. Power-law relaxation in a complex system: omori law after a financial market crash. *Phys Rev E*. (2003) **68**:016119. doi: 10.1103/PhysRevE.68.016119
36. Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat*. (1951) **22**:79–86. doi: 10.1214/aoms/1177729694
37. Bellingeri M, Bevacqua D, Scotognella F, Cassi D. The heterogeneity in link weights may decrease the robustness of real-world complex weighted networks. *Sci Rep*. (2019) **9**:10692. doi: 10.1038/s41598-019-47119-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yang, Ren and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 16 June 2020
doi: 10.3389/fphy.2020.00206

Effective Transfer Entropy Approach to Information Flow Among EPU, Investor Sentiment and Stock Market

Can-Zhong Yao and Hong-Yu Li*

School of Economics and Commerce, South China University of Technology, Guangzhou, China

In contrast to the traditional view that economic policy affects investor sentiment and eventually causes stock price fluctuations, we reveal that investor sentiment is a reflection of both economic policy and stock market information. This article first proposes an improved ETE method with a sliding window. We verify that this new method can capture the dynamic orders effectively by validating this method with the linear TE method. Furthermore, using the improved method, we investigate the strength and direction of information flow among economic policy uncertainty (EPU), investor sentiment and stock market by the novel concept of dynamic effective transfer entropy. The EPU and investor sentiment results show that EPU influenced investor sentiment mainly from August 2015 to June 2016. Among different policies, China's exchange rate reform policy and "circuit-breaker" policy in the stock market played an important role. Moreover, the analysis of sentiment and stock price returns shows that investor sentiment is more a reflection of changes in stock price returns with a 1-month lag order and that the stock market has a significant bargainer effect and a weaker bandwagon effect. Finally, there is no significant information flow transmission relationship between EPU and stock market volatility, indicating that stock market fluctuations are essentially not affected by national policy fluctuations. Although investor sentiment is affected by changes, such as exchange rate reform and stock market policies, many investors do not form consensus expectations.

Keywords: EPU, investor sentiment, stock market, information flow, transfer entropy

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Feng Ma,
Southwest Jiaotong University, China
Peng-Fei Dai,
Tianjin University, China

****Correspondence:***
Can-Zhong Yao
ycz20120911@gmail.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *16 March 2020*
Accepted: *13 May 2020*
Published: *16 June 2020*

Citation:
Yao C-Z and Li H-Y (2020) Effective Transfer Entropy Approach to Information Flow Among EPU, Investor Sentiment and Stock Market. Front. Phys. 8:206. doi: 10.3389/fphy.2020.00206

INTRODUCTION

Transfer entropy arises from the formulation of conditional mutual information. When conditioning on past values of variables, it quantifies the reduction in uncertainty provided by these past values in predicting the dependent variable, which presents a natural way to model statistical causality between variables in multivariate distributions. In the general formulation, transfer entropy is a model-free statistic that is able to measure the time-directed transfer of information between stochastic variables and therefore provides an asymmetric method to measure information transfer.

The information transfer method has been widely used in the finance field. Kwon & Yang [1] employed it to measure the relationship between equities indices, showing that the information transfer was greatest from the US and toward the Asia Pacific region. In particular, the S&P 500 was shown to be the strongest driver of other stock indices. In earlier and somewhat related work, Marschinski and Kantz [2] defined and used effective transfer entropy to quantify contagion in financial markets. Kyrtsou et al. [3] proposed a Granger causality method based on partial

transfer entropy to explore the complex relationships among the S&P 500, VIX and volume. Dimpfl and Peter [4] proposed an appropriate bootstrap to derive confidence bounds and showed in a simulation study that standard linear approaches in economics and finance, such as vector autoregressions and Granger causality tests, are not well-suited to detect information transfer. Garcia-Medina et al. [5] used random matrix theory and information theory to analyze the correlations and flow of information between 64,939 news items from The New York Times and 40 world financial indices over 10 months during the 2015–2016 period. Their results suggested a deep relationship between news and world indices and showed that the news drives world market movement. Sensoy et al. [6] explored the strength of information flow between exchange rates and stock prices by the effective transfer entropy with symbolic encoding methodology. Yang et al. [7] proposed an effective phase transfer entropy method based on the transfer entropy method. These scholars also analyzed the relationship among 9 stock indices from the U.S., Europe and China by using transfer entropy, effective transfer entropy, Rényi transfer entropy, and effective Rényi transfer entropy [8].

The vast influential literature focuses on the correlation between EPU and its impact of investment [9–21]. Many new studies investigated the fluctuation characteristics of China's uncertainty index and its impact on the economy [22–24].

Additionally, many other studies focused on the impact of investor sentiment on stock prices. A study investigating the correlation between investor sentiment and the equity market was conducted following De long's research [25]. Baker and Wurgler [26] investigated investor sentiment using a principal component method. Subsequently, a large number of studies emerged, suggesting that investor sentiment has a significant impact on stock returns [27–40]. In addition, some studies reveal a complex correlation between online investor sentiment and stock market volatility [35, 36, 41, 42].

Although there are many studies on the correlation between EPU and financial or economic indices, there are few studies on the influence of EPU on investor sentiment. There is also a lack of research on the transmission pathways of the EPU, investor sentiment and stock market.

It is widely considered that national economic policies will pass relevant information to influence investor sentiment, affect investment decisions, and ultimately cause stock price fluctuations. However, the real situation may be more complicated especially in China. Motivated by previous studies, we follow the theorizing on information flow to reveal the transmission path of information among these three variables and to provide more references for macro policy makers.

The rest of this paper is organized as follows. In section Data Description, we describe the time series data used in the models. In section Methodology, we introduce the effective transfer entropy (ETE) method. In section An Improved Effective Transfer Entropy Method Based on a Sliding Window, we propose an improved ETE method based on sliding windows. In section Empirical Results, we present the empirical results of EPU, trade and the exchange rate obtained using the improved ETE model. Finally, we provide the conclusion of this paper.

DATA DESCRIPTION

The economic uncertainty index used in this article is an index compiled by Paul Luk's research team [43] from Hong Kong Baptist University. These authors construct the index using words from 10 Hong Kong newspapers, including *Wen Wei Po, Sing Pao, etc.* For each newspaper, the research group counts articles containing Chinese words related to economics, uncertainty and policy to formulate the economic policy uncertainty index. The latest data of the index and related papers can be downloaded from the web: https://economicpolicyuncertaintyinchina.weebly.com/.

The investment sentiment index is constructed with the principal component method based on four underlying proxie, i.e., the relative strength index, the psychological line index, the trading volume and the adjusted turnover rate [40]. The Shanghai Composite Index mostly reflects the performance of the Chinese stock market. The data of this index are from *the China Stock Market & Accounting Research Database.* The data period for all three time series is from February 2005 to May 2019 (**Figure 1**).

METHODOLOGY

We calculate the statistical causality between time series using two different approaches. The first assumes linearity and employs vector autoregressive techniques to estimate the extent to which knowing the driving time series can help predict the dependent series. The second technique compares the difference in mutual information between the independent case and the joint case to describe the success of predicting the dependent series. When predictability is increased by considering the past values of the driving variable, statistical causality is observed.

Linear Causality

We model a time series as autoregressive by expressing its value Y_t at time t as a sum of the contributions over m distinct lagged series using the following linear equation:

$$Y_t = \sum_{k=1}^{m} \beta_k^{(Y)} Y_{t-k} + \varepsilon_t \tag{1}$$

Where $\beta_k^{(Y)}$ is a general coefficient term and ε_t is the residual. Linear regression estimates coefficient parameters $\beta_k^{(Y)}$, which minimize the sum of squared residuals.

To detect whether the values of some second time series X anticipate the future values of Y, we can compare Equation (1) with

$$Y_t = \sum_{k=1}^{m} \beta_k'^{(Y)} Y_{t-k} + \sum_{k=1}^{m} \beta_k'^{(X)} X_{t-k} + \varepsilon_t' \tag{2}$$

We determine that distribution Y is Granger caused by X if the residual in the second regression is significantly smaller than the residual in the first regression. When this distribution holds, then

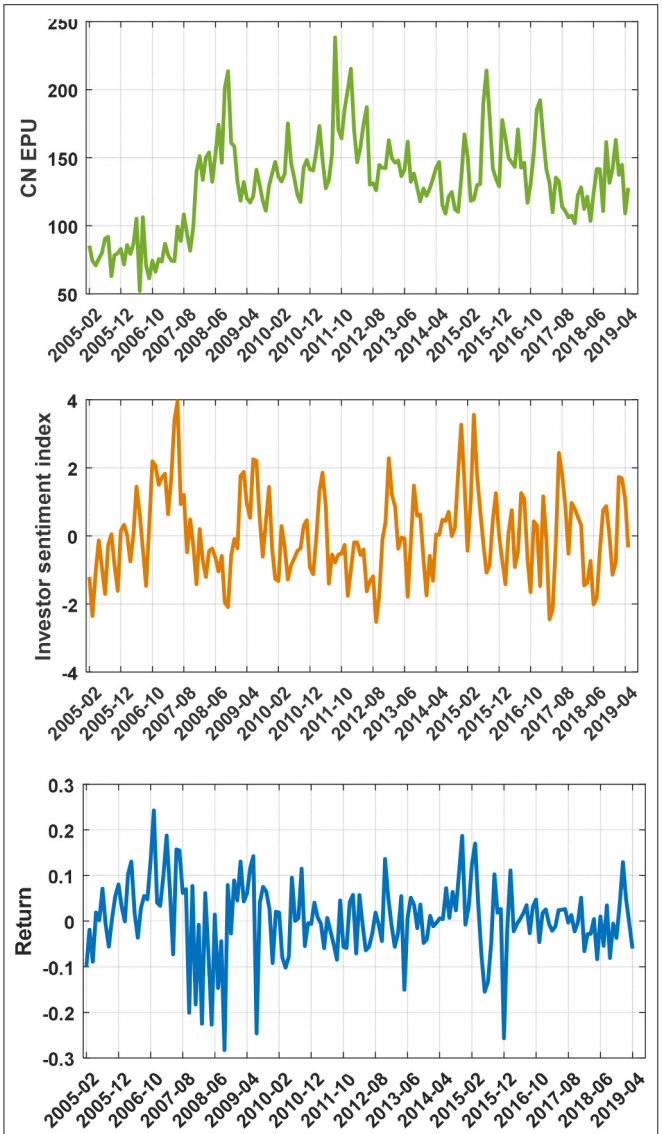


FIGURE 1 | Time series of the CN EPU index, Investor sentiment and Return of the Shanghai Composite Index's closing price.

there must be some information transfer from X to Y. Following Geweke [44], we can represent the information transfer by

$$TE_{X\to Y} = \frac{1}{2}\log\left(\frac{\mathrm{var}(\varepsilon_t)}{\mathrm{var}(\varepsilon_t')}\right) \tag{3}$$

where we adopt the transfer entropy notation (TE) following Barnett et al. [45], whose result shows Granger causality to be equivalent to transfer entropy for multivariate normal distributions.

Non-linear Causality

To detect non-linear causality, we apply an information-theoretic approach. Equation (3) measures the extent to which the additional information in the lagged variable reduces the variance in the model residuals. Transfer entropy extends this concept by considering the uncertainty instead of the variance. Adopting Shannon's measure of information [46], we can express the uncertainty associated with random variable X by

$$H(X) = -\sum_x p(x)\,.\,\log p(x) \tag{4}$$

where $H(x)$ is the Shannon entropy of the distribution, and $p(x)$ represents the probability of $X = x$, which can be conditioned on a second variable to give the conditional entropy:

$$H(Y|X) = H(X,Y) - H(X) \tag{5}$$

When two random variables share information, the mutual information is given by

$$I(X;Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) \tag{6}$$

The entropy of Y conditioned on two variables is

$$H(Y|X,Z) = H(X,Y,Z) - H(X,Z) \tag{7}$$

and the conditional mutual information is therefore

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z) \tag{8}$$

Now, for each lag k, we can describe the information transfer from X_{t-k} to Y_t in terms of the following conditional mutual information:

$$\begin{aligned} TE^{(k)}_{X\to Y} &= I\left(Y_t; X_{t-k}|Y_{t-k}\right) = H\left(Y_t|Y_{t-k}\right) \\ &- H\left(Y_t|X_{t-k}, Y_{t-k}\right) \end{aligned} \tag{9}$$

This equation represents the resolution of uncertainty in predicting Y when considering the past values of both Y and X compared with considering the past values of Y alone.

Considering Equations (5) and (7), we can therefore represent the transfer entropy for a single lag k, which is shown in Equation (9), in terms of four separate joint entropy terms. Following equation (4), these terms may be estimated from the data using a non-parametric density estimation of the probability distributions. For multivariate normal statistics, Equations (9) and (3) coincide [45].

Effective Transfer Entropy (ETE)

It is a feature of the non-parametric estimation of entropy that the absolute scale of the transfer entropy measure has only limited meaning; to detect causality, a relative position must be considered. A simple technique proposed by Marschinski and Kantz [2] is the ETE, derived by subtracting from the observed transfer entropy an average transfer entropy figure calculated over independently shuffled time series, which destroys the temporal order and hence any possible causality. We adopt a shuffling approach producing 50 null-hypothesis transfer entropy values from independently shuffled time series over the same domain containing no causality. By calculating the mean and standard deviation of the shuffled transfer entropy figures,

we estimate the significance of a causal result as the distance between the result and the average shuffled result standardized by the shuffled standard deviation:

$$ETE = TE - \overline{TE}_{shuffle} \tag{10}$$
$$Z := \frac{TE - \overline{TE}_{shuffle}}{\sigma_{shuffle}}$$

where $\overline{TE}_{shuffle}$ is the mean of the shuffled values, and $\sigma_{shuffle}$ is the standard deviation. The shuffling of the time series destroys temporality and should ensure that the mean is approximately zero; therefore, the spread of the data dictates the significance of the result. Assuming that the distribution is close to Gaussian, we can say that a result with $Z > 3$ is roughly in the top 1% of results and hence is comparable to a p-value of 0.01. The nature of the method typically enables clearer significance to be observed with fewer shuffles, even without a strict Gaussian distribution; thus, this method is computationally more attractive than the p-value.

This expression corresponds to the degree to which the result lies in the right tail of the distribution of the zero-causality shuffled samples and hence how unlikely the result is due to chance. Therefore, the Z-score figure represents the significance of the excess transfer entropy in the unshuffled case. We compute the Z-score in Equation (10) for both linear and non-linear results.

AN IMPROVED EFFECTIVE TRANSFER ENTROPY METHOD BASED ON A SLIDING WINDOW

Improved Method Based on a Sliding Window and Comparison With a Traditional Linear Method

Keskin and Aste [47] validated that the non-linear TE method would be useful for detecting a non-linear process. However, the lag order they found was global and unique and thus was unsuitable for capturing the accurate order between two non-stationary series. For non-stationary time series, the data structure changes over time, which means that the causal relationships also evolve dynamically. In addition, due to policy or unexpected events, the causal structure of real financial sequences tends to change over time. Therefore, it would be inaccurate to use a single k to measure global causality.

Considering that the locality of non-stationary data may be stationary or approximately stationary, this paper proposes an improved transfer entropy method based on a sliding window to solve the influence of a non-stationary data structure on traditional transfer entropy. The improved method calculates the transfer entropy as described in section Non-linear Causality but is limited to a certain time segment. Through forward scrolling, the transfer entropy at each time point is obtained, and the causal relationship between the two times series can be revealed. In addition to its ability to capture the structural changes between two time series, the improved method can help us trace the specific time period of the structural change, which cannot be achieved using the traditional linear TE method. We next verify the validity of the algorithm.

First, we generate a time series X following the geometric Brownian motion according to Equation (11) as follows:

$$X_{t+1} = (1 + \mu) X_t + \sigma X_t \eta_t \tag{11}$$

where η_t is a noise obeying the standard normal distribution, $\eta_t \sim N(0, 1)$, and μ and σ represent the drift coefficient and the diffusion coefficient, respectively. Y depends on X, and the equation is constructed as follows (Equation 12):

$$Y_t = (1 - \alpha) X_{t-k} + \alpha X'_{t-k} \tag{12}$$

where X'_{t-k} is another time series generated according to Equation (11). k is the given lag order, and $\alpha \in (0, 1)$ determines the dependence strength between the series Y and X, i.e., the values of the transfer entropy.

Assuming $k = 2, \alpha = 0.5$; $k = 4, \alpha = 0.5$; and $k = 5, \alpha = 0.5$, we can obtain three time series with a length of 200, i.e., $X_t, Y_t^{k=2}$; $X_t, Y_t^{k=4}$ and $X_t, Y_t^{k=5}$ according to Equations (11) and (12).

As shown in **Figure 2**, for a correlation series with a single lag structure, both the traditional transfer entropy, i.e., the linear TE, and the improved TE method can capture the lag order accurately. However, according to the Z-score significance test, we can observe that when the temporal order is destroyed, the linear TE does not show significance in the relevant order; thus, the linear TE method depends on time evolution. As shown in **Figure 3**, the linear TE could only identify the order $k = 4$, which is the highest corresponding transfer entropy value (Z-score indicates that the value is above a significant level). However, the improved TE could detect both $k = 4$ and $k = 5$. Moreover, as shown in **Figure 4**, we can also track the specific time period during which the lead-lag order fluctuates with the improved method.

We reshape $X_t, Y_t^{k=2}$; $X_t, Y_t^{k=4}$ and $X_t, Y_t^{k=5}$ into two new time series X'_t, Y'_t, where $X'_t = [X_t, X_t, X_t]$ and $Y'_t = \left[Y_t^{k=2}, Y_t^{k=4}, Y_t^{k=5}\right]$. These new series show obvious structure fluctuations, and the features are more consistent with the characteristics of real financial data.

Due to the shortcomings of traditional linear methods in revealing dynamic orders, in the empirical analysis in section Empirical Results, we apply the improved transfer entropy to explore the information flow between all sequences. The sliding window length of all structures is 36 months with a forward step size of 1 month.

Comparison With the Granger Causality Test

The Granger causality test is essentially a test used to determine whether a lagging variable can be introduced into an equation containing other variables. If a variable is affected by the lag of other variables, the variables are considered to have

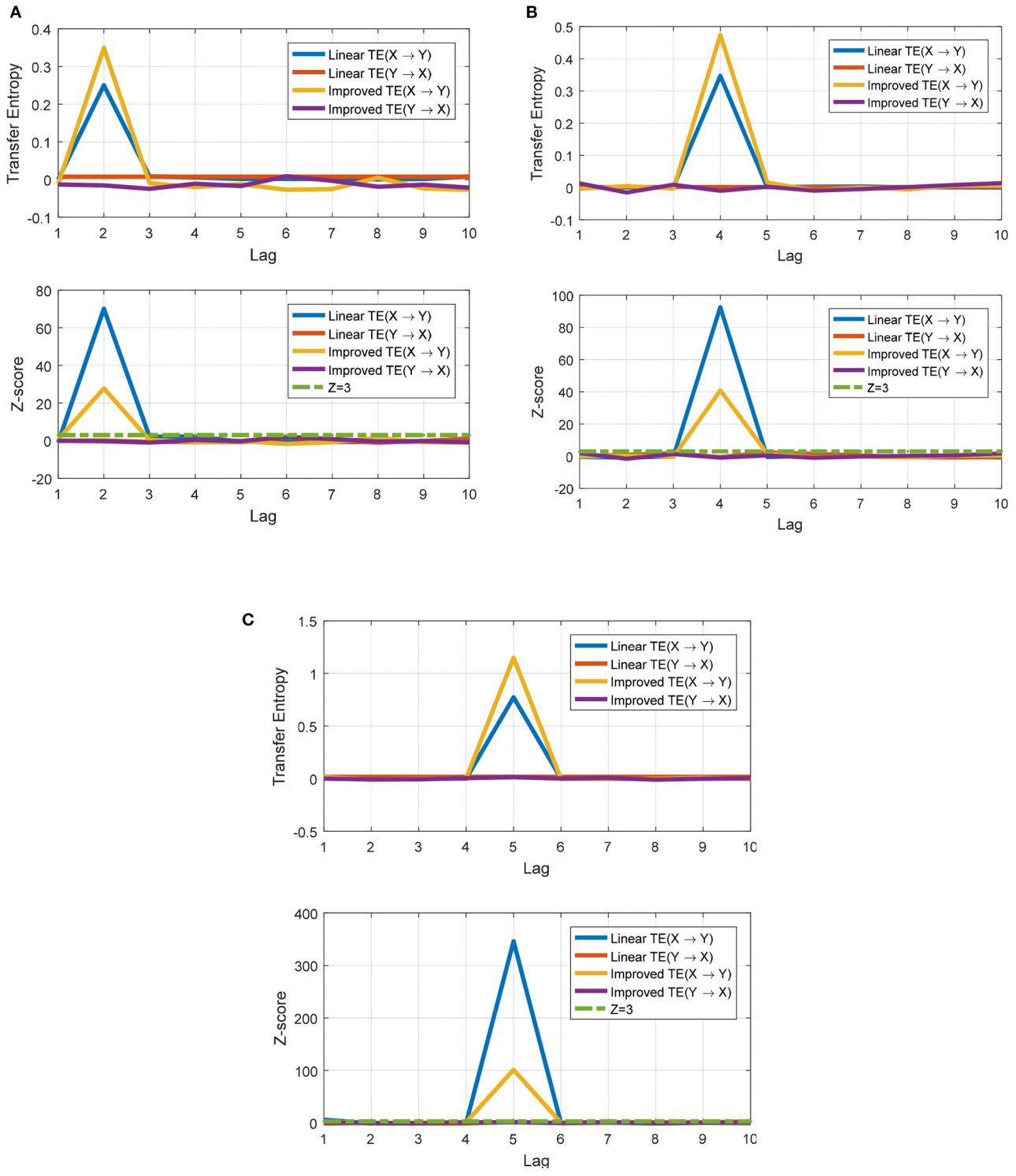


FIGURE 2 | Demonstration that both methods identify the true lag values with maximal transfer entropy. Non-linear transfer entropy is calculated using a quantile-binned histogram, of 6 classes per dimension over 2,500 points. The Z-score for each result is also plotted for both methods. According to the $z > 3$ principle, it can be concluded that for two time series with a single lag order, the two methods can both identify the lag orders accurately. **(A)** $k = 2, \alpha = 0.5$. **(B)** $k = 4, \alpha = 0.5$. **(C)** $k = 5, \alpha = 0.5$.

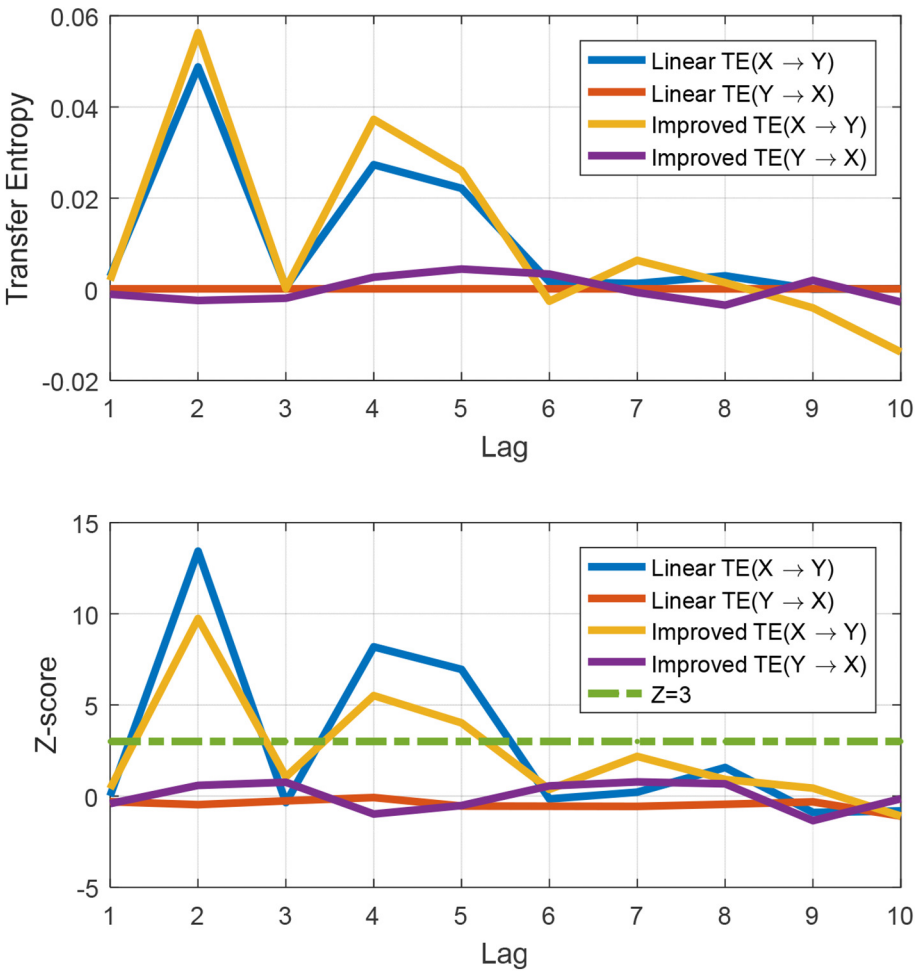


FIGURE 3 | Demonstration that both methods identify the true lag values with maximal transfer entropy. The linear TE could only capture $k = 2$, corresponding to the highest transfer entropy value (Z-score indicates that the value is above a significant level), while the improved TE method could detect $k = 2, 4$, and 5.

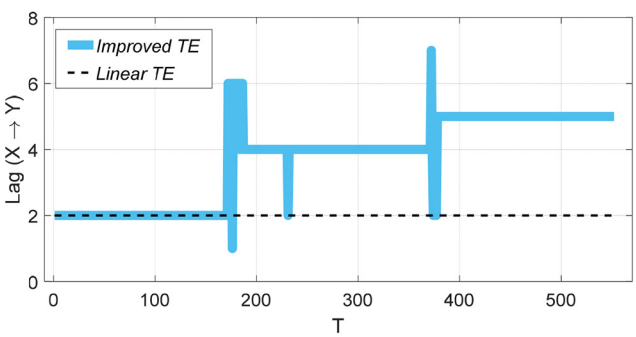


FIGURE 4 | Order identification by the two transfer entropy methods. The dashed line $k = 2$ corresponds to the lag order when the transfer entropy value is the largest in the linear TE method in **Figure 3**.

Granger causality. For the sequences X and Y, using different lag orders, we obtain the causality test results of the two sequences (**Table 1**).

Granger causality is a regression-based interpretation of Wiener's causality definition [48]. In this section, the Granger causality test is employed as a comparison with the improved TE to detect the true lag orders. Following Granger's work [49], we model the Granger causality test with the following two regression equations:

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + u_i \tag{13}$$

$$X_t = \sum_{i=1}^{p} b_i X_{t-i} + \sum_{i=1}^{p} c_i Y_{t-i} + v_t \tag{14}$$

where X denotes the object needed to find the Granger cause, Y denotes the object needed to determine whether it can Granger cause X, and residuals u_t and v_t are assumed to be mutually independent and individually distributed with a zero mean and constant variance. These equations were tested using the following hypothesis:

TABLE 1 | Granger causality test results.

Lag	1	2	3	4	5	6	7	8	9	10
F_test	0.5664	4.0797	2.9568	2.2895	6.6669	5.8331	4.9915	5.0310	4.4082	3.9662
P_val	0.4520	0.0174	0.0319	0.0586	4.76E-06*	6.48E-06*	1.68E-05*	4.75E-06*	1.38E-05*	3E-05*

**p_val < 0.01 indicates that the test result significantly rejects the null hypothesis and that at least one lag variable X Granger causes Y. Therefore, the orders in which X Granger causes Y are 6, 7, 8, 9, and 10.*

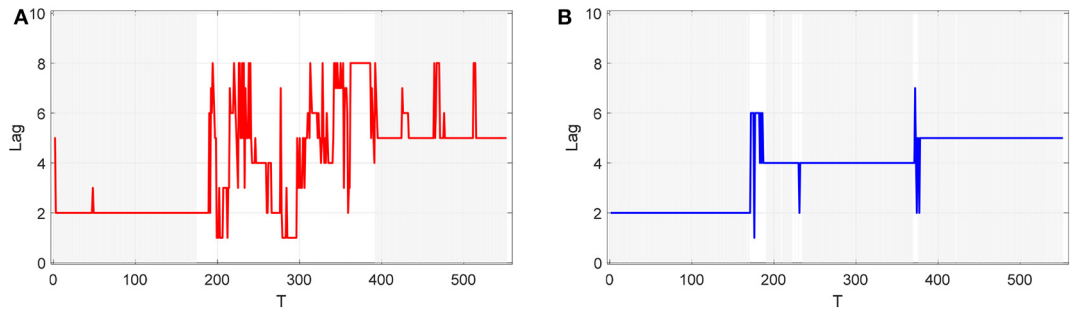


FIGURE 5 | Comparison of the Granger causality method based on a sliding window and the improved TE method. The widow length in both methods is $W = 36$, and the significant level is 1%. The Granger method can clearly identify $k = 2$ but cannot identify $k = 4$, and there is considerable noise interference when identifying $k = 5$. The improved TE method can clearly identify the three orders 2, 4, and 5. **(A)** Granger causality test based on the sliding window method. The gray part indicates that the *p*-value of the *F* statistic is <1%, indicating that the causal relationship is significant in this area. **(B)** The result based on the improved TE method. The gray part indicates that the *Z*-score is higher than 3, which is equivalent to a significance level of $p_val < 0.01$ [section Effective Transfer Entropy (ETE)].

$H_0 : Y$ does not Granger cause $X(c_1 = c_2 = ... = c_p = 0)$.

The $F - test$ can be expressed as follows:

$$F = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(n - 2p - 1)} \sim F(p, n - 2p - 1) \tag{15}$$

where RSS_0 is the residual sum of squares of Equation (13), RSS_1 is the residual sum of squares of Equation (14), n is the number of observations, and p is a lag value. We reject the hypothesis H_0 and accept that Y is a Granger cause of X if and only if $F > F(p, n - 2p - 1)$. The model order p can be determined by minimizing the AIC [50], which is defined as follows:

$$AIC(p) = 2\log(|\sigma|) + \frac{2m^2 p}{\hat{n}} \tag{16}$$

where σ is the estimated noise covariance, m is the dimension of the stochastic process and $\hat{n}$ is the length of the data window used to estimate the model. For example, to detect the causal relationship from exports to US EPU, Y should be set to the exports sequence, while X should be set to the US EPU sequence. In contrast, Y should be set to the US EPU before detecting the causal relationship between US EPU and the exports.

The Granger causality test based on the sliding window method can also obtain the order and significance of two series' correlation. Using Y and X as an example, we elaborate upon the processes of the Granger model estimation within a fixed window as follows:

(a) The maximum of the lag value p is set to a fixed number, such as 10.
(b) By calculating the total AIC of Equations (13) and (14) by traversing the p value from 1 to 10, we obtain the corresponding p of the minimum AIC. The experimental results show that the optimal p is 5.
(c) Equations (13) and (14) are estimated by OLS with $p = 5$.
(d) F and $F(p, n - 2p - 1)$ are calculated according to Equation (15). The results show that $F = 4.0635$ and $F(p, n-2p-1) = 3.8549$ (at the 99% confidence level).
(e) If $F > F(p,n-2p-1)$, we conclude that Y can significantly Granger cause X.
(f) The window is moved forward by a 1-month step, and steps (a–e) are repeated.

Using the process described above, we obtained the Granger causality test results based on a window length $W = 36$ (**Figure 5A**). As shown in **Figure 5A**, although the Granger causality test can identify $k = 2$, it cannot effectively capture the two orders of 4 and 5. Using $k = 4$ cannot pass the significance test; although using $k = 5$ can pass the significance test, there may be other orders, such as $k = 8$. The improved TE method can accurately identify three different orders (**Figure 5B**). In addition, the stage during which the order jumps cannot pass the significance test.

EMPIRICAL RESULTS

Since traditional linear methods cannot identify dynamic orders between time series or track specific lead-lag orders when

structural fluctuations occur, we apply the improved transfer entropy to explore the information flow among EPU, investor sentiment and the stock market.

EPU and Investor Sentiment

Based on the dynamic TE method, we analyze the causal relationship between EPU and investor sentiment. As shown in

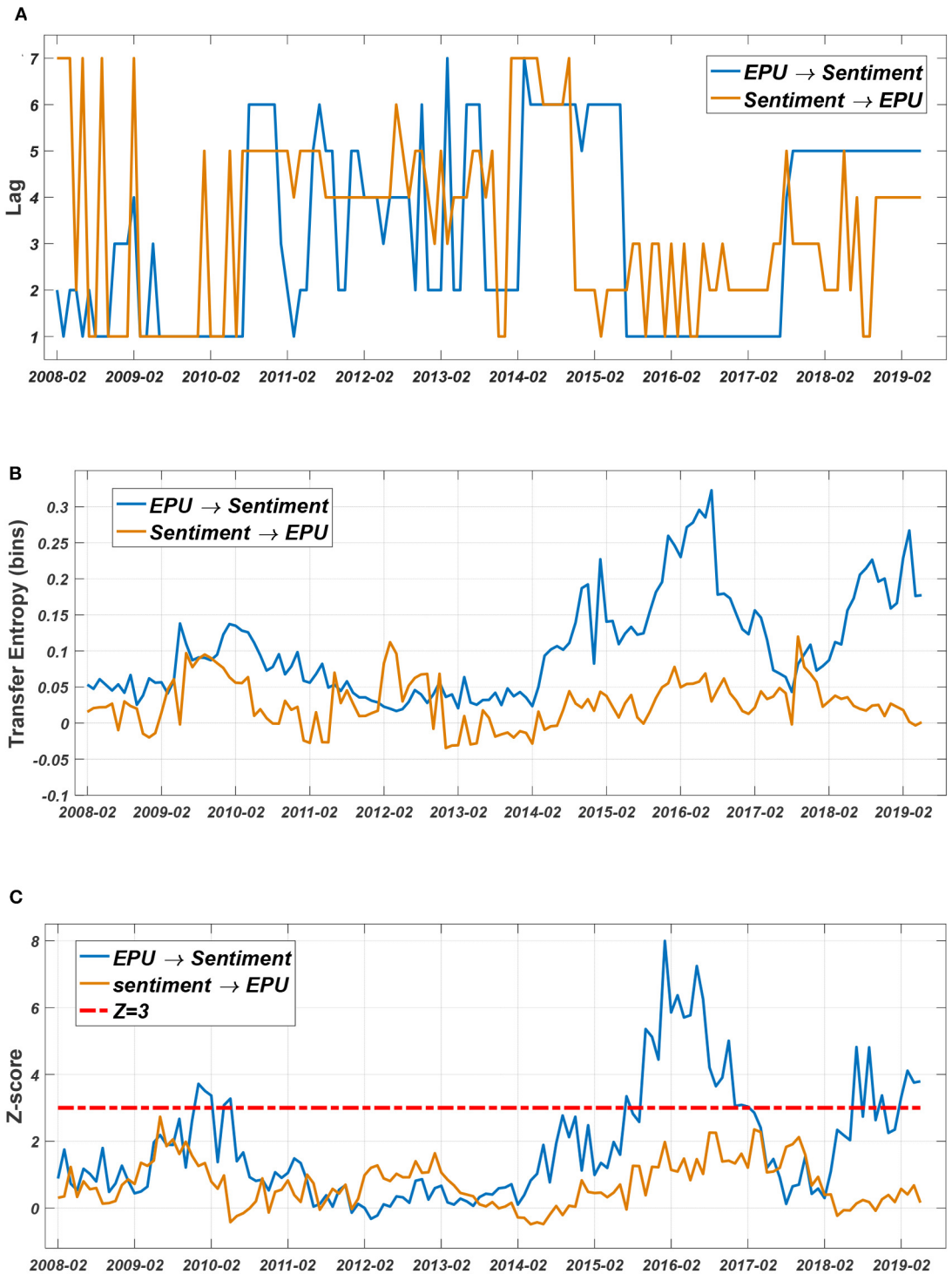


FIGURE 6 | Dynamic Entropy results between EPU and investor sentiment. **(A)** Lag structure. **(B)** TE fluctuation. **(C)** *Z*-score.

Figure 6A, there is an obvious dynamic order in the correlation between EPU and investor sentiment.

As shown in **Figure 6C**, a Z-score > 3 is mainly located in August 2015–June 2016. This means that during this time period, EPU had a significant impact on investor sentiment, and uncertain information about national economic policies significantly affected investor sentiment. From **Figure 6B**, it can also be seen that in this stage, the EPU's transfer entropy to investor sentiment was significantly higher than the impact of investor sentiment on EPU.

The impact on investor sentiment is related to the nature of the policy, i.e., whether the policy is a domestic policy or a foreign policy. During this period, China's economic policy involved the following two important measures: a change in the CNY fixing mechanism and the launch of the "circuit-breaker" mechanism.

On August 11, 2015, the central bank made more reference to the closing price of the previous day in the daily CNY-USD mid-price quotation formation mechanism. This change makes the method of forming the middle price more market-oriented, which more closely reflects the actual supply-demand relationship of the market compared to the previous method.

The circuit-breaker benchmark index is the CSI 300 Index, which uses two thresholds of 5 and 7%. When the CSI 300 Index triggers a 5% breaking threshold, the three exchanges suspend trading for 15 min, and if the 5% is triggered late in the day or 7% is triggered at any time throughout the day, trading is suspended until the market closes. From January 4th to January 7th, the breaking mechanism was implemented for only 3 days, and it became the shortest-lived stock market policy in the history of Chinese securities. This policy uncertainty had a great impact on investor sentiment.

Before August 2015, there were incidents such as the bankruptcy of Lehman Brothers (September 2008), the downgrade of the US sovereign credit rating (August 2011), and the European debt crisis (January 2011–January 2014). However, probably because these events did not occur in China, their impact on consumer sentiment was not significant.

Investment Sentiment Index and the Stock Market

The correlation between sentiment and stock price returns is illustrated in **Figure 7**. As shown in **Figure 7C**, the impact of sentiment on stock returns is non-significant; in contrast, the fluctuation in stock price returns has a significant impact on investor sentiment throughout the time period. This shows that in the Chinese stock market, using emotions to predict changes in stock prices is useless, and investor sentiment is more a lagging reflection of stock price returns. **Figure 7A** shows that the lag time is approximately 1 month. Our results further verify the long-term correlation characteristics suggesting that investor sentiment is mainly affected by fluctuation in the market, which may be related to the existence of cyclical fluctuations in the market and futures arbitrage [40].

The study conducted by Brown and Cliff [29, 51] revealed that the bandwagon effect and bargain shopper effect can offset each other, reducing the predictability of stock returns. The bandwagon effect indicates that higher investor sentiment could increase the stock price, which is reflected in the positive correlation between stock prices and sentiment during the same period; in contrast, the bargain shopper effect indicates that investors optimistically believe the shares at a relatively low price represent a purchase opportunity; therefore, their sentiment negatively changes the returns.

The bargain shopper and bandwagon effect make it difficult to explore the causality between investor sentiment and stock returns. In our analysis results, the bandwagon effect is weaker, and the bargain shopper effect is more significant. The bandwagon effect reflects the herd effect of investors. This effect makes the stock market prone to sudden rises and falls in the short term; it cannot reflect the true value of a company and is not conducive to the healthy and stable development of the stock market.

EPU and Stock Market

If investor sentiment has a significant impact on the stock price, then according to our expectations, national policy information will be transmitted to the stock price through investors' expectations and eventually cause stock price fluctuations; in other words, EPU also has some kind of information transmission relationship with the stock price. However, the results now show that both stock price fluctuations and EPU have an effect on investor sentiment and are not affected by investor sentiment. Therefore, either the stock price fluctuations and EPU have a weak information transmission effect or there is a mutually offsetting effect.

To further verify our assumptions, we explore the information transfer relationship between EPU and stock price returns (**Figure 8**). As shown from the results of **Figure 8C**, there are only a few discontinuous time points with a Z-score > 3 in the entire event period. Overall, the information transmitted by the EPU to the stock market is non-significant; in other words, the EPU has no obvious information transmission relationship with the stock market.

A considerable number of related studies showed that the stock market and EPU are significantly negatively correlated [11, 12, 21, 52]. Regarding the relationship between China's EPU and the stock market, Chen and Chiang [21] also verified that the stock returns in China are negatively correlated with EPU. Notably, the main correlation revealed by Chen and Chiang based on the GARCH method is the overall correlation between sequences. However, we reveal a time-varying relationship between sequences based on non-linear methods. As shown in **Figure 8C**, it can be concluded that in the short term, China's EPU also significantly impacts the stock market during the period from 2011 to 2012 and in 2016, but in the long run, this effect is generally not significant.

DISCUSSION

According to the efficient market hypothesis theory, an efficient market (**Figure 9A**) should reflect all changes in information, including regular investor sentiment changes and shocking policy fluctuations. Therefore, the information flow should flow from

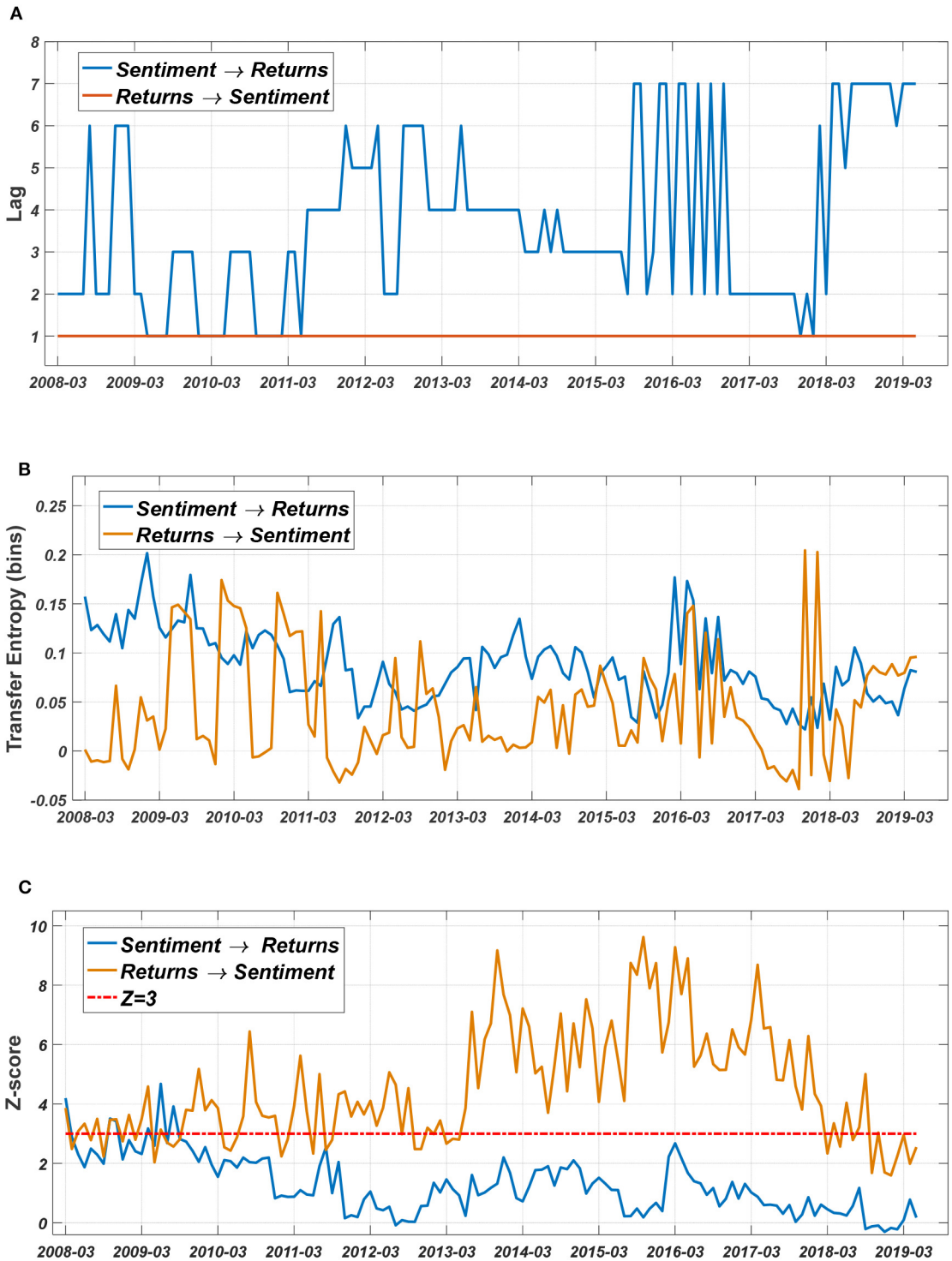


FIGURE 7 | Entropy between investor sentiment and stock market. **(A)** Lag structure. **(B)** TE fluctuation. **(C)** *Z*-score.

the EPU and investor sentiment to the stock market. In addition, since policy shocks often affect sentiment in the short term, information flow should flow from policy to sentiment, but this is uncertain.

The results show that the Shanghai Stock market is not yet an efficient market (**Figure 9B**) and cannot reflect information from regular investment and low-frequency policy shocks. Therefore, investors can reap potential excess

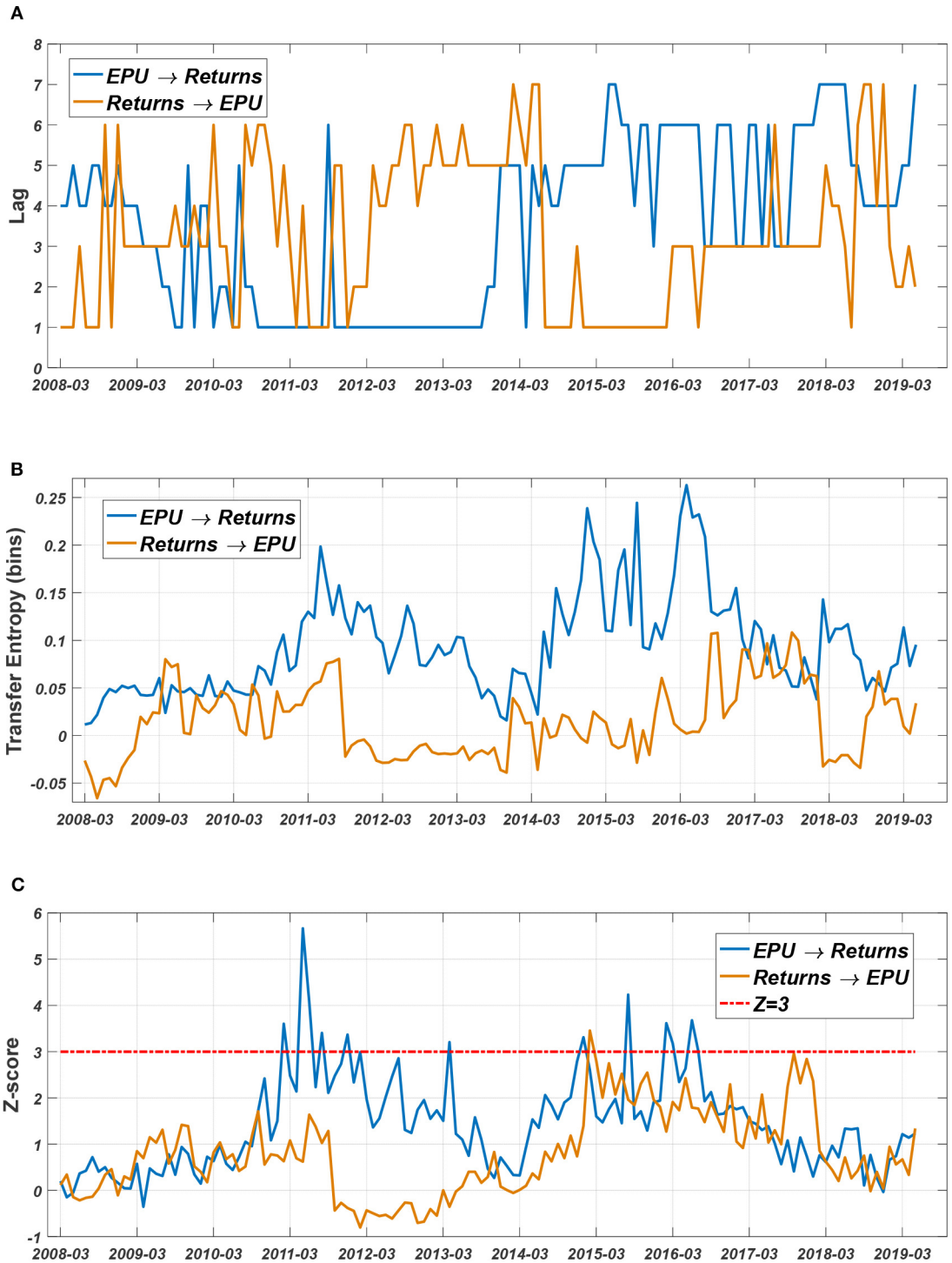


FIGURE 8 | Entropy between EPU and stock market. **(A)** Lag structure. **(B)** TE fluctuation. **(C)** *Z*-score.

profits through operations. Furthermore, the stock market cannot form an effective path to reflect investor sentiment information; thus, in the long run, EPU cannot affect the stock market.

Compared with the market and policy factors, investor sentiment has a certain lag (**Figure 9B**), reflecting the volatility information of the two. Therefore, we should consider policy factors when studying the construction of investor sentiment

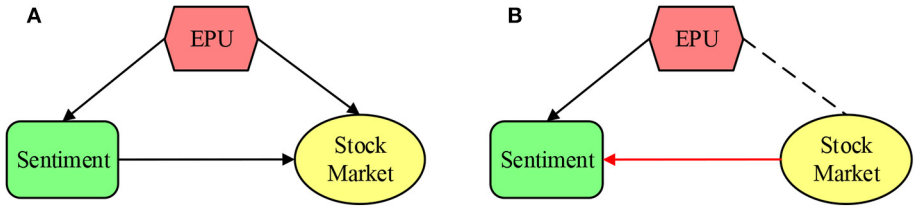


FIGURE 9 | The correlation among CN EPU, investor sentiment and the stock market. **(A)** The efficient market hypothesis. The stock market may be able to effectively reflect information regarding conventional investment and policy shocks. **(B)** China's stock market is an inefficient market, and stock market volatility is an important factor affecting emotional volatility.

indicators, which is rarely investigated in research concerning the factors affecting investor sentiment.

CONCLUSION

A widely accepted fact is that economic policy affects investor sentiment and will be ultimately reflected in the stock market through investment decisions, causing stock price volatility. Therefore, is this really the case?

Since traditional linear methods cannot identify the dynamic orders between time series and are unable to track specific lead-lag orders when structural fluctuations occur, we proposed an improved transfer entropy method based on a sliding window. By comparing with the linear ETE method and Granger causality method, we verify the effectiveness of the improved method. The main advantages of this methodology are the easy implementation-interpretation by non-parametricity to capture the non-linear dynamics and the point in time when the structure changes. Therefore, this method is considered a nice and promising alternative to the standard measures. We further employ this improved method to examine the information flow among EPU, investor sentiment and stock market.

The results of the information flow analysis of EPU and investment sentiment show that EPU influenced investor sentiment mainly from August 2015 to June 2016. Among different policies, China's exchange rate reform policy and "circuit-breaker" policy have played an important role. For other time periods, there are also points in time when policies were highly uncertain, such as the bankruptcy of Lehman Brothers (September 2008), the downgrade of the US sovereign credit rating (August 2011), and the European debt crisis (January 2011–January 2014). However, likely because these events did not occur in China, their impact on consumer sentiment was non-significant.

The analysis of the information flow between sentiment and stock price returns shows that the impact of sentiment on returns is non-significant, while the fluctuation in stock price returns has a significant impact on investor sentiment. Therefore, using emotions to predict changes in stock prices is valueless. Investor sentiment is more a reflection of changes in stock price returns with a 1-month lag order. The results show that in the Chinese stock market, the bargainer effect is more significant and the bandwagon effect is weaker.

There is no direct information flow from EPU to stock market, and according to our previous analysis, there is no indirect information flow through which EPU transmits information to the stock market through investor sentiment. Therefore, stock market fluctuations are basically not affected by national policy fluctuations. Although investor sentiment is affected by changes such as exchange rate reform and stock market policies, this effect is reflected only at the emotional level. Many investors can digest and neutralize extreme emotions. Therefore, a final consensus is not easy to form.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

All authors analyzed and discussed the results, and contributed to the draft paper and revisions.

FUNDING

This work was supported by the National Social Science Foundation of China (Grant No. 19BJL086), Natural Science Foundation of Guangdong Province of China (Grant Nos. 2019A1515010471, 2017A030313396), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No. 17YJAZH109), Fundamental Research Funds for the Central Universities (Grant Nos. 2019MS082, 2020ZDPY19), and Guangzhou National Innovation-oriented City Development Research Center (Grant No. 2017IC02).

ACKNOWLEDGMENTS

We would like to acknowledge the reviewers for their constructive comments.

REFERENCES

1. Kwon O, Yang J-S. Information flow between stock indices. *EPL*. (2008) **82**:68003. doi: 10.1209/0295-5075/82/68003
2. Marschinski R, Kantz H. Analysing the information flow between financial time series. *Eur Phys J B*. (2002) **30**:275–81. doi: 10.1140/epjb/e2002-00379-2
3. Kyrtsou C, Kugiumtzis D, Papana A. Further insights on the relationship between SP500, VIX and volume: a new asymmetric causality test. *Europ J Finance*. (2019) **25**:1402–19. doi: 10.1080/1351847X.2019.1599406
4. Dimpfl T, Peter FJ. Group transfer entropy with an application to cryptocurrencies. *Phys A Stat Mech Appl*. (2019) **516**:543–51. doi: 10.1016/j.physa.2018.10.048
5. García-Medina A, Sandoval L, Bañuelos EU, Martínez-Argüello AM. Correlations and flow of information between the New York times and stock markets. *Phys A Stat Mech Appl*. (2018) **502**:403–15. doi: 10.1016/j.physa.2018.02.154
6. Sensoy A, Sobaci C, Sensoy S, Alali F. Effective transfer entropy approach to information flow between exchange rates and stock markets. *Chaos Soliton Fract*. (2014) **68**:180–5. doi: 10.1016/j.chaos.2014.08.007
7. Yang PB, Shang PJ, Lin AJ. Financial time series analysis based on effective phase transfer entropy. *Phys A Stat Mech Appl*. (2017) **468**:398–408. doi: 10.1016/j.physa.2016.10.085
8. He JY, Shang PJ. Comparison of transfer entropy methods for financial time series. *Phys A Stat Mech Appl*. (2017) **482**:772–85. doi: 10.1016/j.physa.2017.04.089
9. Chung SL, Hung CH, Yeh CY. When does investor sentiment predict stock returns? *J Empir Finance*. (2012) **19**:217–40. doi: 10.1016/j.jempfin.2012.01.002
10. Dakhlaoui I, Aloui C. The interactive relationship between the US economic policy uncertainty and BRIC stock markets. *Int Econ*. (2016) **146**:141–57. doi: 10.1016/j.inteco.2015.12.002
11. Arouri M, Estay C, Rault C, Roubaud D. Economic policy uncertainty and stock markets: long-run evidence from the US. *Finance Res Lett*. (2016) **18**:136––41. doi: 10.1016/j.frl.2016.04.011
12. Antonakakis N, Chatziantoniou I, Filis G. Dynamic co-movements of stock market returns, implied volatility and policy uncertainty. *Econ Lett*. (2013) **120**:87–92. doi: 10.1016/j.econlet.2013.04.004
13. Antonakakis N, André C, Gupta R. Dynamic co-movements between economic policy uncertainty and housing market returns. *J Real Estate Portfol Manage*. (2015) **21**:53–60. doi: 10.5555/1083-5547-21.1.53
14. Wisniewski TP, Lambe BJ. Does economic policy uncertainty drive CDS spreads? *Int Rev Finan Anal*. (2015) **42**:447–58. doi: 10.1016/j.irfa.2015.09.009
15. Li XM, Peng L. US economic policy uncertainty and co-movements between Chinese and US stock markets. *Econ Model*. (2017) **61**:27–39. doi: 10.1016/j.econmod.2016.11.019
16. Hammoudeh S, Kim WJ, Sarafrazi S. Sources of fluctuations in islamic US, EU, and Asia equity markets: the roles of economic uncertainty, interest rates, and stock indexes. *Emerg Mark Financ Tr*. (2015) **52**:1–15. doi: 10.1080/1540496X.2014.998561
17. Sarwar G, Khan W. The effect of US stock market uncertainty on emerging market returns. *Emerg Mark Financ Tr*. (2017) **53**:1796–811. doi: 10.1080/1540496X.2016.1180592
18. Tsai IC. The source of global stock market risk: a viewpoint of economic policy uncertainty. *Econ Model*. (2017) **60**:122–31. doi: 10.1016/j.econmod.2016.09.002
19. Li XL, Balcilar M, Gupta R, Chang T. The causal relationship between economic policy uncertainty and stock returns in China and India: evidence from a bootstrap rolling window approach. *Emerg Mark Financ Tr*. (2016) **52**:674–89. doi: 10.1080/1540496X.2014.998564
20. Yao CZ, Liu C, Ju WJ. Multifractal analysis of the WTI crude oil market, US stock market and EPU. *Phys A Stat Mech Appl*. (2020) **550**:124096. doi: 10.1016/j.physa.2019.124096
21. Chen X, Chiang TC. Empirical investigation of changes in policy uncertainty on stock returns—evidence from China's market. *Res Int Bus Finance*. (2020) **53**:101183. doi: 10.1016/j.ribaf.2020.101183
22. Dai PF, Xiong X, Zhou WX. Visibility graph analysis of economy policy uncertainty indices. *Phys A Stat Mech Appl*. (2019) **531**:121748. doi: 10.1016/j.physa.2019.121748
23. Jiang YH, Zhu ZX, Tian GY, Nie H. Determinants of within and cross-country economic policy uncertainty spillovers: evidence from US and China. *Finance Res Lett*. (2019) **31**:195–206. doi: 10.1016/j.frl.2019.08.004
24. Chen L, Du Z, Hu Z. Impact of economic policy uncertainty on exchange rate volatility of China. *Finance Res Lett*. (2020) **32**:101266. doi: 10.1016/j.frl.2019.08.014
25. De Long JB, Andrei S, Lawrence HS, Robert JW. Noise trader risk in financial markets. *J Polit Econ*. (1990) **98**:703–38. doi: 10.1086/261703
26. Baker M, Wurgler J. Investor sentiment and the cross-section of stock returns. *J Finance*. (2006) **61**:1645–80. doi: 10.1111/j.1540-6261.2006.00885.x
27. Fisher KL, Statman M. Investor sentiment and stock returns. *Financ Anal J*. (2000) **56**:16–23. doi: 10.2469/faj.v56.n2.2340
28. Baker M, Stein JC. Market liquidity as a sentiment indicator. *J Financial Mark*. (2004) **7**:271–99. doi: 10.1016/j.finmar.2003.11.005
29. Brown GW, Cliff MT. Investor sentiment and asset valuation. *J Bus*. (2005) **78**:405–40. doi: 10.1086/427633
30. Baker M, Wurgler J. Investor sentiment in the stock market. *J Econ Perspect*. (2007) **21**:129–52. doi: 10.3386/w13189
31. Kaplanski G, Levy H. Sentiment and stock prices: the case of aviation disasters. *J Finan Econ*. (2010) **95**:174–201. doi: 10.1016/j.jfineco.2009.10.002
32. Corredor P, Ferrer E, Santamaria R. Investor sentiment effect in stock markets: stock characteristics or country-specific factors? *Int Rev Econ Finance*. (2013) **27**:572–91. doi: 10.1016/j.iref.2013.02.001
33. Yang C, Gao B. The term structure of sentiment effect in stock index futures market. *N Amer J Econ Finance*. (2014) **30**:171–82. doi: 10.1016/j.najef.2014.09.001
34. Yang C, Zhou L. Investor trading behavior, investor sentiment and asset prices. *N Amer J Econ Finance*. (2015) **34**:42–62. doi: 10.1016/j.najef.2015.08.003
35. Yao CZ, Sun BY, Lin JN. A study of correlation between investor sentiment and stock market based on copula model. *Kybernetes*. (2017) **46**:550–71. doi: 10.1108/K-10-2016-0297
36. Renault T. Intraday online investor sentiment and return patterns in the U.S. stock market. *J Bank Finance*. (2017) **84**:25–40. doi: 10.1016/j.jbankfin.2017.07.002
37. Xu Q, Wang L, Jiang C, Zhang X. A novel UMIDAS-SVQR model with mixed frequency investor sentiment for predicting stock market volatility. *Expert Syst Appl*. (2019) **132**:12–27. doi: 10.1016/j.eswa.2019.04.066
38. Mascio DA, Fabozzi F. Sentiment indices and their forecasting ability. *Int J Forecast*. (2019) **38**:257–76. doi: 10.1002/for.2571
39. David EA, Michael M, Abhay KS. Daily market news sentiment and stock prices. *Appl Econ*. (2019) **51**:3212–35. doi: 10.1080/00036846.2018.1564115
40. Yao CZ, Li HY. Time-varying lead–lag structure between investor sentiment and stock market. *N Amer J Econ Finance*. (2020) **52**:101148. doi: 10.1016/j.najef.2020.101148
41. Corea F. Can twitter proxy the investors' sentiment? The case for the technology sector. *Big Data Res*. (2016) **4**:70–74. doi: 10.1016/j.bdr.2016.05.001
42. Ballinari D, Behrendt S. Structural breaks in online investor sentiment: a note on the nonstationarity of financial chatter. *Finance Res Lett*. (2020). doi: 10.1016/j.frl.2020.101479. [Epub ahead of print].
43. Huang Y, Luk P. Measuring economic policy uncertainty in China. *China Econ Rev*. (2020) **59**:101367. doi: 10.1016/j.chieco.2019.101367
44. Geweke JF. Measures of conditional linear dependence and feedback between time series. *J Amer Statistical Assoc*. (1984) **79**:907–15. doi: 10.1080/01621459.1984.10477110
45. Barnett L, Barrett A, Seth A. Granger causality and transfer entropy are equivalent for gaussian variables. *Phys Rev Lett*. (2009) **103**:238701. doi: 10.1103/PhysRevLett.103.238701
46. Shannon C. A mathematical theory of communication. *Bell Syst Tech J*. (1948) **27**:379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x
47. Keskin Z, Aste T. Information-theoretic measures for non-*linear* causality detection: application to social media sentiment and cryptocurrency prices. *arXiv:1906.05740v2*. (2019).
48. Wiener N. *The Theory of Prediction. Modern Mathematics for Engineers*. New York, NY: McGraw-Hill Book Company Inc (1956).

49. Granger C. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. (1969) **37**:424–38. doi: 10.2307/1912791
50. Akaike H. Fitting autoregressive models for prediction. *Ann Inst Stat Math*. (1969) **21**:243–7 doi: 10.1007/BF02532251
51. Brown GW, Cliff MT. Investor sentiment and the near-term stock market. *J Empir Finance*. (2004) **11**:1–27. doi: 10.1016/j.jempfin.2002.12.001
52. Alexopoulos M, Cohen J. The power of print: uncertainty shocks, markets, and the economy. *Int Rev Econ Finance*. (2015) **40**:8–28. doi: 10.1016/j.iref.2015.02.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yao and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 26 June 2020
doi: 10.3389/fphy.2020.00180

Is Export a Probe for Domestic Production?

Francesco Saltarelli[1†], *Valeria Cimini*[1,2†], *Andrea Tacchella*[3,4], *Andrea Zaccaria*[3*] and *Matthieu Cristelli*[3]

[1] Dipartimento di Fisica, "Sapienza" Università di Roma, Rome, Italy, [2] Dipartimento di Scienze, Università degli Studi Roma Tre, Rome, Italy, [3] Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche, UOS "Sapienza", Rome, Italy, [4] European Commission, Joint Research Centre (JRC), Seville, Spain

Recent works leverage export data to assess country production structure and ultimately country relative competitiveness. These works mostly rely only on the exported part of the total country output for reasons of data availability, homogeneity, and quality. Here we use the World Input-Output Database (WIOD), which offers cross-country harmonized data that accounts both for domestic production and export, to investigate to what extent export is a proxy for domestic production. We find that export mirrors remarkably well domestic production for manufacturing sectors or sectors related to physical goods. Conversely, this relation fades away for service related sectors. We found those relations consistently across most of the 40 countries for which data are available.

Keywords: economic complexity, services, industrial production, WIOD, economic fitness

OPEN ACCESS

Edited by:
Damien Challet,
Université Paris-Saclay, France

Reviewed by:
Matúš Medo,
University of Electronic Science and Technology of China, China
Alessandro Pluchino,
University of Catania, Italy

***Correspondence:**
Andrea Zaccaria
andrea.zaccaria@cnr.it

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *28 January 2020*
Accepted: *27 April 2020*
Published: *26 June 2020*

Citation:
Saltarelli F, Cimini V, Tacchella A, Zaccaria A and Cristelli M (2020) Is Export a Probe for Domestic Production? Front. Phys. 8:180. doi: 10.3389/fphy.2020.00180

1. INTRODUCTION

The last decades have witnessed the building up of the awareness that economic thinking must embrace new paradigms in order to properly tackle the challenges set by the complex and adaptable nature of economic systems [1–3]. This shift has acted as a breeding ground for cross-disciplinary economics and finance theories and has led to a number of flourishing works bridging several fields ranging from network domain to complexity science. To illustrate a small fraction of these approaches we refer for instance to the complex linkage between micro and macro economic fluctuations [4], the non-trivial topology of World Trade Web [5–7], modeling of the inter-bank network [8] to assess financial systemic risk [9, 10], technological and scientific progress modeling [11, 12], and complex firm diversification trajectories [13, 14].

In this work we focus our attention on one of these novel branches, typically known under the name Economic Complexity [15–25], which discusses the determinants of country development and growth [26–35] in a radically new way. Most of the empirical economic literature has tried to explain development pattern differences directly acting and measuring the underlying drivers [35, 36]: the countries' endowments or capabilities which range from *expected* factors as investments, education, etc., to very exotic ones including genetic diversity factors [37, 38].

However, the design of these studies and their general setups often reflect the general vision of economies as complicated rather than complex (adaptive) systems. This means that these empirical analyses tend to look at very limited channels of interaction suggesting direct and simple cause-effects (or in-out effects) [1]. This general frame for the empirical search of development determinants faces critical issues when the systems are increasingly complex and adaptive because internal feedback tends to break *in-out* schemes and lead to the emergence of collective and evolutionary behavior.

Economic Complexity reverses this perspective and starts from the final outputs in order to explain the root of country competitiveness and consequent growth trajectories. It indeed wants to infer country competitiveness from economic outputs, specifically from cross country output differences as they reflect cross country endowment differences which encode relative country strength [16, 18, 19, 39–42]. Conceptually speaking, these approaches are close to PageRank methodology: extracting information from network topology in order to measure nodes' features.

The economic output which is typically leveraged to measure differences of production across countries is countries' export basket, which is a subset of the total output of an economic system. Domestic production represents the remaining part of the economic output of a country. Exports are the preferred output in order to evaluate cross-country differences in terms of productive capabilities as they occur mainly on a competitive basis. A country exporting a product is likely signaling a competitive advantage and proving that country owns the capabilities required to produce that specific product. In addition, exports also offer a number of auxiliary features which make them an ideal candidate for these analyses:

- Export datasets are harmonized across countries, being the result of data collection from customs offices. This means that all countries identify and define in the same way a specific product, making export baskets, once suitably normalized, comparable across countries;
- The value of the flows is often doubly reported, by the exporter and by the importer, allowing to correct many errors and to de-bias inaccurate reporters;
- They are available continuously since the sixties [43, 44];
- They are available up to a very disaggregated level. Considering Harmonized System, products are hierarchically organized using different levels of aggregation identified by the number of digits used for the product code. For instance, 2 digits codes refer to about one hundred aggregated sectors, while 4 digits codes identify more than one thousand different products. Exports are available up to 6 digits. As a reference, 8–10 digits levels specify the level at which single firms compete (i.e., at those levels if two firms produce the same product, they are likely direct competitors). Exports are then available, consistently for all countries, just one level of aggregation higher than the level setting firm competition [44].

In this work we want to address the relationship that exists between export and domestic production of countries. In particular we want to understand to what extent export flows are mirroring production and therefore they can be used as a good proxy to decode the complexity of domestic production. In particular, we want to understand if there exist sector-wise patterns of variability of the probing power of export flows.

Unfortunately, from an operational point of view, we do not have direct, reliable, highly disaggregated, consistent cross-country datasets tracing the structure of internal production, differently from what is available in the bilateral trade network. However, we are still able to test the relation at a more aggregate level. In order to test the validity of this assumption, we leverage a type of data made recently available by a number of scholars. The dataset we will refer in this work is the so-called WIOD [45, 46]. This dataset extends the original Leontief input-output approach, which is usually provided for internal intra-sector flows, at a global scale (further details are provided in the next section). In WIOD we have access to the input and the output flows for 34 sectors, due to both domestic and import/export contribution, for a limited but significant set of 40 countries covering more than 85% of the world GDP in 2008. Additionally, the trades due to the remaining non-covered part of the world are estimated and included in an additional "country" called "Rest of the World" (RoW). We design a number of tests to statistically assess the probing power of export flows along the two possible cross-sections we can explore: first fixing sectors and then fixing countries.

Our results can be summarized as follows:

i) At an aggregate level, exports are a good proxy for internal production for manufacturing sectors and sectors delivering physical goods.
ii) The relationship between internal production and export fades away for service related sectors. This highlights differences between products and services and shows that services exports might not have the same meaning of *tangible* good related exports. This questions approaches aiming to achieve a straightforward extension to the service domain of cross country export differences by treating these class of activities as an extra set of products [47, 48].

We found those relationships consistently for the countries considered and discuss the exceptions in the remainder of the paper.

The paper is structured as follows: in section 2 we present the results of our research. Particularly, section 2.1 describes how we calculate the internal production and the export for each sector of each country considered. We then analyze those data in sections 2.2 and 2.3, respectively, sector by sector and country by country. We conclude in section 3 discussing our findings and presenting an outlook for our work. Finally, section 4 provides technical details on the statistical methods used. In the **Supplementary Information**, we provide further results and analyses supporting the main findings of the paper.

2. RESULTS

2.1. Assessing Domestic Production and Export

The Input-Output analysis, formalized by Leontief [49], provides a picture of the inter-industrial relationships. This kind of analysis gives a matrix representation of the interactions between industrial sectors of a country. The model considers an exchange economy divided into a certain number of industrial sectors in which the output from a sector becomes an input for another. In this way, it is possible to see how much each sector depends upon the others. The idea of Dietzenbacher et al. [45, 46] was to expand the Leontief's approach to world trades so they created the World Input-Output Database (WIOD), in which there are flows,

quantified in current US dollars, exchanged between industrial sectors relative to several countries of the world. The WIOD contains annual time-series of WIOT, collected for a period of 17 years ranging from 1995 to 2011.

From each one of the WIOT we created a network (whose properties have been studied in [50]) as in **Figure 1**. We distinguish three different types of links: (i) self-link, representing the inputs that an industry takes from itself (colored dashed lines in the figure); (ii) link between the same industrial sector in two different countries (gray dotted line in the figure); (iii) link between different industrial sectors in the same country (colored solid line) and among countries (gray solid line). Self-links are mainly due to aggregated industry classification [50] and often represent a large amount of the total sector input/output. We neglect this data together with any link connecting the same industrial sector across countries. Therefore, we keep only links represented by solid lines in **Figure 1**. For each industrial sector s of the country c we define the *internal flow* I_{sc} as the sum of the output flows toward industrial sectors of the same country. Similarly, we define the *export flow* E_{sc} as the sum of the output flows toward industrial sectors belonging to other countries. The sum of I_{sc} and E_{sc} gives the total output flow of country c, industrial sector s.

Internal and export flows show high variability in terms of volume from country to country. A better parameter to estimate the importance of an industrial sector is the *share* with respect to the country's overall internal production or export. Hence, for each industrial sector of each country, we define an internal share i_{sc} and an export share e_{sc}. The former reflects the importance of that sector relative to the country's internal economy while the latter reflects its importance relative to the country's export. Shares are defined as:

$$i_{sc} = \frac{I_{sc}}{\sum_{s'} I_{s'c}}$$

$$e_{sc} = \frac{E_{sc}}{\sum_{s'} E_{s'c}}$$

where s' runs over all the 34 industrial sectors.

The main goal of this work is to measure the similarity and the similarity's statistical significance of domestic and export shares sector-wise and country-wise.

2.2. Sector-Wise Analysis

Let us first consider the sector-wise similarity. We thus want to measure sector-by-sector whether domestic shares mirror export shares for the available countries. Being n the number of countries, we define $\mathbf{d}_s = \{i_{sc^1}, i_{sc^2}, ..., i_{sc^n}\}$ the vector specifying the domestic shares of a product across countries and $\mathbf{ex}_s = \{e_{sc^1}, e_{sc^2}, ..., e_{sc^n}\}$ the vector of the corresponding export shares. We measure the *per sector* similarity as the sample Pearson correlation of the vectors $\mathbf{d}_s$ and $\mathbf{ex}_s$. The limited number of countries ($n = 41$) and the consequent limited statistics make a robust statistical validation of the measured correlation essential. We then require strategies in order to exclude that the sample correlation we observe is associated with a vanishing correlation for the underlying population, i.e., $\rho = 0$: where ρ is the population correlation coefficient. We will denote population correlation with Greek letters while sample correlation by Roman letters. The statistical validation of correlation can be achieved using different strategies; we will perform the most common ones and discuss the similarities of results witnessing the robustness of our basket of analyses. In detail:

- **Mitigation of outliers' role (analysis I)**: to study the typical range of variability of the observed sample correlation coefficient between the domestic and the export shares, we develop a bootstrapping procedure. Unfortunately domestic and export shares occasionally show a broad distribution and therefore we may occasionally fall into an outlier-type regime for some sectors. We then devise a procedure to mitigate the effect of outliers to test the robustness of our findings. The procedure combines a modified bootstrap with a permutation test and it is easily described by means of a concrete example. In **Figure 2A**, we show the scatter plot of the internal shares and export shares, i.e., $\mathbf{d}_s$ and $\mathbf{ex}_s$, for two sectors (namely Electrical and Optical Equipment and Inland Transport). Each point in the graph represents a country. The sample correlation coefficient of these data is calculated through a *modified* bootstrapping to mitigate the possible effects of outliers. We essentially re-sample many subsets of the original pairs (further details are provided in Methods section). This permits to evaluate the typical range of variability of the correlation coefficients as shown by the histogram in **Figure 2B**. We define the sample correlation coefficient r for this sector as the average of the data in this histogram (pointed out by the vertical dotted black line in the same figure panel). To assess the significance of the obtained r we develop a p-value analysis: for each data subset extracted during the bootstrap we calculate the p-value as the results of a permutation test (see section Methods for further details). We then construct the cumulative distribution function of the obtained p-values, shown in **Figure 2C**. A significant correlation is usually attested by a low p-value. This translates in a p-values' cumulative distribution approaching 1 for small p-values. In the examples shown in **Figure 2** this is the case for the "Electrical and Optical equipment" industrial sector, while it is not the case for "Inland transport." We set a threshold $T = 0.15$ to define a sector correlated or not. If the 70th percentile of the p-values distribution is below T then the sector is said to be correlated otherwise it is not. We marked in panel (c) of the same figure the 70th percentile of the data by a dotted black line and the 0.15 threshold by a dashed red line. We see that for the "Electrical and Optical Equipment" sector the 70th percentile of the data is well below the threshold. This means that the internal share $\mathbf{d}_s$ and $\mathbf{ex}_s$ of this sector are significantly correlated as measured by our definition of statistical significance. Vice versa for "Inland Transport" the 70th percentile of the data is greater than the threshold meaning a lack of a significant correlation.
- **Confidence level assessment via simple bootstrapping (analysis II)**: to visualize the comparison of the sample

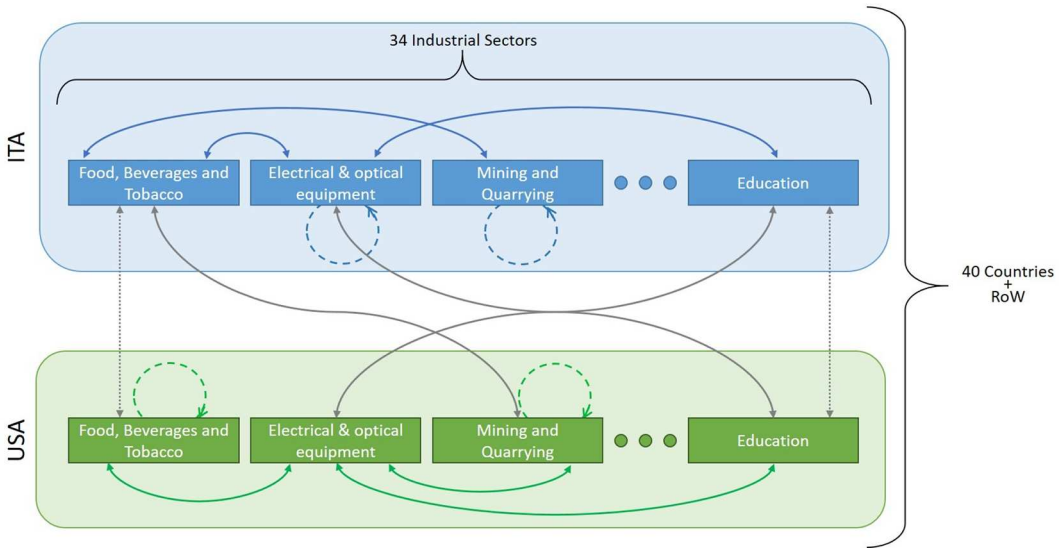


FIGURE 1 | Schematic of the world input-output tables. The WIOT contain data for the domestic production and the export of 34 industrial sectors in 40 countries, plus a model for the remaining countries (named Rest of the World—RoW). In this Figure, for simplicity, we show two countries Italy (ITA) and USA with some industrial sectors. Gray lines represent import/export flows, colored lines trades internal to a country, dashed lines self-consumption of a sector, and dotted lines trades between the same industrial sector of two different countries.

correlation confidence level, we also performed a standard bootstapping procedure. We perform a sampling with replacement of n pairs from the original pairs defining our sample and, by repeating several times this procedure, we can estimate the distribution characterizing the sample correlation variability.

- **Permutation test (analysis III)**: to compare the sample's correlation information with a null model we perform a permutation test shuffling $\mathbf{d}_s$ (or alternatively $\mathbf{ex}_s$) and subsequently measuring the correlation and repeat several times this procedure in order to build the ensemble corresponding to the null case we want to exclude, i.e., the zero correlation scenario. A slightly different way to estimate the sample correlation distribution is to generate n pairs of uncorrelated (normal) random numbers, measure the correlation and repeat the procedure several times. Both procedures allow to define a p-value for the observed sample correlation under the null hypothesis $\rho = 0$. In this work we will provide both approaches.
- **Fisher transformation (analysis III)**: a different approach consists in characterizing the statistics of the sample correlation provided the value of the population correlation ρ. Unfortunately this statistics is Gaussian only for zero population correlation preventing the use of t-statistics. However, it has been shown that the statistics of a non-linear transformation of the sample correlation r - the Fisher transformation - is approximately Gaussian. In detail the variable $F(r) = 1/2[\ln((1+r)/(1-r))]$ approximately follows a Gaussian distribution with mean $\mu_\rho = F(\rho)$ and variance $\sigma_\rho^2 = 1/(n-3)$ where n is the sample size. It follows that the p-value of the sample correlation r under the null hypothesis $\rho = 0$ can be retrieved from the z-score $z = (F(r) - \mu_\rho)/\sigma_\rho = F(r)\sqrt{n-3}$.

2.2.1. Sector Analysis I: Outlier Mitigation

In **Figure 3**, we present the 70th percentile p-value for all the sectors in the years from 1996 to 2011. They are sorted by the p-value in 2011 and the sector names belonging to the services [51] are in bold text. We identify sectors for which the correlation is validated and sectors for which is not. Visually, we see that the service sectors are mostly at the bottom of the figure and they present a large p-value for most of the years analyzed. This reflects the fact that for those factor there is not a statistically significant correlation between the domestic production and the export. The three analyses underline the same trend in terms of validated and non-validated sectors (see **Supplementary Information** for detailed graphs).

In general, a clear clustering is present between two categories of sectors:

- *Sectors showing a statistically significant correlation*: "Wood and Products of Wood and Cork," "Agriculture, Hunting, Forestry and Fishing," "Textiles and Textile Products," "Mining and Quarrying," "Leather, Leather and Footwear," "Pulp, Paper, Printing and Publishing," "Basic Metals and Fabricated Metal," "Electrical and Optical Equipment," "Post and Telecommunication." We note that, with the only exception of "Post and Telecommunication," all these sectors belong to the manufacturing and raw materials industries.
- *Sectors not showing a significant correlation*: "Inland Transport," "Health and Social Work," "Public Admin and Defense; Compulsory Social Security," "Air Transport," "Other Supporting and Auxiliary Transport Activities; Activities of Travel Agencies," "Electricity, Gas, and Water

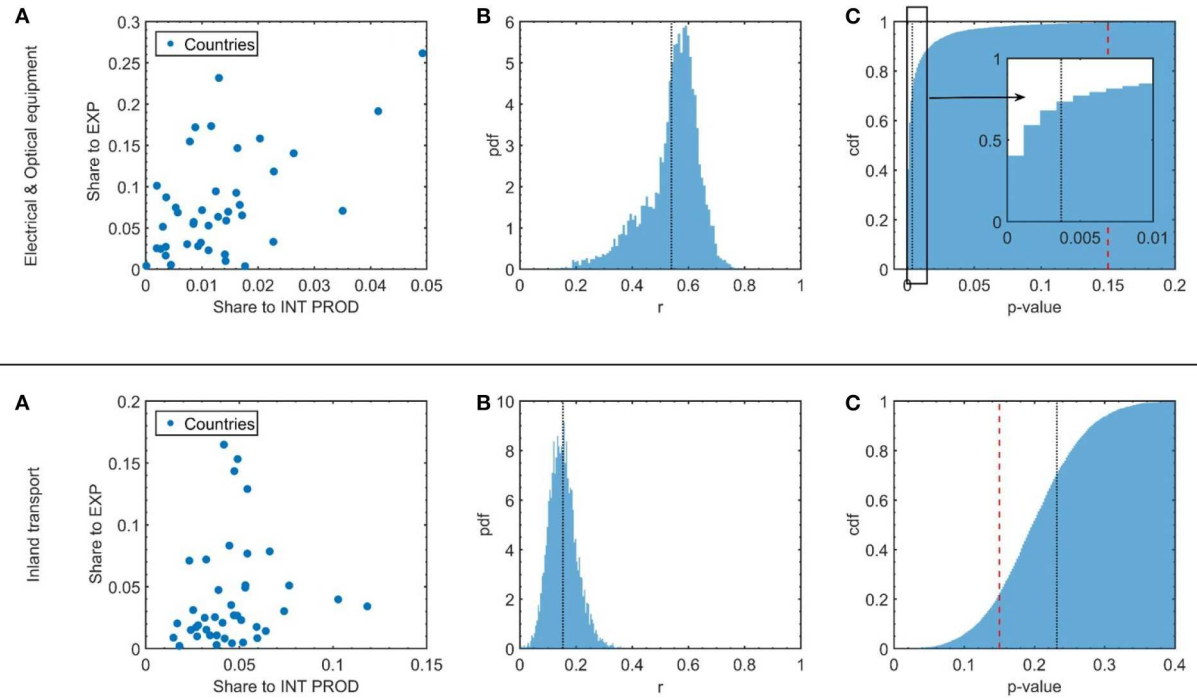


FIGURE 2 | Correlation assessment. The first row refers to a sector where output shares toward export and toward internal production are highly correlated; the second row, a sector for which the correlation is weak. For each row, we present: **(A)** the scatter plot of countries' export share vs. countries' internal-production share; **(B)** the histogram of the bootstrap replications and **(C)** the cumulative distribution of the *p*-values relative to the bootstrap replications in **(B)**, obtained through data reshuffling. The red dashed line represents the threshold we use to define correlations.

Supply," "Hotels and Restaurants," "Sale, Maintenance and Repair of Motor Vehicles and Motorcycles; Retail Sale of Fuel," "Retail Trade, Except of Motor Vehicles and Motorcycles; Repair of Household Goods," "Water Transport," "Real Estate Activities," "Wholesale Trade and Commission Trade, Except of Motor Vehicles and Motorcycles," "Financial Intermediation." Those sectors, all belonging to services, present a *p*-value higher than T for most of the years considered in this analysis.

The general ordering of the sectors in terms of the significance of the measured correlation and in particular the different behavior for manufacturing and service sectors is robust across all the years available and not specific of a limited time period. However, few exceptions and trends can be spotted. A more explicit visualization of the evolution of significance in time is provided in **Figure S3** where we show the time evolution of the 70th percentile *p*-value from 1996 to 2011 for each sector. We identify a temporal trend for some industrial sectors. In particular "Food, Beverages and Tobacco," "Coke, Refined Petroleum and Nuclear Fuel," "Chemicals and Chemical Products," "Machinery, Nec," and "Transport Equipment" show an increase in the correlation significance between internal share and export share in the period considered. On the contrary, the industrial sectors "Other Non-Metallic Mineral," "Manufacturing, Nec; Recycling," "Electricity, Gas and Water Supply," "Sale, Maintenance and Repair of Motor Vehicles and Motorcycles; Retail Sale of Fuel" and "Retail Trade, Except of Motor Vehicles and Motorcycles; Repair of Household Goods" show a clear worsening of the correlation significance between the two quantities with time. Several factors may share a role in shaping the similarity between internal and export production. As a general consideration, an increasing correlation may be a signature for an increasing globalization and a reduction of trade barriers. However, underpinning the origin of the forces underlying these trends is beyond the scope of this work.

The general ordering of sectors by statistical significance induced by the different tests proposed exhibits minor differences. However, two common features are shared by all analyses:

1. Statistical significance of domestic and export shares similarity tends to increase in time. We argue this may due to increasing trade liberalization and openness together with more integrated global value chain;
2. Two groups of sectors emerge in a consistent way in the period under investigation. One group, composed of sectors related to the manufacturing and raw material industries, present a significant correlation between the domestic output and the export. This correlation is instead not significant for another group of sectors composed of service-related sectors.

2.2.2. Sector Analysis II: Simple Bootstrap Results

The correlation coefficient between domestic and export shares of the statistically validated sectors are typically observed in the

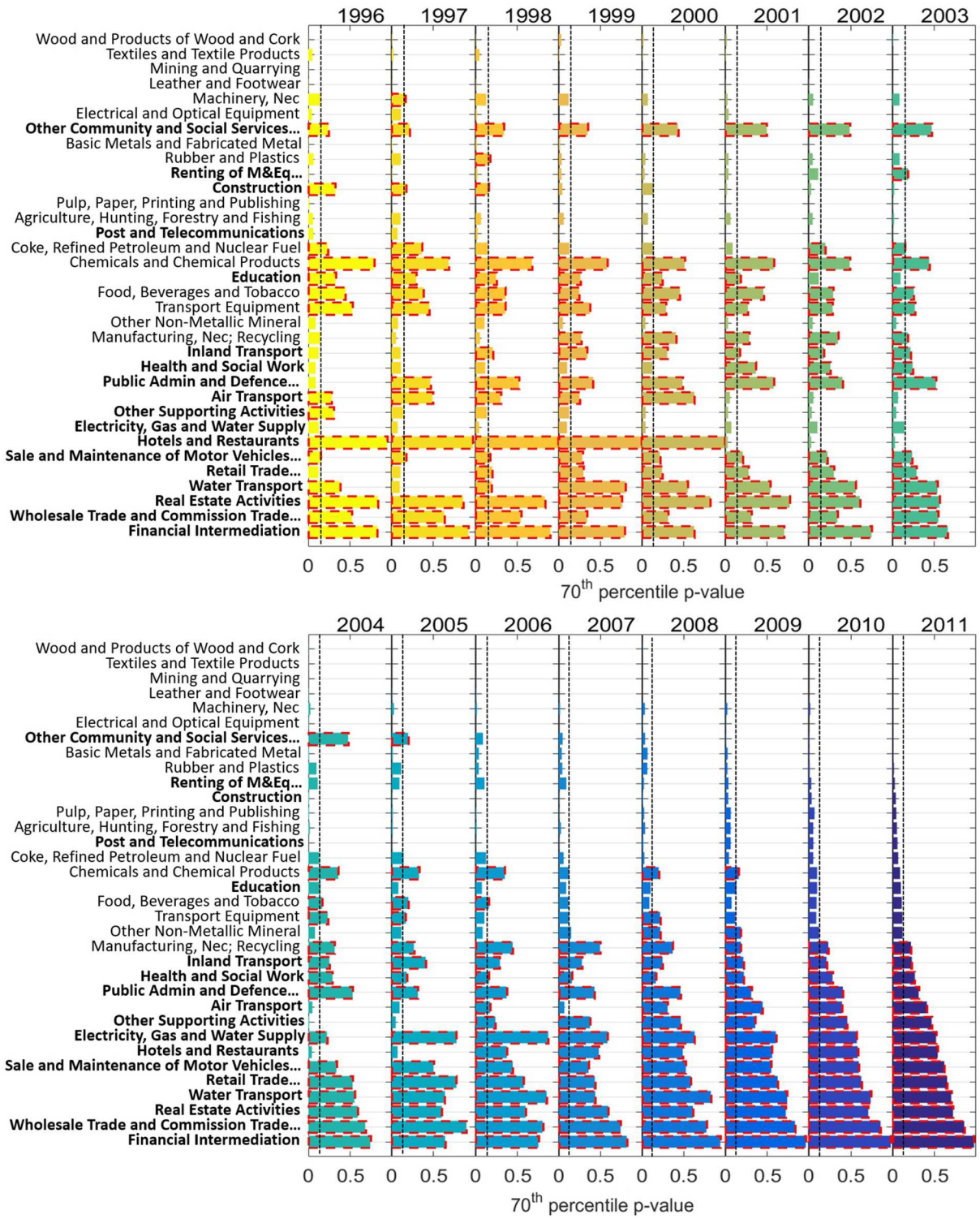


FIGURE 3 | Industrial sectors' correlation. Bars represent the p-values' 70th percentile. Data are sorted according to 2011 p-values. A clear clustering is present: service-related sectors (in bold) do not present a statistically validated correlation between the domestic output and the export. Sectors belonging to the manufacturing and raw material industries have a low p-value hence a robust correlation. Vertical black dashed lines represent the threshold ($T = 0.15$) to define correlated sectors. with red edges have a p-value larger than T. Sectors are sorted according to their 2011 p-value.

range $0.2 - 0.9$ as shown in **Figure S2** and tend to increase over time. Interestingly negative coefficients are usually never statistically validated.

2.2.3. Sector Analysis III: Permutation Test and Fisher Transformation Results

In **Figure S1**, we show two Yes-No grids summarizing the statistical validation of the sample correlations we measure for the 34 sectors available, respectively, for the two permutation tests we propose and the Fisher transformation. Sectors are ordered according to decreasing p-value in 2011. Ordering is similar but few differences apply. A green plain dot corresponds to $p \leq 0.05$, empty red squares to non-validated sectors. As a first result, both strategies provide essentially the same results and, more interestingly, we observe that non-validated sectors tend to be service-related sectors. Detailed p-value tables are provided in the **Supplementary Information**.

2.3. Country-Wise Analysis

So far we have seen that industrial sectors can be approximately classified into two groups on the basis of the statistical significance of the correlation between domestic and export shares. *On average*, export is a good proxy for domestic production for manufacturing sectors. Let us now consider the second cross-section analysis we are interested in: the *country-wise* analysis. Specifically, we want to investigate whether there exist country-specific patterns for the relationship between export and domestic production. In this section we will deal directly with flows (I_{sc} and E_{sc}) instead of shares since we do not have scaling issues between flows from different countries, as in the previous section.

We estimate the statistical robustness of per country correlations with the same methods we used for sector-wise analyses, namely permutation tests, Fisher transformation, bootstrap and the test with mitigation of outliers' effects. Referring to **Figure 2** again, the general framework is similar: we still consider scatter plots as in panel (a) but now they have $\log(I_{sc})$ and $\log(E_{sc})$ on the horizontal and vertical axis, respectively, i.e., the internal flows of sectors and the export flow of the same sectors for a specified country.

In this section, we will run all the statistical tests proposed with two setups: excluding and keeping those sectors which are discarded by the outlier mitigated test discussed in the previous section. Discarded sectors can be retrieved year-by-year from **Figure 3** (they are identified by a red dashed edge).

2.3.1. Country Analysis I: Outlier Mitigation

We define as $p^{all}_{c,yr}$ and $p_{c,yr}$ the 70th percentile p-value calculated with all sectors and with only the validated sectors, respectively. We present the results of this analysis in **Figure 4** for the years ranging from 1996 to 2011. Solid black lines represent $p^{all}_{c,yr}$ while colored bars represent $p_{c,yr}$. The color of the bar is:

- Light green if $p^{all}_{c,yr} > T$ and $p_{c,yr} < T$, i.e., the correlation is statistically significant only excluding the discarded sectors in the sector-wise analysis;
- Dark green if $p^{all}_{c,yr} < T$ and $p_{c,yr} < T$, i.e., in both cases the correlation is statistically significant. However, we note that $p^{all}_{c,yr}$ tends to be always larger than $p_{c,yr}$, therefore the case excluding services tends to be more significant from a statistical point of view;
- Dark red if $p^{all}_{c,yr} > T$ and $p_{c,yr} > T$, i.e., the correlation is not statistically significant in both cases;
- And light red if $p^{all}_{c,yr} < T$ and $p_{c,yr} > T$, i.e., excluding the discarded sectors decrease the statistical significance of the correlation between internal flows and export flows.

A visual inspection of **Figure 4**—countries are ordered with respect to $p_{c,yr}$ in 2011—reveals that the vast majority of countries show a notable increase in the significance and of the correlation itself after the removal of the non-validated sectors in the sector-wise analysis. This visually corresponds to the fact that empty bars are larger than colored ones for almost all countries. For instance, in year 2011 only 24 countries out of 41 have validated correlation coefficients including all sectors, after removing not-validated sectors, only 3 countries (i.e., France, Romania, and Taiwan) are not validated as statistically correlated in the country-wise analysis.

The second main observation revealed by the visual inspection of **Figure 4** is the presence of a well-defined temporal trend which sees the growth of the number of validated correlation coefficients between export and internal production during the considered period. We have already identified this trend in the sectors' analysis (**Figure 3**) considering that the 70th percentile of the p-value is overall lower in the last years compared to the previous. However, in this perspective the country-wise analysis is a more suitable playground to look at structural changes of trades (**Figure 4**). We observe that there is a clear increase of green bars over time. Light red bars completely disappear after 2008 and as mentioned in the last year available the correlation is validated for 37 countries out of 40.

This implies that country's specific patterns are disappearing and export is a good probe for internal output for the majority of the countries we can test. A tentative explanation of this behavior can be rooted in the evolution and the rise of world trades due to the globalization process and to the reduction of trade barriers in the period studied. In particular starting from 2008 a very high correlation between export and internal production is present for the vast majority of countries taken under exam.

Interestingly, most of the countries for which the correlation fails to be validated can be easily interpreted. Starting with persistent red light bars which are, in the perspective of the previous section, the most surprising cases, we find for instance Luxembourg which is indeed an economy traditionally dominated by services. We also find Italy but, as argued in Di Clemente et al. [13], Italy's economic system has evolutionary features which are peculiar and may affect the internal output. We do not have instead obvious interpretation of Brazil's behavior in the late nineties and Japan's one in the early 00s. Turkey and India trends toward an increasing correlation underpin their rising economic trajectories which is leading both countries to be pivotal nodes in the trade network.

A persistent anomaly with respect to the observed positive trend is represented by Romania where not only the correlation is lacking for all the years considered but also removing the

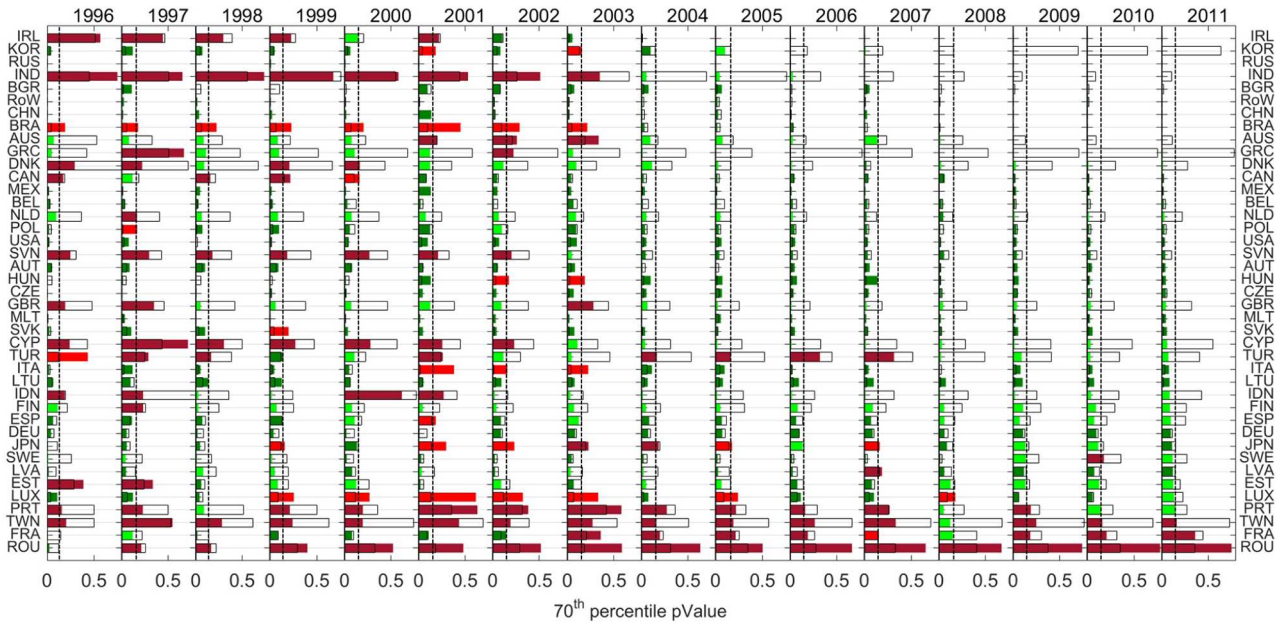


FIGURE 4 | Internal output and export correlation on countries. Bars represent the 70th percentile *p*-values calculated on countries. These are calculated considering all the sectors (solid black lines) and excluding non-correlated industrial sectors (colored bars). We refer to the text for bar color scheme.

non-correlated sectors worsens the situation. Regarding France and Taiwan, in some years the correlation is missing but still we find an improvement by removing the selected sectors. France's trade network appear to have specific features since also in [18] some anomalies have been detected.

2.3.2. Country Analysis II: Simple Bootstrap Results

The correlation coefficient between domestic ad export shares of the statistically validated countries are typically observed in the range $0.0 - 0.8$ as shown in **Figure S5**. The red bands correspond to the case with all sectors while blue bands to the case with validated sectors only.

2.3.3. Country Analysis III: Permutation Test and Fisher Transformation Results

In **Figure S4**, we show two Yes-No grids summarizing the statistical validation of the sample correlations we measure for the 41 countries available, respectively, for the two permutation tests we propose and the Fisher transformation. In all cases we provide the results keeping and discarding non-validated sectors. Countries are ordered according to decreasing *p*-value in 2011. As for sectors, a green plain dot corresponds to $p \leq 0.05$, empty red squares to non-validated countries. As a first result, both strategies provide essentially the same results, a country validated by the permutation tests is also validated by Fisher test. However, major differences apply when we discard non-validated sectors (the small symbols aside larger ones represent the results in this latter case). Discarding non-validated sectors we observe that an increasing trend of validated countries occur and the majority of countries is validated in 2011. Detailed *p*-value tables are provided in the **Supplementary Information**.

3. DISCUSSION

World Input-Output tables allow us to investigate, at an aggregate level, the relationship between the two parts of the economic output of a country: export and domestic production. The former part can be leveraged as a proxy for cross country production differences in order to assess country competitiveness. So a natural question arises, namely to what extent the fully competition-driven part of a country output, the export, is mirroring the domestic production network features and whether significant differences apply between the two parts. Input-Output tables allow making a substantial direct comparison as they provide sector output flows broken down into domestic and foreign contributions. The relation holds country-wise, even if few exceptions exist, as in the case of Romania.

The main finding is instead the existence of a sector-wise pattern of validity of statistical equivalence between domestic and export-destined production. While export mirrors domestic production structure for manufacturing sectors, the relationship fades away for services sectors. This implies that services export cannot be interpreted as in the case of manufacturing or physical goods: on the contrary, services are economic products characterized by an elusive and subtle nature, which shares features of both products and endowments/capabilities.

We point out, however, that this result does not necessarily question a straightforward extension of country competitiveness measures to services [47, 48], by simply making use of data on international trade. Indeed, services are very different in nature from manufacturing, and are far less tradable; this shows up in the results of the present work. However, the economic complexity framework tries to track the hidden capabilities of

countries, and these could emerge in a clearer way by looking to exports than to internal production, given the fact that the international competition plays a major role in the former.

In any case, this analysis is setting constraints and caveats on the general meaning of services export on a competitive basis. Services are economic activities for which geographical localization is often hard. For some of these activities the concept of localization is likely ill-defined, as in the case, for instance, of strategic consulting firms whose teams and project operate worldwide.

The results also provide a perspective to reconcile manufacturing and services sectors in order to join the two dimensions. Starting from those few countries for which export and domestic services are correlated one should first understand at an aggregate level how these two parts are mutually linked. Then, by segmenting countries on the basis of the domestic services diversity similarity, we can try to extend the mapping provided for those countries where the relationship holds to the countries belonging to the same cluster but for which there is a missing correlation between domestic services and services export. Provided in this way a scheme to estimate how to reconcile export and domestic services diversity and a re-scaling profile for each country, this mapping can be eventually extended at a disaggregate level.

4. METHODS

4.1. Datasets

We used data extracted from World Input Output Tables (WIOT) [45, 46]; they consist in 17 different tables, one for each year from 1995 to 2011. The structure of the table is a matrix that lists economic sectors associated to countries, in the same sequence, both vertically and horizontally. Values on the column represent inputs for the industrial sector and the country at the beginning of the column, expressed in monetary value; while the values on the row represent outputs from the sector and the country at the beginning of the row. Thus, any sector can be analyzed in terms of the direction and amount of its inputs and outputs. We used only the information relative to the fluxes exchanged between industrial sectors of all the countries considered in the database, which covers 27 European countries and 13 other major countries in the world. The 40 countries considered cover more than 85% of world gross domestic product (GDP) in 2008. A model for the Rest of the World (RoW), which accounts for the remaining 15% of world GDP, is used to predict the remaining trades. Each table contains fluxes in current US dollars between 35 industrial sectors. Fluxes both inside the same economy and toward foreign economies are reported. We use only data for 34 sectors since 'Private Households with Employed Persons' has often null input or output. WIOT also provides data for the final demand, government expenditures, depreciation of capital, taxes, etc. However, we do not use these data for a two-fold reason. On one hand, we are interested in the inter-industrial trades. On the other hand, by performing an analysis of competitiveness for countries as in Tacchella et al. [19] using export flows derived from WIOD dataset, we obverse that correlation with the results of the same analysis run on bilateral trade flows is higher when we remove final consumption, especially when services are included in the analysis. This again points in the direction of a non-trivial relationship between domestically-consumed and exported services.

4.2. Sector Names

Throughout the paper we used shortened versions of the WIOT sectors' names. In **Supplementary Information**, we provide the mapping of those shortened names with WIOD ones.

4.3. Correlation Significance Assessment for Sectors: Outlier Mitigation (Analysis I)

Our aim is to study the correlation between the internal production of a country and its export. We define these quantities correlated if the p-value of the correlation coefficients' distribution is lower then $T = 0.15$. We can in this way exclude having an accidental correlation between internal production and export. As a first step, we need a method that allows eliminating outliers from our data set in a systematic way, so that they do not influence the value of the correlation coefficient. For this reason, we perform a bootstrap using only 80% of data, randomly drawn, and we calculate the correlation coefficient only on these data. We repeat this operation 2, 500 times, in this way it is possible to build an empirical distribution measuring the typical range of the correlation coefficients (as shown in **Figure 3**). In order to assess the statistical robustness of the correlation coefficients, for each bootstrapped subsets we calculated the p-value (this means we now have 2,500 p-values). Each p-value is estimated by reshuffling bootstrapped subset data 5,000 times and by calculating the percentile corresponding to the correlation coefficient of the bootstrapped subset with respect to the correlation distribution obtained from this random ensemble. If the 70th percentile of the p-values distribution is below T then the sector is said to be correlated, otherwise it is not.

It is worth noticing that this approach is robust against noise thanks to the bootstrapping and the calculation of p-values on the bootstrapped data. This is a necessity when dealing with this kind of data, which naturally present outliers and a component of noise.

4.4. Correlation Significance Assessment for Countries: Outlier Mitigation (Analysis I)

When we study the correlation between internal production and export relative to each country we deal directly with fluxes instead of shares. Indeed, in this case, we do not mix up data from different countries. Eventually the values that we take for the comparison are the log of the internal flux $\log(I_{sc})$ and the log of the export flux $\log(E_{sc})$.

The procedure we adopted to establish the correlation is exactly the same used for products. We obtain the p-value relative to the 70th percentile of the distribution if its lower than T for that country export is a good probe of internal production otherwise it is not.

DATA AVAILABILITY STATEMENT

WIOD dataset is publicly available and can be downloaded from the website (www.wiod.org/home).

AUTHOR CONTRIBUTIONS

MC, AT, and AZ developed the logic for these ideas and designed the analysis. MC, FS, and VC performed the numerical simulations and experiments. All five authors were involved in writing the paper. All authors developed the formal analysis and explored the data.

FUNDING

CNR Progetto di Interesse CRISISLAB covered the salaries of MC, AT, and AZ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

MC, AT, and AZ thank CNR Progetto di Interesse CRISIS LAB for supporting their research activity.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00180/full#supplementary-material

REFERENCES

1. Bouchaud J-P. Economics needs a scientific revolution. *Nature*. (2008) **455**:1181. doi: 10.1038/4551181a
2. Romer P. The trouble with macroeconomics. *Am Econ*. (2016).
3. Hommes CH. Financial markets as nonlinear adaptive evolutionary systems. *Quant Fin*. (2001) **1**:149–67. doi: 10.1080/713665542
4. Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A. The network origins of aggregate fluctuations. *Econometrica*. (2012) **80**:1977–2016. doi: 10.3982/ECTA9623
5. Squartini T, Garlaschelli D. Analytical maximum-likelihood method to detect patterns in real networks. *N J Phys*. (2011) **13**:083001. doi: 10.1088/1367-2630/13/8/083001
6. Garlaschelli D, Loffredo MI. Structure and evolution of the world trade network. *Phys A Stat Mech Appl*. (2005) **355**:138–44. doi: 10.1016/j.physa.2005.02.075
7. Fagiolo G, Reyes J, Schiavo S. World-trade web: Topological properties, dynamics, and evolution. *Phys Rev E*. (2009) **79**:036115. doi: 10.1103/PhysRevE.79.036115
8. Cimini G, Squartini T, Gabrielli A, Garlaschelli D. Estimating topological properties of weighted networks from limited information. *Phys Rev E*. (2015) **92**:040802. doi: 10.1103/PhysRevE.92.040802
9. Battiston S, Puliga M, Kaushik R, Tasca P, Caldarelli G. Debtrank: Too central to fail? Financial networks, the fed and systemic risk. *Sci Rep*. (2012) **2**:541. doi: 10.1038/srep00541
10. Poce G, Cimini G, Gabrielli A, Zaccaria A, Baldacci G, Polito M, et al. What do central counterparty default funds really cover? A network-based stress test answer. *J Netw Theory Fin*. (2018) **4**:43–57. doi: 10.21314/JNTF.2018.047
11. Farmer JD, Lafond F. How predictable is technological progress? *Res Policy*. (2016) **45**:647–65. doi: 10.1016/j.respol.2015.11.001
12. Pugliese E, Cimini G, Patelli A, Zaccaria A, Pietronero L, Gabrielli A. Unfolding the innovation system for the development of countries: coevolution of science, technology and production. *Sci Rep*. (2019) **9**:16440. doi: 10.1038/s41598-019-52767-5
13. Di Clemente R, Chiarotti GL, Cristelli M, Tacchella A, Pietronero L. Diversification versus specialization in complex ecosystems. *PLoS ONE*. (2014) **9**:e112525. doi: 10.1371/journal.pone.0112525
14. Pugliese E, Napolitano L, Zaccaria A, Pietronero L. Coherent diversification in corporate technological portfolios. *PLoS ONE*. (2019) **14**:e223403. doi: 10.1371/journal.pone.0223403
15. Teece DJ, Rumelt R, Dosi G, Winter S. Understanding corporate coherence: theory and evidence. *J Econ Behav Organ*. (1994) **23**:1–30. doi: 10.1016/0167-2681(94)90094-9
16. Hausmann R, Hwang J, Rodrik D. What you export matters. *J Econ Growth*. (2007) **12**:1–25. doi: 10.1007/s10887-006-9009-4
17. Hidalgo CA, Klinger B, Barabási A-L, Hausmann R. The product space conditions the development of nations. *Science*. (2007) **317**:482–7. doi: 10.1126/science.1144581
18. Hidalgo C. A, Hausmann R. The building blocks of economic complexity. *Proc Natl Acad Sci USA*. (2009) **106**:10570–5. doi: 10.1073/pnas.0900943106
19. Tacchella A, Cristelli M, Caldarelli G, Gabrielli A, Pietronero L. A new metrics for countries' fitness and products' complexity. *Sci Rep*. (2012) **2**:723. doi: 10.1038/srep00723
20. Zaccaria A, Cristelli M, Tacchella A, Pietronero L. How the taxonomy of products drives the economic development of countries. *PLoS ONE*. (2014) **9**:e113770. doi: 10.1371/journal.pone.0113770
21. Cristelli M, Tacchella A, Pietronero L. The heterogeneous dynamics of economic complexity. *PLoS ONE*. (2015) **10**:e0117174. doi: 10.1371/journal.pone.0117174
22. Pugliese E, Zaccaria A, Pietronero L. On the convergence of the fitness-complexity algorithm. *Eur Phys J Spcl Top*. (2016) **225**:1893–911. doi: 10.1140/epjst/e2015-50118-1
23. Zaccaria A, Cristelli M, Kupers R, Tacchella A, Pietronero L. A case study for a new metrics for economic complexity: the Netherlands. *J Econ Interact Coord*. (2016) **11**:151–69. doi: 10.1007/s11403-015-0145-9
24. Pugliese E, Chiarotti GL, Zaccaria A, Pietronero L. Complex economies have a lateral escape from the poverty trap. *PLoS One*. (2017) **12**:e0168540. doi: 10.1371/journal.pone.0168540
25. Cimini G, Gabrielli A, Labini FS. The scientific competitiveness of nations. *PLoS ONE*. (2014) **9**:e113470. doi: 10.1371/journal.pone.0113470
26. Smith A, Nicholson JS. *An Inquiry Into the Nature and Causes of the Wealth of Nations*. London: T. Nelson and Sons (1887).
27. Ricardo D. *Principles of Political Economy and Taxation*. London: G. Bell and sons (1891).
28. Schumpeter JA. *Business Cycles*. Vol. 1. New York, NY: Cambridge Univ Press (1939).
29. Aghion P, Howitt P. A model of growth through creative destruction. *Econometrica*. (1992) **60**:323–51. doi: 10.2307/2951599
30. Romer PM. Endogenous technological change. *J Polit Econ*. (1990) **98**:S71–S102. doi: 10.1086/261725
31. Heckscher E. F, Ohlin BG. *Heckscher-Ohlin Trade Theory*. Cambridge: The MIT Press (1991).
32. Leontief W. Domestic production and foreign trade; the American capital position re-examined. *Proc. Am. Philos. Soc*. (1953) **97**:332–49.
33. Leontief W. Factor proportions and the structure of American trade: further theoretical and empirical analysis. *Rev. Econ Stat*. (1956) **38**:386–407. doi: 10.2307/1926500
34. Solow RM. A contribution to the theory of economic growth. *Q J Econ*. (1956) **70**:65–94. doi: 10.2307/1884513
35. Barro RJ. *Economic Growth in a Cross Section of Countries*. Technical Report. National Bureau of Economic Research (1989).
36. Sbardella A, Pugliese E, Zaccaria A, Scaramozzino P. The role of complex analysis in modelling economic growth. *Entropy*. (2018) **20**:883. doi: 10.3390/e20110883

37. Ashraf Q, Galor O. The 'out of Africa' hypothesis, human genetic diversity, and comparative economic development. *Am Econ Rev*. (2013) **103**:1–46. doi: 10.1257/aer.103.1.1
38. D'Alpoim Guedes J, Bestor TC, Carrasco D, Flad R, Fosse E, Herzfeld M, et al. Is poverty in our genes? A critique of Ashraf and Galor, "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development". American Economic Review (Forthcoming). *Curr Anthropol*. (2013) **54**.1:71–9.
39. Lall S, Weiss J, Zhang J. The 'sophistication' of exports: a new trade measure. *World Dev*. (2006) **34**:222–37. doi: 10.1016/j.worlddev.2005.09.002
40. Cristelli M, Gabrielli A, Tacchella A, Caldarelli G, Pietronero L. Measuring the intangibles: A metrics for the economic complexity of countries and products. *PLoS ONE*. (2013) **8**:e70726. doi: 10.1371/journal.pone.0070726
41. Tacchella A, Cristelli M, Caldarelli G, Gabrielli A, Pietronero L. Economic complexity: conceptual grounding of a new metrics for global competitiveness. *J Econ Dyn Control*. (2013) **37**:1683–91. doi: 10.1016/j.jedc.2013.04.006
42. Annunziata MA, Petri A, Pontuale G, Zaccaria A. How log-normal is your country? An analysis of the statistical distribution of the exported volumes of products. *Eur Phys J Spcl Top*. (2016) **225**:1985–95. doi: 10.1140/epjst/e2015-50320-7
43. Feenstra RC, Lipsey RE, Deng H, Ma A. C, Mo H. *World Trade Flows: 1962-2000*. Technical Report. National Bureau of Economic Research (2005). doi: 10.3386/w11040
44. Gaulier G, Zignago S. Baci: international trade database at the product-level (the 1994–2007 version). *CEPII Working Paper 2010–23*. (Paris) (2010). doi: 10.2139/ssrn.1994500
45. Dietzenbacher E, Los B, Stehrer R, Timmer M, De Vries G. The construction of world input-output tables in the wiod project. *Econ Syst Res*. (2013) **25**:71–98. doi: 10.1080/09535314.2012.761180
46. Timmer MP, Dietzenbacher E, Los B, Stehrer R, Vries GJ. An illustrated user guide to the world input-output database: the case of global automotive production. *Rev Int Econ*. (2015) **23**:575–605. doi: 10.1111/roie.12178
47. Stojkoski V, Utkovski Z, Kocarev L. The impact of services on economic complexity: service sophistication as route for economic growth. *PLoS ONE*. (2016) **11**:161633. doi: 10.1371/journal.pone.0161633
48. Zaccaria A, Mishra S, Cader M. Z, Pietronero L. Integrating services in the economic fitness approach. *World Bank Policy Res Work Pap*. (2018). doi: 10.1596/1813-9450-8485
49. Leontief WW. *Input-Output Economics*. New York, NY: Oxford University Press (1986).
50. Cerina F, Zhu Z, Chessa A, Riccaboni M. World input-output network. *PLoS ONE*. (2015) **10**:e0134025. doi: 10.1371/journal.pone.0134025
51. Timmer M, Erumban AA, Gouma R, Los B, Temurshoev U, de Vries GJ, et al. *The World Input-Output Database (WIOD): Contents, Sources and Methods*. Institue for International and Development Economics (2012). Available online at: https://ideas.repec.org/p/lnz/wpaper/20120401.html

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Saltarelli, Cimini, Tacchella, Zaccaria and Cristelli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 31 July 2020
doi: 10.3389/fphy.2020.00296

Grandpa, Grandpa, Tell Me the One About Bitcoin Being a Safe Haven: New Evidence From the COVID-19 Pandemic

Ladislav Kristoufek [1,2]*

[1] Faculty of Social Sciences, Institute of Economic Studies, Charles University, Prague, Czechia, [2] The Czech Academy of Sciences, Institute of Information Theory and Automation, Prague, Czechia

Bitcoin being a safe-haven asset is one of the traditional stories in the cryptocurrency community. However, during its existence and relevant presence, i.e., approximately since 2013, there has been no severe situation on the financial markets globally to prove or disprove this story until the COVID-19 pandemic. We study the quantile correlations of Bitcoin and two benchmarks—the S&P 500 and VIX—and make comparison with gold as the traditional safe-haven asset. The Bitcoin safe haven story is shown and discussed to be unsubstantiated and far-fetched, while gold comes out as a clear winner in this contest even when a broader cryptocurrency index (CRIX) is considered.

Keywords: Bitcoin, safe haven, extreme events, COVID-19, coronavirus

OPEN ACCESS

Edited by:
Siew Ann Cheong,
Nanyang Technological University,
Singapore

Reviewed by:
Chi Xie,
Hunan University, China
Di Jin,
Tianjin University, China
Aviral Kumar Tiwari,
Rajagiri Business School, India

***Correspondence:**
Ladislav Kristoufek
lk@fsv.cuni.cz

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *22 April 2020*
Accepted: *30 June 2020*
Published: *31 July 2020*

Citation:
Kristoufek L (2020) Grandpa, Grandpa, Tell Me the One About Bitcoin Being a Safe Haven: New Evidence From the COVID-19 Pandemic. Front. Phys. 8:296. doi: 10.3389/fphy.2020.00296

1. INTRODUCTION

The history of Bitcoin is tightly connected to its detachment and independence from the standard financial markets and the proclaimed properties that should make it serve as "digital gold" [1]. An important implication of such status is Bitcoin potentially being a safe-haven asset either in addition to or as a replacement of gold itself, which has served as such for decades [2]. A safe-haven asset is an asset in which capital can take refuge when other assets are in distress. The distress situation of the other assets is a clear distinction from being a good diversifier, which traditionally leads to a low or even negative correlation with other assets in the Markowitz portfolio construction logic [3]. An asset might be considered a safe haven if its correlation with other assets during turbulent periods is lower (or at least not higher) than during calm periods [4–7].

The safe-haven status of Bitcoin is one of its cornerstones and narratives in the financial part of the crypto-community, and lately, it has been a popular topic in the scientific literature as well [8–12]. However, its validity had been, by definition, very difficult to properly discuss and test, as empirical tests had lacked the essential part of the safe-haven definition—financial markets in distress. As Bitcoin was developed in 2008 and 2009 [13], its first legendary pizza transaction took place in March 2010, and it gained larger public attention only by 2013, still mostly due to its controversial aspects (such as the Mt. Gox collapse, the darknet, and the Silk Road), it avoided the most turbulent times of the global financial crisis. Further, it took until the middle of 2016 for the Bitcoin markets to reach a stable daily traded volume of more than $100 million. To illustrate the historical perspective, **Figure 1** shows the S&P 500 standardized daily logarithmic returns back to the beginning of 1946 where we find critical historical events with episodes of numerous negative returns of more than five historical standard deviations (the series is demeaned and standardized by the historical mean and standard deviation of the dataset between January 1, 1946, and April

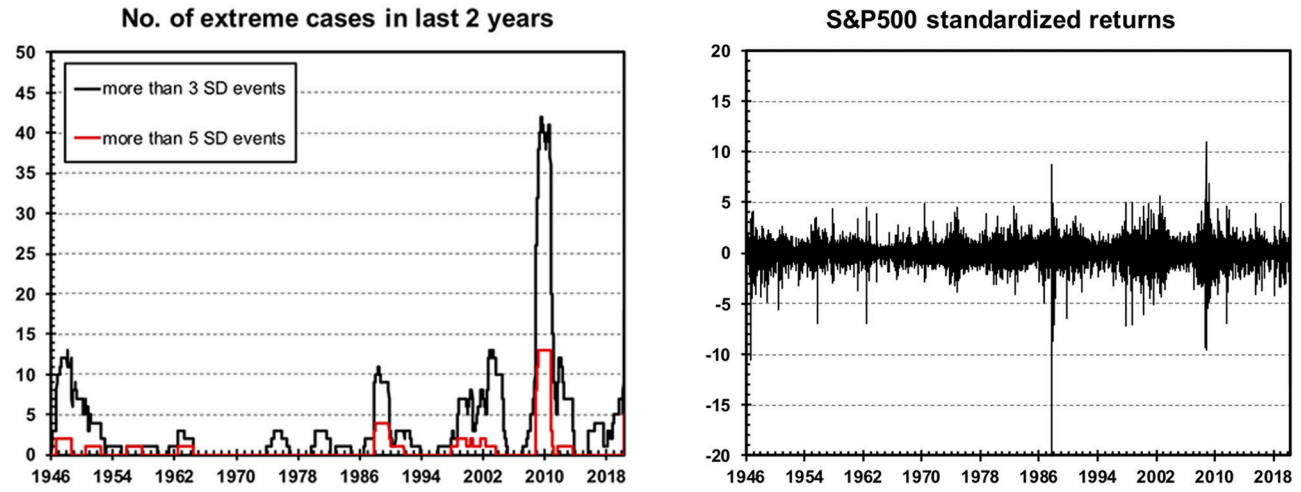


FIGURE 1 | Historical extreme events of the S&P 500. (Left) Logarithmic returns of the S&P 500 between January 1, 1946, and April 17, 2020, demeaned and standardized by the historical mean and standard deviation over the whole examination period. (Right) Number of extreme returns over 3 or 5 standard deviations. Cumulative count on a rolling window of 500 days is shown.

17, 2020). To put the extreme events into a better perspective, the right panel of **Figure 1** shows a number of extreme events above three and five standard deviations on a sliding window of two trading years (500 trading days). There, we see that since 1987, there have been only a few periods of time without these 5-SD critical events. Yet, one of these periods has taken place between 2013 and 2020, i.e., the period of Bitcoin's existence with some palpable trading volume and usage. Only in the days of March 2020 did the financial markets experience severe losses due to fear and uncertainty connected with the COVID-19 (coronavirus disease 2019, caused by the SARS-CoV-2 virus) pandemic, which rapidly spread globally.

Even though the spread of the virus had been assumed possibly to be locally contained, its unprecedented spread caused pervasive panic in global society, which quickly translated to sell-outs and havoc on the financial markets. Purely statistically (and perhaps cynically) speaking, this creates a unique opportunity to test the safe-haven properties of Bitcoin and compare it with gold as the traditional safe haven of choice.

2. RESULTS

We study the interconnection between Bitcoin (BTC) and two benchmarks—the Standard & Poor's 500 (S&P 500) index as a representative of the global financial markets and the CBOE Volatility Index (VIX) as a measure of market uncertainty. We use publicly available data from Yahoo Finance, and we also utilize the Bitcoin prices provided there (these reflect CoinMarkepCap.com data), which restricts the analysis to start from September 16, 2014. The ending is April 17, 2020. As Bitcoin is traded on a 24/7 basis and stocks are not, we use the close-close logarithmic returns[1] (rather than open-close) to include the weekend movements of Bitcoin. This gives us 1,405 daily observations.

As the safe-haven property is similar to being a diversifier, i.e., having a low correlation with other assets, but only during critical times, we approach it from a simple perspective of examining correlations between Bitcoin and the other two assets—the S&P 500 and VIX—during critical events. We treat the critical events as rarely occurring, negative events, i.e., events in the (very) low quantiles of the distribution of the baseline asset. For this purpose, we utilize the quantile correlation [14]. For statistical validity, we estimate the quantile correlation coefficient on 1,000 bootstrapped samples (resampling the time index with a replacement) so that we can present not only a point estimate but also confidence intervals.

In **Figure 2**, we see the quantile correlations between BTC and the S&P 500 (left) and between BTC and the VIX (right). The quantile here represents the conditional quantile of the latter asset in the pair, i.e., either the S&P 500 or the VIX. We find that BTC is a good diversifier with respect to the S&P 500 in calm and bullish times, i.e., in the bulk of the distribution and more generally from quantile 0.2 upwards, with correlations very close to zero and the 90 % confidence intervals including the zero correlation. For the very low quantiles below 0.1, the correlation increases up to more than 0.1, precisely to 0.13 for the lowest analyzed quantile of 0.01. The combination of low quantiles of the S&P 500 and a positive correlation signals that BTC drops together with the stock market if the situation is critical. Note that the size of the correlation is still quite low but is well above the levels during the calmer periods and significantly different from zero for the lowest quantiles. For the VIX, which represents the overall mood on the market and expected future

[1]It is certainly up for discussion whether to use returns for the VIX as well. We considered this possible issue and performed the analysis on both the levels and returns of the VIX index. The results are qualitatively the same. Note that the distinction between logarithmic and original series plays no role here, as we apply a quantile-based method (and logarithm is a monotonous transformation).

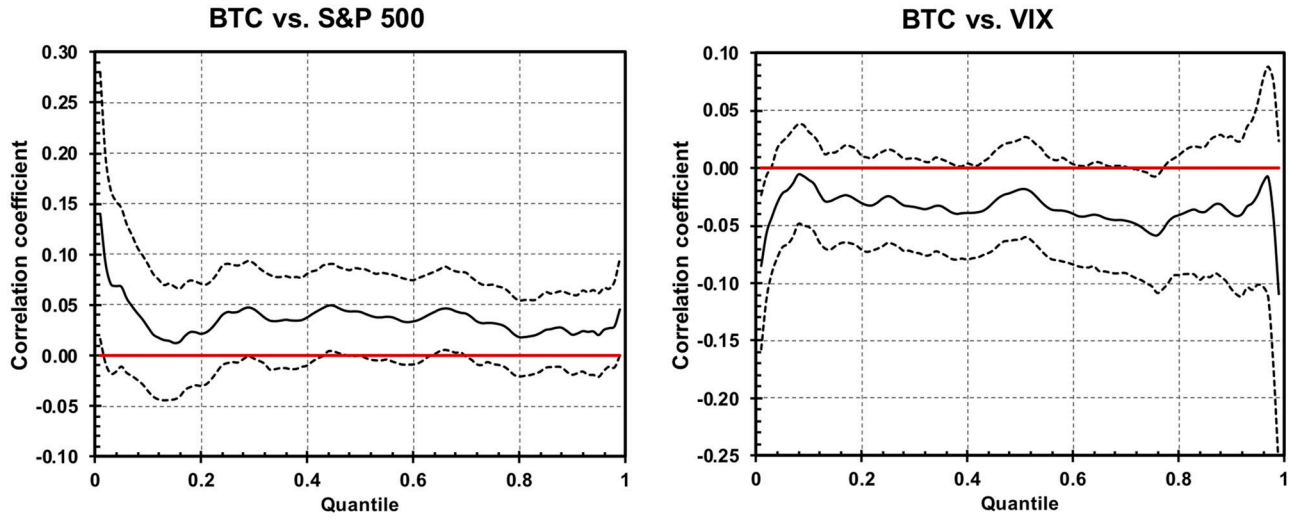


FIGURE 2 | Quantile correlations for Bitcoin. (Left) Quantile correlations between Bitcoin and the S&P 500 index. The quantiles on the *x*-axis are with respect to the S&P 500 index. The low quantiles show the extreme negative events. Black bold curve shows the mean value of 1,000 bootstrapped estimates. The dashed curves show the 90 % confidence intervals based on the bootstrapped estimates. (Right) Quantile correlations between Bitcoin and the VIX index. The quantiles on the *x*-axis are with respect to the VIX index. The high quantiles show the periods of high uncertainty. The other notation holds.

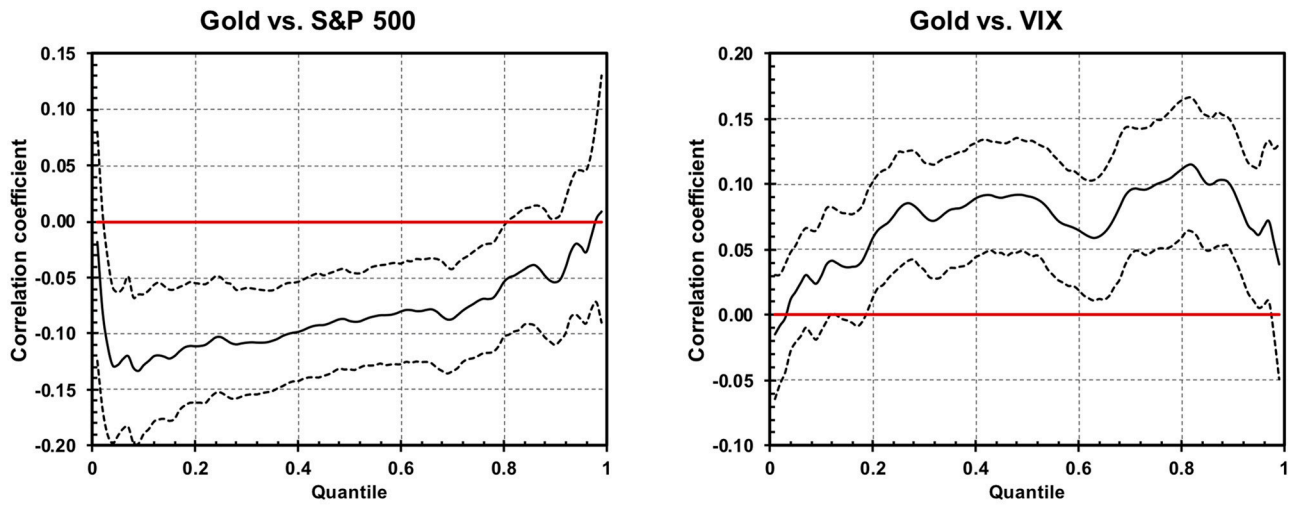


FIGURE 3 | Quantile correlations for gold. (Left) Quantile correlations between gold and the S&P 500 index. (Right) Quantile correlations between gold and the VIX index. The notation from **Figure 2** holds.

uncertainty, we need to look at the high quantiles, as it holds that the higher the VIX is, the higher the uncertainty. For a safe-haven asset, we would expect a low or positive correlation, at least in these high quantiles, or ideally positive correlations for all quantiles. We observe a similar picture as for the S&P 500 case, as the correlation is very close to zero for most situations, but it drops markedly for the times of high uncertainty, which is not a desirable sign for a safe haven.

Comparing the results to the traditional safe haven of gold (**Figure 3**), we see a different picture. In the bulk of the distribution, gold is negatively correlated with the S&P 500, and even though its correlation increases during extreme negative events, its estimate still remains below the zero correlation (statistically speaking not different from zero). With respect to the VIX, gold is positively correlated with it in the bulk of the distribution, and even though its correlation decreases for the most uncertain periods, it still remains above zero. Both of these attributes are the ones we would expect for a safe haven asset, albeit ideally in a more pronounced manner.

3. DISCUSSION AND CONCLUSIONS

The COVID-19 pandemic is the first global economic and financial earthquake that has taken place during the existence and actual use and wider knowledge of Bitcoin, which made it

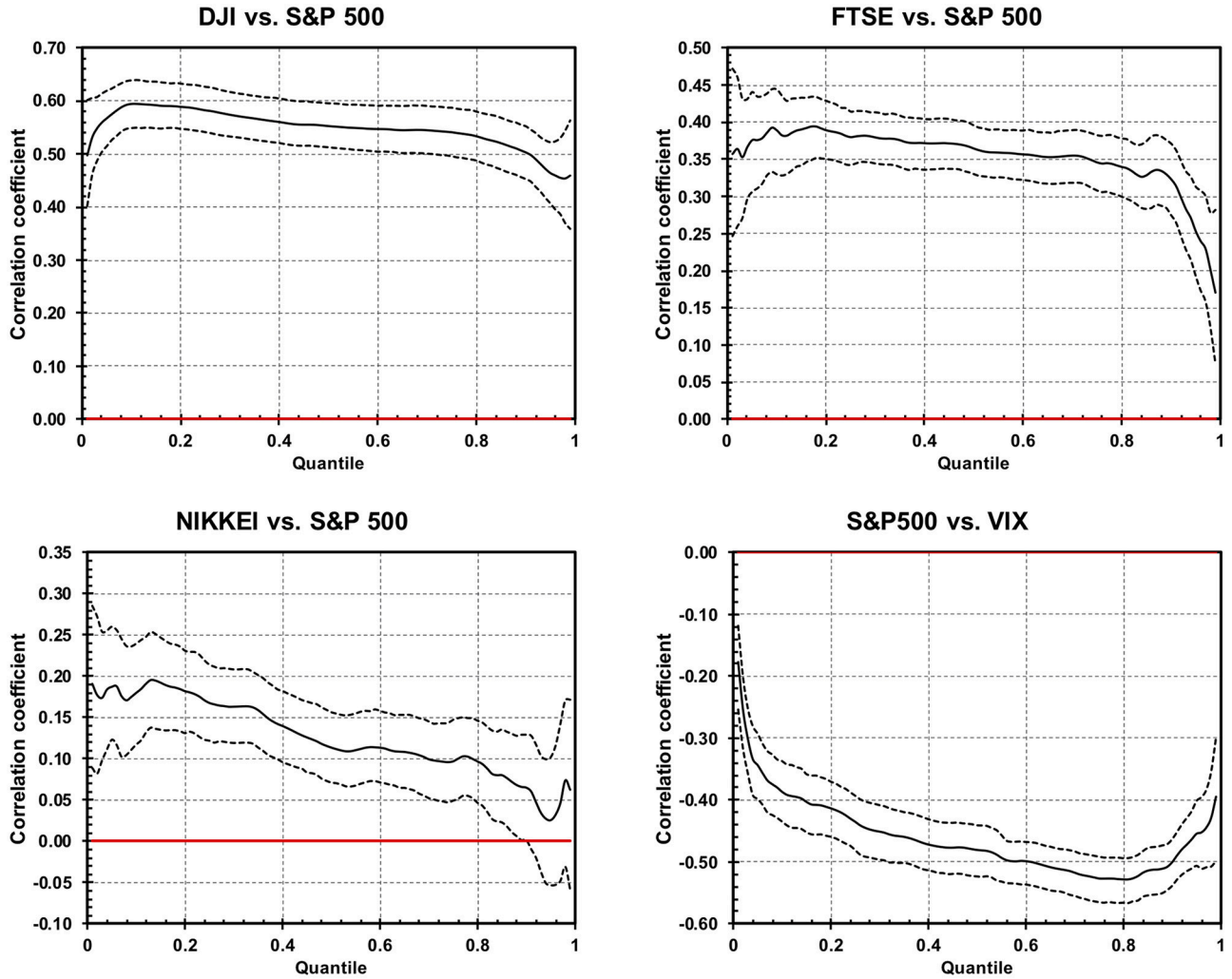


FIGURE 4 | Quantile correlations of the S&P 500 with other assets. The notation from **Figure 2** holds.

possible to put the claims of Bitcoin being a safe-haven asset to an actual empirical examination. We study the quantile correlations between Bitcoin and a pair of global financial benchmarks—the S&P 500 index as the stock market benchmark and the VIX index as a measure of uncertainty and future expectations. What we find is that Bitcoin can easily be considered as a good diversifier, as its correlation with the S&P 500 is close to zero for most of the quantiles. However, its correlation increases markedly during turbulent periods of the S&P 500. The mirror result is observed for its relationship with the VIX index, as the correlation remains close to zero for most quantiles again but drops for the most uncertain times. However, even the extreme-quantile correlations between Bitcoin and either the S&P 500 or the VIX still remain rather low (in absolute terms), and one needs a comparison to fairly comment on its safe-haven properties.

The first comparison is at hand—to gold. This has been presented in the main Results section, but it needs to be stressed that gold shows favorable properties with respect to portfolio diversification utility compared to Bitcoin. It shows negative correlations with the S&P 500 for the bulk of the distribution. The correlations grow for higher quantiles (even though they do not cross to positive ones), i.e., more bullish periods, which is again beneficial. And even though the correlation increases for the lowest quantiles, i.e., the most extreme negative cases, it still collapses to zero, not higher. In addition, we have the connection to the VIX, where gold is again favored in most portfolio-related aspects. We see positive correlations for the bulk of the distribution, i.e., if uncertainty increases, the price of gold increases as well, and for extreme cases, even though the correlation drops, it still remains positive. Therefore, even if we forget about other issues connected to Bitcoin (such as low liquidity, exchange risk, and various legal and accounting/tax issues [15–19]), it does not outperform gold in any important aspect as a safe-haven asset.

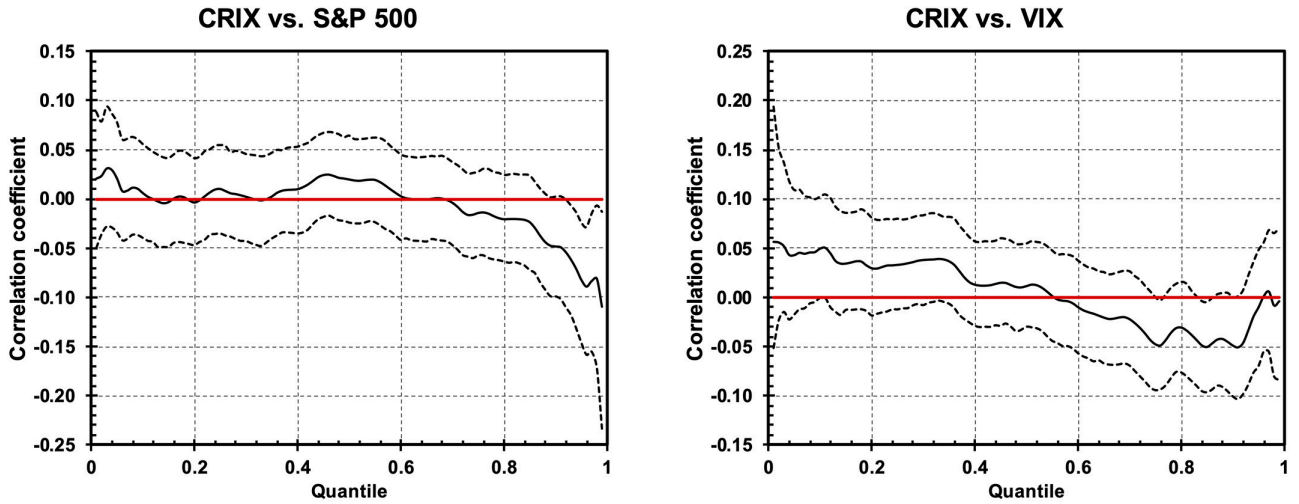


FIGURE 5 | Quantile correlations for CRIX with the S&P 500 (left) and the VIX (right). The notation from **Figure 2** holds.

The second comparison is to other stock indices, mostly to get the correct grasp of the scale of the correlations presented above. In **Figure 4**, we show the quantile correlations of the S&P 500 with the VIX and three other stock indices- the Dow Jones Industrial Average (DJI), Footsie 100 (FTSE), and NIKKEI 225 (NIKKEI)—for the same period of time. There are several interesting observations. First, even for the pair of the S&P 500 and the DJI, the two main US stock indices (in addition to the NASDAQ), the tails correlations are not as strong as one might expect—around 0.5 for both sides of the extreme cases. Second, not surprisingly, the S&P 500 is strongly connected to the VIX. But again, its connection weakens for the extreme cases, more markedly for the calmer periods. Third, the markets are not very correlated during the extremely positive movements of the S&P 500 index, where we find the quantile correlations fall to very low values for both the FTSE and the NIKKEI. And fourth, BTC shows similar properties to the NIKKEI, showing mild correlations for the whole spectrum of quantiles, with slightly higher correlations for the extreme negative movements and practically zero correlation for the extremely positive movements. To be fair, BTC still shows more favorable low-quantile correlations than the NIKKEI does, but not by much.

The last perspective and comparison we provide is to a more general cryptocurrency index. Here, we utilize the CRIX index[2], which is constructed as a capitalization-weighted price index [20]. Currently (mid-2020), it contains around 70% Bitcoin, 10% Ethereum, 5% Ripple, 2.5% Bitcoin Cash and Tether, between 1 and 2% Bitcoin SV, Litecoin, Binance Coin, and EOS, and below 1% OKB (OKEx exchange coin). The quantile correlations between the logarithmic returns of the CRIX index and the two baseline series—the S&P 500 and VIX—are shown in **Figure 5**. Even though the index is majorly formed by Bitcoin, the connection to the stock markets is quite distinct. For the S&P 500 index, we find quantile correlations of practically zero ranging from the lowest quantiles up to around quantile 0.7, where the correlation starts decreasing, and it reaches around −0.1 for the highest quantiles. The diversified cryptocurrency index thus does not follow the stock market index in negative events, which is a good sign for the identification as a safe haven (even though it actually goes against the stock market in the extreme positive events). These results are mostly confirmed for the dynamics between the CRIX and VIX indices, where we observe mildly positive correlations for the lowest quantiles of the VIX (calm periods) that decrease with increasing quantile. Above quantile 0.7, the quantile correlations are of a similar level (around −0.05) as for the relationship between Bitcoin and the VIX index. However, for the most extreme cases, the quantile correlation between CRIX and VIX grows to zero, whereas for Bitcoin, we found a rather sharp drop to negative correlations. The results for the CRIX index are thus more favorable for the safe-haven label when compared to Bitcoin alone. A diversified portfolio of cryptocurrencies has more desirable properties, both as a diversifier (generally lower correlations for all quantiles compared to Bitcoin alone) and as a safe haven (lower correlations with the stock index in the lowest quantiles and correlations closer to zero for the turbulent periods measured by the VIX index). Nevertheless, both of these features are more profoundly represented by gold even compared to the CRIX index.

Overall, we argue that the claim of Bitcoin being a safe haven and an alternative to gold or even being the "digital gold" seems unsubstantiated and far-fetched. This is true even if a broader cryptocurrency index is considered. We do not, however, want to discredit Bitcoin in this aspect completely, as the COVID-19 pandemic and the financial market turmoil induced by it are only the first real tests to its status. In addition, the potential safe-haven properties of Bitcoin and cryptocurrencies in general

[2]The index is described in detail at http://thecrix.de. The volatility CRIX index (VCRIX), in a way parallel to the VIX index, is also available there.

are certainly not the only factor making cryptocurrencies attractive and sought after. Nevertheless, at this point and with respect to the safe haven contest, gold emerges as a clear winner.

DATA AVAILABILITY STATEMENT

The datasets analyzed in this study are publicly available at Yahoo Finance and the CRIX-VCRIX website.

AUTHOR CONTRIBUTIONS

LK contributed to all parts of the presented research.

FUNDING

This work was supported from the Charles University PRIMUS program (project PRIMUS/19/HUM/17) and the Czech Science Foundation (project 20-17295S) was highly appreciated.

REFERENCES

1. Courtois NT, Grajek M, Naik R. Optimizing SHA256 in Bitcoin mining. In: Kotulski Z, Kdiezopolski B, Mazur K, editors. *Cryptography and Security Systems*. Lublin: Springer (2014). p. 131–44. doi: 10.1007/978-3-662-44893-9_12
2. O'Connor FA, Lucey BM, Batten JA, Baur DB. The financial economics of gold–a survey. *Int Rev Financ Anal*. (2015) **41**:186–205. doi: 10.1016/j.irfa.2015.07.005
3. Markowitz HM. Portfolio selection. *J Finance*. (1952) **7**:77–91. doi: 10.1111/j.1540-6261.1952.tb01525.x
4. Baur DG, Lucey BM. Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Financ Rev*. (2010) **45**:217–29. doi: 10.1111/j.1540-6288.2010.00244.x
5. Beckmann J, Berger T, Czudaj R. Does gold act as a hedge or a safe haven for stocks? A smooth transition approach. *Econ Model*. (2015) **48**:16–24. doi: 10.1016/j.econmod.2014.10.044
6. Klein T. Dynamic correlation of precious metals and flight-to-quality in developed markets. *Finance Res Lett*. (2017) **23**:283–90. doi: 10.1016/j.frl.2017.05.002
7. Klein T, Pham Thu H, Walther T. Bitcoin is not the new gold–a comparison of volatility, correlation, and portfolio performance. *Int Rev Financ Anal*. (2018) **59**:105–16. doi: 10.1016/j.irfa.2018.07.010
8. Bouri E, Molnar P, Azzi G, Roubaud D, Hagfors LI. On the hedge and safe haven properties of Bitcoin: is it really more than a diversifier? *Finance Res Lett*. (2017) **20**:192–8. doi: 10.1016/j.frl.2016.09.025
9. Wang G, Tang Y, Xie C, Chen S. Is Bitcoin a safe haven or a hedging asset? Evidence from China. *J Manag Sci Eng*. (2019) **4**:173–88. doi: 10.1016/j.jmse.2019.09.001
10. Smales LA. Bitcoin as a safe haven: is it even worth considering? *Finance Res Lett*. (2019) **63**:322–30. doi: 10.2139/ssrn.3204237
11. Selmi R, Mensi W, Hammoudeh S, Bouoiyour J. Is Bitcoin a hedge, a safe haven or a diversifier for oil price movements? A comparison with gold. *Energy Econ*. (2018) **74**:787–801. doi: 10.1016/j.eneco.2018.07.007
12. Urquhart A, Zhang H. Is Bitcoin a hedge or safe haven for currencies? An intraday analysis. *Int Rev Financ Anal*. (2019) **63**:49–57. doi: 10.1016/j.irfa.2019.02.009
13. Nakamoto S. *Bitcoin: A Peer-to-Peer Electronic Cash System* (2008). Available online at: https://bitcoin.org/bitcoin.pdf
14. Li G, Li Y, Tsai CL. Quantile correlations and quantile autoregressive modeling. *J Am Stat Assoc*. (2015) **110**:246–61. doi: 10.1080/01621459.2014.892007
15. Koutmos D. Liquidity uncertainty and Bitcoin's market microstructure. *Econ Lett*. (2018) **172**:97–101. doi: 10.1016/j.econlet.2018.08.041
16. Gandal N, Hamrick JT, Moore T, Oberman T. Price manipulation in the Bitcoin ecosystem. *J Monet Econ*. (2018) **95**:86–96. doi: 10.1016/j.jmoneco.2017.12.004
17. Schilling L, Uhlig H. Some simple Bitcoin economics. *J Monet Econ*. (2019) **106**:16–26. doi: 10.1016/j.jmoneco.2019.07.002
18. Wu Y, Luo A, Xu D. Identifying suspicious addresses in Bitcoin thefts. *Digit Invest*. (2019) **31**:200895. doi: 10.1016/j.fsidi.2019.200895
19. Al-Yahyaee KH, Mensi W, Ko HU, Yoon SM, Kang SH. Why cryptocurrency markets are inefficient: the impact of liquidity and volatility. *North Am J Econ Finance*. (2020) **52**:101168. doi: 10.1016/j.najef.2020.101168
20. Trimborn S, Härdle WK. CRIX an index for cryptocurrencies. *J Empir Finance*. (2018) **49**:107–22. doi: 10.1016/j.jempfin.2018.08.004

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kristoufek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 04 August 2020
doi: 10.3389/fphy.2020.00291

Evaluation of the Heterogeneous Spatial Distribution of Population and Stores/Facilities by Multifractal Analysis

Mariko I. Ito and Takaaki Ohnishi*

Graduate School of Artificial Intelligence and Science, Rikkyo University, Tokyo, Japan

The spatial distribution of a population is not homogeneous—some areas attract many residents, while others do not. The spatial distribution of stores and facilities that have been coevolving with that of the population is also heterogeneous. Previous studies have shown that multifractality is a characteristic of the spatial distribution of a population, as well as other quantities associated with the urban system. We found that stores/facilities belonging to some categories also exhibit multifractality in spatial distribution. We quantified the spatial distributions of the population and stores/facilities in each category by multifractal analysis and compared their multifractal properties. Multifractal measures that reflect the heterogeneity of the densities in each location were able to capture additional features that cannot be seen when only the box-counting dimension was observed. Further, high concentrations of stores/facilities in categories relating to professional or commercial businesses were observed, consistent with previous studies on the scaling law, another pattern observed in urban morphology. We discuss the implications of the multifractal properties on the arrangement of locations where stores/facilities are concentrated. We believe that multifractal analysis is a powerful tool for the quantification of spatial distributions and expect that our interpretation of multifractal measures will stimulate further investigations into urban morphology.

Keywords: multifractal analysis, spatial distribution, population, stores/facilities, agglomeration

OPEN ACCESS

Edited by:
Gabjin Oh,
Chosun University, South Korea

Reviewed by:
Gui-Quan Sun,
North University of China, China
Jiuchuan Jiang,
Nanjing University of Finance and Economics, China

****Correspondence:***
Mariko I. Ito
marikoito.rnu1@gmail.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *14 April 2020*
Accepted: *26 June 2020*
Published: *04 August 2020*

Citation:
Ito MI and Ohnishi T (2020) Evaluation of the Heterogeneous Spatial Distribution of Population and Stores/Facilities by Multifractal Analysis. Front. Phys. 8:291. doi: 10.3389/fphy.2020.00291

1. INTRODUCTION

Fractality has been observed in various spatial distributions relating to the morphology of cities, e.g., population, buildings, land price, and street networks [1–10]. Fractality is represented by the nature where the mass (e.g., the population) in a region exhibits power law dependence on the size of the region. The power-law exponent is called the fractal dimension [11]. The abovementioned spatial distributions also exhibit heterogeneity in fractality in the sense that the locally measured fractal dimension around each spot diverges in the structure, known as multifractality [4, 12, 13].

The location of stores/facilities should depend on and, in turn, affect the spatial distribution of the population. People may choose the location of their settlement based on the availability and variety of stores/facilities. On the other hand, companies may invest in the construction of stores/facilities to secure employers and customers [14]. It is known that the nature of agglomeration of stores/facilities depends on their type. For example, previous studies showed that facilities relating to businesses offering professional services tend to concentrate in areas with a

large population [14–16]. One of these studies evaluated the concentration by investigating the scaling law between the population and densities of facilities, showing that the number of facilities in a city increases with a power of the population in the city [16]. The scaling law is a universal pattern observed in urban morphology, e.g., accessibility, road surface, and crime [16–21]. It should be interesting to see whether the agglomeration of stores/facilities exhibits multifractality, another urban-related feature.

If the spatial distribution of stores/facilities exhibits multifractality, multifractal analysis can be performed on the structure and is a strong tool to quantify the feature of it. In multifractal analysis, the resulted curve exhibits the correspondence between the locally measured fractal dimension, called *singularity strength* α, and the fractal dimension of the arrangement of spots that exhibit the singularity strength α, called *spectrum* $f(\alpha)$ [4, 13].

Besides multifractal analysis, various methods have been applied to evaluate spatial distributions including that of the population and stores/facilities. Average nearest neighbor (ANN), which is the averaged distance to the nearest point, is an indicator of the spatial clustering [22–24]. The distribution of nearest neighbor distance is also used to evaluate the qualitative feature of spatial structure [24, 25]. On the other hand, when grid lines are drawn on the space, probability of finding neighboring cells both of which are occupied is often calculated to assess the degree of clustering [26, 27]. Regarding the industrial coagglomeration between two industries, the extent to which the facilities of these industries are in the same region is evaluated [14, 28]. Compared to these methods, an advantage of multifractal analysis should be that it can demonstrate both local and global features of the spatial distribution [9]. By multifractal analysis, the strength of local concentration can be captured by the singularity strength. The global view of the arrangement of spots with each level of concentration, on the other hand, is incorporated into the spectrum.

In this study, we aim to (1) determine if the spatial distributions of stores/facilities in various categories exhibit multifractality, and if so, (2) determine the characteristics captured by the multifractal properties of each spatial distribution.

We chose the largest metropolitan area in Japan, the Kantō area, as the object of our analysis. We investigated the multifractal properties of the spatial distributions of the population and stores/facilities in Kantō. Our analysis showed that the spatial distribution of stores/facilities in some categories exhibit multifractality, as does that of the population. Though these spatial distributions are on the same geographical substrate, their multifractal properties are significantly different from each other.

This paper is organized as follows. The principles of fractal geometry, multifractality, and the generalized dimension, as well as the methods of our analysis are presented in section 2. We show the results of multifractal analysis of the spatial distributions in section 3. We also compare the multifractal properties of these spatial distributions and highlight the information extractable from the multifractal measures of the spatial distributions. We discuss the results in section 4.

2. MATERIALS AND METHODS

2.1. Data

Data for the spatial distributions of the population and stores/facilities were extracted from the Japanese 100-Meter Estimated Mesh Data of the 2015 National Censuses (Zenrin Marketing Solutions Co., Ltd.) and the Corporate Telephone Directory Database Telepoint with Coordinates (Telepoint Pack! provided by ZENRIN Co., Ltd.) of 2017, respectively. The former dataset contains data on the estimated population in each mesh. The length of each side of a mesh is ∼100 m, while the exact size is 3 s in the latitude direction and 4.5 s in the longitude direction. The latter dataset contains the geospatial information of each store/facility. Stores/facilities are categorized hierarchically. In the largest classification, which we adopted, there are 39 categories as shown in **Table 1**.

The analyzed area is a part of Kantō in Japan, that includes the capital, Tokyo, and a major industrial area, the Keihin industrial area. The range of the latitude is from 35°29′54″ to 35°55′30″ and that of the longitude is from 139°16′52.5″ to 139°55′16.5″ (**Figure 1**). There are $2^9 \times 2^9$ (= 262,144) meshes inside this region.

Table 1 shows a summary of the data analyzed in this study. The total population (the total number of stores/facilities) in the analyzed area and the number of non-empty meshes are shown in the third and fourth columns, respectively. Here, the meshes/boxes with non-zero populations (stores/facilities) are called *support*. The maximum and the mean population (number of stores/facilities) in a mesh on the support are shown from the fifth to the sixth columns.

Figure 2A shows the spatial distribution of the population in the analyzed area. The logarithm of the proportion of the population in each mesh to the total population is represented by the heatmap. The other panels (B–L) show the spatial distributions of stores/facilities in 11 categories.

2.2. Multifractality

We briefly introduce the concepts of fractal geometry and multifractality. There are several ways of defining fractals and multifractals; we present one of them here [9, 11–13, 29]. Additionally, we limit our explanation to structures embedded in $\mathbb{R}^2$, while higher dimensions of fractal and multifractal structures can be generally defined.

When a mass ($m(\varepsilon)$) composing a structure within a region of size ε increases with ε according to $m(\varepsilon) \sim \varepsilon^D$, the structure is regarded as having fractal characteristics. Here, the "size" is, for example, the length of a side if the region is a square. The exponent D is the fractal dimension of the structure. A more precise definition of the fractal dimension of a structure X is the one by the following box-counting method [11]. Let us assume that a structure X is covered with boxes of size ε. Let $N(\varepsilon)$ be the minimum number of such boxes required to cover the structure.

TABLE 1 | Data summary. For each category, the ID and the name of the category are shown in the first and second columns, respectively.

ID	Category	Total	# Support	Max	Mean
0*	Population (POP)	21,963,431	175,255	5,884	125.323
1	Fisheries, Agriculture, and Forestry	1,024	830	4	1.234
2	Mining	207	147	18	1.408
3*	Construction and Civil (CC)	101,349	43,284	139	2.341
4	Foods	10,934	6,806	210	1.607
5	Textiles and Apparels	10,111	6,380	42	1.585
6	Pulp and Paper	3,559	2,654	16	1.341
7	Chemicals and Pharmaceutical	5,374	3,262	84	1.647
8	Oil and Coal Products	4,376	3,240	91	1.351
9	Rubber Products	1,970	1,573	28	1.252
10	Glass and Ceramics Products	3,124	2,032	28	1.537
11	Iron and Steel	2,362	1,757	135	1.344
12	Non-ferrous Metals	3,654	2,760	51	1.324
13*	Metal Products (MP)	18,606	7,132	91	2.609
14	Machinery	24,745	15,694	60	1.577
15	Electric Appliances	8,806	6,250	27	1.409
16	Transportation Equipment	1,424	1,082	19	1.316
17	Precision Instruments	10,332	6,676	49	1.548
18	Other Products	27,707	15,107	55	1.834
19	Commercial Services	23,574	12,662	570	1.862
20*	Financing Business (FIN)	12,684	7,116	76	1.782
21*	Real Estate (RE)	38,051	17,850	64	2.132
22	Land Transportation	838	654	15	1.281
23	Marine Transportation	1,193	606	19	1.969
24	Air Transportation	438	223	29	1.964
25*	Warehousing and Harbor Transportation (WH)	12,171	8,180	68	1.488
26	Information and Communication	22,311	9,408	92	2.371
27	Electric Power and Gas	2,401	1,883	16	1.275
28	Professional Services	30,734	13,707	64	2.242
29*	Sports Facilities (SF)	5,863	4,326	18	1.355
30	Sporting-goods Stores	2,337	1,761	14	1.327
31*	Amusement, Eating, and Drinking (AED)	103,993	30,891	213	3.366
32*	Resort (RES)	13,401	8,709	24	1.539
33*	Hospitals and Welfare (HW)	75,265	34,918	40	2.155
34	Supermarkets	13,699	7,641	232	1.793
35*	Living-related services (LS)	185,824	53,808	215	3.453
36	Automobiles	24,422	13,824	49	1.767
37*	School Education (SE)	35,977	20,200	87	1.781
38	Public Agencies	15,481	6,378	347	2.427
39	Others	14,815	7,953	61	1.863

The star beside the ID indicates that the spatial distribution of that category exhibits multifractality. The star-marked categories also have their abbreviation enclosed in brackets. The total population (the total number of stores/facilities) in the analyzed area is shown in the third column. The number of meshes that is not empty is shown in the fourth column. These non-empty areas are called support. The maximum and the mean population (number of stores/facilities) in a mesh on the support are shown in the fifth and the sixth columns, respectively.

Then the box-counting dimension D is defined as:

$$D = -\lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log \varepsilon}. \tag{1}$$

We introduce multifractality. We again consider a set X and a function μ on X that gives a quantity, such as the density, at each point $x \in X$. Let us assume that X is divided by square boxes that have the same size ε. For the i-th box of size ε, $C_{i,\varepsilon}$, the value $P_{i,\varepsilon}$ is called the *probability measure* on the box:

$$P_{i,\varepsilon} = \frac{\int_{C_{i,\varepsilon}} \mu(x)dx}{\int_X \mu(x)dx}. \tag{2}$$

If $P_{i,\varepsilon}$ and ε have the following power-law relationship for any i:

$$P_{i,\varepsilon} \sim \varepsilon^{\alpha}, \text{ as } \varepsilon \to 0, \tag{3}$$

FIGURE 1 | Map of the analyzed area (1:1,000,000 INTERNATIONAL MAP, Geospatial Information Authority of Japan).

then fractality can be seen around each point of X. Here we assume that the exponent α diverges in X and let $N(\varepsilon, \alpha)$ be the number of boxes that satisfy $P_{i,\varepsilon} \sim \varepsilon^{\tilde{\alpha}}$ where $\tilde{\alpha} \in [\alpha, \alpha + \Delta\alpha]$. If $N(\varepsilon, \alpha)$ decreases with ε as

$$N(\varepsilon, \alpha) \sim \rho(\alpha)\varepsilon^{-f(\alpha)}\Delta\alpha, \tag{4}$$

the set X can be regarded as having a multifractal structure. Exponent α, which can be regarded as the local fractal dimension, is called the singularity strength. On the other hand, exponent $f(\alpha)$ stands for the box-counting dimension of the set of points with the singularity strength α. This dimension $f(\alpha)$ is called the spectrum. In this paper, the curve of $(\alpha, f(\alpha))$ is called the *multifractal curve*. **Figure 3** is an example of the relationship between α and $f(\alpha)$ in the spatial distribution of the population (**Figure 2A**). Each panel in **Figure 3** shows the units that exhibit the singularity strength α within the range shown on each panel. We derived the singularity strength of each unit by estimating the exponent in the following relationship $P_{i,\varepsilon} \sim \varepsilon^{\alpha}$ (see Equation 3). The box-counting dimension of the arrangement of units with the singularity strength α, $f(\alpha)$, was also derived based on Equation (4) when the power-law relationship in the equation holds. For example, units with the singularity strength α within $[1.85, 1.90)$ are rare. Also, the arrangement of such units is curve-like (i.e., a one-dimensional shape) and exhibits a low box-counting dimension $f(\alpha) \sim 1.60$. On the other hand, the arrangement of units with the singularity strength α within $[2.05, 2.10)$ spans a two dimensional region and exhibits a high box-counting dimension that is about 2.

2.3. Generalized Dimension

We introduce the generalized dimension and explain the relationship between the singularity strength, the spectrum, and the generalized dimension [29, 30]. The q-th generalized dimension D_q is defined as follows. First, we define τ_q as

$$\tau_q = \lim_{\varepsilon \to 0} \frac{\log \sum_i P_{i,\varepsilon}^q}{\log \varepsilon}. \tag{5}$$

Then the generalized dimension D_q for $q \neq 1$ is defined as

$$D_q = \frac{1}{q-1}\tau_q. \tag{6}$$

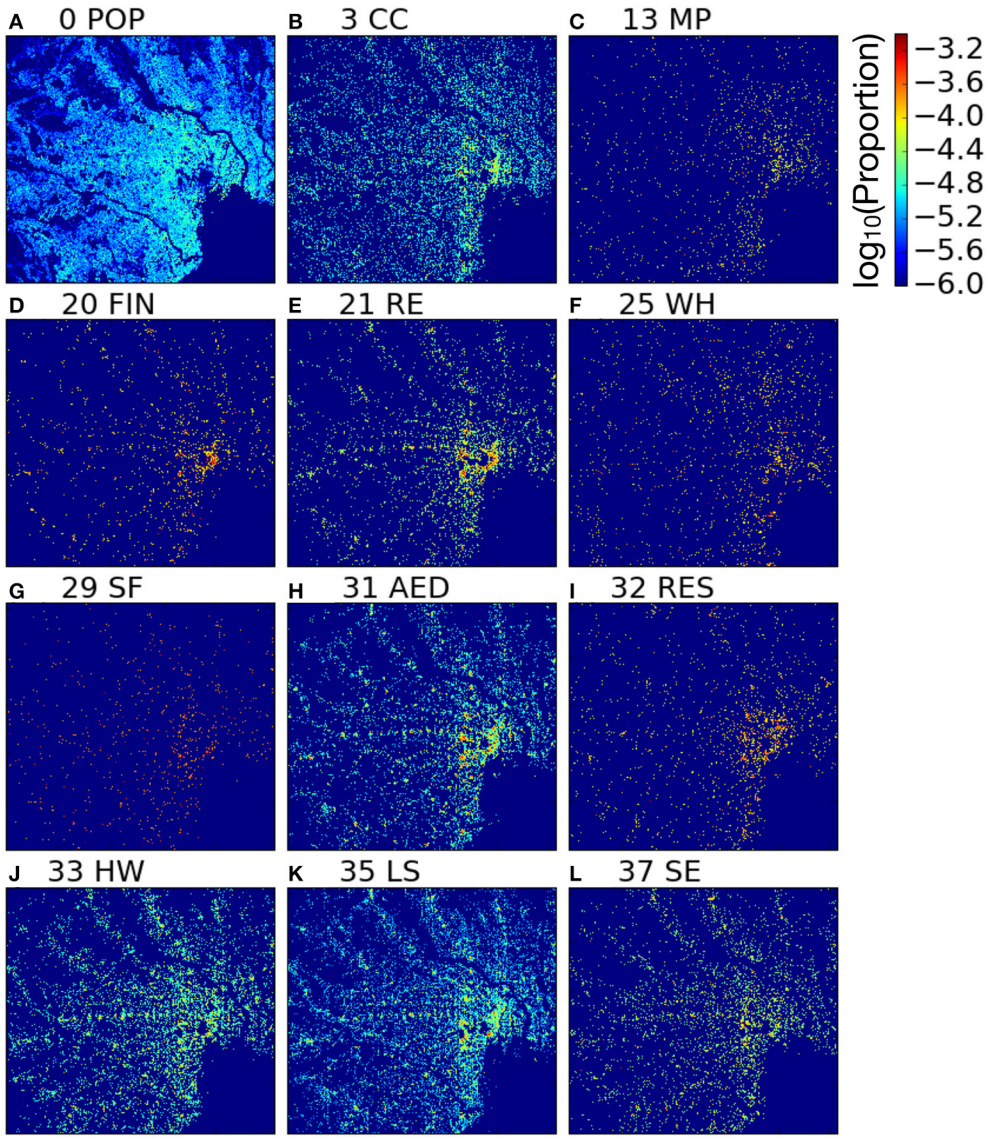


FIGURE 2 | Spatial distributions of the population and stores/facilities in the analyzed area. **(A)** The spatial distribution of the population. **(B–L)** The spatial distributions of stores/facilities in 11 categories: **(B)** 3CC (Construction and Civil); **(C)** 13MP (Metal Products); **(D)** 20FIN (Financing Business); **(E)** 21RE (Real Estate); **(F)** 25WH (Warehousing and Harbor Transportation); **(G)** 29SF (Sports Facilities); **(H)** 31AED (Amusement, Eating, and Drinking); **(I)** 32RES (Resort); **(J)** 33HW (Hospitals and Welfare); **(K)** 35LS (Living-related Services); **(L)** 37SE (School Education). For the other categories, the spatial distributions are shown in **Figures S1, S2**. The color stands for the value of $\log_{10}$ [(the population in the mesh)/(the total population)] for each mesh in **(A)**. The number of stores/facilities is used instead of the population in **(B–L)**. Each mesh is a 100-m mesh as described in section 2.

In the case of $q = 1$,

$$D_1 = \lim_{\varepsilon \to 0} \frac{\sum_i P_{i,\varepsilon} \log P_{i,\varepsilon}}{\log \varepsilon}. \tag{7}$$

In the summation on the right-hand side of each of Equations (5), (7), let the i-th term be summed when the i-th box is not empty, i.e., $P_{i,\varepsilon} \neq 0$. Here, the generalized dimension is equal to the box-counting dimension of the support D when q is zero.

It is known that the following values of α_q and $f(\alpha_q)$ give the approximation of the pair of α and $f(\alpha)$ for each q:

$$\alpha_q = \frac{d\tau_q}{dq}, \tag{8}$$

$$f(\alpha_q) = \alpha_q q - \tau_q. \tag{9}$$

2.4. Method of Analysis

In this study, we are concerned with finite morphologies that do not contain infinitesimal structures as the smallest unit of our data is a 100-m mesh. Therefore, we cannot rigorously calculate

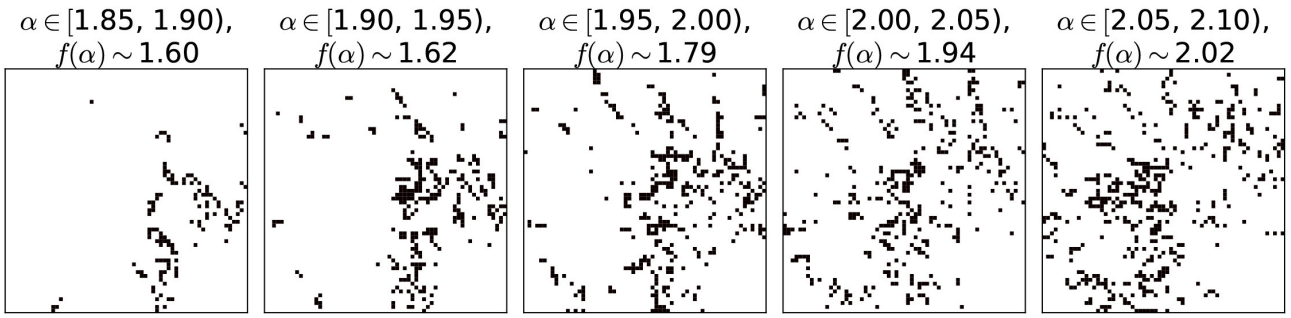


FIGURE 3 | Example of the singularity strength α and the spectrum $f(\alpha)$. In each panel, units that have the singularity strength α within the range mentioned at the top of the panel are shown. Here the size of each unit is $2^3/2^9$ on one side, and α is derived based on Equation (3). Also, the spectrum $f(\alpha)$ is derived based on Equation (4), only when the power-law relationship in Equation (4) holds.

the generalized dimension, singularity strength, and spectrum. Instead, we consider the range of q and ε in which the structure can be regarded as exhibiting (multi)fractality, by evaluating the range of q and ε in which the following relationship holds:

$$\sum_i P_{i,\varepsilon}^q \sim \varepsilon^d. \tag{10}$$

Thus, we examine the linearity of the relationship between $\log \sum_i P_{i,\varepsilon}^q$ and $\log \varepsilon$ by the frequently used method [3, 4, 7, 8, 31–34]. Then we regard d as τ_q by Equation (5) when $\log \sum_i P_{i,\varepsilon}^q \sim d \log \varepsilon + (\text{const.})$.

In our analysis, we assign the size of one side of a 100-m mesh to $\varepsilon = 1/2^9$. Grid lines are drawn on the analyzed area so that it is covered by non-overlapping boxes with the same size [3, 7]. As the size of boxes, we consider $\varepsilon = 1/2^9, 2/2^9, 2^2/2^9, ..., 1$. We define the probability measure $P_{i,\varepsilon}$ on the i-th box with the size ε as the proportion of the population (the number of stores/facilities) in the i-th box to the total population (the total number of stores/facilities).

Subsequently, for the population, we evaluate the linearity of the relationship between $\log \sum_i P_{i,\varepsilon}^q$ and $\log \varepsilon$ for various ranges of ε and q as shown in **Figure 4**. Regarding the criterion for this linearity, we examined whether or not the coefficient of determination of the linear regression, R^2, exceeds 0.99. A linear relationship was not observed when q takes a negative value and when the range of ε includes values $< 2^3/2^9$. Consequently, we considered the range of ε from $2^3/2^9$ to $2^9/2^9$, and the range of q from 0 to 20. We performed a multifractal analysis on these ranges of ε and q for the spatial distribution of the population. Furthermore, for the spatial distribution of stores/facilities in each category, we examined multifractality on the same ranges of ε and q as that for the population. Plots of $\log \sum_i P_{i,\varepsilon}^q$ against $\log \varepsilon$ for all categories of stores/facilities are shown in **Figures S3, S4**. The star marker was added beside the ID in the first column of **Table 1** if the category showed multifractality in the spatial distribution. Also, the spatial distribution of such categories are shown in **Figures 2B–L**. In this figure and in the following discussion, abbreviations are used for these categories—the abbreviation of each category is enclosed by brackets after the name of the category in the second column of **Table 1**.

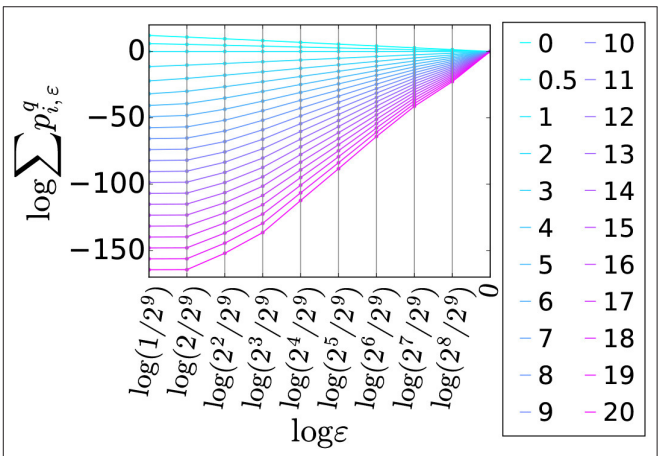


FIGURE 4 | $\log \sum_i P_{i,\varepsilon}^q$ vs. $\log \varepsilon$ for each q. The color of the plots correspond to the value of q in the legend.

We obtained τ_q as the slope of the linear regression of $\log \sum_i P_{i,\varepsilon}^q$ by $\log \varepsilon$, and derived D_q from Equations (6, 7). To derive the singularity strength α_q and the spectrum $f(\alpha_q)$, we used the following formulae:

$$\alpha_q = \lim_{\varepsilon \to 0} \frac{\sum_i \mu_{i,\varepsilon,q} \log P_{i,\varepsilon}}{\log \varepsilon} \tag{11}$$

$$f(\alpha_q) = \lim_{\varepsilon \to 0} \frac{\sum_i \mu_{i,\varepsilon,q} \log \mu_{i,\varepsilon,q}}{\log \varepsilon}, \tag{12}$$

where $\mu_{i,\varepsilon,q} = P_{i,\varepsilon}^q / \sum_j P_{j,\varepsilon}^q$. These formulae can be directly derived from Equation (5), (8), and (9) [12, 35, 36]. We obtained α_q by the linear regression between $\sum_i \mu_{i,\varepsilon,q} \log P_{i,\varepsilon}$ and $\log \varepsilon$, and obtained $f(\alpha_q)$ by the linear regression between $\sum_i \mu_{i,\varepsilon,q} \log \mu_{i,\varepsilon,q}$ and $\log \varepsilon$.

3. RESULTS

3.1. Density of Units and Multifractality

As we briefly mentioned in section 2.4, multifractality could not be observed for the spatial distribution of the population when the value of q is negative. Knowledge of the relationship between the densities in a unit, the value of q, and the multifractal measures is helpful for interpreting the results of the multifractal analysis. In Equation (5) for τ_q, a larger value of q corresponds to a greater contribution of the boxes with large probability measures $P_{i,\varepsilon}$ to the sum. Therefore, the boxes with high densities are significantly incorporated into the calculation of the q-th generalized dimension D_q when the value of q is large, see Equation (6). Note that no differences in density is considered when $q = 0$. The pair of the singularity strength and the spectrum also has a relationship with q by Equations (8), (9). Therefore, the multifractal measures of D_q, α_q, and $f(\alpha_q)$ for a large q, significantly reflect those units with high densities. The fact that multifractality was only observed in the spatial distribution of the population with positive q values, indicates that multifractality cannot be observed in sparsely populated areas.

3.2. Generalized Dimension

Figure 5 shows the q-th generalized dimension vs. q for the population and stores/facilities in four categories: 20FIN, 25WH, 31AED, and 33HW. The generalized dimension for stores/facilities in the other categories is shown in **Figure S5**. As mentioned before, the generalized dimension for $q = 0$, D_0, is equal to the box-counting dimension of the support. For each category, the value of D_0 is close to 2, which is the dimension of $\mathbb{R}^2$ to which each spatial distribution is embedded. The rate of decline in the generalized dimension with q significantly varies between the categories. The generalized dimension for 20FIN drops dramatically with q, while that of the population decreases minimally with q. In the cases of 25WH and 33HW, the decline of the generalized dimension is milder than that of 20FIN and 31AED. Considering the relationship between the value of q and the densities in the boxes, $|\Delta D_{20}| := |D_{20} - D_0|$ can be one of the indicators for the strength of the heterogeneity in spatial density distribution, and 20FIN and 31AED can be regarded as having strong heterogeneity in spatial distributions.

3.3. Multifractal Curve

The multifractal curves are shown in **Figure 6A**. Before we investigate each multifractal curve, let us revisit the relationship between the densities in a unit and the value of q, and the singularity strength α_q [9]. Recall that each point on the multifractal curve can be derived for each q (Equations 11, 12). Generally speaking, in each multifractal curve, the greater is the value of q, the lower is the value of α_q, i.e., the plot of a multifractal curve proceeds to the left as the value of q increases. Therefore, by the relationship between the value of q and the densities in the units, a unit with a high density tends to have a small singularity strength—we can observe this by comparing **Figure 2A** with **Figure 3**. An interpretation is

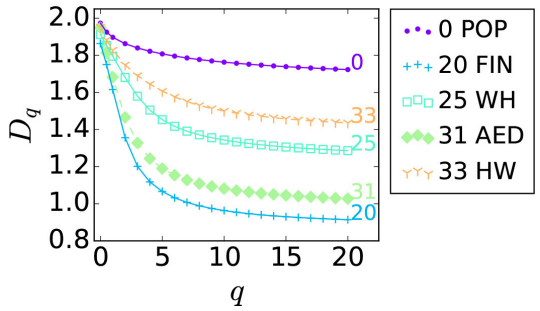


FIGURE 5 | q-th generalized dimension vs. q for the population and stores/facilities in four categories: 20FIN, 25WH, 31AED, and 33HW. The generalized dimension for stores/facilities in the other categories are shown in **Figure S5**.

that around a unit with a density significantly higher than the surroundings, the probability measure $P_{i,\varepsilon}$ does not increase rapidly by expanding the area, and vice versa (see Equation 3). **Figure 7** explains this interpretation. Let us assume that each of the gray units has a constant density. Only the i-th unit is filled in the case of panel (A), while all the units are filled in the case of panel (B). The singularity strength of the i-th unit in panel (A) is 0, since the probability measure of the square boxes emphasized by the thick line, that expands around the i-th unit, does not increase with the size ε of the box and remains at the same value. On the other hand, the probability measure on these boxes changes according to the square of ε, and the singularity strength of the i-th unit is 2, in panel (B). Though panels (A) and (B) are extreme examples, these diagrams suggest that a unit with a density significantly higher than the surrounding units can have a low singularity strength.

Therefore, in **Figure 6A**, multifractal curves with a significantly low singularity strength α_q, such as 20FIN and 31AED, indicate the existence of units with a density much higher than the surroundings. Also, the spectrum $f(\alpha_q)$ is quite low for small singularity strengths α_q in the cases of 20FIN and 31AED. Recall that the spectrum $f(\alpha_q)$ represents the box-counting dimension of the arrangement of units with the singularity strength α_q. The pair of small α_q and $f(\alpha_q)$, thus, is an indication that there are a few isolated units, each with an extremely high density.

Multifractal curves contain information not only at small singularity strengths but also across the entire range. The multifractal curve of 13MP has similar α_q and $f(\alpha_q)$ values with that of 25WH, 29SF, 35LS, and 37SE when α_q is quite small. However, in the mid-range of α_q, the value of $f(\alpha_q)$ for 13MP is significantly greater than that of the others. Therefore, in the case of 13MP, units with a mid-range density (compared to the surroundings) remain concentrated compared to the case of the other categories. This may correspond to the yellowish units gathered on the right of panel (C) in **Figure 2**. On the other hand, a feature of the multifractal curve of the

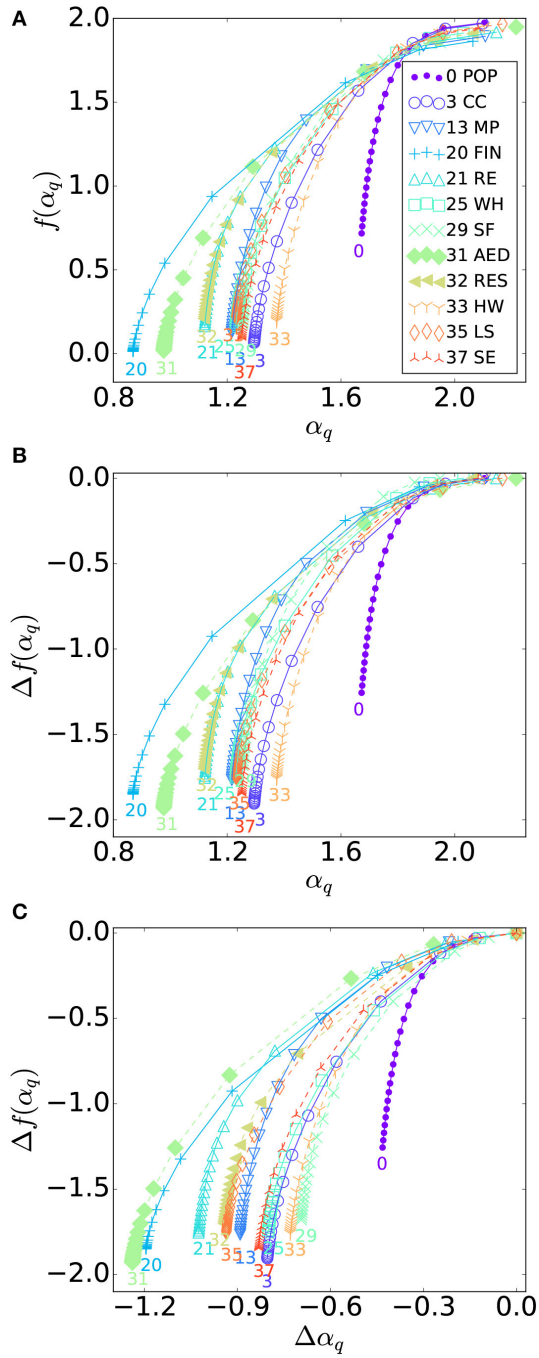


FIGURE 6 | (A) Multifractal curve. The pair of the singularity strength (abscissa) and the spectrum (ordinate), $(\alpha_q, f(\alpha_q))$, for each q of the population and stores/facilities. Each multifractal curve is also separately shown in **Figure S6**. **(B)** $\Delta f(\alpha_q)$ vs. α_q, where $\Delta f(\alpha_q) := f(\alpha_q) - f(\alpha_0)$. **(C)** $\Delta f(\alpha_q)$ vs. $\Delta\alpha_q$, where $\Delta\alpha_q := \alpha_q - \alpha_0$.

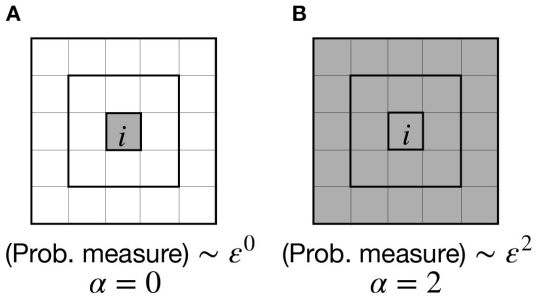


FIGURE 7 | Schematic image of the singularity strength and the densities in a unit. Each of the gray-colored units has the same density, i.e., has the same probability measure ρ. **(A)** Only the i-th unit has non-zero density and the other units are empty. The probability measure of the box remains ρ, even when the size of the box ε increases as emphasized by the thick line. Therefore, the singularity strength of the i-th unit is 0. **(B)** All units are filled. The probability measure of the box is ρ, 9ρ, and 25ρ, for the smallest, the second smallest and the largest boxes (emphasized by the thick line), respectively. Thus, the singularity strength of the i-th unit is 2.

population is that neither α_q nor $f(\alpha_q)$ declines rapidly with q. This corresponds to the following features of the spatial distribution of the population: The densities in the units does not widely diverge and the arrangement of the units with each density does not change dramatically with the density (**Figure 2A**).

Figure 6B shows the plots of $\Delta f(\alpha_q) := f(\alpha_q) - f(\alpha_0)$ against α_q. Recall that $f(\alpha_0)$ is the box-counting dimension of the support. In the value of $f(\alpha_0)$, the differences in the densities in the units is not incorporated. Thus, the vertical axis $\Delta f(\alpha_q)$ represents the degree of decline in the spectrum from the box-counting dimension of the support when the differences in the densities in the units is gradually emphasized. Another intuitive meaning of $\Delta f(\alpha_q)$ is the nature of the difference in arrangement between the panels in **Figure 3**. In the case of 31AED, in which the pair of a small α_q and $f(\alpha_q)$ exists, we observe that the value of $\Delta f(\alpha_q)$ is also small for small α_q. This indicates that the arrangement of units with small α_q is quite sparse compared to the support. Therefore, we can infer the following characteristics of the spatial distribution of 31AED: There are a few isolated locations where stores/facilities in 31AED are extremely concentrated, while covering a vast region.

Figure 6C shows plots of $\Delta f(\alpha_q)$ against $\Delta\alpha_q := \alpha_q - \alpha_0$. The horizontal axis shows that how the singularity strength declines as q increases from the one calculated under the situation where differences in the densities in the units were not taken into account. The decline of both $\Delta\alpha_q$ and $\Delta f(\alpha_q)$ with q of 29SF is milder than the other categories excepting the population, suggesting that the spatial distribution of stores/facilities is homogeneous in 29SF. Specifically, by comparing with **Figure 6A**, the value of α_q is small across the range of q including $q = 0$, in the case of 29SF. This may represent the sparsely but relatively uniform scattering of stores/facilities in 29SF (**Figure 2G**). Additionally, though the difference in the multifractal curves of 21RE and 32RES is ambiguous for large q in **Figure 6A**, we can observe that the singularity strength of 21RE has a wider range than 32RES (**Figure 6C**). When we restrict our observation to the

concentrated units, the arrangement of the units in 21RE has a similar box-counting dimension to that of 32RES. On the other hand, in 21RE, we should be able to observe local regions where the density increases mildly with the size of the region, that may not be observed in 32RES.

Finally, **Figure 8** shows a summary of the multifractal properties of the spatial distributions of the population and stores/facilities for all categories that have multifractality in the ranges of ε and q tested in this study. In both panels, the horizontal axis shows $f(\alpha_0)$, i.e., the box-counting dimension of the support. In $f(\alpha_0)$, the differences in the densities in the units is not incorporated—only whether or not each unit is empty is taken into account. The box-counting dimension of the support, of course, can capture a feature of each spatial distribution. The value of $f(\alpha_0)$ for each category is near two, that is the dimension of $\mathbb{R}^2$, but it diverges a little. We can see that $f(\alpha_0)$ tends to be small when the spatial distribution is sparse, e.g., 20FIN and 29SF, in (**Figures 2D,G**, **8**).

On the other hand, the vertical axes in **Figure 8** show the values that incorporate the differences in the densities in the units. In **Figure 8A**, the vertical axis shows the difference between the largest and the smallest singularity strength $|\Delta\alpha_{20}|$. The color of each marker represents the value of $|\Delta D_{20}|$, for each category. The brighter marker color corresponds to the larger value of $|\Delta D_{20}|$. In **Figure 8B**, the vertical axis shows the difference between the largest and the smallest spectrum $|\Delta f(\alpha_{20})|$. The color of each marker again represents the value of $\Delta\, D_{20}$.

The values of $|\Delta\alpha_{20}|$ and $|\Delta f(\alpha_{20})|$ for some categories can diverge even when $f(\alpha_0)$ takes almost the same value. For example, $|\Delta\alpha_{20}|$ of 29SF is much smaller than that of 20FIN, while $f(\alpha_0)$ of both of these categories are around 1.86. This result indicates that both 29SF and 20FIN have sparse spatial distributions, but the heterogeneity of the densities in the units for 20FIN is stronger than that of 29SF—this can be observed in **Figures 2D,G**. In **Figure 8A**, the plots of 3CC and 33HW are nearer that of the population than the others. This suggests that the nature of the spatial distribution in these categories is similar to that of the population when we consider the large region covered by stores/facilities and the small differences in the relative densities in the units.

4. DISCUSSION

In this study, we evaluated whether the spatial distributions of the population and stores/facilities exhibit multifractality. Multifractality in the spatial distribution of the population has been demonstrated in previous studies [3, 5, 10]; we also demonstrated this result in the Kantō area, Japan. However, we were not able to observe multifractality in the population distribution for negative values of q. In the previous studies that evaluated multifractality in city morphologies, the authors carefully examined the range of q in which multifractality was observed [4, 7, 8]. These studies showed that multifractality can be observed at both positive and negative values of q. As we mentioned above, positive (the negative) values of q corresponds to boxes with high (low) densities. Therefore, previous studies observed multifractality in both densely and sparsely distributed regions. However, they also showed asymmetry in the positive and negative ranges of q and discussed the structural differences in dense/sparse regions. In our case, the range of q for multifractality suggests that the sparsely populated region does not have a structure characterized by multifractals, attributable to the geographical characteristics of the examined region in this study. The examined region includes the mountainous areas in the upper and left sides of each panel in **Figure 2** as it represents the general feature of the Japanese terrain. It also contains Tokyo bay in the lower-right corner of each panel (**Figure 2**). Along Tokyo bay, there are rich residential areas as well as plenty of facilities in various industries and numerous stores. This examined region should be an interesting object to investigate considering these geographical features. However, complex substrates may restrict the range in which multifractality appears.

We also investigated which category of stores/facilities shows multifractality for the same ranges of q and ε as that for the population. Stores/facilities in some categories also exhibit multifractality in the spatial distribution, but the determined multifractal measures significantly depend on the category. Diverging multifractal properties can reflect qualitative differences in the spatial distributions—stores/facilities are sparsely and uniformly scattered in some categories, while others are centralized, e.g., the stations in some categories. Importantly, our analysis showed that the box-counting dimension performs poorly in capturing qualitative diversities in the spatial distributions of these categories. The box-counting dimension captures the arrangement of the units that are not empty. On the other hand, multifractal measures can represent the arrangement of units with a certain density. Multifractal curves can indicate, for example, the existence of units with comparably high densities by evaluating the range of the singularity strength α, and the spatial distribution of such units by the spectrum $f(\alpha)$.

The spatial distribution of the population can be characterized by the high box-counting dimension of the support and the homogeneity measured by the generalized dimension and the multifractal curve. The population is distributed across all the regions examined in this study, while the densities in a unit do not vary significantly. In addition, we will observe a similar spatial distribution, even when we change the filter on units according to these densities. Considering its high box-counting dimension and the homogeneity seen in the range of the singularity strength, the spatial distribution of stores/facilities in 33HW exhibits similar features with that of the population. These stores/facilities cover a large area and the heterogeneity of the densities in the units is low. On the other hand, in the cases of 20FIN and 31AED, the singularity strength and the spectrum sharply decline with q. This shape of the multifractal curves indicates a strong centralization of stores/facilities to a few locations. Some of these

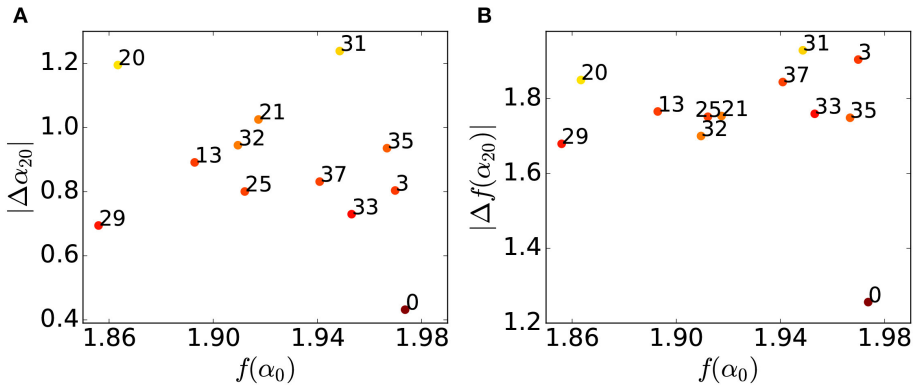


FIGURE 8 | Summary of multifractal properties. **(A)** The difference between the largest singularity strength α_0 and the smallest one α_{20}, $|\Delta\alpha_{20}|$, against the box-counting dimension of the support $f(\alpha_0)$. **(B)** The difference between the largest spectrum $f(\alpha_0)$ and the smallest one $f(\alpha_{20})$, $|\Delta f(\alpha_{20})|$, against the box-counting dimension of the support $f(\alpha_0)$. In both panels, the number beside each plot shows the category. The color of each plot represents the value of $|\Delta D_{20}|$. The brighter color corresponds to the larger $|\Delta D_{20}|$.

concentrated locations presumably correspond to large stations in the capital.

In addition to multifractality of the spatial distribution, the scaling law is also a universal pattern observed in city morphology [16–21, 37, 38]. For example, the population X and a quantity Y related to the city morphology have the relationship of $Y \sim X^{\beta}$. The large scaling exponent β in this relationship, i.e., the super-linear increase of Y with X, indicates a strong concentration of these urban-related objects to locations with a large population. Previous studies also found the scaling law in (industrial) agglomerations and showed that facilities, outputs, and jobs concentrated stronger in cities with a large population when these objects are in a category associated with professional and complex skills or with commercial facilities than when the type is relatively primitive or public-related [16, 19–21]. Many of our results are consistent with these previous studies. Our multifractal analysis indicates the centralization of facilities in 20FIN, which is a category related to services requiring professional skills and frequent communication with customers [14]. Also, 31AED, a category related to commercial activities, showed a strong concentration. The concentration of objects related to construction and healthcare was shown to be mild in previous studies, which is also consistent with the mild decline of the singularity strength with q in our results (3CC and 33HW).

While consistency between multifractality and the scaling law exists, an advantage of multifractal analysis should be the richness of information in the result. As we discussed so far, we can quantify the nature of the divergence of the densities in each location and the spatial distribution of each density, by multifractal analysis. For example, for the centralization of stores/facilities in 20FIN and 31AED, multifractal properties further explain the following difference. Considering the large range in the singularity strength and the high box-counting dimension, we can expect to see stores/facilities in 31AED everywhere, with some centralized locations. On the other hand, facilities in 20FIN are encountered only in concentrated locations, which is represented by the overall small singularity strength. Therefore, the results in our analysis exhibit not only the existence of concentrated areas but also the various state of concentration. In this study, we attempted to interpret the multifractal curves that correspond to qualitatively diverging spatial distributions. We hope that our discussion will contribute to future investigations on spatial distributions by multifractal analysis. Additionally, the characteristic of concentration of stores/facilities in each category, which was revealed in this study, should be considered in the actual urban design. For example, the stores/facilities in 31AED is expected to have a tendency to concentrate strongly. Such a tendency of agglomeration should be taken into account in advance when it is required to avoid an extreme concentration of buildings in a landscape.

The temporal development of various urban morphologies, e.g., the spatial distributions of streets and buildings, have been discussed in previous studies [2, 4, 7–9]. Some of them revealed that the spatial distribution was developed to the packed state and to exhibit features close to monofractals [2, 4]. We are also interested in how these developments depend on the category in which the stores/facilities belong. As a future perspective, the comparison of such developments is possible by quantification with multifractal analysis. Furthermore, we expect that the classification of cities is possible by comparing such developments between cities.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because these are available under the contract. Requests to access these datasets should be directed to Zenrin Co., Ltd., https://www.zenrin.co.jp/product/category/gis/contents/teldt/index.html for Corporate Telephone Directory Database Telepoint with Coordinates (Telepoint Pack!) and to Zenrin Marketing Solutions Co., Ltd., https://www.zenrin-ms.co.jp/gis_marketing/database/statistics/census_100m_mesh/ for Japanese 100-Meter Estimated Mesh Data of the 2015 National Censuses.

AUTHOR CONTRIBUTIONS

MI and TO contributed the conception and design of the study, performed the statistical analysis, and wrote the sections of the manuscript. MI wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by JSPS KAKENHI Grant Number JP16H02872.

ACKNOWLEDGMENTS

This work was supported by Joint Research Program No. 674 at CSIS, UTokyo [Telepoint Pack DB February 2017 (Yellow pages) by ZENRIN Co. Ltd.].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00291/full#supplementary-material

REFERENCES

1. Batty M. New ways of looking at cities. *Nature.* (1995) **377**:574–574. doi: 10.1038/377574a0
2. Encarnação S, Gaudiano M, Santos FC, Tenedório JA, Pacheco JM. Fractal cartography of urban areas. *Sci Rep.* (2012) **2**:527. doi: 10.1038/srep00527
3. Appleby S. Multifractal characterization of the distribution pattern of the human population. *Geogr Anal.* (1996) **28**:147–60. doi: 10.1111/j.1538-4632.1996.tb00926.x
4. Murcio R, Masucci AP, Arcaute E, Batty M. Multifractal to monofractal evolution of the London street network. *Phys Rev E.* (2015) **92**:062130. doi: 10.1103/PhysRevE.92.062130
5. Ozik J, Hunt BR, Ott E. Formation of multifractal population patterns from reproductive growth and local resettlement. *Phys Rev E.* (2005) **72**:046213. doi: 10.1103/PhysRevE.72.046213
6. Ariza-Villaverde AB, Jiménez-Hornero FJ, De Ravé EG. Multifractal analysis of axial maps applied to the study of urban morphology. *Comput Environ Urban Syst.* (2013) **38**:1–10. doi: 10.1016/j.compenvurbsys.2012.11.001
7. Chen Y, Wang J. Multifractal characterization of urban form and growth: the case of Beijing. *Environ Plann B Plann Des.* (2013) **40**:884–904. doi: 10.1068/b36155
8. Hu S, Cheng Q, Wang L, Xie S. Multifractal characterization of urban residential land price in space and time. *Appl Geogr.* (2012) **34**:161–70. doi: 10.1016/j.apgeog.2011.10.016
9. Salat H, Murcio R, Yano K, Arcaute E. Uncovering inequality through multifractality of land prices: 1912 and contemporary Kyoto. *PLoS ONE.* (2018) **13**:e0196737. doi: 10.1371/journal.pone.0196737
10. Chen Y, Zhou Y. Multi-fractal measures of city-size distributions based on the three-parameter Zipf model. *Chaos Solit Fract.* (2004) **22**:793–805. doi: 10.1016/j.chaos.2004.02.059
11. Peitgen HO, Jürgens H, Saupe D. *Chaos and Fractals: New Frontiers of Science*. New York, NY: Springer Science & Business Media (2006).
12. Meakin P. *Fractals, Scaling and Growth Far From Equilibrium*, Vol. 5. Cambridge: Cambridge University Press (1998).
13. Salat H, Murcio R, Arcaute E. Multifractal methodology. *Phys A.* (2017) **473**:467–87. doi: 10.1016/j.physa.2017.01.041
14. Diodato D, Neffke F, O'Clery N. Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *J Urban Econ.* (2018) **106**:1–26. doi: 10.1016/j.jue.2018.05.002
15. O'Clery N, Heroy S, Hulot F, Beguerisse-Diaz M. Unravelling the forces underlying urban industrial agglomeration. *arXiv.* (2019) 1903.09279.
16. Balland PA, Jara-Figueroa C, Petralia SG, Steijn MP, Rigby DL, Hidalgo CA. Complex economic activities concentrate in large cities. *Nat Hum Behav.* (2020) **4**:248–54. doi: 10.1038/s41562-019-0803-3
17. Batty M. A theory of city size. *Science.* (2013) **340**:1418–9. doi: 10.1126/science.1239870
18. Bettencourt LM, Lobo J, Helbing D, Kühnert C, West GB. Growth, innovation, scaling, and the pace of life in cities. *Proc Natl Acad Sci USA.* (2007) **104**:7301–6. doi: 10.1073/pnas.0610172104
19. Cottineau C, Hatna E, Arcaute E, Batty M. Diverse cities or the systematic paradox of urban scaling laws. *Comput Environ Urban Syst.* (2017) **63**:80–94. doi: 10.1016/j.compenvurbsys.2016.04.006
20. Um J, Son SW, Lee SI, Jeong H, Kim BJ. Scaling laws between population and facility densities. *Proc Natl Acad Sci USA.* (2009) **106**:14236–40. doi: 10.1073/pnas.0901898106
21. Youn H, Bettencourt LM, Lobo J, Strumsky D, Samaniego H, West GB. Scaling and universality in urban economic diversification. *J R Soc Interface.* (2016) **13**:20150937. doi: 10.1098/rsif.2015.0937
22. Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS ONE.* (2012) **7**:e31743. doi: 10.1371/journal.pone.0031743
23. Scott LM, Janikas MV. Spatial statistics in arcgis. In: Fischer M, Getis A, editors. *Handbook of Applied Spatial Analysis*. Berlin; Heidelberg: Springer (2010). p. 27–41. doi: 10.1007/978-3-642-03647-7_2
24. Guo D, Zhou H, Zou Y, Yin W, Yu H, Si Y, et al. Geographical analysis of the distribution and spread of human rabies in china from 2005 to 2011. *PLoS ONE.* (2013) **8**:e72352. doi: 10.1371/journal.pone.0072352
25. Carlow G, Zinke-Allmang M. Self-similar spatial ordering of clusters on surfaces during Ostwald ripening. *Phys Rev Lett.* (1997) **78**:4601. doi: 10.1103/PhysRevLett.78.4601
26. Riitters K. Pattern metrics for a transdisciplinary landscape ecology. *Landsc Ecol.* (2019) **34**:2057–63. doi: 10.1007/s10980-018-0755-4
27. Saeki K, Sasaki A. The role of spatial heterogeneity in the evolution of local and global infections of viruses. *PLoS Comput Biol.* (2018) **14**:e1005952. doi: 10.1371/journal.pcbi.1005952
28. Ellison G, Glaeser EL, Kerr WR. What causes industry agglomeration? Evidence from coagglomeration patterns. *Am Econ Rev.* (2010) **100**:1195–213. doi: 10.1257/aer.100.3.1195
29. Jiang ZQ, Xie WJ, Zhou WX, Sornette D. Multifractal analysis of financial markets: a review. *Rep Prog Phys.* (2019) **82**:125901. doi: 10.1088/1361-6633/ab42fb
30. Stanley HE, Meakin P. Multifractal phenomena in physics and chemistry. *Nature.* (1988) **335**:405–9. doi: 10.1038/335405a0
31. Saa A, Gascó G, Grau J, Antón J, Tarquis A. Comparison of gliding box and box-counting methods in river network analysis. *Nonlin Process Geophys.* (2007) **14**:603–13. doi: 10.5194/npg-14-603-2007
32. Sun X, Chen H, Wu Z, Yuan Y. Multifractal analysis of Hang Seng index in Hong Kong stock market. *Phys A.* (2001) **291**:553–62. doi: 10.1016/S0378-4371(00)00606-3
33. Torre I, Losada JC, Heck R, Tarquis A. Multifractal analysis of 3D images of tillage soil. *Geoderma.* (2018) **311**:167–74. doi: 10.1016/j.geoderma.2017.02.013
34. Grau J, Méndez V, Tarquis AM, Diaz MC, Saa A. Comparison of gliding box and box-counting methods in soil image analysis. *Geoderma.* (2006) **134**:349–59. doi: 10.1016/j.geoderma.2006.03.009

35. Chhabra A, Jensen RV. Direct determination of the f (α) singularity spectrum. *Phys Rev Lett.* (1989) **62**:1327–30. doi: 10.1103/PhysRevLett.62.1327
36. Chhabra AB, Meneveau C, Jensen RV, Sreenivasan K. Direct determination of the f (α) singularity spectrum and its application to fully developed turbulence. *Phys Rev A.* (1989) **40**:5284–94. doi: 10.1103/PhysRevA.40.5284
37. Nomaler Ö, Frenken K, Heimeriks G. On scaling of scientific knowledge production in us metropolitan areas. *PLoS ONE.* (2014) **9**:e110805. doi: 10.1371/journal.pone.0110805
38. Bettencourt LM. The origins of scaling in cities. *Science.* (2013) **340**:1438–41. doi: 10.1126/science.1235823

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ito and Ohnishi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 10 August 2020
doi: 10.3389/fphy.2020.00230

Rational Belief Bubbles

Hyun-U Sohn[1]* *and Didier Sornette*[1,2]*

[1] Chair of Entrepreneurial Risks, ETH Zurich, Zurich, Switzerland, [2] Swiss Finance Institute, Zurich, Switzerland

We propose an extension of the class of rational expectations bubbles (REBs) to the more general rational beliefs setting of [1, 2]. In a potentially non-stationary but stationarizable environment, among a heterogenous population of agents, it is possible to hold more than one "rational" expectation. When rational but diverse beliefs converge ("correlated beliefs"), they do not cancel each other out in aggregate anymore. This can make them an object of rational speculation. Accounting for the fact that market efficiency has an intrinsic time-dimension, we show that diverse but correlated beliefs can thus account for speculative bubbles, without the need for irrational agents or limits to arbitrage. Many of the shortcomings of REBs that make rational bubbles implausible can be overcome once we relax the ergodicity requirement. In particular, we argue that the hitherto unexplained "bubble component" of REBs corresponds to the extension of the state space in [3].

Keywords: asset pricing, bubbles, efficient markets, rational expectations, rational beliefs, aggregation, heterogeneous expectations, correlated beliefs

JEL Classification: B20, D53, D58, D81, D83, D84, D85, E13, G00

OPEN ACCESS

Edited by:
Siew Ann Cheong,
Nanyang Technological University,
Singapore

Reviewed by:
Satyam Mukherjee,
Indian Institute of Management
Udaipur, India
Yuji Aruka,
Chuo University, Japan

***Correspondence:**
Hyun-U Sohn
hsohn@pm.me
Didier Sornette
dsornette@ethz.ch

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *05 February 2020*
Accepted: *28 May 2020*
Published: *10 August 2020*

Citation:
Sohn H-U and Sornette D (2020) Rational Belief Bubbles. Front. Phys. 8:230. doi: 10.3389/fphy.2020.00230

1. INTRODUCTION

In the following, we provide some addenda and modifications to the work of Kurz and Motolese [3]. We show that their asset pricing model encompasses asset price bubbles, even though they did not intend for it. The significance of this lies in that theirs is an equilibrium model with perfectly rational agents in a frictionless competitive market—the economic equivalent of Euclidean geometry with perfect circles. The only other Euclidean account of bubbles we are aware of is the class of Rational Expectations bubbles (REBs) [4–6]. All other bubble theories rely on either irrational agents or frictions of some sort [see e.g., [7–9] for surveys of the bubble literature].

Just as lines and circles are useful abstractions even though they do not exist in the real world [10, 11], the "purity" of the setting in which bubbles are theorized is of practical import: If frictions are a necessary condition of bubbles, then the policy implication is that the regulator should focus on perfecting markets, pushing ever further the "financialization" of the economy [but see 12, 13]. If it is irrational behavior that causes them, then bubbles may only push the system to more efficient capital allocations [14, 15] or to new learning experiences [16, 17]. Either way, we should then expect to see bubbles never more than once in a generation (but see [18, 19]).

If, by contrast, bubbles can arise even in a perfectly rational, friction-free world, then they should be seen as an inherent feature of market economies, not as temporary aberrations. The regulatory imperative would be to guard, not to "lean," against them—the difference between building dykes and trying to suppress the tides. The problem is that REBs as

the only such account thus far are unexplained sunspot equilibria with strict conditions. For example, they must be present from the beginning of time[1], cannot be negative, and, due to explosive conditional expectations, eventually dominate the economy [20–22]. Hence the preference in the literature for irrational or frictional explanations, which are more plausible at the cost of being less general (see however [9]).

The reduced form of Kurz and Motolese's [3] model corresponds to that of a noisy rational expectations equilibrium [23–25] but the model is derived under rational beliefs [1, 2, 26]. The rational beliefs framework widens the scope of expectational "rationality"[2] from ergodic to non-ergodic, possibly non-stationary economic environments. As a consequence, rational beliefs are diverse—that is, diverse as a feature, not as a bug[3]. The persistent diversity of expectations adds a strategic element to the decision calculi of agents, enlarging the state space of the model by an endogenous variable called "market belief." We contend that this endogenous state variable, which arises from the diversity of beliefs, corresponds to the unexplained "bubble component" in REBs.

There are two classical views on diverse beliefs. One is encapsulated by [28, 29], whose report from a weight-judging competition is the classic example of the "wisdom of crowds"[4]. The opposite view was set forth by [30, ch. 12] who said that, if agents disagreed too much about the future, speculation—betting on others's beliefs—would drive a wedge between price and value. The former leads to Hayek and the efficient market hypothesis (EMH); the latter underpins Minsky's financial instability hypothesis (FIH), and thereby most of the literature on bubbles outside of REBs[5].

What makes the EMH and FIH seem so irreconcilable is that something is missing from the classical accounts, namely the dependence structure of beliefs. It is interesting to note therefore that both noisy rational expectations and the rational beliefs framework do take it into account but in opposite ways. Grossman [32], Hellwig [24], and Diamond and Verrecchia [25] all require independence between private signals (which implies independent beliefs) lest the aggregate signal lose its sufficiency[6] and prices lose their informativeness. By contrast, Kurz and Motolese [3] require dependence between beliefs in order for the endogenous state variable to appear in the equilibrium price. One might say there is really only one model once we move up a level of abstraction: In the corner cases, when individual beliefs are either independent or comonotonic, the equilibrium price conforms to rational expectations; when they are (imperfectly) dependent, the market extends its state space through the endogenous market belief variable.

Accepting diverse beliefs as not the exception but the rule, and shifting the focus to their dependence structure instead, also changes our view of the role they play in the bubble literature thus far. The generic argument is that the broader the range of expectations, the more "confused" or "fantastical" about the future investors are, the more opportunity for speculation, and the bigger the bubble[7]. Historically, though, bubbles seem to have been associated with a reduction in diversity rather than an expansion of it [33]. People are literally "buying into" a, necessarily common, bubble narrative (see also [34]). It is interesting to note therefore that "correlated beliefs" in [3] actually refers to correlated *innovations* to individual beliefs. This means that, as the market belief variable appears, belief *states* converge, lending additional support to our contention that the endogenous state variable is the hitherto unexplained "bubble component" of rational bubbles.

Our approach is different from the idea of chaotic equilibrium cycles (see e.g., [35]). In the later works based on equilibrium cycles theory, the existence of a chaotic equilibrium growth may be derived, and "equilibrium bubbles" associated with transient excursions with nonlinear reversal (crashes) can be observed [36]. In contrast to the small number of "degrees of freedom" involved in the chaotic equilibrium growth models, our framework is based on a social graph representing a large number of individual agents with different but correlated beliefs.

Before we can re-interpret Kurz and Motolese's [3] market belief variable in this way, we need to add some detail. First, section 2 exemplifies the role of correlated beliefs and thus prepares for the subsequent developments. Then, we clarify what we mean when we speak of a bubble. After a quick summary of [3] in section 3, section 4 shows how their model relates to the definition we propose in [37]. Next, for market belief to explain the dynamic appearance of bubbles, we would like to be able to switch it "on" and "off" at will[8]. Kurz and Motolese [3] basically assume that it is always "on," hence have little to say about the dependence structure that gives rise to it, other than that "due to correlation across agents, the law of large numbers is not operative." But not only are there different variants of the law of large numbers (LLN), some LLNs *can* operate on correlated variates. We need more specificity about how much (or what kind of) dependence is necessary for market belief to emerge. Ideally, we would find a precise threshold in the parameter space fixing the dependence structure.

[1] If a REB exists at any time $t > 0$, then it must also exist at $t - 1$.

[2] To be distinguished from choice-theoretic rationality. See [27] for a critique of the terminology.

[3] One way to think about it is that a non-stationary environment underdetermines the set of admissible, or rational, expectations (an expectation is inadmissible, or irrational, if it leads to systematic mistakes). If rational expectations are the fixed-point solution of a function, rational beliefs are the fixed-set solution of a correspondence.

[4] Asked to guess the weight of an ox, none of the 787 individual estimates from the attending crowd were correct but the "middlemost" (presumably the median in today's terminology; the OED defines the word as "That [. . .] in the very middle, or nearest the middle.") came within less than one percent of the correct value (1,207 vs. 1,198 lbs.), beating all but a handful of presumably lucky individuals.

[5] As Brunnermeier and Oehmke [31] point out, "much of the theoretical literature on bubbles can be seen as an attempt to formalize this [Minsky's] narrative."

[6] [32, lemma 1].

[7] We identified the general mechanism behind this way of thinking as a form of explicit symmetry breaking, a notion borrowed from theoretical physics, in [8]. Breaking the symmetry inherent in a diverse set of beliefs explicitly (as opposed to spontaneously) also explains the need for frictions in these models.

[8] Recall that REBs, by contrast, cannot be restarted once they are deflated.

Section 5 introduces a simple graph model for the dependence of beliefs. This not only reduces the dimensionality of the problem but also creates a partition in the parameter space. The phase transition between small and giant components in the graph then effectively serves as the on/off switch for the market belief variable. The idea is to "rotate" the problem from one of identically distributed but correlated belief innovations to an independent but not identically distributed setting. This makes the problem accessible to LLNs of differently-sized variables [38, 39] via the eigenvalue distribution of the correlation matrix. In section 6, we then use linear aggregation [40, 41] to change the endogenous state variable from short- to long-memory. This "dramatizes" the emergence of bubbles. Section 7 concludes with a few suggestions regarding policy implications and empirical applications.

2. THE ROLE OF CORRELATED BELIEFS: AN EXAMPLE

We begin with a re-examination of the introductory example of [42]. The chief distinction between rational belief bubbles and other heterogeneous belief bubbles [43] is our focus on the dependence structure, or correlation, of beliefs rather than their diversity. Lest the main idea get lost in technical details later, the simpler, stylized setting serves to highlight this distinction and provides some intuition about it before we embark on the main argument.

There is an asset that pays dividends d, which are iid with $d \sim N(y, 1/a)$. Disregarding learning effects over time, each agent receives a signal $x_i = d + \varepsilon_i$ composed of a public element d and a private element $\varepsilon_i \sim N(0, 1/b)$. Expectations are heterogeneous because agents mix the common prior with the signals they receive individually:

$$\mathbb{E}^i(d) = \frac{ay + bx_i}{a+b} \tag{1}$$

Aggregated over the set of investors I, these private expectations yield an average expectation of

$$\bar{\mathbb{E}}(d) = \int_I \mathbb{E}^i \tag{2}$$

$$= \frac{ay + bd}{a+b} \tag{3}$$

Let agent i form a second-order belief

$$\mathbb{E}^i(\bar{\mathbb{E}}(d)) = \frac{ay + b\mathbb{E}^i(d)}{a+b} \tag{4}$$

$$= \frac{ay + b\left(\frac{ay+bx_i}{a+b}\right)}{a+b} \tag{5}$$

$$= \left(1 - \left(\frac{b}{a+b}\right)^2\right) y + \left(\frac{b}{a+b}\right)^2 x_i \tag{6}$$

Aggregating again and comparing to Equation (3) now shows that the average expectations operator is not a "normal" expectations operator. It violates the law of iterated expectations:

$$\bar{\mathbb{E}}(\bar{\mathbb{E}}(d)) = \left(1 - \left(\frac{b}{a+b}\right)^2\right) y + \left(\frac{b}{a+b}\right)^2 d \neq \bar{\mathbb{E}}(d) \tag{7}$$

This violation in turn leads to systematic pricing errors in the market. Insofar as the relevant expectations operator is furnished by "the average investor," one can see this heuristically by expanding the basic asset pricing equation

$$p_t = \bar{\mathbb{E}}_t(m_{t+1}(p_{t+1} + d_{t+1})) \tag{8}$$

$$= \bar{\mathbb{E}}_t(m_{t+1})\left[\bar{\mathbb{E}}_t(\bar{\mathbb{E}}_{t+1}(m_{t+2}x_{t+2})) + \bar{\mathbb{E}}_t(d_t + 1)\right] \tag{9}$$

$$= \text{etc.} \tag{10}$$

where m_t is the stochastic discount factor. A failure of the law of iterated expectations thus also means a breakdown of the basic asset price equation. Allen et al. [42] essentially blame the disagreement induced by the private signals for this failure: "if there is differential information between investors so that there is some role for the average expectations about payoffs, the folding back of future outcomes to the present cannot easily be achieved." This is akin to Keynes's position.

Indeed, if one eliminates the private signals from the model [which amounts to taking the limit $1/a \to 0$ in expression (1)], the law of iterated expectations starts to work again: $\mathbb{E}^i(d) = \bar{\mathbb{E}}(d) = \bar{\mathbb{E}}(\bar{\mathbb{E}}(d)) = y$. It thus appears that investor disagreement introduced some sort of noise that prevented the law from operating. In reality, though, it was not the disagreement that induced the failure. If we go to the other extreme and eliminate the public signal instead: $x_i = y + \varepsilon_i$, the information agents receive is still differential and still induces diversity of beliefs $\mathbb{E}^i(d) = x_i$. Yet aggregating over the set I of agents now yields $\bar{\mathbb{E}}(d) = \int_I x_i = y = \mathbb{E}^i(\bar{\mathbb{E}}(d)) = \bar{\mathbb{E}}(\bar{\mathbb{E}}(d))$. Suddenly, we find ourselves in Galton's world!

The reason the law of iterated expectations for the average expectations operator failed is that the expectations were not only diverse or differential but also correlated. And the reason for this is that the signals x_i were coordinated by the public element d:

$$\mathbb{C}\text{ov}(\mathbb{E}^i(d), \mathbb{E}^j(d)) = \mathbb{E}((x_i - y)(x_j - y)) \tag{11}$$

$$= \mathbb{E}((d - y + \varepsilon_i)(d - y + \varepsilon_j)) \tag{12}$$

$$= \mathbb{E}((d-y)^2 + (d-y)\varepsilon_i + (d-y)\varepsilon_j + \varepsilon_i\varepsilon_j) \tag{13}$$

$$= \frac{1}{a} \tag{14}$$

$$> 0 \tag{15}$$

To summarize, the (potential) wedge between price and value associated with the failure of basic asset pricing relations and identified by the failure of the law of iterated expectations was caused *not* by the diversity of beliefs but by the correlation between them.

3. ASSET PRICING UNDER RATIONAL BELIEFS

Kurz and Motolese [3] present an infinite-horizon, discrete-time equilibrium asset pricing model. In order to be self-contained, we provide a brief summary[9].

Let $(d_t)_t^\infty$ be an exogenous random sequence of payoffs of a risky asset. This random sequence has a true probability which is possibly non-stationary, unknown, and unknowable but assumed to be weak asymptotic mean stationary (WAMS). The agents have a large sample of historical data $d_t, t = -1, -2, \ldots$ at their disposal from which they infer a unique empirical probability measure. This empirical measure generates data according to

$$d_{t+1} = \lambda_d d_t + \rho_{t+1}^d \tag{16}$$

$$\rho_{t+1}^d \sim N(0, \sigma_d^2) \tag{17}$$

This is the common reference point for all agents $i \in I = \{1, 2, \ldots, n\}$, where n is the number of agents in the system. The law of motion (16) fixes the set of rational beliefs.

Subjective beliefs d_t^i about the fundamentals d_t are formed by augmenting the stationary measure with an individual belief state g_t^i:

$$d_{t+1}^i = \lambda_d d_t + \lambda_d^g g_t^i + \rho_{t+1}^{id} \tag{18}$$

The individual beliefs are assumed to be rational, which means that g_t^i fluctuates around 0:

$$g_{t+1}^i = \lambda_Z g_t^i + \rho_{t+1}^{ig} \tag{19}$$

$$\rho_{t+1}^{ig} \sim N(0, \sigma_g^2) \tag{20}$$

The innovations ρ_{t+1}^{ig} to the individual belief states are coupled by a correlation matrix $\mathbf{\Sigma}^g$.

As diverse expectations are readily ascertained by agents, the state space is "expanded" by the average state of belief, dubbed "market belief" Z_t:

$$Z_t = \frac{1}{n} \sum_i^n g_t^i \tag{21}$$

This is basically a measure of market sentiment, with $Z_t > 0$ indicating that agents on average expect temporarily higher than normal payoffs (and vice versa). Market belief or sentiment Z_t evolves according to

$$Z_{t+1} = \frac{1}{n} \sum_i \lambda_Z g_t^i + \frac{1}{n} \sum_i \rho_{t+1}^{ig} \tag{22}$$

$$= \lambda_Z Z_t + \rho_{t+1}^Z \tag{23}$$

Since $\mathbf{\Sigma}^g$ may be time-varying, Z_t may be non-stationary. It is, however, also assumed to be WAMS, yielding a stationary representation with

$$\rho_{t+1}^Z \sim N(0, \sigma_Z^2) \tag{24}$$

By assumption,

$$\begin{pmatrix} \rho_{t+1}^d \\ \rho_{t+1}^Z \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0, \end{matrix} \begin{bmatrix} \sigma_d^2 & 0 \\ 0 & \sigma_Z^2 \end{bmatrix} = \tilde{\Sigma} \right), \text{ iid} \tag{25}$$

Given this setup, there exists a unique equilibrium map $d \times Z \to p$ from the state space to prices.

Proposition 3.1. *Under the conditions described above, there exists a unique equilibrium price map*

$$p_t = a_d d_t + a_Z Z_t + P_0 \tag{26}$$

from the state space of fundamentals d_t and market belief Z_t to prices p_t.

Proof: See [3, theorem 2]. □

Our main interest lies in this map. Can we use market belief or sentiment Z_t to explain bubbles?

We examine the origin as well as nature of correlations between individual beliefs that can lead to a natural explanation of bubbles in this rational belief framework. We use a formalism in terms of processes, which is coarse-grained compared with the field of strategic noncooperative games, for which it has been proved that there exists two classes of correlated equilibria, exogenous, and endogenous [44, 45]. Our derivation of correlated equilibria performed in section 6 does not require the game theoretical mathematics and can proceed more intuitively and transparently.

4. BUBBLES VS. EFFICIENT MARKETS

In contrast to topics of similar standing, bubbles are remarkably contentious even on an existential level. This is because even after centuries the term is still not well-defined. The problem stems from the use of "fundamental value" in many definitions—a highly contingent concept of its own. In our view, it is fine to construct a theory of value; it is a stretch to build definitions on top of it. This is why we proposed, in [37], to re-define bubbles not in "money-space" where value lives but in the time-domain where the efficient market lives. Before presenting the idea, we provide a brief review of various bubble definitions that have been proposed in the literature to contrast with our present approach.

4.1. Bubble Definitions in the Literature

4.1.1. Statistical Definitions

As a first group of bubble definitions, there are those that focus on the price trajectory or other observables such as trading volume, without reference to theoretical notions like fundamental value. For example, Kindleberger and Aliber [46] regard as bubbles "any upward price movement over an extended period of 15 to 40 months that then implodes." Exchanging the specification of the time-horizon for a size requirement, as it were, Goetzmann [47] defines bubbles as a doubling in the market price followed by a 50% fall[10]. Presumably, then, bubbles cannot occur in fixed

[9]For a quick overview of rational belief theory, we refer to its entry in Wikipedia which one of us (HS) contributed. For a survey, see [26].

[10]One may recall here that [48] defined efficient markets as ones "in which price is within a factor of 2 of value. [. . .] The factor of 2 is arbitrary, of course."

income or other markets where there is a natural upper bound on the market price! The fund manager GMO proposes that bubbles occur "when prices rise two standard deviations above their norm" [49, 50]. This is more flexible than an absolute size requirement but, alas, opens a whole other can of worms, like estimation issues, ergodicity assumptions, or the question if the second moment even exists for a given asset.

Brock, as cited in [51], defines bubbles as "a monotonically increasing sequence of prices." Hüsler et al. [52] and Leiss et al. [53] cite super-exponential growth rates[11] as the hallmark of a bubble. This chimes with [46] in that it also implies an unsustainable price path but differs in that it does not require an "implosion" or market crash.

What the definitions in this category have in common is that they neither imply nor necessitate a mispricing *per se*. They focus on the observable (the price series) and do not mix theoretical concepts into the definition. In particular, there is no notion of value here. This is an appealing feature for a definition, as *explanandum* and *explanans* then are clearly separated from each other. Bubbles, defined like this, can be tested without the problem of the joint hypothesis. On the downside, insofar as a definition depends on the full path, including a crash at the end, it can be guilty of *post hoc ergo propter hoc* in practice. Insofar as the theoretical underpinning is lacking, the definitions in this category can also be too broad in scope: Empirically, too many price series can fit a statistical bubble definition without necessarily corresponding to our intuition of what a bubble "should" be. For example, an interest rate sensitive stock might follow a rate cycle up "over an extended period of fifteen to forty months" only to then "implode" upon the revelation of a criminal investigation. Few would characterize this as a bubble. Context, as it were, *is* important.

4.1.2. Comparative Definitions

As a second category, there are bubble definitions based on comparisons, usually between price and some notion of value. For instance, the *New Palgrave Dictionary of Economics* defines bubbles as "asset prices that exceed an asset's fundamental value" [54]. Bland as it may appear, this excludes the possibility of negative bubbles, a significant restriction to make *by definition*, as it were. Temin and Voth [55] by contrast identify bubbles as "periods of substantial mispricing" which allows for undervaluations as well as overvaluations but adds a size requirement ("substantial"). Levine et al. [56] define bubbles as simply a "misfit between the market price and the true value of an asset" with no such qualification. This lack of specificity makes it hard to see where the line between excess volatility and bubbles should be drawn. The point is not to niggle or read too much into what may have been intended as merely passing remarks in a much longer work. It is to show that just because a definition is done casually does not mean it has no consequences—especially when we have to relate different studies to each other.

Apart from direct appeals to value, comparisons can also refer, more obliquely, to the information sets on which "true value" is presumably based. For instance, Blanchard and Watson [5] define bubbles as price movements which are "unjustified by information available at the time." More emphatically, Asness [57] demands that the term should apply only when "no reasonable future outcome can justify" the price. This seems to posit a range of admissible price paths, defining bubbles negatively, or by exclusion.

For all their differences, comparative definitions always require a theory of asset pricing, if only implicitly, for a notion of what the correct price is supposed to be. This is their Achilles' heel and the chief criticism of efficient market proponents. For example, Santos and Woodford [58] compare the market price of an asset to the state-price weighted sum of its real payoffs, while [59] uses the realized return on an asset over a sufficiently long time after trading. Different studies can thus agree, in general terms, to define bubbles as a divergence of price from value *and still disagree* over whether a particular price series is a bubble or not. This can make it all seem a bit arbitrary.

4.1.3. Detailed Definitions

A third group of definitions goes beyond the perceived gap between price and value by tying it to specific explanations. For example, Kirman and Teyssière [60] require that the gap between price and value be "endogenous, i.e., not directly produced by exogenous shocks." In other words, the mispricing must arise in a certain way in order for it to count as a bubble. Brunnermeier and Oehmke [31] concur that "not every temporary mispricing can be called a bubble." In particular, it has to arise "because investors believe they can sell the asset at an even higher price to some other investor in the future," so for them the speculative motive is essential. Roubini [61] even introduces a policy dimension by distinguishing between "endogenous" and "exogenous" bubbles, where the former are bubbles whose "probability and size can be affected by monetary policy" while the latter cannot[12]. As an extreme example of the involute nature of the definitions in this category, let us quote from [63]:

> I would say that a speculative bubble is a peculiar kind of fad or social epidemic that is regularly seen in speculative markets; not a wild orgy of delusions but a natural consequence of the principles of social psychology coupled with imperfect news media and information channels. […] I offered a definition of bubble that I think represents the term's best use: A situation in which news of price increases spurs investor enthusiasm which spreads by psychological contagion from person to person, in the process amplifying stories that might justify the price increases and bringing in a larger and larger class of investors, who, despite doubts about the real value of an investment, are drawn to it partly through envy of others' successes and partly through a gambler's excitement.

Basically the obverse to our first category, it is not surprising then to find that detailed definitions tend to be too narrow in scope. Would a bubble that arose by a different mechanism, or in a market in which the proposed mechanism does not apply, also be a "bubble"? For example, would a "political bubble" [64] not

[11] faster than exponential growth, or growth rates that themselves grow.

[12] A similar but more general argument, less focused on monetary policy, has been put forth in [62].

count as a bubble to [31]? Or if it did, doesn't this mean that there must exist a less restrictive superset of bubbles, of which the two variants (political vs. speculative) are but particular cases? And if not, how are we to relate the results and policy implications of different studies to each other? Would a bubble indicator constructed for, say, speculative bubbles still be expected to detect politically driven ones?

The above quote also illustrates that the more detailed a definition, the more likely it is to mesh the notion of bubbles with behavioral assumptions or market frictions. Arguably it is this that makes bubbles such a loaded term. With respect to recessions, inflation or unemployment, the debates may be vigorous but at least their subjects are accepted. By contrast, bubbles remain "existentially controversial." Perhaps this is because the more detailed a definition, the more it acts as a Trojan horse: the mere use of the term may already admit of assumptions one does not wish to make. It is thus that the rejection of behavioral hypotheses or doubt about the effectiveness of monetary policy may lead one to reject the concept of bubbles, almost as an unintended side effect. For the sake of discussion, we should therefore move away from such evocative definitions toward greater formalism and pithiness. In the words of Brock [65], "for the quality of a theory to improve over time, definitions must become more rigorous and less ambiguous."

4.2. Market Efficiency and Time-Dependence of Market Efficiency

It remains commonly overlooked that the concept of market efficiency has a time-dimension. Markets essentially transform informational inputs, modeled by a filtration $(\mathcal{F}_t)_{t=0}^{\infty}$, into price signals $(p_t)_{t=0}^{\infty}$. Markets thus act as a map $\Delta\mathcal{F} \rightarrow \Delta p$ from news to price changes. Market attributes are naturally defined in terms of these primitives. Eliding the discount factor for simplicity, the efficiency of markets has been characterized by the martingale property (cf. [66–68]), where

$$\mathbb{E}_t(p_{t+1}|\mathcal{F}_t) = p_t \tag{27}$$

A market is efficient *relative to the news process* $(\mathcal{F}_t)_{t=t_0}^{\infty}$ iff the map $\Delta\mathcal{F} \rightarrow \Delta p$ produces a martingale. This means that price is an unbiased predictor and that an efficient market does not allow trading profits based on the current information set [69].

But just as efficiency can only be defined relative to an information set, it also requires a time-scale. It is implicitly understood in Equation (27) that the time step 1 from t to $t+1$ is the relevant time-scale. That is, if we take t_0 to be the present, Equation (27) can be written like

$$\mathbb{E}_{t_0}(p_{t_0+nt}|\mathcal{F}_{t_0}) = p_{t_0} \tag{28}$$

with $n = 1$ and the understanding that the martingale condition holds for $n \in \mathbb{N}$. But any given discrete-time price process can be seen as merely a sampling from an underlying continuous-time process, which could have been sampled at a different rate or frequency, say τ or T with $0 < \tau < t < T$. That a process is a martingale on one time-scale neither necessitates nor implies that it is one on another[13]. This opens the possibility that a market is efficient on one time-scale but inefficient on another.

Such a disjunction between time-scales can be supported empirically[14] as well as theoretically, from reading [72] "in reverse": To recap his argument, as long as liquidity is not infinite and there is a strictly positive bid-ask spread $s > 0$ in the market, successive price changes Δp will exhibit serial dependence and the martingale property will not hold. Adapting his notation, let those price changes be measured at the time-scale $\tau < t$, i.e., $\Delta p_{2\tau} = p_{2\tau} - p_\tau$, to make the connection to our discussion clearer. The bid-ask spread induces an asymmetry in the price path at the scale τ (see **Figure 1**): If the last transaction was conducted at the bid, then the next move can only be up (by the spread s) or 0. If the last transaction was conducted at the ask, then the next move can only be down (by s) or 0. One time-step further, the situation is reversed. If the last move was up or 0 (down or 0), then the next move can only be down or 0 (up or 0). The bid-ask spread thus introduces a serial dependence into successive price movements that is not compatible with the martingale condition of an efficient market.

At the same time, over a sufficient number n of time-steps τ, the transition probabilities converge to a (symmetric) steady state. This means that for $t \geq n\tau$, with n sufficiently large, the effect of the bid-ask spread (or, by extension, other microstructural factors) "washes out": Measured on the micro-scale τ, the process exhibits serial dependence; measured on the macro-scale $t \geq n\tau$, the price process can conform to the martingale property again.

Let us illustrate this phenomenon analytically with a toy model, the two-step random walk in [73], a special case of the class of persistent random walks [cf. [74], section 5.2]. Let $\Delta p \in \{U, D\}$ for up = +1, down = -1. Define π_{UU} as the joint probability that the price goes up twice in a row; π_{UD} as the probability that an up move is followed by a down move; and π_{DD}, π_{DU} as the probabilities of down-down and down-up moves. Let $\pi_{UU} = 1/6, \pi_{DU} = \pi_{UD} = 1/3, \pi_{DD} = 1/6$. Suppose the last move was up and start at time t_0 with $p_{t_0} = 100$. Then

$$\begin{aligned}
\mathbb{E}(p_{t_0+\tau}|\mathcal{F}_{t_0}) &= \mathbb{E}(p_{t_0+\tau}|\text{up}) && (29)\\
&= 100 + \pi_{U|U} \times 1 + \pi_{D|U} \times (-1) && (30)\\
&= 100 + \frac{1}{3} - \frac{2}{3} && (31)\\
&\neq 100 && (32)
\end{aligned}$$

where $\pi_{U|U}, \pi_{D|U}$ are the corresponding conditional probabilities. That is, one time-step forward, this two-step random walk is not a martingale. However, if we perform the

[13] In [37], we use so-called n-step or persistent random walks as an example. Another way to see this is by letting the expectation cycle through a periodic function, e.g. a sine-curve: $x_t = \sin(t) + \epsilon_t$ with $\epsilon \sim N(0, \sigma^2)$. Then $\mathbb{E}_t(x_{t+2\pi n}) = x_t$. As long as we sample at the frequency of $(2\pi)^{-1}$, the series looks like a martingale but at any other frequency, it is not.

[14] See for example [70] or, more plastically, the case study of [71]. The point being that a market, no matter how efficient, always needs *some* time to digest information.

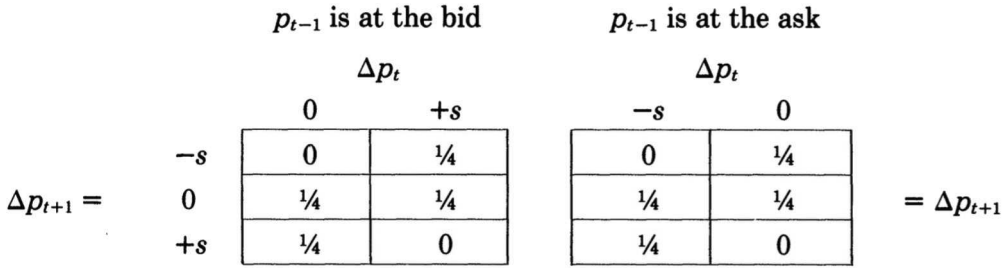


FIGURE 1 | Table of transition probabilities, conditional on the last transaction having been conducted at the bid or at the ask price, adapted from [72, p. 1129].

same calculation two time-steps forward,

$$\mathbb{E}(p_{t_0+2\tau}|\mathcal{F}_{t_0}) = \mathbb{E}(p_{t_0+2\tau}|\text{up}) \tag{33}$$
$$= 100 + \pi_{UU|U} \times 2 + \pi_{DD|U} \times (-2) \tag{34}$$
$$= 100 + \frac{1}{3} - \frac{1}{3} \tag{35}$$
$$= 100 \tag{36}$$

The reason is that the memory gets lost at the time-scale $t = 2\tau$,

$$\pi_{UU|U} = \frac{\pi_{UU\wedge U}}{\pi_U} \tag{37}$$
$$= \frac{\pi_{UU} \times \pi_U}{\pi_U} \tag{38}$$
$$= \pi_{UU} \tag{39}$$

As a result, even though the same price process exhibits serial correlation at the scale τ, it conforms to the martingale property at the scale $t = 2\tau$.

4.3. Bubbles as Elongations of Characteristic Time-Scales of Markets

We call the time-scale at which the martingale property holds the characteristic time-scale of a market. It is inversely proportional to the speed with which a market can "digest" information. Not only do different markets process information at different speeds, the same market can slow down or speed up over time. Projected onto the time-line, bubbles appear as regimes in which the speed of a market slows down significantly from normal.

***Definition** 4.1.* Given a market that is efficient relative to $\mathcal{F}$ at the time-scale t, a t-bubble occurs when the price process changes such that the martingale condition $\mathbb{E}_{t_0}(p_{t_0+T}|\mathcal{F}_{t_0}) = p_{t_0}$ now only holds at time-scales $T > t$. As a boundary case, we include regimes where $T = \infty$ or the condition never holds.

Colloquially, we may call t-bubbles simply "bubbles" so long as it is clear that the notion only makes sense when set in relation to the characteristic time-scale t of the market. Conversely, a bubble is a slowdown in the map $\Delta\mathcal{F} \to \Delta p$ from the "normal speed" $\sim 1/t$ of the market to $\sim 1/T$—a sort of "informational constipation" if you will, as the "digestion" of news becomes slower[15].

Our re-definition of bubbles is sufficiently general so as to be compatible with most of the existing definitions in the literature. The lengthening of the time-scale only serves to create space for a variety of bubble dynamics "in-between" the points at which the martingale property is restored. The general principle is to eliminate (the conditions for) the bubble from a model and inspect the time-scale t at which the market in the model is efficient. If the bubble component has a finite survival time, this can be taken as a lower bound for T. For example, under the limited arbitrage argument of [78], the duration of the bubble is finite with a survival time of $\bar{\tau}$ (in their notation). Without the bubble, the market's characteristic time-scale is t; with a bubble, it slows down to $1/\bar{\tau}$.

To sum up, market efficiency has a time-dimension. It is therefore not enough to speak of a market as efficient. In addition to the news process $(\mathcal{F}_t)_{t=0}^{\infty}$ relative to which efficiency is defined, one also needs to state *at which time-scale* efficiency is supposed to hold. The time it takes a market to fully absorb an information increment $\Delta\mathcal{F}$ can be random but has a characteristic scale, in the sense that it fluctuates within certain bounds or that its mean is defined. In the following, we will take this characteristic time-scale of a market as a given[16].

4.3.1. Application to REBs

Blanchard [4] constitutes an interesting example because even under a bubble the price path still follows a martingale. In equilibrium, the probability of a crash is supposed to exactly balance the added growth factor of the bubble component $b_t = p_t - \bar{p}$, where $\bar{p}$ is the fundamental value, or

$$\mathbb{E}_t(b_{t+1}|b_t > 0) = b_t \tag{40}$$

if we elide the discount factor for simplicity. That is, the price would simply incorporate the bubble component via

$$\mathbb{E}(p_{t+1}|\mathcal{G}_t) = \bar{p} + b_t \tag{41}$$

[15] An incarnation of this is provided by the Hawkes self-excited conditional Poisson process, which has been used to characterize the level of endogeneity or reflexivity of financial markets [75–77]. As the "branching ratio" approaches 1, the market becomes more and more endogenous and the response time to shocks diverges, so that news have a longer and longer lived impact.

[16] It is also possible, though, to conceive of financial markets in which the mean time to digest news diverges. This could occur, for instance, when the absorption time is distributed according to a power law in the tail with tail exponent <1. As many response functions are power laws in the time domain with small exponent, this is indeed an interesting possibility. In this case, the market would never be efficient even at arbitrarily large time-scales.

Note however that $\mathcal{G} = (\mathcal{G}_t)_{t=0}^{\infty}$ is generated by fundamental variables *as well as* the bubble component $b = (b_t)_{t=0}^{\infty}$. But definition 4.1 relies on an efficient market as a benchmark. Therefore, the bubble according to 4.1 cannot be defined relative to $\mathcal{G}$. Instead, we must introduce a "copy" of the market, a hypothetical market in which all the elements are the same (agents and their preferences, assets, institutions, etc.) *except* the information process, which must exclude the bubble component b_t. Denoting this filtration $\mathcal{F} = (\mathcal{F}_t)_{t=0}^{\infty}$,

$$\mathbb{E}(p_{t+1}|\mathcal{F}_t) = \bar{p} \tag{42}$$

$$\neq \bar{p} + b_t \tag{43}$$

That is, relative to the efficient market, the bubble component introduces an estimation or valuation error which survives with probability π and collapses with probability $1-\pi$ in [4]. It thus has an expected length of $\pi/(1-\pi)$ time-steps, or $T \geq \pi/(1-\pi)$ time steps. For example, if $\pi = .95$, then $T \geq 19$ time steps. As the probability $(1-\pi)$ of a crash approaches zero, $T \to \infty$ in the limit.

4.3.2. Application to [3]

In accordance with definition 4.1, we must first specify an efficient benchmark against which a bubble can be defined. In [3], the natural benchmark is a market driven exclusively by the fundamentals (d_t), or the price map $p_t = P_0 + a_d d_t$ (see theorem 3.1). This relationship holds when the correlation of beliefs is sufficiently low for the law of large numbers to apply and market belief to vanish. Since the stable distribution of the dividend process is persistent (cf. Equation 16), we have

$$\mathbb{E}_t(p_{t+n}|d_t) = P_0 + a_d \lambda_d^n d_t \tag{44}$$

This means that any dividend shock will wash out exponentially fast, or that the market is (approximately) efficient at a time-scale T, where $T = \{n \in \mathbb{N} : \lambda_d^n < \varepsilon\}$ for any chosen tolerance $\varepsilon > 0$. For the sake of argument, let us fix $\lambda_d = 1/4$ and $\varepsilon = 1/100$. Then the market is efficient at the time-scale of 4 time steps.

If market belief Z_t enters the equation because "due to correlation across agents, the law of large numbers is not operative," ([3], p. 301) there are two scenarios. One, it has the same, or lower, persistence with $0 < \lambda_Z \leq 1/4$ in Equation (23). Then market belief increases the variance of the process but does not change the efficient time-scale. This was the original focus of [3]. Two, $\lambda_Z > \lambda_d$. Suppose, for instance, that $\lambda_Z = 3/4$. Now the innovations ρ_{t+1}^Z to market belief Z_t affect the expected price (under the stable distribution) for $T = 17$ periods.

So far, we haven't done anything yet. It is simply an observation that if $\lambda_Z > \lambda_d$, market belief Z_t potentially slows down the market or "interferes" with the fundamentals d_t in the equilibrium price equation. This conforms to our re-definition of bubbles in terms of time-scales rather than price-value divergences but there is nothing in the original work of Kurz and Motolese [3] that would motivate a particular parameterization.

In the following, we now propose two extensions or modifications to the model: First, to find an "on/off" switch for market belief Z_t in the parameter space of the dependence structure of beliefs, we propose to generate $\mathbf{\Sigma}^g$ from a "social graph." This reduces the dimensionality of the problem and at the same time provides a natural partition of the parameter space. Then, to emphasize or perhaps motivate $\lambda_Z > \lambda_d$, we use linear aggregation [40, 41] to move the model beyond a simple parametrization. The aggregation of short-memory correlated belief transitions can lead to a long-memory series. The resulting divide between fast-moving fundamentals d_t and slow-moving market belief Z_t magnifies the bubble effect outlined above.

5. HOW MUCH DEPENDENCE IS ENOUGH?

This section discusses the first modification to the model of [3], which consists in introducing a "social graph" to motivate the existence of a non-trivial correlation matrix $\mathbf{\Sigma}^g$ of the innovations ρ_{t+1}^{ig} to the individual belief states defined in Equation (19). In other words, we propose to generate $\mathbf{\Sigma}^g$ from a "social graph," which represents the investor population. For simplicity, we use the Erdős-Renyi random graph model $G(n,p)$, where n is the number of agents (or vertices) in the system and $p \in [0,1]$ is the probability with which a random pair of agents i,j is linked by an edge $\{i,j\}$. The edge is represented by a set as opposed to a tuple as we take the graph to be undirected. To keep things tractable, we also do not assign weights to links or distinguish between different degrees of correlation.

Whenever agents (or their nodes) are connected in the graph, we take their beliefs to be correlated[17]. Moreover, we let their beliefs be correlated by a constant $\bar{\rho} \in (0,1]$, thereby abstracting not only from directions of influence (by the undirectedness of the graph) but also from variations in the degree of influence[18]. In our model, the pertinent fact is that (or whether) there exists *some* level of mutual influence or communication, regardless of its direction or strength. Formally, $\mathbf{\Sigma}^g$ is constructed via the rule $\forall i \neq j \in \{1,\ldots,n\}$,

$$\sigma_{ij}^g = \sigma_{ji}^g = \begin{cases} \bar{\rho}, & \text{if } i,j \text{ are connected;} \\ 0, & \text{otherwise} \end{cases} \tag{45}$$

where σ_{ij}^g is the (i,j)-th component of $\mathbf{\Sigma}^g$. Since the social graph of the agents is random, so is $\mathbf{\Sigma}^g$.

A group of agents that is connected to each other, or whose beliefs are correlated, is called a clique or "component" of the

[17]For those unfamiliar with the terminology of graphs, note that there is a difference between two agents being "linked" by an edge and being "connected" by a path. Vertices i,j in the graph are said to be linked if the graph contains an edge $\{i,j\}$ between them. For vertices to be connected, it suffices that there exist a path between them. Agents that are linked are also connected but agents may be connected without being linked. For example, if agents i and j are linked by $\{i,j\}$ and agents j and k are linked by $\{j,k\}$ but there is no edge between agents i and k, agents i and k are still considered connected via the path $\{\{i,j\},\{j,k\}\}$.

[18]That $\bar{\rho}$ is chosen positive stems from the fact that sufficiently large groups of agents cannot be all negatively correlated. Suppose, for example, that there are three agents i,j,k all perfectly *negatively* correlated. If i becomes more optimistic, j must become more pessimistic, which means k must become more optimistic. But this belies the negative correlation between i and k. The same principle holds with less-than-perfect (negative) dependence in larger groups.

graph. One can distinguish between two kinds of cliques or components. If the size of a clique does not scale with the total number of agents in the system, or is independent of n, then we call it a "small component." If a clique is such that it becomes larger the larger becomes n, it is called a "giant component." This difference in the scaling behavior of cliques has an important consequence for $\mathbf{\Sigma}^g$.

By construction, the correlation of beliefs is transitive: if agents i, j have correlated beliefs and agents j, k also have correlated beliefs, then so do agents i, k. Graphically, this means that every clique of agents can, after appropriate reshuffling of the indices, be represented by a block matrix in $\mathbf{\Sigma}^g$. The size of a clique is the number of nodes or agents contained in it. If a clique of correlated agents has size m, say, then the corresponding block matrix in $\mathbf{\Sigma}^g$ is of $m \times m$ dimension. Obviously, if we pick any off-diagonal entry σ_{ij}^g from such a block, its value is $\bar{\rho}$; if we pick an entry of $\mathbf{\Sigma}^g$ that is not part of a clique, then its value is zero. But what about a random pick from $\mathbf{\Sigma}^g$, i.e., the unconditional expectation $\mathbb{E}(\sigma_{ij}^g)$ or average correlation in the system as a whole?

The average correlation of beliefs depends on whether the agents are organized into many small separate components or into a giant connected component. The intuition is that components or cliques of correlated agents in the original space act like "composite agents" when transposed into eigenspace[19]. The law of large numbers then acts, or not, on these "composite agents," which are independent but differently-sized, instead of the actual agents, which are equally-sized but correlated. A system with 1,000 isolated agents behaves essentially the same as a system with 2,000 agents who are organized into cliques of 2.

Small components do not scale with n. As the graph increases from, say, 100 to 1,000 agents, there are more cliques but of the same size (or size distribution) as before. A giant component, by contrast, scales with the system size n. Its size is a constant fraction $S \in [0, 1]$ of n. As a giant connected component arises in the social graph of agents (or their beliefs), the weight of the non-zero correlations in $\mathbf{\Sigma}^g$ remains a constant fraction of the total, and the average correlation of beliefs in the system becomes positive.

Formally, the question is how the average correlation of beliefs behaves as the system becomes large, or

$$\frac{\sum_i \sum_{j \neq i} \sigma_{ij}^g}{n(n-1)} \to \mathbb{E}(\sigma_{ij}^g) \tag{46}$$

as $n \to \infty$. Let $c = (n-1)p$ be the mean degree of the graph[20]. The average size of the small components is

$$R = \frac{2}{2 - c - cS} \tag{47}$$

The fraction S of the n agents or vertices contained in a giant connected component depends on c and is the solution or fixed point of

$$S = 1 - \exp\{-cS\} \tag{48}$$

[19]This is the basic idea of principal component analysis [79].

[20]We skip calculations that do not add to the main point and refer interested readers to [80] instead.

When $c < 1$, $S = 0$ or there is no giant component in the graph. This yields an average correlation coefficient of

$$\frac{2(n-1) \times \bar{\rho} + (n-2)(n-1) \times 0}{n(n-1)} = \frac{2\bar{\rho}}{n} \tag{49}$$

As the system size n increases, the average correlation tends to 0, and in the limit, we can take $\mathbb{E}(\sigma_{ij}^g) = 0$.

When $c \geq 1$, a giant component emerges in the graph and the system behaves differently. The reason is that unlike the small components, the giant component scales with the system size n. Then the average correlation of beliefs is

$$\frac{(S \times n)(S \times n - 1)\bar{\rho}}{n(n-1)} \to S^2 \bar{\rho} \tag{50}$$

as $n \to \infty$ which is strictly positive.

Our use of the Erdős-Renyi random graph theory has been made for the sake of simplicity and tractability. It will be interesting in the future to extend our treatment to the case of Barabasi-Albert networks and other random geometric graphs. However, our main point on the emergence of a non-zero average correlation of beliefs will not be changed. What will be modified are the specific control parameters and conditions under which a non-zero average correlation of beliefs emerges. Furthermore, the application to real-world social networks is reported to a future work.

6. THE EMERGENCE OF BUBBLES

We now make the following modification to [3]:

Axiom 6.1. *Let the agents have coefficients λ^{ig} in (19). In particular, we assume that the coefficients λ^{ig} are drawn from a family of absolutely continuous distributions with support $[0, 1)$ with density*

$$f(\lambda^{ig}; b) \sim c_b(1 - \lambda^{ig})^b \text{ as } \lambda^{ig} \to 1 \text{ from below} \tag{51}$$

with parameters $-1/2 < b < 0$ and $0 < c_b < \infty$.

The density in the axiom is only specified for values close to 1, so this is a flexible semiparametric specification. The only hard requirement is that the coefficients cannot be bounded away from 1 (although they can never attain it). Agents now differ from each other w.r.t. the persistence of their belief states g_t^i. This means that agents differ in their trading horizons: Short-term day traders form subjective expectations that flit around much faster than long-term investors who form expectations over multi-year, even decade-long horizons.

As a result of this positive average correlation in the system, market belief Z_t will assume long-memory and a price bubble develops, as we now show.

Proposition 6.1. *Suppose that $\mathbb{E}(\sigma_{ij}^g) > 0$ or that the average correlation of beliefs in $\mathbf{\Sigma}^g$ is positive. Then there exists a common component or "representative belief" $\pi_t \sim N(0, \vartheta)$ iid in the*

innovations and the transition functions (19) of agent states of belief can be rewritten in terms of this representative belief to

$$g^i_{t+1} = \lambda^{ig} g^i_t + s_i \pi_{t+1} + \varepsilon_{i,t+1} \tag{52}$$

As a consequence, the aggregate market belief Z_t assumes long memory persistence.

Proof: Since $\mathbf{\Sigma}^g$ is symmetric and positive definite, we can perform a spectral decomposition

$$\mathbf{\Sigma}^g = \mathbf{S}\Theta\mathbf{S}' \tag{53}$$

where $\mathbf{S}$ is the matrix of eigenvectors, Θ is the n-dimensional matrix with the eigenvalues $\vartheta_i, i = 1, \ldots, n$ on the diagonal, $\mathbf{SS}' = \mathbf{I}_n$, the n-dimensional identity matrix, $\vartheta_i \in \mathbb{R}^+$, and $\vartheta_i \neq \vartheta_j, \forall i \neq j$.

According to [81], if $\mathbb{E}(\sigma^g_{ij}) > 0$, which we have shown above to be the case when a giant connected component arises in the social graph of the agents, then the biggest eigenvalue, ϑ_1, is distributed according to

$$\vartheta_1 \sim N\left((n-1)\mathbb{E}(\sigma^g_{ij}) + 1 + \frac{\mathbb{V}\mathrm{ar}(\sigma^g_{ij})}{\mathbb{E}(\sigma^g_{ij})}, 2\,\mathbb{V}\mathrm{ar}(\sigma^g_{ij})\right) \tag{54}$$

For correlation matrices, one needs to ensure positive definiteness of $\mathbf{\Sigma}^g$ and [82] show that the result remains valid under suitable restrictions on the support of the off-diagonal entries for large sample correlation matrices.

Given that $\sum_i \vartheta_i = \mathrm{trace}(\mathbf{\Sigma}^g) = n$, all other eigenvalues $\vartheta_{i>1}$ are therefore constrained to be of the order $(1 - \mathbb{E}(\sigma^g_{ij}))$. That is, the larger the system, the greater the dominance of the largest eigenvalue, ϑ_1, over all others, $\vartheta_{i>1}$. Furthermore, as the variance, unlike the mean, in (54) does not scale with n, this dominance also becomes more certain as n grows large.

We exploit this by rewriting the innovations in individual states of belief in terms of a factor model

$$\rho^{ig}_t = s_{i1}\pi^1_t + \varepsilon^i_t \tag{55}$$

where π^1_t is the first principal component

$$\pi^1_t = \mathbf{S}'_1 \boldsymbol{\rho}^g_t \tag{56}$$

and "factor loading" s_{i1} is the $(i, 1)$-th component of eigenvector matrix $\mathbf{S}$ or the i-th component of the first eigenvector $\mathbf{S_1}$. Since π^1_t is simply a linear combination of (multivariate) Gaussians with mean 0 and variance σ^2_g, it is itself $\sim N(0, \vartheta_1)$. The belief transitions (19) now take the form

$$g^i_{t+1} = \lambda^{ig} g^i_t + s_i \pi_{t+1} + \varepsilon_{i,t+1} \tag{57}$$

$$= \frac{s_i}{1 - \lambda^{ig} L}\pi_{t+1} + \frac{1}{1 - \lambda^{ig} L}\varepsilon_{i,t+1} \tag{58}$$

where L is the lag operator.

Hidden in the $\varepsilon^i_t = (s_{i2}, s_{i3}, \ldots, s_{in}) \cdot (\pi^2_t, \pi^3_t, \ldots, \pi^n_t)'$ is the variation of all the other principal components in eigenspace. Since the system is orthogonal, all the summands are independent from each other. Again, each principal component $\pi^k_t, k = 2, \ldots, n$ is nothing but a different linear combination of the original innovations $\boldsymbol{\rho}^g_t$ (Equation 56), hence $\mathbb{E}(\pi^k_t) = 0$. $\sum_{k=2}^n s^2_{ik}\mathbb{V}\mathrm{ar}(\pi^k_t) = \sum_{k=2}^n s^2_{ik}\vartheta_k < \infty$ because the total variance is finite. Therefore, the series ε^i_t converges to zero as $n \to \infty$ [83, thm. 22.6].

It follows that we can neglect the idiosyncratic terms and rewrite individual beliefs solely in terms of the common component

$$g^i_{t+1} = \frac{s_i}{1 - \lambda^{ig} L}\pi_{t+1} \tag{59}$$

Accordingly, market belief (21) now takes the form

$$Z_{n,t} = \frac{1}{n}\sum_i \frac{s_i}{1 - \lambda^{ig} L}\pi_t \tag{60}$$

If we expand this expression,

$$Z_{n,t} = \frac{1}{n}\sum_i^n \lambda^{ig} g^i_{t-1} + \frac{1}{n}\sum_i^n s_i \pi_t \tag{61}$$

$$= \frac{1}{n}\sum_i^n \lambda^{ig}(\lambda^{ig} g^i_{t-2} + s_i \pi_{t-1}) + \frac{1}{n}\sum_i^n s_i \pi_t \tag{62}$$

$$= \sum_{k=0}^{\infty}\left(\frac{1}{n}\sum_{i=1}^n s_i (\lambda^{ig})^k\right)\pi_{t-k} \tag{63}$$

A stochastic process has long memory if its spectral density is of the form $L(\omega)|\omega|^{-2d}$, for some slowly varying function L at zero and $d \in (0, \frac{1}{2})$ [84]. Zaffaroni [41] shows that, for coefficients λ^{ig} distributed according to expression (51) with $b > -1/2$, the coefficients in (63) converge

$$\mu_k = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n s_i (\lambda^{ig})^k \tag{64}$$

with

$$\mu_k \sim a k^{-(b+1)} \tag{65}$$

for some constant a as $k \to \infty$, and that the limit process

$$Z_t = \sum_{k=0}^{\infty}\mu_k \pi_{t-k} \tag{66}$$

has a spectral density $\sim c\omega^{-2d}$ if $b < 0$, for some constant c and $d = -b$. □

For example, what is the probability that the equilibrium price will lie at least one standard deviation of Z_t above P_0 for a period of 100 days? To isolate the effect of long memory on the bubble probability, we normalize the variance of Z_t to a constant 1 for

all t and calculate $\mathbb{P}(\sum_{t=1}^{100} Z_t > 100) = 1 - \mathbb{P}(\sum_{t=1}^{100} Z_t \leq 100)$. Recalling that Z_t is a Gaussian process, we note that

$$\sum_{t=1}^{100} Z_t \sim N\left(0, \mathbf{1}'\mathbf{\Sigma}_{100}\mathbf{1}\right) \tag{67}$$

where $\mathbf{1}$ is the 100×1 vector of ones, $\cdot'$ indicates the transpose, and $\mathbf{\Sigma}_{100}$ is the 100×100-dimensional covariance matrix of $(Z_1, \ldots, Z_{100})$ with (i,j)-entries σ_{ij}^Z.

Case 1: Independence. Under time independence, i.e., if Z_t were iid, the entries σ_{ij}^Z in $\mathbf{\Sigma}_{100}$ are equal to 0 for $i \neq j$ and 1 for $i = j$. Then

$$\sum_{t=1}^{100} Z_t \sim N(0, 100) \tag{68}$$

and the probability $\mathbb{P}(\sum_{t=1}^{100} Z_t > 100) \simeq 0$.

Case 2: Short memory. Under the setup of [3], we have exponentially decaying autocorrelations and the (i,j)-th entry of $\mathbf{\Sigma}_{100}$ is

$$\sigma_{ij}^Z = \lambda_Z^{|i-j|} \tag{69}$$

This corresponds to the yellow line in **Figure 2**. For comparison purposes, we choose a value of $\lambda_Z \simeq .77$, which yields

$$\sum_{t=1}^{100} Z_t \sim N(0, 739) \tag{70}$$

and the probability $\mathbb{P}(\sum_{t=1}^{100} Z_t > 100) \simeq 0.0001173$ or less than 1 in 8,500.

Case 3: Long memory. By contrast with the previous cases, we see a significant probability of bubbles arising under the long memory specification of the previous section, with coefficients as in Equation (65). The (i,j)-th entry of $\mathbf{\Sigma}_{100}$ is now

$$\sigma_{ij}^Z = \sum_{t=0}^{\infty} \mu_t \mu_{t+|i-j|} \tag{71}$$

$$= \sum_{t=0}^{\infty} c t^{-(b+1)} c(t + |i-j|)^{-(b+1)} \tag{72}$$

Figure 2 plots the first 100 autocorrelations for a parametrization of $b = -1/4$ or $\lambda_Z \simeq .77$ (both yielding the same first-order autocorrelations).

The slowly decaying off-diagonal entries of $\mathbf{\Sigma}_{100}$ lead to a much higher dispersion of the sum,

$$\sum_{t=1}^{100} Z_t \sim N(0, 3133) \tag{73}$$

and the probability $\mathbb{P}(\sum_{t=1}^{100} Z_t > 100) \simeq 0.037$ or about 1 in 27. *Ceteris paribus*, introducing long memory increased the chances of a bubble by two orders of magnitude.

7. CONCLUSION

We mentioned in the introduction that the theoretical setting in which bubbles are conceived has practical implications. If bubbles can arise, as we have argued, even under idealized circumstances—in a world of perfectly rational agents acting

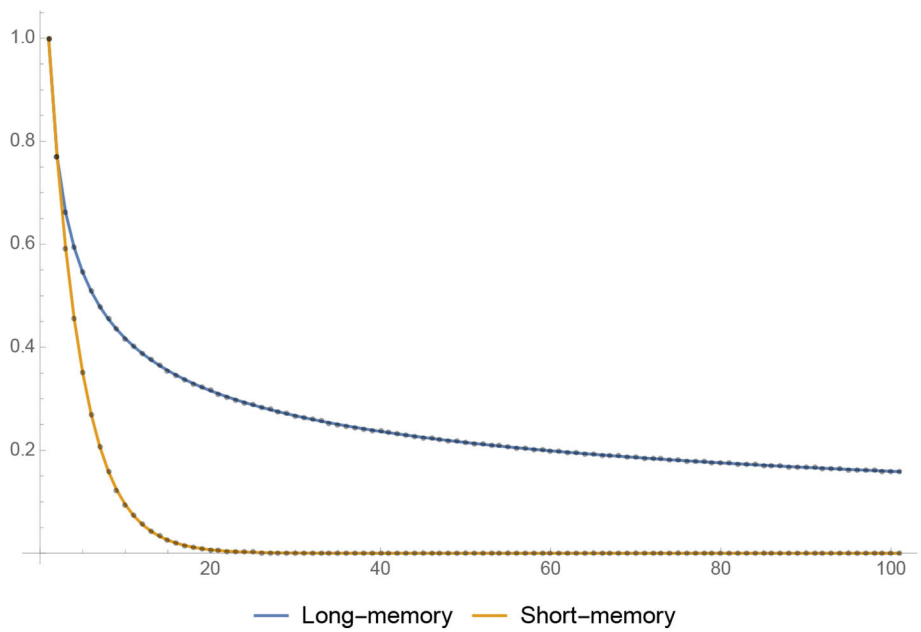


FIGURE 2 | First 100 autocorrelations of the original, short-memory (yellow) and the modified, long-memory (blue) processes, where the parameters are chosen such that the first-order autocorrelation is equal.

in markets with no frictions or "limits of arbitrage"—then they should be recognized as a general and system-immanent feature of market economies. Continuing to view them as aberrations, due to specific faults or circumstances, implies that one can "lean against" or eradicate them by addressing said faults or circumstances. This strikes us as an attempt to suppress the tides. We favor building dykes instead, guarding against their inevitable recurrence by increasing the robustness of the system.

It is also not clear that bubbles are necessarily "bad." Indeed, another way to look at them is as the free-market alternative of industrial policy, i.e. as a spontaneous and decentralized way to achieve coordination, instead of the controlled and centralized approach favored in statist economies. This would be another argument against efforts to suppress bubbles. Alas, it is undeniable that bubbles *can* have bad consequences, particularly if they end in "crashes" or sudden ruptures which can destabilize the system at large.

To make a rational policy tradeoff, we therefore need two further elements, apart from the bubble itself: One, a model of what might be termed "rational belief crashes," noting that there is no violent ending inherent in rational belief bubbles themselves. This conforms to the view explored here that bubbles and crashes are separate events and require separate theories. A bubble does not *have* to end in a crash (it can deflate gently), just as an asset price can crash without a bubble (adverse news, e.g., in the form of a lawsuit, may arrive). The second element that is needed is a welfare analytical model of how rational belief bubbles and crashes fit into an economy with a production and a banking sector.

In terms of empirical applications, we would like to see an augmentation of current sentiment indicators with cross-sectional quantities [recall 32, lemma 1]. This could lead to a "real-time" bubble indicator based on disaggregated, contemporaneous expectations. As far as we are aware, current bubble indicators, which are used to monitor financial markets and systemic risk, are based on historical data and/or estimates of fundamental value.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR'S NOTE

An earlier preliminary version of this manuscript has been released as a pre-print on July 2018 at http://ssrn.com/abstract=3176871 with the present authors (HS-U and DS). The present extended version has been released as a pre-print at http://ssrn.com/abstract=3530895 with the same authors [85].

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We thank (in alphabetical order) Thorsten Hens, Mordecai Kurz, Maurizio Motolese, Hiroyuki Nakata, and Carsten K. Nielsen, and seminar participants at the BIS for helpful comments. Financial support from the Institute for New Economic Thinking (INET) is gratefully acknowledged.

REFERENCES

1. Kurz M. On the structure and diversity of rational beliefs. *Econ Theor*. (1994) **4**:877–900. doi: 10.1007/BF01213817
2. Kurz M. On rational belief equilibria. *Econ Theor*. (1994) **4**:859–76. doi: 10.1007/BF01213816
3. Kurz M, Motolese M. Diverse beliefs and time variability of risk premia. *Econ Theor*. (2011) **47**:293–335. doi: 10.1007/s00199-010-0550-1
4. Blanchard OJ. Speculative bubbles, crashes and rational expectations. *Econ Lett*. (1979) **3**:387–89. doi: 10.1016/0165-1765(79)90017-X
5. Blanchard O, Watson MW. Bubbles, rational expectations and financial markets. In: Wachtel P, editor. *Crises in the Economic and Financial Structure*. Lexington, MA: Lexington Books (1982) **5**:295–315. doi: 10.3386/w0945
6. Tirole J. Asset bubbles and overlapping generations. *Econometrica*. (1985) **53**:1499–528. doi: 10.2307/1913232
7. Kaizoji T, Sornette D. Bubbles and crashes. In R. Cont, editor, *Encyclopedia of Quantitative Finance*. Hoboken, NJ: Wiley. (2010).
8. Sohn H, Sornette D. *Bubbles and Symmetry Breaking: A Review, Working Paper*. (2017).
9. Schatz M, Sornette D. Inefficient bubbles and efficient drawdowns in financial markets. *Int J Theor Appl Fin*. (2020). Available online at: https://ideas.repec.org/p/chf/rpseri/rp1849.html
10. Mandelbrot B. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science*. (1967) **156**:636–8. doi: 10.1126/science.156.3775.636
11. Mandelbrot BB. *The Fractal Geometry of Nature*. New York, NY: W.H. Freeman (1982).
12. Shin HS. Securitisation and financial stability. *Econ J*. (2009) **119**:309–32. doi: 10.1111/j.1468-0297.2008.02239.x
13. Brock WA, Hommes CH, Wagener FOO. More hedging instruments may destabilize markets. *J Econ Dyn Control*. (2009) **33**:1912–8. doi: 10.1016/j.jedc.2009.05.004
14. Farmer JD, Lo AW. Frontiers of finance: evolution and efficient markets. *Proc Natl Acad Sci USA*. (1999) **96**:9991–2. doi: 10.1073/pnas.96.18.9991
15. Evstigneev IV, Hens T, Schenk-Hoppé K. Evolutionary finance. In: T. Hens and K. Schenk-Hoppé, editors. *Handbook of Financial Markets: Dynamics and Evolution*. North-Holland; Amsterdam. (2009).
16. Camerer CF. Taxi drivers and beauty contests. *Eng Sci*. (1997) 11–9.
17. Dufwenberg M, Lindqvist T, Moore E. Bubbles and experience: an experiment. *Am Econ Rev*. (2005) **95**:1731–7. doi: 10.1257/000282805775014362
18. Sornette D, Cauwels P. 1980-2008: The Illusion of the Perpetual Money Machine and what it bodes for the future. *Risks*. (2014) **2**:103–31. doi: 10.3390/risks2020103
19. Reid J, Nicol C, Burns N, Chanda S. *The Next Financial Cri 2017*. Deutsche Bank Markets Research.

20. Diba BT, Grossman HI. On the inception of rational bubbles. *Q J Econ*. (1987) **102**:697–700. doi: 10.2307/1884225
21. Diba BT, Grossman HI. The theory of rational bubbles in stock prices. *Econ J*. (1988) **98**:746–54. doi: 10.2307/2233912
22. Benhabib J, Farmer REA. Indeterminacy and sunspots in macroeconomics. In: Taylor JB, Woodford M, editors. *Handbook of Macroeconomics*. 1st ed. Amsterdam: Elseiver (1999). doi: 10.1016/S1574-0048(99)01009-5
23. Grossman SJ, Stiglitz JE. On the impossibility of informationally efficient markets. *Am Econ Rev*. (1980) **70**:393–408.
24. Hellwig MF. On the aggregation of information in competitive markets. *J Econ Theor*. (1980) **22**:477–98. doi: 10.1016/0022-0531(80)90056-3
25. Diamond DW, Verrecchia RE. Information aggregation in a noisy rational expectations economy. *J Fin Econ*. (1981) **9**:221–35. doi: 10.1016/0304-405X(81)90026-X
26. Motolese M, Nielsen CK. Rational beliefs theory: a review. *Rivista Internazionale di Scienze Sociali*. (2007) 293–326.
27. Kurz M. Symposium: on the role of market belief in economic dynamics, an introduction. *Econ Theor*. (2011) **47**:189–204. doi: 10.1007/s00199-010-0560-z
28. Galton F. Vox populi. *Nature*. (1907) **75**:450–1. doi: 10.1038/075450a0
29. Galton F. *Memories of My Life*. North Yorkshire: Methuen & Co. (1908). doi: 10.5962/bhl.title.28398
30. Keynes JM. *The General Theory of Employment Interest and Money*. London: Palgrave Macmillan. (1936).
31. Brunnermeier MK, Oehmke M. Bubbles, financial crises, and systemic risk. In: Constantinides GM, Harris M, Stulz RM, editors. *Handbook of the Economics of Finance*. Vol. 2B. *Handbooks in Economics*. Amsterdam: Elseiver (2013). doi: 10.1016/B978-0-44-459406-8.00018-4
32. Grossman SJ. On the efficiency of competitive stock markets where traders have diverse information. *J Fin*. (1976) **31**:573–85. doi: 10.1111/j.1540-6261.1976.tb01907.x
33. Chancellor E. *Devil Take the Hindmost: A History of Financial Speculation*. New York, NY: Plume. (1999).
34. Shiller RJ. Narrative economics. *Am Econ Rev*. (2017) **107**:967–1004. doi: 10.1257/aer.107.4.967
35. Boldrin M, Nishimura K, Shigoka T, Makoto Y. Chaotic equilibrium dynamics in endogenous growth models. *J Econ Theor*. (2001) **96**:97–132. doi: 10.1006/jeth.2000.2677
36. Corcos A, Eckmann JP, Malaspinas A, Malevergne Y, Sornette D. Imitation and contrarian behavior: hyperbolic bubbles, crashes and chaos. *Quant Fin*. (2002) **2**:264–81. doi: 10.1088/1469-7688/2/4/303
37. Sohn HU, Sornette D. Bubbles as violations of efficient time-scales. (2017).
38. Bingham NH, Rogers LCG. Summability Methods and almost-sure convergence. In: Bellow A, Jones RL, editors. *Almost Everywhere Convergence II*. Cambridge, MA: Academic Press. (1991). p. 69–83. doi: 10.1016/B978-0-12-085520-9.50012-5
39. Stout WF. *Almost Sure Convergence*. Cambridge, MA: Academic Press (1974).
40. Granger CWJ. Long memory relationships and the aggregation of dynamic models. *J Econ*. (1980) **14**:227–38. doi: 10.1016/0304-4076(80)90092-5
41. Zaffaroni P. Contemporaneous aggregation of linear dynamic models in large economies. *J Econ*. (2004) **120**:75–102. doi: 10.1016/S0304-4076(03)00207-0
42. Allen F, Morris S, Shin HS. Beauty contests and iterated expectations in asset markets. *Rev Fin Stud*. (2006) **19**:161–77. doi: 10.1093/rfs/hhj036
43. Xiong W. Bubbles, crises, and heterogeneous beliefs. In: Fouque JP, Langsam JA, editors. *Handbook on Systemic Risk*. Cambridge, UK: Cambridge University Press (2013). p. 663–713. doi: 10.1017/CBO9781139151184.033
44. Kono N. Noncooperative game in cooperation: reformulation of correlated equilibria. *Kyoto Econ Rev*. (2008) **77**:107–25. Available online at: https://www.jstor.org/stable/43213362
45. Kono N. Noncooperative game in cooperation: reformulation of correlated equilibria (II). *Kyoto Econ Rev*. (2008) **78**:1–18. Available online at: https://www.jstor.org/stable/43213369
46. Kindleberger CP, Aliber RZ. *Manias, Panics, and Crashes: A History of Financial Crises*. 5th Edn. (2005). doi: 10.1057/9780230628045
47. Goetzmann WN. Bubble investing: learning from history. In: Chambers D, Dimson E, editors. *Financial Market History: Reflections on the Past for Investors To*. Charlottesville, Va: CFA Institute. (2017). pp. 149–69.
48. Black F. Noise. *J Fin*. (1986) **41**:529–43. doi: 10.1111/j.1540-6261.1986.tb04513.x
49. Mackintosh J. Market mania: a 'how-to' guide to hot assets. *Fin Times*. (2014).
50. Buttonwood. Time and tide. *The Economist*. (2017).
51. Veres B. Woody Brock's challenge to Krugman and the Keynesians. *Advisor Perspectives*. (2013). Available online at: https://www.advisorperspectives.com/newsletters13/pdfs/Woody_Brocks_Challenge_to_Krugman_and_the_Keynesians.pdf
52. Hüsler A, Sornette D, Hommes CH. Super-exponential bubbles in lab experiments: evidence for anchoring over-optimistic expectations on price. *J Econ Behav Organ*. (2013) **92**:304–16. doi: 10.1016/j.jebo.2013.06.005
53. Leiss M, Nax HH, Sornette D. Super-exponential growth expectations and the global financial crisis. *J Econ Dyn Control*. (2015) **55**:1–13. doi: 10.1016/j.jedc.2015.03.005
54. Brunnermeier MK. Bubbles. In: Durlauf SN, Blume LE, editors. *The New Palgrave Dictionary of Economics*. 2nd ed. London: Palgrave Macmillan. (2008). doi: 10.1057/978-1-349-95121-5_44-2
55. Temin P, Voth HJ. Riding the South Sea bubble. *Am Econ Rev*. (2004) **94**:1654–68. doi: 10.1257/0002828043052268
56. Levine SS, Apfelbaum EP, Bernard M, Bartelt VL, Zajac EJ, Stark D. Ethnic diversity deflates price bubbles. *Proc Natl Acad Sci USA*. (2014) **111**:18524–9. doi: 10.1073/pnas.1407301111
57. Asness CS. My top 10 peeves. *Fin Anal J*. (2014) **70**:1–9. doi: 10.2469/faj.v70.n1.2
58. Santos M, Woodford M. Rational asset pricing bubbles. *Econometrica*. (1997) **65**:19–57. doi: 10.2307/2171812
59. Siegel JJ. What is an asset price bubble? An operational definition. *Eur Fin Manage*. (2003) **9**:11–24. doi: 10.1111/1468-036X.00206
60. Kirman AP, Teyssière G. Bubbles and long-range dependence in asset prices volatilities. In: Hommes CH, Ramer R, Withagen CA, editors. *Equilibrium, Markets, and Dynamics: Essays in Honour of Claus Wedepohl*. Hoboken, NJ: Wiley. (2002). doi: 10.1007/978-3-642-56131-3_22
61. Roubini N. Why central banks should burst bubbles. *Int Fin*. (2006) **9**:87–107. doi: 10.1111/j.1468-2362.2006.00032.x
62. Johansen A, Sornette D. Shocks, Crashes and bubbles in financial markets. *Brussels Econ Rev*. (2010) **53**:201–53.
63. Shiller RJ. Speculative asset prices. *Am Econ Rev*. (2014) **104**:1486–517. doi: 10.1257/aer.104.6.1486
64. McCarty N, Poole KT, Rosenthal H. *Political Bubbles: Financial Crises and the Failure of American Democracy*. Princeton, NJ: Princeton University Press (2013). doi: 10.1515/9781400846399
65. Brock HW. Resolving the market efficiency paradox. *SED: Strategic Economic Decisions*. (2014). Available online at: http://www.sedinc.com/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/user_upload/reports/March_2014/March_2014_SED_PROFILE.pdf&t=1592843788&hash=b35ee8273298892f2132c4862380abd7
66. Samuelson PA. Proof that properly anticipated prices fluctuate randomly. *Indus Manage Rev*. (1965) **6**:41–9.
67. Samuelson PA. Proof that properly discounted present values of assets vibrate randomly. *Bell J Econ Manage Sci*. (1973) **4**:369–74. doi: 10.2307/3003046
68. LeRoy SF. Efficient capital markets and martingales. *J Econ Literat*. (1989) **27**:1583–621.
69. Jensen MC. Some anomalous evidence regarding market efficiency. *J Finan Econ*. (1978) **6**:95–101. doi: 10.1016/0304-405X(78)90025-9
70. Barany E, Beccar Varela MP. Long correlations applied to the study of memory effects in high frequency (tick) data, the dow jones index, and international indices. In: Viens FG, Mariani MC, Florescu I, editors. *Handbook of Modeling High-Frequency Data in Finance*. Hoboken, NJ: Wiley. (2012). doi: 10.1002/9781118204580.ch6
71. Maloney MT, Mulherin JH. The complexity of price discovery in an efficient market: the stock market reaction to the challenger crash. *J Corpor Finan*. (2003) **9**:453–79. doi: 10.1016/S0929-1199(02)00055-X
72. Roll R. A simple implicit measure of the effective bid-ask spread in an efficient market. *J Finan*. (1984) **39**:1127–39. doi: 10.1111/j.1540-6261.1984.tb03897.x
73. D. Sornette and A. Arneodo. Chaos, pseudo-random number generators and the random walk problem. *J Phys*. (1984) 45:1843. doi: 10.1051/jphys:0198400450120184300

74. Rudnick J, Gaspari G. *Elements of the Random Walk: An introduction for Advanced Students and Researchers*. Cambridge, UK: Cambridge University Press (2004). doi: 10.1017/CBO9780511610912
75. Filimonov VA, Sornette D. Quantifying reflexivity in financial markets: towards a prediction of flash crashes. *Phys Rev E*. (2012) **85**:056108. doi: 10.1103/PhysRevE.85.056108
76. Filimonov V, Bicchetti D, Maystre N, Sornette D. Quantification of the high level of endogeneity and of structural regime shifts in commodity prices. *J Int Money Finan*. (2014) **42**:174–92. doi: 10.1016/j.jimonfin.2013.08.010
77. Wheatley S, Wehrli A, Sornette D. The endo-exo problem in high frequency financial price fluctuations and rejecting criticality. *Quant Finan*. (2019) **19**:1165–78. doi: 10.1080/14697688.2018.1550266
78. Abreu D, Brunnermeier MK. Bubbles and crashes. *Econometrica*. (2003) **71**:173–204. doi: 10.1111/1468-0262.00393
79. Jolliffe IT. *Principal Component Analysis*. New York, NY: Springer Series in Statistics. 2nd ed. Springer (2002).
80. Newman MEJ. *Networks: An Introduction*. Oxford, UK: Oxford University Press (2010).
81. Füredi Z, Komlós J. The eigenvalue of random symmetric matrices. *Combinatorica*. (1981) **1**:233–41.
82. Malevergne Y, Sornette D. Collective origin of the coexistence of apparent random matrix theory noise and of factors in large sample correlation matrices. *Phys A*. (2004) **331**:660–8. doi: 10.1016/j.physa.2003.09.004
83. Billingsley P. *Probability and Measure*. 3rd edition. New York, NY: Wiley-Interscience (1995).
84. Beran J, Feng Y, Ghosh S, Kulik R. *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Berlin: Springer (2013).
85. Sohn H, Sornette D. *Rational Belief Bubbles*. Swiss Finance Institute Research Paper No. (2020) 20-05. Available online at: https://ssrn.com/abstract=3530895.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Sohn and Sornette. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 21 August 2020
doi: 10.3389/fphy.2020.00319

Confucianism and Firm Performance

Lisong Wang[1,2], *Yongjie Zhang*[1], *LiNa Wang*[3]* *and Jingkai Fu*[4]

[1] College of Management and Economics, Tianjin University, Tianjin, China, [2] School of Foreign Language and Literature,Tianjin University, Tianjin, China, [3] Department of Finance, School of Economics, Tianjin University of Finance and Economics Pearl River College, Tianjin, China, [4] Peddie School, Hightstown, NJ, United States

This paper examines the impact of Confucianism on firm performance, taking Chinese listed companies from 2000 to 2018 as the research object. The results show that (1) Confucianism provides legitimacy for a company's profit-seeking behavior and therefore helps to improve firm performance; and (2) Confucianism can effectively improve the efficiency of supervision mechanisms but weaken the marginal contribution of incentive mechanisms to financial performance. This paper provides empirical evidence for the influence of Confucianism on firm performance, broadening the understanding of the role of informal institutions in company finance and enriching the theory of "culture and finance."

Keywords: confucianism, firm performance, corporate governance, culture, supervision mechanisms

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Mei Xue,
Boston College, United States
Khalil Jebran,
Dongbei University of Finance and Economics, China

***Correspondence:**
LiNa Wang
wanglina1008@163.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *13 April 2020*
Accepted: *10 July 2020*
Published: *21 August 2020*

Citation:
Wang L, Zhang Y, Wang L and Fu J (2020) Confucianism and Firm Performance. Front. Phys. 8:319. doi: 10.3389/fphy.2020.00319

INTRODUCTION

In recent years, some studies have shown that in addition to economic factors affecting firm performance, non-economic factors such as culture also play an important role in firm performance [1, 2]. For China, a transition economy with imperfect institutions and still less efficient law enforcement, informal systems such as culture may play a more important role in firm performance. For the Chinese society, Confucianism is the most far-reaching informal system; it shapes the spirit of Chinese enterprises and may provide important spiritual support in the process of China's modernization. Confucianism is now once again faced with the possibility of rejuvenation in China even after having gone through a century of obstruction, which will absolutely have an effect on all aspects of China's economy [3]. Therefore, by examining the impact of Confucianism on firm performance, this paper attempts to discover how the informal system affects microenterprises and thus boosts China's economy.

It is not common to see empirical research on these issues because there are a series of difficulties and controversies that arise in measuring the impact of culture on economic behavior. Currently, the general methods used to measure culture's effect mainly include the Hofstede index [4], which focuses on the influence of cultures from different countries on individual behavior. This index excludes "benevolence," a main concept of Confucianism, so it is difficult to use the index to objectively reflect the influence of Confucianism on management behavior. In recent years, many scholars have started to use historical data to explain the impact of culture and institutions on financial performance [5, 6]. Confucianism, a mainstream ideology of ancient China, has accumulated for a long time and has gradually become the criterion the public uses to think about and judge individual behavior. Therefore, it is basically feasible to use historical data to measure Confucianism. After a comprehensive consideration of the above factors, we chose the local records of the Qing Dynasty as the basic data source and found proxy variables that indirectly measure Confucianism. Then, the

financial data from A-share listed companies from 2000 to 2012 were selected as the sample for the research.

First, this paper studies the impact of Confucianism on firm performance, showing that the greater the impact of Confucianism on a company, the higher its profitability. This statement implies that Confucianism not only legalizes a company's profit-seeking behavior but also contributes to its capital appreciation and wealth creation. Second, the supervision and incentive measures adopted by a company for the purpose of improving profitability are all affected by Confucianism. The empirical results show that Confucianism is conducive to improving the efficiency of supervision while weakening the marginal contribution of the incentive mechanisms to corporate performance.

This paper has two main research contributions. First, it attempts to combine empirical scientific methods with Confucianism and establish a research framework, striving to find a feasible path for empirical research on Confucianism. Second, this paper studies the impact of Confucianism on corporate performance by considering the combination of the reality of Confucianism and corporate governance. The research conclusions are of great theoretical significance not only for academic circles seeking to rediscover Confucianism but also for listed companies in the Chinese cultural context seeking to choose appropriate governance mechanisms and government departments to improve their supervision system.

This paper is arranged as follows: the second part is a literature review and provides the research hypothesis; the third part describes the measurement of variables and sample selection; the fourth part provides the empirical research results; the fifth part presents the robustness test; and the last part provides the conclusion.

LITERATURE REVIEW AND RESEARCH HYPOTHESIS

Literature Review

As shown in the literature on corporate finance, culture is defined as the concept and related values of stabilizing intergeneration inheritance through ethnic, religion, and social groups [7]. Byod and Richerson [8] defined culture as "the knowledge, values and other factors that are passed down from generation to generation through education and imitation, and can influence behavior," and claimed that culture has the following functions: (1) culture provides language-based information processing methods; and (2) the ethics derived from a culture help to reduce the cost of contract implementation and can effectively reduce the externality caused by the free rider problem. Guiso et al. [6] claimed that the intergenerational culture would affect the level of trust an individual has with other partners, thus affecting that individual's behavior and financial performance. As a long-term and stable informal institutional arrangement in a specific region, culture may affect the quality of information disclosure [9–11], tax behavior [12], business decision-making [6, 13], and investment behavior [14, 15]. Similarly, firm performance is also affected by external cultural factors.

Franke et al. [16] conducted an empirical study on cultural and financial performance in 18 countries from 1965 to 1987. The results show that power distance is positively correlated with firm performance, while individualism is negatively correlated with firm performance. Gorodnichenko and Roland [17] used per capita GDP to measure financial performance. The empirical results show that individualism is positively correlated with corporate performance. The above empirical results imply that it is difficult to use the Hofstadter index. In other words, the regression results depend on indicators to measure corporate performance, which makes the results unstable. Doney et al. [18] studied the Hofstadter index and trust level and found that the individualism orientation means that the trust process involves rational calculations, and collectivism means that trust is more predictable. Therefore, from the perspective of trust and firm performance, both individualistic orientation and collectivistic orientation may produce trust and affect corporate performance. Furthermore, the empirical results of Fidrmuc and Marcus [9] show that there is a significant positive correlation between the individualism orientation and corporate performance.

In Chinese history, Confucianism was put forward by Confucius in the Spring and Autumn period and then inherited and developed by Mencius and others in the Warring States period. After Emperor Wu of the Han Dynasty "ousted hundreds of schools and respected Confucianism alone," the combination of Confucianism and feudal monarchy established a set of social ethics and moral norms and became a tool for governors. For a long time, Confucianism has been deeply rooted in Chinese society and become the common people's inadvertently held ethics. Although the relationship between culture and corporate performance has been discussed, there are few studies on Confucianism and corporate performance.

Research Hypothesis

Traditional Chinese culture, for which Confucianism is the core, has a long history of thousands of years, and it has gradually become an indispensable component of the Chinese humanistic environment. Companies operating in this environment may be affected by Confucianism. Confucianism does not reject wealth and even has positive significance for building wealth. Therefore, the Confucian concept of wealth and the method used for wealth building may affect the company's management behavior and decision-making, so they have a positive significance for improving firm performance. Therefore, the following research hypotheses are proposed:

Hypothesis 1: With all other conditions unchanged, the greater the influence of Confucianism, the higher the profitability of the company.

Prior studies show that formal and informal institutions both impact corporate financial behavior [14]. Jebran et al. [19] examined the monitoring role of institutional ownership in the relationship between social trust and corporate financial behavior. They found that institutional ownership, as a formal governance system, weakens the influence of social trust (informal institution) on corporate financial behavior.

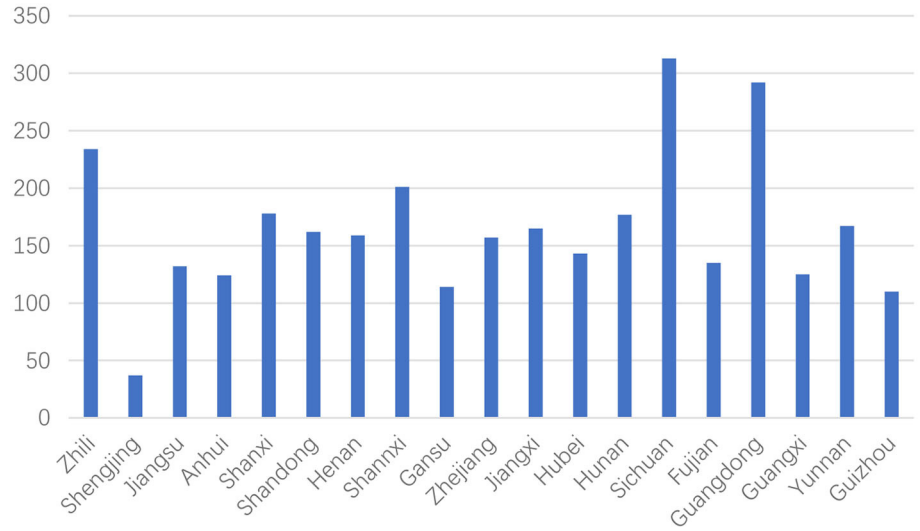


FIGURE 1 | Number of Confucian schools in the provinces of the Qing Dynasty. This figure plots the number of Confucian schools in the provinces of the Qing Dynasty. The horizontal axis represents the provinces, and the vertical axis represents the number of Confucian schools.

In addition, [20] showed that foreign qualified institutional investors can effectively monitor firms.

We conjecture that the influence of Confucianism on firm performance will be weaker when a firm's internal governance mechanism is stronger. We argue that institutional ownership, an internal governance system, can moderate the relationship between Confucianism and firm performance. We assume that greater institutional ownership indicates the use of effective internal governance mechanism and thus informal institutions, such as Confucianism, are less likely to influence corporate decisions. Thus, the following research hypotheses are proposed:

Hypothesis 2: With all other conditions unchanged, the negative association between Confucianism and firm performance is attenuated by institutional ownership.

VARIABLE MEASUREMENT AND SAMPLE SELECTION

Independent Variable Measurement

As we all know, the most important feature of Confucianism is that it seeks to improve human behavior through ethical education. Confucian education in the Ming and Qing Dynasties gradually shifted to secularization and popularization. The academy was an important place for Confucian education in the Ming Dynasty. During the Qing Dynasty, official schools and academies laid the foundation for the popularization and spread of Confucian education. In addition to the above historical factors, the choice of attending a Confucian school as a proxy variable was made in light of the following considerations. First, North [21] generalized theory of the system into theory of property rights, theory of the state, and theory of ideology and believed that the role of education was to "instill a set of values repeatedly." Weber [22] claimed that the level of Confucian education was closely related to the quality of education and pointed out that the difference in the distribution of educational institutions represents the strength of Confucian influence.

As mentioned above, Confucian schools are an important source of Confucianism and have gradually developed a common understanding with officials and intellectuals over a long period of time. Therefore, it is impossible to directly measure the influence of Confucianism; however, by using school choice as a proxy variable it is possible to measure the influence of Confucianism. Kwok and Tadesse [23] summarized institutional theory as ownership theory, state theory and ideology theory and believed that the social function of the education system is connected with the inculcation of values. In other words, consistent with the basic theory of institutional economies, we choose Confucian schools as the proxy variable of Confucian influence. Specifically, we choose the local chronicles of the Qing Dynasty as the data source and record the number of official schools (prefectural, state, and county) and academies within the jurisdiction of the provinces, counties, and counties according to the provincial administrative regions of the Qing Dynasty and compare them with those in the jurisdiction of the People's Republic from 1796 to 1840 AD. The statistical results are shown in **Figure 1**. During this timeframe, 3,284 Confucian educational institutions operated within the scope of the Qing government.

When measuring the influence of culture on a region, usually, two models are used. One is a regional model, and the other is a distance model. Currently, the widely used model is the distance model [24–27]. The greatest advantage of the distance model is that it can use company-level data to determine whether the explanatory variables represent the research object itself, eliminating the influence of other factors. Wines and Napier [28] claimed that the distance model has more advantages than the regional model when considering the

company or individual level. John et al. [29] used this model to study investment decisions and dividend policies. After a comprehensive consideration of the above factors, we chose the distance model to measure the proxy variable.

The main calculation steps are as follows: First, the longitude and latitude of the listed company are determined by using through the Google Map icon; second, the longitude and latitude of ancient schools (including official schools and colleges) are determined. Finally, we calculate the distance between every listed company and every ancient school according to their respective longitudes and latitudes, following Equations (1–3) shown below.

(i) We define the longitude and latitude of ancient schools (listed companies) as ω_S and φ_S (ω_F and φ_F), respectively. The central angle (θ) is calculated via the following Equation (1):

$$\cos\theta = \sin\varphi_S \times \sin\varphi_F + \cos\varphi_S \times \cos\varphi_F \times (\omega_S - \omega_F) \quad (1)$$

(ii) We calculate the arc length per radian using the following Equation (2):

$$\mathrm{Rad} = \frac{40075.04}{360} \times \frac{180}{\pi} \quad (2)$$

(iii) Note that the distance between two points equals the length of the *minor arc* across the surface of the earth. Therefore, we calculate the distance between the location of every ancient school and listed company using Equation (3), which is a well-known equation used for GIS, as shown below:

$$\mathrm{Distance} = \mathrm{rad} \times \left(\frac{\pi}{2} - \arctan\left(\frac{\cos\theta}{\sqrt{1-\cos^2\theta}}\right)\right) \quad (3)$$

Finally, 100, 200, and 300 km are utilized as the distance criteria or upper limits to calculate the number of ancient schools and then define the variables School_100, School_200, and School_300, respectively.

Control Variables

In this paper, the explained variate is firm performance, which is measured by ROA (return on assets). ROA is calculated by dividing EBIT by average total assets and reflects the economic surplus contributed by the company to stakeholders such as shareholders, creditors, and the government.

Based on the studies of Chen et al. [30] and Jebran et al. [19], we add the following control variables to the regression model: board size, board independence, management shareholding, leverage, firm size, liquidity, fixed assets, CEO duality, GDP, and the number of universities. GDP per capita is used to control for the economic development level of the place (province) where the company is registered and is calculated as the logarithm of GDP per capita (lnGDP). Second, the higher education level of Puritans has a larger direct impact on financial performance than ideology [31]. This paper controls for the influence of local higher education by using the number of 211 project universities in the company's registered place (province). Finally, the influence

TABLE 1 | Definitions and calculations of the variables.

Variable name	Variable definition	Description
ROA	Return on assets of the company	Earnings before interest and taxes/average total assets
Tobin Q	The ratio of market value to replacement value	Market value/total assets
INST	Holding proportion of institutional shareholder	Data from CSMAR database
Board_size	Broad size	Logarithm of the number of directors
Board_ind	Broad independence	The ratio of independent directors to all directors
Man_share	Holding proportion of management	Data from CSMAR database
Lev	Debt to asset ratio	Data from CSMAR database
Firm size	Firm size, measured by natural log of total assets	Data from CSMAR database
Liquidity	The average daily return to volume in the past month	Data from CSMAR database
Fixed assets	Fixed assets that are held for the production of goods or services and have a useful life of more than one fiscal year	Data from CSMAR database
Duality	An indicator variable that is equal to 1 if the CEO and the chairman of the board are the same person, and 0 otherwise	Data from CSMAR database
School_100	Number of schools within 100 km from the registered address of the company	Measure according to local chronicle
School_200	Number of schools within 200 km from the registered address of the company	
School_300	Number of schools within 300 km from the registered address of the company	
lnGDP	The logarithm of the per capita GDP in the registered province	Statistics yearbook (Website of National Bureau of Statistics) http://www.stats.gov.cn/
University	Numbers of Project 211 universities in the registered province	Website of Ministry of Education http://www.moe.gov.cn/

This table provides the definitions and calculations used for the variables. The first column is the variable name, the second column is the variable definition, and the third column is the variable description.

TABLE 2 | Descriptive statistics.

Variable name	Sample observation	Mean	Standard deviation	Minimum	Median	Maximum
School_100	20,121	0.122	0.049	0	0.132	0.223
School_200	20,121	0.229	0.103	0	0.242	0.505
School_300	20,121	0.224	0.095	0	0.243	0.501
ROA	20,121	0.053	0.052	−0.363	0.027	0.831
Tobin Q	20,121	3.231	2.478	0.435	3.125	23.742
INST	20,121	0.352	0.167	0.013	0.345	0.865
Board_size	20,121	2.008	0.196	1.055	2.132	2.677
Board_ind	20,121	0.324	0.103	0	0.232	0.600
Man_share	20,121	0.053	0.145	0	0	0.687
Lev	20,121	0.423	0.187	0	0.422	0.978
Firm size	20,121	21.234	25.673	19.231	21.537	24.435
Liquidity	20,121	2.219	7.234	0.122	1.343	65.443
Fixed assets	20,121	15.324	17.342	13.326	15.568	19.763
Duality	20,121	0.145	0.327	0	0	1
lnGDP	20,121	10.122	0.758	0.767	9.967	10.345
University	20,121	5.656	5.764	5.456	3.879	21.994

This table shows the descriptive statistics of the variables. The first column is the variable name, and the second column is the sample observation.

of industry and time also need to be considered [13, 14]; thus, industry and year control variates are added to the regression. Financial data related to corporate performance and governance are all extracted from the CSMAR database. The data sources and calculation methods are shown in **Table 1**.

Sample Selection

The sample selected for this empirical study is companies that issued shares and were listed on the Shanghai or Shenzhen Stock Exchanges from 2000 to 2018. After downloading information from the CSMAR database, 22,111 primary samples were obtained, among which were 342 financial industry samples in total, 676 samples with operating incomes less than or equal to 0 or insolvent (leverage >100%), and 972 samples with missing data. After eliminating the above 1990 samples, 20,121 valid samples were obtained, accounting for more than 91% of the primary samples. See **Table 2** for the descriptive statistics of the research variates.

Model

To test our hypothesis, we consider that firm performance is a function of Confucianism and other control variables. The regression equation is given as follows:

$$ROA_{it} = \beta_0 + \beta_1 * School_num_{it} + \sum_{q=2}^{m} \beta_q * \left(qth\ ControlVariable_{it}\right) + Industry + Year + \varepsilon_{it} \quad (4)$$

where School_num represents the number of schools within 100 kilometers, 200 kilometers, and 300 kilometers of the registered address of firm i at time t. We use control variables, mainly including the shareholding ratio of the institutional shareholders (INST), board size (Board_size), board independence (Board_ind), management shareholding ratio (Man_share), debt-to-asset ratio (Lev), firm size (Firm size), liquidity (Liquidity), fixed assets (Fixed assets), CEO duality (Duality), GDP (lnGDP), and the number of universities (University). ε is the regression residual.

To test Hypothesis 2, the following regression equation is used in this paper:

$$ROA_{it} = \beta_0 + \beta_1 * School_num_{it} * INST_{it} + \sum_{q=2}^{m} \beta_q^* \left(qth\ ControlVariable_{it}\right) + Industry + Year + \varepsilon_{it} \quad (5)$$

where $INST_{it}$ denotes institutional ownership; the interaction term, $School_num_{it} * INST_{it}$ captures the moderating effect of institutional ownership; and $ControlVariable_{it}$ denotes a set of control variables.

THE EMPIRICAL RESEARCH RESULTS

Table 3 reports the regression results. The explanatory variable in the first column is School_100, which is the number of schools within 100 kilometers of the registered address of the firm. The regression results show that the coefficient of this variable is 0.0232, and it is significant at the 1% confidence level. The explanatory variables in the second and third columns are School_200 and School_300, and the regression coefficients are 0.0132 and 0.0123, respectively; these are significant at the 1% confidence level. The above regression results indicate that as Confucian influence increases, firm performance is enhanced. The empirical results support Hypothesis 1.

Table 4 reports the regression results. In the first column, the regression coefficient of the explanatory variable INST is

TABLE 3 | The impact of Confucianism on firm performance.

School_100	0.0232***		
	(3.98)		
School_200		0.0132***	
		(3.43)	
School_300			0.0123***
			(3.32)
INST	0.0354***	0.0381***	0.0351***
	(13.15)	(13.09)	(13.05)
Board_size	0.0135***	0.0135***	0.0135***
	(7.23)	(7.22)	(7.22)
Board_ind	0.0131**	0.0146**	0.0162**
	(1.97)	(2.06)	(2.05)
Man_share	0.0306***	0.0312***	0.0318***
	(9.67)	(10.02)	(10.03)
Lev	−0.0682***	−0.0665***	−0.0698***
	(−27.52)	(−27.03)	(−27.35)
Firm size	0.2313***	0.2412***	0.2431***
	(4.23)	(4.33)	(4.54)
Liquidity	0.0081	0.0103	0.0114
	(0.33)	(0.41)	(0.42)
Fixed assets	0.1543	0.1632	0.1521
	(1.34)	(1.43)	(1.25)
Duality	0.0213	0.0204	0.0254
	(0.87)	(0.86)	(0.91)
lnGDP	0.0056***	0.0055***	0.0052***
	(5.28)	(5.53)	(5.25)
University	−0.0003***	−0.0003***	−0.0003***
	(−4.43)	(−4.34)	(−4.76)
Constant	−0.0010	−0.0040	−0.0011
	(−0.10)	(−0.38)	(−0.11)
Industry	Control	Control	Control
Year	Control	Control	Control
R^2 (%)	11.89	11.98	11.87
Sample capacity	20,121	20,121	20,121
F statistic	81.97***	82.94***	82.96***

This table shows the impact of Confucianism on firm performance. The dependent variable is ROA, and the main independent variable is School_num (including School_100, School_200, and School_300). The variable definitions are presented in ***Table 1****. We control for industry effects and year effects. The figures in parentheses are t-statistics. **, *** indicate significance at the 5, and 1% confidence levels, respectively.*

TABLE 4 | Moderating effect of institutional ownership.

School_100*INST	0.0489***		
	(2.62)		
School_200*INST		0.0257**	
		(2.54)	
School_300*INST			0.1182***
			(3.50)
INST	0.0316***	0.0317***	0.0299***
	(8.59)	(8.62)	(8.32)
Board_size	0.0156***	0.0135***	0.0153***
	(7.24)	(7.12)	(7.21)
Board_ind	0.0153*	0.0155**	0.0151**
	(1.92)	(1.98)	(1.99)
Man_share	0.0312***	0.0321***	0.0320***
	(10.09)	(10.11)	(10.12)
Lev	−0.0675***	−0.0675***	−0.0677***
	(−27.02)	(−27.04)	(−27.09)
Firm size	0.2254***	0.2389***	0.2129***
	(4.26)	(4.45)	(4.09)
Liquidity	0.0102	0.0113	0.0093
	(0.45)	(0.48)	(0.34)
Fixed assets	0.1432	0.1427	0.1321
	(1.21)	(1.19)	(1.02)
Duality	0.0243	0.0231	0.0245
	(0.86)	(0.81)	(0.88)
lnGDP	0.0052***	0.0056***	0.0052***
	(5.46)	(5.76)	(5.34)
University	−0.0003***	−0.0003***	−0.0004***
	(−4.42)	(−4.31)	(−4.46)
Constant	−0.0011	−0.0028	−0.0011
	(−0.10)	(−0.29)	(−0.12)
Industry	Control	Control	Control
Year	Control	Control	Control
R^2 (%)	11.87	11.76	11.98
Sample capacity	20,121	20,121	20,121
F statistic	82.55***	82.54***	82.65***

This table reports the moderating effect of institutional ownership on the relation between Confucianism and firm performance. The variable definitions are presented in ***Table 1****. We control for industry effects and year effects. The figures in parentheses are t-statistics. *, **, *** indicate significance at the 10, 5, and 1% confidence levels, respectively.*

0.0316 ($t = 8.59$), and the regression coefficient of the cross-multiplying term of School_100 and INST is 0.0489 ($t = 2.62$). The above results imply that as the shareholding ratio of the institutional shareholder increases, firm performance increases, which means that the supervision of the institutional shareholder can effectively improve firm performance. The significant positive correlation between the cross-multiplying term of School_100 and INST and the explanatory variable shows that as Confucian influence increases, institutional shareholder supervision's marginal contribution to financial performance is enhanced. In conclusion, the regression results provided in **Table 4** show that the supervision measures taken by the firm's institutional shareholder to improve financial performance are justified according to Confucianism, and Confucianism can effectively improve the efficiency of supervision.

ROBUSTNESS TEST

Alternative Proxy of Confucianism

To ensure the robustness of the regression results, we conducted the following robustness test; that is, we selected the number of Confucian schools in the provincial administrative areas to replace the corresponding explanatory variables mentioned above and performed another regression. In addition, referring to the studies of Du [32], Chen et al. [30], Jebran et al. [19], and Chen et al. [30], we measured Confucianism by the distance between a firm's registered address and 7 Confucian centers in China. We calculated the geographical-proximity-based Confucianism, $CONF_N$

$$CONF_N = \frac{MaxDIS_N - DIS_N}{MaxDIS_N - MinDIS_N}$$

TABLE 5 | Robustness test: independent variables School_pro and CONF.

School_pro	0.0362*** (6.35)			
School_pro*INST		0.0819*** (6.12)		
$CONF_N$			0.1534** (2.34)	
$CONF_N^*$INST				0.3212** (2.56)
INST	0.0369*** (13.25)	0.0272*** (8.21)	0.0432*** (14.43)	0.0532*** (15.23)
Board_size	0.0153*** (7.19)	0.0159*** (7.50)	0.0231*** (5.21)	0.0224*** (4.51)
Board_ind	0.0135* (1.78)	0.0142* (1.87)	0.0231** (2.31)	0.0257*** (2.58)
Man_share	0.0303*** (9.55)	0.0309*** (9.66)	0.0231*** (4.87)	0.0203*** (4.43)
Lev	−0.0673*** (−26.87)	−0.0656*** (−26.32)	−0.0532*** (−17.73)	−0.0521*** (−16.33)
Firm size	0.2413*** (4.23)	0.2132*** (3.81)	0.2768*** (5.42)	0.2543*** (5.05)
Liquidity	0 (0.12)	0.0023 (0.14)	0.0124 (0.54)	0.0321 (1.32)
Fixed assets	0.1922* (1.68)	0.2065* (1.73)	0.2013* (1.73)	0.1987* (1.66)
Duality	0.0312 (0.95)	0.0324 (1.01)	0.0214 (0.75)	0.0321 (0.87)
lnGDP	0.0072*** (7.29)	0.0073*** (7.30)	0.0123*** (6.62)	0.0121*** (6.53)
University	−0.0003*** (−3.24)	−0.0005*** (−3.25)	−0.0032*** (−4.65)	−0.0023*** (−3.32)
Constant	−0.0231** (−2.31)	−0.0157* (−1.65)	−0.0356*** (−3.32)	−0.0457*** (−4.25)
Industry	Control	Control	Control	Control
Year	Control	Control	Control	Control
R^2 (%)	12.09	12.13	9.09	12.12
Sample capacity	20,121	20,121	20,121	20,121
F statistic	83.79***	81.15***	74.21***	81.59***

This table shows the results of the robustness test. The dependent variable is ROA, and the main independent variables are School_pro and CONF. The variable definitions are presented in ***Table 1****. We control for industry effects and year effects. The figures in parentheses are t-statistics. *, **, *** indicate significance at the 10, 5, and 1% confidence levels, respectively.*

TABLE 6 | Robustness test: dependent variable Tobin Q.

School_100	2.3213*** (3.32)		
School_200		2.7625*** (3.54)	
School_300			2.3342*** (3.26)
INST	0.5634*** (9.21)	0.5472*** (8.09)	0.4281*** (6.05)
Board_size	0.1321*** (4.24)	0.1353*** (4.65)	0.1224*** (4.17)
Board_ind	0.6534 (1.55)	0.4623 (1.16)	0.4322 (1.12)
Man_share	0.1282*** (4.03)	0.1321*** (4.21)	0.1223*** (3.86)
Lev	−0.2721*** (−17.24)	−0.2323*** (−15.03)	−0.2431*** (−16.32)
Firm size	0.4323*** (3.24)	0.4873*** (3.54)	0.4313*** (3.27)
Liquidity	0.0112 (0.32)	0.0134 (0.35)	0.0105 (0.32)
Fixed assets	0.2341 (1.20)	0.2512 (1.53)	0.2132 (1.12)
Duality	0.1321 (0.37)	0.1043 (0.26)	0.1546 (0.51)
lnGDP	0.0323*** (4.12)	0.0344*** (4.53)	0.0422*** (5.21)
University	−0.0011*** (−3.21)	−0.0013*** (−3.34)	−0.0010*** (−3.06)
Constant	−0.0234 (−0.12)	−0.0543 (−0.28)	−0.0312 (−0.21)
Industry	Control	Control	Control
Year	Control	Control	Control
R^2 (%)	14.23	14.28	13.67
Sample capacity	20,121	20,121	20,121
F statistic	88.21***	86.44***	85.45***

This table shows the results of robustness test. The dependent variable is Tobin Q, and the main independent variable is the School_num (including School_100, School_200, and School_300). Variable definitions are presented in ***Table 1****. We also control the industry effect and the year effect. Figures in parentheses are t-statistics. *** are significant at 1% confidence level.*

where N denotes 7 Confucianism centers; DIS_N denotes the distance between the firm and Confucianism center N; $MaxDIS_N$ and $MinDIS_N$ represent the maximum and minimum values, respectively, of DIS_N for all firms by year. The regression results are shown in **Table 5**.

The explanatory variable in columns (1) and (2) of **Table 5** is the total number of schools in the provincial administrative areas. By observing the regression results, it is not difficult to find that the number of schools in the provincial administrative areas School_pro is significantly positively correlated with the dependent variable and passes the significance test at the 1% confidence level. Similarly, the cross-product terms of School_pro and supervision mechanism are significantly positively correlated with the explained variables, and both pass the significance test the 1% confidence level. In addition, we measure Confucianism by the distance between the registered address of the company and the seven Confucian centers in China. The regression results are shown in columns (3) and (4) of **Table 5** and are similar to the previous regression results. $CONF_N$ is significantly positively correlated with the dependent variable. Similarly, the cross-product terms of $CONF_N$ and INST are significantly positively correlated with the dependent variable.

Alternative Proxy of Firm Performance

To ensure the robustness of the model, we use Tobin Q as a substitute of ROA and repeat the above regression to measure the influence of Confucianism on firm performance. The regression results are shown in **Table 6**.

The regression results showed that the coefficients of the independent variables (school_100, shool_200, and school_300)

are all significant at the 1% confidence level, indicating that firm performance increases with the increasing influence of Confucianism; this result is similar to the results shown in **Table 3**.

We also carried out additional robustness tests by (1) replacing the explained variable ROA with the industry-adjusted ROA, that is, considering the influence of industry factors; and (2) considering the impact of the financial crisis; that is, the timeframe of the study was divided into two periods, 2000–2007 and 2008–2012. The results of these tests are very similar to the previous regression results. Due to space limitations, we do not report the results of these tests.

CONCLUSION

This paper examines the impact of Confucianism on firm performance, taking Chinese listed companies from 2000 to 2018 as the research object. The following conclusions are obtained. (1) Confucianism provides legitimacy for the company's profit-seeking behavior and therefore helps to improve firm performance. (2) Confucianism can effectively improve the efficiency of supervision mechanisms but weakens the marginal contribution of incentive mechanisms to financial performance. This paper provides empirical evidence on the influence of Confucianism on firm performance, broadening the understanding of the role of informal institutions in company finance and enriching theory on "culture and finance."

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was partly supported by Philosophy and Social Science Planning Project of Tianjin (Grant No. TJYY19XSX-017) and the National Natural Science Foundation of China (71771170).

REFERENCES

1. Licht AN, Goldschmidt C, Schwartz SH. Culture rules: the foundations of the rule of law and other norms of governance. *J Compar Econ.* (2007) **35**:659–88. doi: 10.1016/j.jce.2007.09.001
2. Li K, Griffin D, Yue H, Zhao L. How does culture influence corporate risk-taking. *J Corporate Finance.* (2013) **23**:1–22. doi: 10.1016/j.jcorpfin.2013.07.008
3. Hofstede G. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations.* London: SAGE Publications Inc. (2001).
4. Hofstede G. *Culture's Consequences: International Differences in Work-Related Values*. Los Angeles, CA: SAGE Publications Inc. (1980).
5. La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny R. Law and finance. *J Polit Econ.* (1998) **106**:1113–55. doi: 10.1086/250042
6. Guiso L, Zingale L, Sapicnz P. Social capital as good culture. *J Eur Econ Assoc.* (2008) **6**:295–320. doi: 10.1162/JEEA.2008.6.2-3.295
7. Hope OK. Firm-level disclosures and the relative roles of culture and legal origin. *J Int Financial Manage Account.* (2003) **14**:218–48. doi: 10.1111/1467-646X.00097
8. Boyd R, Richerson PJ. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press (1985).
9. Fidrmuc JP, Marcus J. Culture, agency costs, and dividends. *J Compar Econ.* (2010) **38**:321–39. doi: 10.1016/j.jce.2010.04.002
10. Kung JK, Ma C. Can cultural norms reduce conflicts? Confucianism and peasant rebellions in Qing China. *J Dev Econ.* (2014) **111**:132–49. doi: 10.1016/j.jdeveco.2014.08.006
11. Braguinsky S, Mityakov S. Foreign corporations and the culture of transparency: evidence from Russian administrative data. *J Financial Econ.* (2015) **117**:139–64. doi: 10.1016/j.jfineco.2013.02.016
12. DeBacker J, Heim BT, Tran A. Importing corruption culture from overseas: evidence from corporate tax evasion in the United States. *J Financial Econ.* (2015) **117**:122–38. doi: 10.1016/j.jfineco.2012.11.009
13. Hilary G, Hui KW. Does religion matter in corporate decision making in America? *J Financial Econ.* (2009) **93**:455–73. doi: 10.1016/j.jfineco.2008.10.001
14. Du X. Does confucianism reduce board gender diversity? Firm-level evidence from China. *J Business Ethics.* (2016) **136**:399–436. doi: 10.1007/s10551-014-2508-x
15. Cardoso FM, Meloni S, Santanche A, Moreno Y. Topical alignment in online social systems. *Front Phys.* (2019) **7**:58. doi: 10.3389/fphy.2019.00058
16. Franke RH, Hofstede G, Bond MH. Cultural roots of economic performance: a research note. *Strat Manage J.* (1991) **12**:165–73. doi: 10.1002/smj.4250120912
17. Gorodnichenko Y, Roland G. *Culture, Institutions· and the Wealth of Nations.* NBER Working Paper, No. 16368 (2010).
18. Doney PM, Cannon JP, Mullen MR. Understanding of the influence of national culture on the development of trust. *Acad Manage Rev.* (1998) **23**:601–20. doi: 10.5465/amr.1998.926629
19. Jebran K, Chen S, Ye Y, Wang C. Confucianism and stock price crash risk: evidence from China. *North Am J Econ Finance.* (2019) **50**:100995. doi: 10.1016/j.najef.2019.100995
20. Cao C, Xia C, Chan KC. Social trust and stock price crash risk: evidence from China. *Int Rev Econ Finance.* (2016) **46**:148–65. doi: 10.1016/j.iref.2016.09.003
21. North DC. *Structure and Change in Economic History*. Beijing: The Commercial Press (2002).
22. Weber M. *The Religion of China*. Wilmington, IL: The Free Press (1951).
23. Kwok CCY, Tadesse S. National culture and financial systems. *J Int Business Stud.* (2006) **37**:227–47. doi: 10.1057/palgrave.jibs.8400188
24. Grinblatt M, Keloharju M. How distance, language, and culture influence stockholdings and trades. *J Finance.* (2001) **56**:1053–73. doi: 10.1111/0022-1082.00355
25. Almog A, Bird R, Garlaschelli D. Enhanced gravity model of trade: reconciling macroeconomic and network models. *Front Phys*. (2019) **7**:55. doi: 10.3389/fphy.2019.00055
26. Kashyap R. The perfect marriage and much more: combining dimension reduction, distance measures and covariance. *Phys A Stat Mech Appl.* (2019) **536**:120938. doi: 10.1016/j.physa.2019.04.174
27. Jiang ZQ, Zhou WX. Multifractal detrending moving average cross-correlation analysis. *Phys Rev E.* (2011) **84**:016106. doi: 10.1103/PhysRevE.84.016106
28. Wines WA, Napier NK. Toward an understanding of cross-cultural ethics: a tentative model. *J Business Ethics.* (1992) **11**:831–41. doi: 10.1007/BF00872361

29. John K, Knyazeva A, Knyazeva D. Does geography matter? Firm location and corporate payout policy. *J Financial Econ.* (2011) **101**:533–51. doi: 10.1016/j.jfineco.2011.03.014
30. Chen S, Ye Y, Jebran K. Does confucianism reduce corporate over-investment? Evidence from China. *Asia Pac J Financial Stud.* (2019) **48**:210–35. doi: 10.1111/ajfs.12254
31. Becker SO, Woessmann BL. Was weber wrong? a human capital theory of protestant economic history. *Q J Econ.* (2009) 124:531–96.
32. Du X. Does confucianism reduce minority shareholder expropriation? Evidence from China. *J Business Ethics.* (2015) **132**:661–716. doi: 10.1007/s10551-014-2325-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wang, Zhang, Wang and Fu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

MINI REVIEW
published: 25 August 2020
doi: 10.3389/fphy.2020.00323

A Perspective on Correlation-Based Financial Networks and Entropy Measures

Vishwas Kukreti[1], *Hirdesh K. Pharasi*[2*], *Priya Gupta*[3] *and Sunil Kumar*[4*]

[1] School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India, [2] Instituto de Ciencias Fisicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico, [3] Atal Bihari Vajpayee School of Management & Entrepreneurship, Jawaharlal Nehru University, New Delhi, India, [4] Department of Physics, Ramjas College, University of Delhi, New Delhi, India

In this mini-review, we critically examine the recent work done on correlation-based networks in financial systems. The structure of empirical correlation matrices constructed from the financial market data changes as the individual stock prices fluctuate with time, showing interesting evolutionary patterns, especially during critical events such as market crashes, bubbles, etc. We show that the study of *correlation-based networks* and their evolution with time is useful for extracting important information of the underlying market dynamics. Also, we present our perspective on the use of recently-developed entropy measures, such as *structural entropy* and *eigen-entropy*, for continuous monitoring of correlation-based networks.

Keywords: econophysics, random matrix theory, correlation, networks, minimum spanning trees, clustering, financial networks

OPEN ACCESS

Edited by:
Takayuki Mizuno,
National Institute of Informatics, Japan

Reviewed by:
Michele Bellingeri,
University of Parma, Italy
Yuyao Wang,
Nanjing University of Science and Technology, China

***Correspondence:**
Sunil Kumar
du.sunil@gmail.com;
skumar@ramjas.du.ac.in
Hirdesh K. Pharasi
hirdeshpharasi@gmail.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *20 April 2020*
Accepted: *13 July 2020*
Published: *25 August 2020*

Citation:
Kukreti V, Pharasi HK, Gupta P and Kumar S (2020) A Perspective on Correlation-Based Financial Networks and Entropy Measures. Front. Phys. 8:323. doi: 10.3389/fphy.2020.00323

1. INTRODUCTION

Network science [1–4] has emerged as an important tool for studying different complex phenomena– spread of infectious diseases [5, 6], economic production [7], construction of robust sustainable infrastructure and technological networks [8], processing human information [9], innovation diffusion [10], detection of financial crashes [11–13], etc. In this mini-review, we focus on the role of network science in understanding complex financial markets. Our aims are two-fold: (i) To uncover the structure of the complex interactions among stocks at a particular period of time (static picture) through *correlation-based networks*, where the *nodes* represent the stocks in the financial market, and the *links* represent the interaction strengths of co-movements of stocks (as measured by correlations). For this purpose, one starts with computing the cross-correlations among stock price returns and then constructs any of the correlation-based networks– Minimum Spanning Tree (MST) [14, 15], Threshold Network [16], Planar Maximally Filtered Graph (PMFG) [17], etc. Using these networks, one can identify stocks (or sectors) that are in the "core" or "periphery" [18], as well as study their hierarchy/importance of the different stocks driving the market fluctuations. The correlations among stocks change with time, and the underlying dynamics of the market produces very intriguing and correlation structures. Temporal networks are those networks in which links are time dependent [19] and are useful for studying systems in which connections change or evolve with time. Correlation-based networks in the stock market are therefore temporal networks, because their links (constructed from correlation values) change or evolve with time. The understanding of the stock market dynamics can be very important for practical applications like portfolio optimization, risk management, etc. (ii) To continuously

monitor (dynamic picture) the health and fragility of the financial market. The market index, which is a weighted arithmetic mean of the prices of selected stocks in the market, reflects the performance of the market and assists agents in comparing the current price levels relative to past prices. The daily index return is the difference of the logarithmic values of the index (at market closure) over a period of 1 day. Thus, for the purpose of continuous monitoring of the financial market, we study the temporal evolution of the market index returns along with two *entropy measures*, structural entropy [20] and eigen-entropy [21]. This becomes very useful and necessary for measuring the systemic risk, market regulation and predicting downturns or crashes [22], since there often exist sizable fluctuations during crashes and bubbles.

2. CORRELATION-BASED NETWORKS

Mantegna first studied the hierarchical structures of correlation-based networks in financial markets [14, 15]. Later, similar studies of correlation-based networks were made (see, e.g., [23–25]). These correlation-based networks provided easy visual representations of multivariate time series and extracted meaningful information about the complex market dynamics. The analysis of evolution of correlation-based networks provides an understanding of the underlying market trends, especially during periods of crisis [16]. For the construction of a correlation-based network to represent N stocks in a financial market in a time-epoch ending on date τ, one begins with the correlation matrix, $\boldsymbol{C}(\tau)$, and uses a transformation to construct a distance matrix, $\boldsymbol{D}(\tau) = \sqrt{2(1 - \boldsymbol{C}(\tau))}$ (for mathematical details, see **Supplementary Material**).

2.1. Minimum Spanning Tree

MST is constructed by using the distances d_{ij}'s, which represent the interaction strengths (correlations) between pairs of stocks $i, j = 1, \ldots, N$ in a market for a specific time window, such that all N nodes (stocks) are connected with exactly $N - 1$ edges under the constraint that total distance is minimum [25–27]. Algorithms due to Kruskal or Prim are generally utilized to obtain MST from a distance matrix. For a non-degenerate distance matrix, the MST is uniquely determined. Two of the main advantages of MST are that: (i) it produces a network structure without putting any arbitrary threshold, and (ii) it has property of inherent hierarchical clustering. There have been many papers with applications of MST in equity markets [16, 18], currency exchange rates [28], global foreign exchange dynamics [29]. MST is useful for studying the taxonomy or the sector classification [30], with potential applications in portfolio optimization. Researchers have also carried out analysis of dynamical correlations using MST [24]. Among disadvantages, there is the fact that the order and classification of nodes in a cluster of MST is not robust, and often sensitive to minor changes in correlations or spurious correlations. Therefore, for improvement of results, either noise suppression techniques like Random Matrix Theory (RMT) [31] and power mapping [13] have been used, or alternative algorithms such as PMFG, Triangulated Maximally Filtered Graph, Average Linkage Minimum Spanning Tree, Directed Bubble Hierarchical Tree [17, 32–35] have been proposed. Instead of using pair-wise Pearson correlations, partial correlations and mutual information have also been explored in some studies [36, 37].

2.2. Threshold Networks

In this approach, an adjacency matrix is constructed by applying a threshold value in the correlation (C_{ij}) or distance (d_{ij}) of the network. It filters out the strongest correlations (or shortest distances) by putting a certain value of threshold and discard the remaining correlations/distances. A high threshold value in the distance gives rise to a completely connected graph (one extreme), while decreasing value of threshold makes the connections less and less, until one gets a null network (at the other extreme). Thus, one can tune the threshold in order to get the desired strength of correlations. For a particular value of threshold, as correlation matrices change with time, the threshold networks also change (see **Supplementary Material**). One drawback of the threshold networks is that we do not get a spanning graph, and therefore, there is a "loss of information"; when we put a threshold value we discard some nodes and edges. Also, threshold networks are found to be very sensitive to the noise (random fluctuations).

2.3. Planar Maximally Filtered Graph

PMFG is a network drawn in a plane, such that there are no intersecting links [17, 38]. If N is total number of stocks, then it contains $3(N - 2)$ links. The PMFG has the advantage that it retains the structure of MST (which contains $N - 1$ links) and provides additional information about the connections [17, 32]. However, PMFG has a disadvantage that there exists a certain arbitrariness in its results, as there is an embedding of data from higher dimension to lower dimension with a zero genus [39]. Recently, PMFG and threshold network have been combined to produce PMFG-based threshold networks [40]. Threshold networks of the financial market are constructed over multi-scale and at multi-threshold [41].

3. ENTROPY MEASURES

As in other domains, entropy has also been used to understand the financial hazards as well as to construct an early warning indicator for predicting systematic risks [42, 43]. Maasoumi and Racine examined the predictability of the market returns using entropy measure and found that it is capable to detect the non-linear dependence within the time series of market returns as well as between returns and other prediction variables obtained from other models [44]. Recently, Ricci curvature and entropy have been used to construct an economic indicator for market fragility and systemic risk [45]. Very recently, Almog *et al.* presented a perspective on the use of entropy measures such as *structural entropy* [20], which is computed from the communities in correlation-based networks. Chakraborti et al. computed the *eigen-entropy* from the eigen-vector centrality of the stocks in the correlation-based network [21]. Below, we discuss the structural entropy [20] and eigen-entropy [21], and compare the two measures.

3.1. Structural Entropy

The concept of structural entropy has resolved the problem of choosing different period of crisis and extracting substantial information from the large network of stock market. The structural entropy measures the amount of heterogeneity of the network nodes with an assumption that more connected nodes share common attributes than others. The authors assume the nature of clusters to be independent sub-units of the network. The process of calculating the structural entropy involves two steps: (i) Calculation of an optimal partition function which places every node in a certain cluster using a community detection algorithm. (ii) Analysing the partition function and extracting the representative value of the diversity level (for mathematical details and schematic diagram, see **Supplementary Material**). The formula for Shannon's entropy is applied: $S(\vec{P}) \equiv -\sum_{i=1}^{M} P_i \ln(P_i)$, in terms of probability vector $\vec{P} \equiv \left[\frac{c_1}{N}, \frac{c_2}{N}, \ldots, \frac{c_M}{N}\right]$, where M is the number of communities and c_i is the size of community i (proportional size of the community in the network).

Structural entropy S of the network provides a way to continuously monitor the state of the network. However, it is sensitive to the choice of community detection algorithm employed in detecting communities. This arbitrariness makes the calculation of entropy dependent on the choice of the user and hence is not unique.

3.2. Eigen-Entropy

Very recently, the concept of eigen-entropy was used in studying financial markets [21]. It is computed from eigen-centrality of the network obtained from the short time series correlation matrices [21, 46]. In order to capture the global feature of the network, every node is ranked by its eigen-centrality (for mathematical details and schematic diagram, see **Supplementary Material**). The similarity of the eigen-centralities (ranks) of the stocks is uniquely measured by eigen-entropy, defined as $H = -\sum_{i=1}^{N} p_i \ln p_i$, where p_i is the eigen-centrality of the i-th node (stock). Higher the similarity of the stock centralities, higher the eigen-entropy.

Empirical correlation matrix of the market may be decomposed in multiple ways. In many papers, it was decomposed into three separated modes, market mode C_M, the group mode C_G and the random mode C_R. However, it is difficult (and somewhat arbitrary) to choose the range of eigenvalues corresponding to the group mode C_G and the random mode C_R, as the boundary is not often distinct. Another way to decompose is to consider the market mode C_M (corresponding to the maximum eigenvalue) and the group-random modes C_{GR} (rest of the eigenvalues), hence without any arbitrariness. C_M&C_{GR} is the preferable decomposition and corresponding eigen-entropy H_M and H_{GR} could be calculated as $\mathbf{A_M} = |C_M|^2$ (matrix element-wise) and $\mathbf{A_{GR}} = |C_{GR}|^2$ (matrix element-wise), respectively. The eigen-entropy computed using above method gives a simple yet robust measure to quantify the randomness of the financial market without using any arbitrary thresholds. Further, Chakraborti et al. [21] used the variables $H - H_M$ and $H - H_{GR}$ to construct a phase space, where the market epochs show phase separation and order-disorder transitions. These results are certainly of deep significance for the understanding of financial market behavior and designing strategies for risk management.

4. EMPIRICAL ANALYSES AND RESULTS

We have analyzed stock prices of the S&P500 USA market for the period of 1985-2016 (for details of data and methodology, see **Supplementary Material**), and made some plots of correlation-based networks as well as entropy measures, as presented below as well as in the **Supplementary Material**. In order to illustrate the usage and concepts of correlation-based networks and entropy measures, we have compared three correlation frames chosen arbitrarily from crash, bubble and normal periods of the market. It may be mentioned that during a market crash there is a sharp fall in the index return and all the stocks start behaving similarly; the whole market begins to act like a single huge cluster or community. During a bubble period, a particular sector gets overpriced or over-performs, causing accentuation of disparities among the various sectors or communities. In both the crash and bubble periods, there are sizable fluctuations (as mentioned earlier in the introduction) and consequently market volatility (see **Supplementary Material** for definition) is higher than the normal period. In the normal or business-as-usual period, there are several distinct sectors performing well, but the market volatility is low.

Figure 1 shows the analysis for three time-epoch of 40 days ending at: (first column) 23/07/1985, (second column) 08/01/2007, and (third column) 17/06/2010. **Figures 1A–C** show the heat-map of correlation matrices at three different periods. It shows the amount of correlation between $N = 194$ stocks of S&P 500 at different time periods: (**Figure 1A**) normal period (23/07/1985), when market behaves normally with low mean correlation between the stocks, (**Figure 1B**) bubble period (08/01/2007), when market experienced an upward drift in price in some sectors only and (**Figure 1C**) crash period (17/06/2010), when the market experienced huge recession. The corresponding MST's are shown in **Figures 1D–F**, which have been generated using the Prim's algorithm. Different colors in MST's correspond to different sectors in the market. The different market structures reflected in the correlation matrix are also visible in the correlation based Threshold Networks **Figures 1G–I** with threshold ($d_{ij} \leq 1$) and PMFG's **Figures 1J–L**.

During the normal phase (**Figures 1A,D,G,J**) the market interactions are well-distributed across the stocks and the mean market correlations are not very high and the volatility is low (see **Supplementary Material**). During the bubble period (**Figures 1B,E,H,K**) certain sectors of stocks are more correlated with each other than the rest of stocks in the market. As visible in **Figure 1H**, few of the stocks are bunched together. This property is pronounced during times when a particular sector experiences a surge, e.g., during the dot-com bubble period, where the IT sector saw a boost but not the entire market. During the crashes (for the list, see **Supplementary Material**), the entire S&P market react in a similar way, which made the stocks in the market extremely correlated with each other (**Figures 1C,F,I,L**).

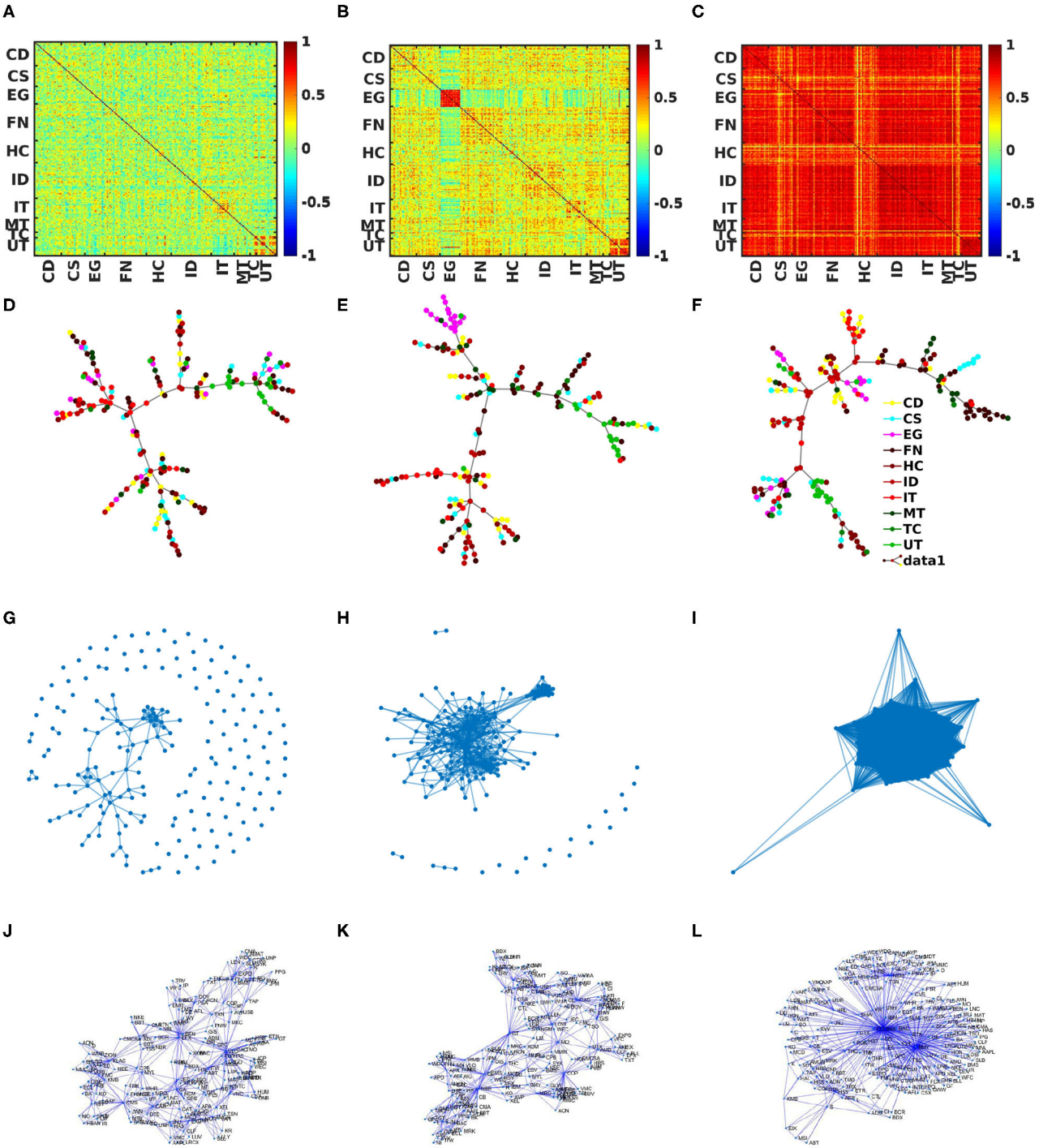


FIGURE 1 | Static correlation-based networks: Analysis of S&P 500 market with 194 stocks (epoch of 40 days) for three different periods: first, second, and third columns are corresponding to 23/07/1985 (normal period), 08/01/2007 (bubble period), and 17/06/2010 (crash period), respectively. **(A–C)** are heat maps of correlation matrices of different periods. Minimum Spanning Trees are shown in **(D–F)**. From **(G–I)**, Threshold Networks at a particular value of threshold. Planar Maximally Filtered Graphs **(J–L)** for three different periods.

Figure 2 shows how the entropy measures may be used for continuous monitoring of the financial markets. **Figures 2A–C** show the evolution of S&P 500 market over a period of 1985 – 2016 for index returns $r(\tau)$, eigen-entropies $H(\tau)$, and structural entropy $S(\tau)$, respectively. Three vertical dashed line are corresponding to epochs ending at 23/07/1985, 08/01/2007,

FIGURE 2 | Continuous monitoring of S&P 500 market with 194 stocks and for a rolling time-epoch of 40 days and shift of 20 days over a period of 1985 – 2016: The logarithmic returns of S&P 500 index is shown in **(A)**. **(B)** Shows the temporal evolution of a new measurement "eigen-entropy" $H(\tau)$, calculated from eigen-vector centralities of correlation matrices. Evolution of structural entropy $S(\tau)$ calculated by using community detection algorithm is shown in **(C)**. The dashed vertical lines are corresponding to different periods (normal, bubble, and crash) whose static results are shown in **Figure 1**.

and 17/06/2010. We find that the Pearson correlation among the two measures $S(\tau)$ and $H(\tau)$ is -0.22, which indicates that the two measures are anti-correlated. The two entropy measures actually capture different aspects of the financial market.

The structural entropy is based on the idea of "structural diversity" in a network, and it was proposed to utilize the number of communities in a system and their corresponding sizes. In a way, the structural entropy tries to capture the amount of heterogeneity of the nodes in the network, with the assumption that nodes which share common attributes belong to the same community [39, 47]. The structural entropy reaches maximum ($\ln N$), when the community structure is heterogeneous– there are N communities of equal size (unity), i.e., each node is assigned to a different community; it reaches minimum (zero), when all the N nodes are assigned to a single community. During a market crash, the market is extremely correlated and all stocks behave in a similar way as if belonging to a single community. Hence, the structural entropy decreases significantly (see also figure in **Supplementary Material**).

The eigen-entropy measures how similar the eigen-centrality ranks of the stocks are. The eigen-entropy reaches its maximum value ($\ln N$), when all the centralities are of similar value, i.e., all the individual nodes have similar rank/importance, such that the variance of the eigen-centralities becomes low. From the return time series point of view, this occurs when all of stock prices are entirely uncorrelated such that the market is totally disordered (or random)—indicative of the lack of any group or sectoral structures, or when the market is extremely correlated such that all the stocks behave in a similar way. During a market crash, the market is extremely correlated and all stocks behave in a similar way and so the eigen-entropy increases significantly.

5. DISCUSSIONS AND CONCLUDING REMARKS

In this review, we have discussed different methods for analysis of static and dynamic correlation-based networks of financial markets, and also studied how entropy measures can be used to identify normal, bubble, and crash periods. Specifically, we have compared the recently developed concepts of structural entropy and eigen-entropy.

It is noteworthy that financial networks are naturally "weighted," as each link bears a numeric value representing the correlation between the nodes (stocks). In a recent paper [48], it has been shown how real weighted network with large link weights heterogeneity may lower robustness in case of nodes/links failure. It would be interesting to see how these methods could be used to increase the robustness in context of financial networks.

We have also seen that many of the correlation-based networks have shown clustering with communities of stocks. Thus, community detection in network science serves as an important technique for extraction of the clustering

information from empirical correlation matrix of a multivariate time series. Several community detection algorithms have been proposed [39, 47, 49]. The problem is that different community detection algorithms yield different results for the same empirical correlation matrix. So, often domain knowledge is required to determine what is a sensible or meaningful community.

Further, we have seen that many of the networks are sensitive to noise or spurious correlations. Properties of random matrices [50] have turned out to be useful in reducing noise and thus understanding dynamics of complex systems [51]. An ensemble of random matrices, also known as stationary or standard random (Gaussian) matrix ensemble [50], introduced by Wigner [52, 53], have been applied to many studies in physics, biology, finance, etc. (see [54] and references therein). The probability distribution of eigenvalues of Wishart orthogonal ensemble (WOE) follows Marčenko-Pastur distribution [55]. The empirical correlation matrix of a complex system is normally compared with WOE [24, 31, 56]. It has been observed from eigenvalues statistics of empirical correlation matrices that the few largest eigenvalues show deviations from the Wishart ensemble. Note that Pearson cross-correlation assumes that the time series are stationary, which are valid for shorter lengths of time series. However, if the number of time series are greater than the lengths of time series, then corresponding empirical correlation matrices are noisy and highly singular. For such short time series, there is a great need of noise suppression in correlation matrix to extract actual correlations. There are different techniques for suppressing the noise in correlation matrix [57–59]. Notably, any empirical correlation matrix of financial market can be decomposed into partial correlations, consisting of market C_M, group C_G and random C_R modes, respectively [60]. It enables us to identify the dominant stocks, sectors and inherent structures of the market. Recently, detailed analyses of the empirical correlation matrices using these approaches have been carried out to understand the complexity in dynamics of stock market [13, 51, 61]. It has been found that during the crisis, the eigenvalue spectrum behaves very differently from one corresponding to a normal period.

Finally, we must mention that the prediction of collapses of financial markets using traditional economic theories has been a disastrous failure. These new and alternate methods have the potential use of continuous monitoring and understanding of the complex structures and dynamics of financial markets. These are a few of the attempts physicists have made for generation of early warning signals for crisis, and these methods can be used for timely intervention.

AUTHOR CONTRIBUTIONS

SK and HP designed the idea, wrote the main manuscript text, and prepared the figures. VK and PG contributed to the literature review. All authors reviewed the manuscript.

ACKNOWLEDGMENTS

The authors are grateful to Anirban Chakraborti, Hrishidev, Suchetana Sadhukhan, Kiran Sharma, and Thomas H. Seligman for their critical inputs. HP is grateful for financial support provided by UNAM-DGAPA and CONACYT Proyecto Fronteras 952. This research was supported in part by the International Centre for Theoretical Sciences (ICTS) during the visits of VK, PG, and SK for participating in the Summer research program on Dynamics of Complex Systems (Code: ICTS/Prog-DCS2019/07). The topic editors are acknowledged for supporting this open access publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00323/full#supplementary-material

REFERENCES

1. Albert R, Barabási AL. Statistical mechanics of complex networks. *Rev Modern Phys*. (2002) **74**:47. doi: 10.1103/RevModPhys.74.47
2. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: structure and dynamics. *Phys Rep*. (2006) **424**:175–308. doi: 10.1016/j.physrep.2005.10.009
3. Newman M. Networks: An Introduction. Oxford: Oxford University Press (2010). doi: 10.1093/acprof:oso/9780199206650.003.0001
4. Barabási AL. *Network Science*. Cambridge: Cambridge University Press (2016).
5. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci USA*. (2009) **106**:21484–9. doi: 10.1073/pnas.0906910106
6. Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. *Science*. (2013) **342**:1337–42. doi: 10.1126/science.1245200
7. Acemoglu D, Azar PD. Endogenous production networks. *Econometrica*. (2020) **88**:33–82. doi: 10.3982/ECTA15899
8. Reis SD, Hu Y, Babino A, Andrade JS Jr, Canals S, Sigman M, et al. Avoiding catastrophic failure in correlated networks of networks. *Nat Phys*. (2014) **10**:762–7. doi: 10.1038/nphys3081
9. Lynn CW, Papadopoulos L, Kahn AE, Bassett DS. Human information processing in complex networks. *Nat Phys*. (2020) 1–9. doi: 10.1038/s41567-020-0985-7
10. Acemoglu D, Akcigit U, Kerr WR. Innovation network. *Proc Natl Acad Sci USA*. (2016) **113**:11483–8. doi: 10.1073/pnas.1613559113
11. Münnix MC, Shimada T, Schäfer R, Leyvraz F, Seligman TH, Guhr T, et al. Identifying states of a financial market. *Sci Rep*. (2012) **2**:644. doi: 10.1038/srep00644
12. Battiston S, Puliga M, Kaushik R, Tasca P, Caldarelli G. Debtrank: too central to fail? Financial networks, the fed and systemic risk. *Sci Rep*. (2012) **2**:541. doi: 10.1038/srep00541
13. Pharasi HK, Sharma K, Chatterjee R, Chakraborti A, Leyvraz F, Seligman TH. Identifying long-term precursors of financial market crashes using correlation patterns. *N J Phys*. (2018) **20**:103041. doi: 10.1088/1367-2630/aae7e0
14. Mantegna RN. Information and hierarchical structure in financial markets. *Comput Phys Commun*. (1999) 121–2:153–6. doi: 10.1016/S0010-4655(99)00302-1
15. Mantegna RN. Hierarchical structure in financial markets. *Eur Phys J B Condens Matter Complex Syst*. (1999) **11**:193–7. doi: 10.1007/s10051005092 9
16. Kumar S, Deo N. Correlation and network analysis of global financial indices. *Phys Rev E*. (2012) **86**:026101. doi: 10.1103/PhysRevE.86.026101

17. Tumminello M, Aste T, Di Matteo T, Mantegna RN. A tool for filtering information in complex systems. *Proc Natl Acad Sci USA*. (2005) **102**:10421–6. doi: 10.1073/pnas.0500298102
18. Sharma K, Gopalakrishnan B, Chakrabarti AS, Chakraborti A. Financial fluctuations anchored to economic fundamentals: a mesoscopic network approach. *Sci Rep*. (2017) **7**:8055. doi: 10.1038/s41598-017-07758-9
19. Holme P, Saramäki J. Temporal networks. *Phys Rep*. (2012) **519**:97–125. doi: 10.1016/j.physrep.2012.03.001
20. Almog A, Shmueli E. Structural entropy: monitoring correlation-based networks over time with application to financial markets. *Sci Rep*. (2019) **9**:1–13. doi: 10.1038/s41598-019-47210-8
21. Chakraborti A, Hrishidev, Sharma K, Pharasi HK. Phase separation and scaling in correlation structures of financial markets. *arXiv [Preprint]*. arXiv:1910.06242v3 (2020). Available online at: https://arxiv.org/abs/1910.06242
22. Battiston S, Farmer JD, Flache A, Garlaschelli D, Haldane AG, Heesterbeek H, et al. Complexity theory and financial regulation. *Science*. (2016) **351**:818–9. doi: 10.1126/science.aad0299
23. Onnela JP, Chakraborti A, Kaski K, Kertiész J. Dynamic asset trees and portfolio analysis. *Eur Phys J B Condens Matter Complex Syst*. (2002) **30**:285–8. doi: 10.1140/epjb/e2002-00380-9
24. Onnela JP, Chakraborti A, Kaski K, Kertesz J, Kanto A. Dynamics of market correlations: taxonomy and portfolio analysis. *Phys Rev E*. (2003) **68**:056110. doi: 10.1103/PhysRevE.68.056110
25. Bonanno G, Caldarelli G, Lillo F, Mantegna RN. Topology of correlation-based minimal spanning trees in real and model markets. *Phys Rev E*. (2003) **68**:046130. doi: 10.1103/PhysRevE.68.046130
26. Bonanno G, Caldarelli G, Lillo F, Micciché S, Vandewalle N, Mantegna RN. Networks of equities in financial markets. *Eur Phys J B*. (2004) **38**:363-371. doi: 10.1140/epjb/e2004-00129-6
27. Tumminello M, Lillo F, Mantegna RN. Correlation, hierarchies, and networks in financial markets. *J Econ Behav Organ*. (2010) **75**:40–58. doi: 10.1016/j.jebo.2010.01.004
28. Jang W, Lee J, Chang W. Currency crises and the evolution of foreign exchange market: Evidence from minimum spanning tree. *Phys A Stat Mech Appl*. (2011) **390**:707–18. doi: 10.1016/j.physa.2010.10.028
29. McDonald M, Suleman O, Williams S, Howison S, Johnson NF. Detecting a currency's dominance or dependence using foreign exchange network trees. *Phys Rev E*. (2005) **72**:046106. doi: 10.1103/PhysRevE.72.046106
30. Onnela JP, Kaski K, Kertész J. Clustering and information in correlation based financial networks. *Eur Phys J B*. (2004) **38**:353–62. doi: 10.1140/epjb/e2004-00128-7
31. Plerou V, Gopikrishnan P, Rosenow B, Amaral LN, Stanley HE. A random matrix theory approach to financial cross-correlations. *Phys A Stat Mech Appl*. (2000) **287**:374–82. doi: 10.1016/S0378-4371(00)00376-9
32. Aste T, Di Matteo T, Hyde S. Complex networks on hyperbolic surfaces. *Phys A Stat Mech Appl*. (2005) **346**:20–6. doi: 10.1016/j.physa.2004.08.045
33. Song WM, Di Matteo T, Aste T. Nested hierarchies in planar graphs. *Discrete App Math*. (2011) **159**:2135–46. doi: 10.1016/j.dam.2011.07.018
34. Song WM, Di Matteo T, Aste T. Hierarchical information clustering by means of topologically embedded graphs. *PLoS ONE*. (2012) **7**:e31929. doi: 10.1371/journal.pone.0031929
35. Tumminello M, Di Matteo T, Aste T, Mantegna RN. Correlation based networks of equity returns sampled at different time horizons. *Eur Phys J B*. (2007) 55:209–17. doi: 10.1140/epjb/e2006-00414-4
36. Kenett DY, Tumminello M, Madi A, Gur-Gershgoren G, Mantegna RN, Ben-Jacob E. Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PLoS ONE*. (2010) **5**:e15032. doi: 10.1371/journal.pone.0015032
37. Kenett DY, Shapira Y, Madi A, Bransburg-Zabary S, Gur-Gershgoren G, Ben-Jacob E. Dynamics of stock market correlations. *AUCO Czech Econ Rev*. (2010) **4**:330–41.
38. Nie CX, Song FT. Relationship between entropy and dimension of financial correlation-based network. *Entropy*. (2018). 20:177. doi: 10.3390/e20030177
39. MacMahon M, Garlaschelli D. Community detection for correlation matrices. *Phys Rev X*. (2015) **5**:021006. doi: 10.1103/PhysRevX.5.021006
40. Nie CX. Dynamics of cluster structure in financial correlation matrix. *Chaos Solitons Fractals*. (2017) doi: 10.1016/j.chaos.2017.05.039
41. Sui G, Li H, Feng S, Liu X, Jiang M. Correlations of stock price fluctuations under multi-scale and multi-threshold scenarios. *Phys A Stat Mech Appl*. (2018) **490**:1501–12. doi: 10.1016/j.physa.2017.08.141
42. Bowden RJ. Directional entropy and tail uncertainty, with applications to financial hazard. *Quant Fin*. (2011) **11**:437–46. doi: 10.1080/14697681003685548
43. Gradojevic N, Caric M. Predicting systemic risk with entropic indicators. *J Forecast*. (2017) **36**:16–25. doi: 10.1002/for.2411
44. Maasoumi E, Racine J. Entropy and predictability of stock market returns. *J Econometr*. (2002) **107**:291–312. doi: 10.1016/S0304-4076(01)00125-7
45. Sandhu RS, Georgiou TT, Tannenbaum AR. Ricci curvature: an economic indicator for market fragility and systemic risk. *Sci Adv*. (2016) **2**:1–10. doi: 10.1126/sciadv.1501495
46. Fan Y, Zeng LL, Shen H, Qin J, Li F, Hu D. Lifespan development of the human brain revealed by large-scale network eigen-entropy. *Entropy*. (2017) **19**:471. doi: 10.3390/e19090471
47. Fortunato S. Community detection in graphs. *Phys Rep*. (2010) **486**:75–174. doi: 10.1016/j.physrep.2009.11.002
48. Bellingeri M, Bevacqua D, Scotognella F, Cassi D. The heterogeneity in link weights may decrease the robustness of real-world complex weighted networks. *Sci Rep*. (2019) **9**:1–13. doi: 10.1038/s41598-019-47119-2
49. Laloux L, Cizeau P, Bouchaud JP, Potters M. Noise dressing of financial correlation matrices. *Phys Rev Lett*. (1999) **83**:1467–70. doi: 10.1103/PhysRevLett.83.1467
50. Mehta ML. *Random Matrices*. Amsterdam: Academic Press (2004).
51. Pharasi HK, Sharma K, Chakraborti A, Seligman TH. Complex market dynamics in the light of random matrix theory. In: Abergel A, Chakrabarti BK, Chakraborti A, Deo N, Sharma K, editors *New Perspectives and Challenges in Econophysics and Sociophysics*. Cham: Springer (2019) p. 13–34. doi: 10.1007/978-3-030-11364-3_2
52. Wigner EP. On the distribution of the roots of certain symmetric matrices. *Ann Math*. (1958) **67**:325–27. doi: 10.2307/1970008
53. Wigner EP. Random matrices in physics. *SIAM Rev*. (1967) **9**:1–23. doi: 10.1137/1009001
54. Guhr T, Müller-Groeling A, Weidenmüller HA. Random-matrix theories in quantum physics: common concepts. *Phys Rep*. (1998) **299**:189–425. doi: 10.1016/S0370-1573(97)00088-4
55. Marčenko VA, Pastur LA. Distribution of eigenvalues for some sets of random matrices. *Math USSR Sbornik*. (1967) **1**:457. doi: 10.1070/SM1967v001n04ABEH001994
56. Kumar S, Deo N. Analysing correlations after the financial crisis of 2008 and multifractality in global financial time series. *Pramana*. (2015) **84**:317–25. doi: 10.1007/s12043-015-0935-6
57. Guhr T, Kälber B. A new method to estimate the noise in financial correlation matrices. *J Phys A Math Gen*. (2003) **36**:3009. doi: 10.1088/0305-4470/36/12/310
58. Vinayak, Schäfer R, Seligman TH. Emerging spectra of singular correlation matrices under small power-map deformations. *Phys Rev E*. (2013) **88**:032115. doi: 10.1103/PhysRevE.88.032115
59. Vinayak, Seligman TH; AIP. Time series, correlation matrices and random matrix models. *AIP Conf Proc*. (2014) **1575**:196–217. doi: 10.1063/1.4861704
60. Sharma K, Shah S, Chakrabarti AS, Chakraborti A. Sectoral co-movements in the Indian stock market: a mesoscopic network analysis. In: Aruka Y, Kirman A, editors *Economic Foundations for Social Complexity Science*. Singapore: Springer (2017) p. 211–38. doi: 10.1007/978-981-10-5705-2_11
61. Pharasi HK, Seligman E, Seligman TH. Market states: a new understanding. *arXiv [Preprint]*. arXiv:2003.07058 (2020). Available online at: https://arxiv.org/abs/2003.07058

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kukreti, Pharasi, Gupta and Kumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 17 September 2020
doi: 10.3389/fphy.2020.00388

Stock-Index Tracking Optimization Using Auto-Encoders

*Chi Zhang, Shuang Liang, Fei Lyu and Libing Fang**

School of Management and Engineering, Nanjing University, Nanjing, China

Deep learning algorithms' powerful capabilities for extracting useful latent information give them the potential to outperform traditional financial models in solving problems of the stock market which is a complex system. In this paper, we explore the use of advanced deep learning algorithms for stock-index tracking. We partially replicate the CSI 300 Index by optimizing with respect to the difference between the returns of the tracking portfolio and the target index. We extract the complex non-linear relationship between index constituents and select a subset of constituents to construct a dynamic tracking portfolio by six well-known auto-encoders (single-hidden-layer undercomplete, sparse, contractive, stacked, denoising, and variational auto-encoders) that have been widely used in contexts other than stock-index tracking. Empirical results show that the auto-encoder-based strategies perform better than conventional ones when the tracking portfolio is constructed with a small number of stocks. Furthermore, strategies based on auto-encoders capable of learning high-capacity encodings of the input, such as sparse and denoising auto-encoders, have even better tracking performance. Our findings offer evidence that deep learning algorithms with explicitly designed hierarchical architectures are suitable for index tracking problems.

Keywords: stock-index tracking, complex system, deep learning, auto-encoders, non-linear relationship

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Dehua Shen,
Tianjin University, China
Elie Bouri,
Holy Spirit University of Kaslik, Lebanon

****Correspondence:***
Libing Fang
lbfang@nju.edu.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *26 June 2020*
Accepted: *10 August 2020*
Published: *17 September 2020*

Citation:
Zhang C, Liang S, Lyu F and Fang L (2020) Stock-Index Tracking Optimization Using Auto-Encoders. Front. Phys. 8:388. doi: 10.3389/fphy.2020.00388

INTRODUCTION

The market index system has evolved with the development of the securities market. Financial products such as index funds, index futures, and index options emerge endlessly, indicating that indexing investment has won the favor of investors, especially institutional investors. Traditional investment based on the analysis of timing and stock fundamentals is an actively managed strategy, whereas indexing investment is passively managed. By constructing a portfolio to track a market index, investors expect to obtain the same return and volatility as the target index, with relatively lower risk and management cost, as well as better liquidity. The choice of how to construct a tracking portfolio (i.e., of an index tracking method) is crucial for the management of index funds, for hedging or arbitrage through index financial derivatives such as index futures, and for maximizing the performance of index investment generally. At the present time, the tracking methods utilized with stock index funds are fairly homogeneous but the tracking errors differ significantly. Therefore, there is great value in attempting to improve index tracking technology. In recent years, the rapid development of computer technologies and the discipline of quantitative finance especially make it possible to propose more effective index tracking methods.

The many index tracking strategies that have been put forward in theory and practice can be divided into full replication strategies and optimization strategies [1]. In the full replication method, all the constituent securities of the target index are purchased and allocated the same weights

that they have in the index. Although full replication is easy to manage and operate and is highly consistent with the target index, it has many unavoidable defects. Its large portfolio size brings high transaction costs and large tracking errors [2]; some of the constituent securities may not be traded due to liquidity problems; the adverse effects of individual securities cannot be avoided; etc. In the optimization method, the historical data of the components are analyzed and a suitable number of assets for inclusion in the tracking portfolio are selected with the help of advanced algorithms. Thus, fewer securities are required to achieve the purpose of indexing investment [3]. Compared with full replication, the optimization method can significantly reduce management costs and increase tracking efficiency, advantages which have made it the focus of much current academic research.

Among the most widely applied approaches for selecting a subset of constituent stocks are market-value ranking, weight ranking, liquidity ranking [4], correlation coefficient ranking, random sampling, stratified sampling [5], and genetic algorithms [6]. However, these established stock selection approaches fail to collect and utilize adequately historical information about constituent stocks, target indexes, and the correlations between them. Therefore, it is necessary to develop new techniques.

The goal of index tracking is to make the return of the tracking portfolio as close as possible to the return of the target index. There are two main indicators used to evaluate the performance of index tracking: the standard deviation of the difference between the return of the tracking portfolio and that of the benchmark index [7] and the square root of the second-order moment of the difference [8]. There are also other, less common metrics for measuring tracking errors, such as Mean Absolute Deviation (MAD), Maximum Absolute Deviation (Max), Mean Absolute Downside Deviation (MADD), and Downside Maximum Absolute Deviation (DMax) [8]. The objective function can be constructed by minimizing one of the tracking errors defined above; the weight allocations of the tracking portfolio can then be obtained. When the tracking error is defined as the square root of second-order moment of the return difference, minimizing it requires a quadric programming model, and therefore its optimal solution can be found by best linear unbiased estimation (BLUE) [9], a standard econometric method. We will use this model to construct a tracking portfolio.

Since Markowitz [10] first proposed the mean-variance model, the measurement of index tracking errors and optimal replication methods have generated an extensive literature. For example, Roll [11] studies partial replication of the index by optimizing with respect to the volatility of the tracking error based on Markowitz's mean-variance model. Ammann and Tobler [12] present four suitable decompositions of tracking error variance. Dunis and Ho [13] introduce the concept of co-integration into the problem of index tracking optimization and obtain good tracking performance. Chiam et al. [14] build a multi-objective evolutionary system that can simultaneously optimize tracking performance and transaction cost to track the index. Filippi et al. [15] focuses on the problem of index tracking with consideration of the expected excess return, using a bi-objective approach.

Machine learning algorithms have made dramatic progress over the past four decades, and applications for them have been found in various disciplines, including financial asset management. The tools of machine learning have notable advantages in solving asset management problems. Asset managers can use machine learning techniques to identify underlying assets by discovering new patterns in a complex system and immediately make investment decisions based these insights. Further, machine learning algorithms enable new forms of data, such as data in graphic and sound formats, to be used as input to models, helping investment managers better analyze the market trend. In addition, machine learning algorithms may also reduce the negative impact of human subjective biases on investment decisions. Consequently, a growing body of research takes advantage of machine learning algorithms to study asset management or index tracking. Focardi and Fabozzi [16] propose to use clustering for constructing index tracking portfolios. They cluster co-integrated stocks based on Euclidean distances between stock price series and select one stock from each cluster to include in the tracking portfolio. Yang et al. [17] study the index-tracking problem by applying a support-vector machine model. Their empirical results show the model performs robustly on tracking the Hang Seng Index (HSI). Jeurissen and Berg [18] use a hybrid genetic algorithm, where each chromosome represents a subset of the stocks, to address the problem of stock index tracking by partial replication. A backpropagation-based neural network has been built by Zorin and Borisov [19] to form full replication of the stock index (although the tracking performance is not as good as expected). Fernández and Gómez [20] propose a heuristic solution for the portfolio selection problem based on the Hopfield network, but their results demonstrate no superiority over other heuristic models. By analyzing data from the Brazilian stock market, Freitas et al. [21] find a neural network model that outperforms the Markowitz's mean-variance model in portfolio optimization. Chen et al. [22] propose a flexible neural tree ensemble model to predict the NASDAQ-100 and S&P CNX NIFTY stock indexes, achieving reliable forecast performance. Wu et al. [23] use the non-negative-lasso method to fit and predict the CSI 300 Index with short-selling constraints; the results indicate that non-negative lasso can achieve a small tracking error.

Recently, with the rapid development of deep-learning technology, methods based on artificial intelligence have enjoyed unprecedented popularity [24]. One approach involves applying deep learning algorithms to the problem of index replication since the stock market is a complex system. A portfolio construction approach based on deep learning is first proposed in academia by Heaton et al. [25]. Ouyang et al. [26] have subsequently expanded this framework by including a dynamic asset-weight calculation method and implemented this model to track the HSI. However, their optimized asset weights may become negative, contrary to traditional asset allocation implementations. In order to accomplish partial replication, both Heaton et al. [25] and Ouyang et al. [26] select stocks by measuring the Euclidean distance between the original returns and the reconstructed returns of the index components using auto-encoders, which are the core elements of their frameworks.

Kim and Kim [27] argue that such an asset selection criterion is artificial. They modify it by constructing an auto-encoder in

such a way that the deepest hidden layer has only one node (a proxy for the market index) and measuring the similarity of this latent representation to individual asset returns. We disagree with this approach. If an auto-encoder uses non-linear activation functions, then the deepest latent representations are non-linear combinations of the input original asset returns and capture some complex abstract features of the market index. Although these features can represent the market index, it is generally difficult to find their corresponding economic meanings. The candidate asset returns' similarities to these abstract features are not equal or even related to their similarities to the target index returns. A selection criterion based on this measure would therefore seem to be meaningless. Moreover, the extremely contractive structure of the auto-encoder with a single-node deepest latent layer may result in excessive loss of input information. None of the above three papers [25–27] suggests that the index tracking approach based on deep learning algorithms can outperform traditional index tracking techniques. Evidence is needed that deep learning is sufficiently advanced to handle index tracking problems. Moreover, various auto-encoders with more complex structures and better properties have been developed; it is reasonable to ask whether they can improve the performance of stock selection.

Based on the framework proposed by Heaton et al. [25], this paper investigates the applications of various auto-encoder deep-learning architectures in selecting representative stocks from the index constituents. The stocks are also selected by measuring the Euclidean distance between the original returns and the reconstructed ones. We then build dynamic tracking portfolios with the selected stocks to partially replicate the return of the index and evaluate their tracking performances. This article differs from Heaton et al. [25] and other related papers in several respects. First, we examine the effectiveness not only of the single-hidden-layer undercomplete auto-encoder but also of five other auto-encoders widely used in academe and industry, including the stacked auto-encoder and the denoising auto-encoder. Second, we propose a method for constructing dynamic tracking portfolios. The weights of the stocks in the tracking portfolio are calculated and adjusted periodically. This is more feasible and appropriate for practical indexing investment than what is done in other deep-learning methods. Third, we introduce two conventional stock selection strategies (weight ranking and market-value ranking) in addition to the strategies implemented by auto-encoders. The tracking performances of all these strategies in selecting various numbers of stocks are contrasted to confirm the advantages of applying auto-encoders.

The rest of the paper is organized as follows: section Methodology outlines the related algorithms and how they will be implemented. Section Empirical Analysis details our experimental setups for index tracking and presents the empirical results and discussion. Section Conclusions concludes the paper.

METHODOLOGY

Stock Selection Using Auto-Encoders

Auto-encoders are a special case of feedforward neural networks [28]. They are generally used for dimensionality reduction and feature extraction. Recently, they have also been employed as generative models to produce, for example, pictures. Unlike other feedforward neural networks, auto-encoders use unsupervised learning; their task is to copy the input to the output. An auto-encoder is composed of an encoder and a decoder. In **Figure 1**, $\boldsymbol{x}$ represents the input data; $f(\boldsymbol{x})$ represents the encoder, forming a hidden layer $\boldsymbol{h}$ that discovers some latent state representation of the input; and $g(\boldsymbol{h}) = g(f(\boldsymbol{x}))$ represents the decoder, which produces a reconstruction $\boldsymbol{x}'$. In general, the learning process of an auto-encoder can be described as minimizing the reconstruction error $\mathcal{L}(\boldsymbol{x}, g(f(\boldsymbol{x})))$, which is defined as the difference between $\boldsymbol{x}$ and $\boldsymbol{x}'$. The output of an auto-encoder is worthless if it is simply a copy of the input. Auto-encoders are prevented from replicating the input completely by imposing constraints on the hidden layers, such as limiting the number of hidden units and adding regularizers, so that latent attributes of the input data can be learned and described.

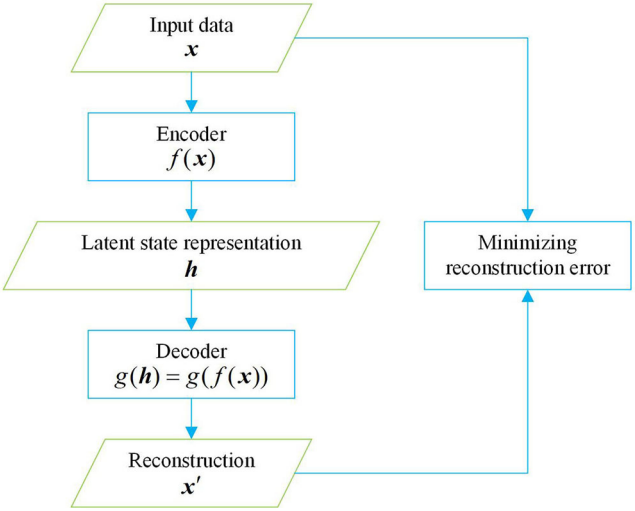


FIGURE 1 | General processing flow of an auto-encoder.

A common way to obtain useful features from an auto-encoder is to require the dimension of $\boldsymbol{h}$ to be smaller than $\boldsymbol{x}$. An auto-encoder with this bottleneck structure is called an undercomplete auto-encoder. Consider first a single-hidden-layer undercomplete auto-encoder that contains one hidden layer with five neurons, consistent with Heaton et al. [25]. Its architecture is shown in **Figure 2**. Given a training batch $D = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(m)}\}$ containing m samples, the input of a single-hidden-layer undercomplete auto-encoder is $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^\top \in \mathbb{R}^n$, a vector representing n index component stock returns on a certain trading day. Similarly, the output is $\boldsymbol{x}' = [x'_1, x'_2, \ldots, x'_n]^\top \in \mathbb{R}^n$. The input $\boldsymbol{x}$ is mapped to $\boldsymbol{h}$ which is a vector of hidden units through the encoder. The subsequent decoder maps $\boldsymbol{h}$ to the output vector $\boldsymbol{x}'$ to reconstruct $\boldsymbol{x}$. The two steps can be written

$$\boldsymbol{h} = f(\boldsymbol{W}_1^\top \boldsymbol{x} + \boldsymbol{b}_1), \tag{1}$$

$$\boldsymbol{x}' = \boldsymbol{W}_2 \boldsymbol{h} + \boldsymbol{b}_2, \tag{2}$$

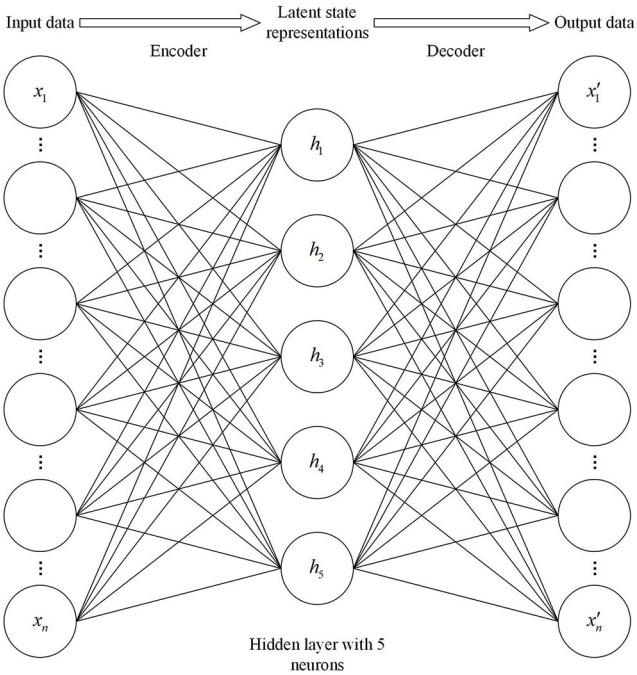


FIGURE 2 | Architecture of the single-hidden-layer undercomplete auto-encoder.

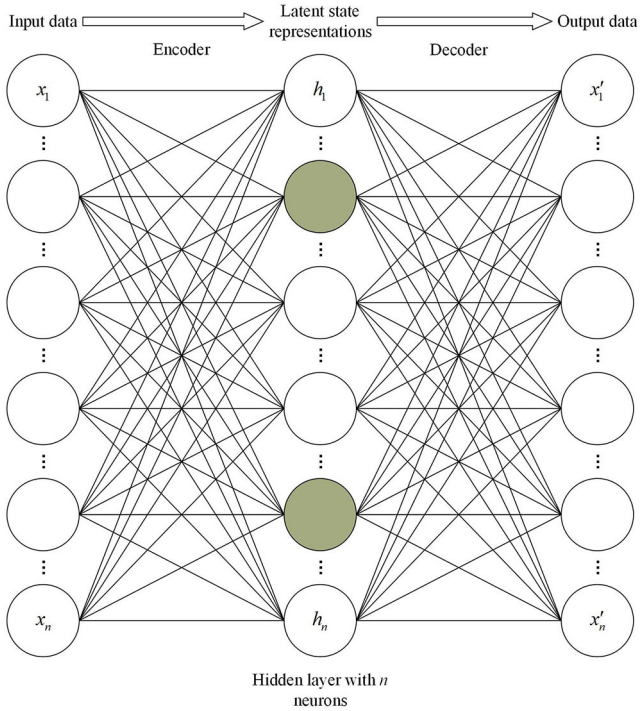


FIGURE 3 | Architecture of the sparse auto-encoder.

where $\boldsymbol{W}_1$, $\boldsymbol{W}_2$ represent the weights of a linear transformation; $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ are the biases; and $f(\cdot)$ is an activation function. Frequently used activation functions are sigmoid $(1/(1+e^{-x}))$ [29, 30], hyperbolic tangent $(\tanh(x))$ [31], or rectified linear units (ReLU) $(\max\{0, x\})$ [32–34]. In this paper, $f(\cdot)$ is set to be a ReLU function, because ReLU solves the gradient vanishing problem (in the positive interval) with a high speed of convergence and calculation compared to other activation functions. When the activation functions are linear and the loss function is the mean squared error, the action of the single-hidden-layer undercomplete auto-encoder is equivalent to Principal Component Analysis (PCA) [35]. In addition, we do linear transformation other than use non-linear activation functions on the output layer to make the output zero-centered. The characteristics of the output are thereby kept consistent with the input data.

The network of the single-hidden-layer undercomplete auto-encoder is trained by minimize the reconstruction error $\mathcal{L}(\boldsymbol{x}, \boldsymbol{x}')$, i.e., the two-norm difference between the input vector and the output vector:

$$\min_{\boldsymbol{W}_1,\boldsymbol{W}_2,\boldsymbol{b}_1,\boldsymbol{b}_2} \sum_{i=1}^{m} \mathcal{L}(x^{(i)}, x'^{(i)}) = \min_{\boldsymbol{W}_1,\boldsymbol{W}_2,\boldsymbol{b}_1,\boldsymbol{b}_2} \sum_{i=1}^{m} \left\| x^{(i)} - x'^{(i)} \right\|_2. \quad (3)$$

Back-propagation is used for the solution of Equation (3), with the popular gradient descent optimization algorithm called Adaptive Moment Estimation (Adam) [36]. (Unless otherwise stated, in the constructions of other auto-encoder models in this paper, the designs of the input and output vectors, the activation functions of the hidden layers, the loss functions, and the parameter-optimization algorithms are consistent with those of the single-hidden-layer undercomplete auto-encoder).

We already know that undercomplete auto-encoders can learn the most significant features of data distribution. However, if these auto-encoders are given too much capacity, they cannot learn any useful information. Regularized auto-encoders can solve this problem by imposing particular forms of regularization on the networks in order to encourage the models to have better generalization abilities rather than limiting their capacity. Sparse auto-encoders [37, 38] are a common kind of regularized auto-encoders. A sparse auto-encoder suppresses the activation of most neurons in the hidden layer by adding a sparsity penalty in the loss function, thereby providing another method of knowledge compression without reducing the number of nodes in the hidden layer. The architecture of the sparse auto-encoder applied in this paper is shown in **Figure 3**. The hidden layer has the same dimension as the input and output layers. The light-colored circles in the hidden layer represent suppressed neurons, while the dark-colored circles represent activated neurons. Since the activation of neurons is data-driven, the sparse auto-encoder can obtain specific feature representations for different input data. The network's capacity is limited to prevent excessive memorizing of input data, while the capacity to extract data features is not limited. There are two common ways of constructing the sparsity penalty: L1 regularization [39] and Kullback–Leibler (KL) divergence [40]. In this paper, we use L1 regularization. The loss function for training our sparse auto-encoder is

given by

$$Loss = \mathcal{L}(\boldsymbol{x}, \boldsymbol{x}') + \lambda \|\boldsymbol{h}\|_1, \tag{4}$$

where the second term penalizes the output value of the hidden layer, scaled by a tuning parameter λ.

We also consider another regularized auto-encoder, the contractive auto-encoder [41], which is designed to make the learned feature representation insensitive to small changes around the training examples. This is accomplished by penalizing instances where a small change in the input results in a large change in the encoding space. Thus, the loss function is

$$Loss = \mathcal{L}(\boldsymbol{x}, \boldsymbol{x}') + \lambda \|\nabla_x \boldsymbol{h}\|_F^2, \tag{5}$$

where the penalty term is the squared Frobenius norm (sum of squared elements) of the Jacobian matrix for the hidden layer outputs with respect to the input observations. Although the contractive auto-encoder regularization criterion is trivial to calculate in the case of a single hidden layer auto-encoder, it becomes much more difficult in the case of deeper auto-encoders. Therefore, the contractive auto-encoder used in this paper adopts the same structure as the single-hidden-layer undercomplete auto-encoder mentioned above. Since we employ ReLU as the activation function on the hidden layer, the regularization criterion can be given the following analytical form:

$$\begin{aligned} \|\nabla_{\boldsymbol{x}} \boldsymbol{h}\|_F^2 &= \sum_{i,j} \left(\frac{\partial h_j}{\partial x_i} \right)^2 \\ &= \sum_j \phi^2(z_j) \sum_i \left(W_{ji}^\top \right)^2, \end{aligned} \tag{6}$$

$$\phi(z_j) = \phi\left(\sum_i W_{ij} x_i + b_j\right) = \begin{cases} 1, \text{ if } z_j \geq 0, \\ 0, \text{ otherwise.} \end{cases} \tag{7}$$

Auto-encoders are not required to be composed of a single-layer encoder and a single-layer decoder. In fact, deep auto-encoders yield much better compression than corresponding shallow auto-encoders [42]. The general method for training a deep auto-encoder consists of training a stack of shallow auto-encoders so as to pretrain the deep architecture. For this reason, deep autoencoders are also called stacked auto-encoders. The stacked auto-encoder employed in this paper is built with the structure shown in **Figure 4**, where the numbers of hidden layers and neurons in each layer are set by trial and error.

Till now, the input and output of the auto-encoders we have introduced are identical. Such models may not perform well on a testing set where the testing and training data do not exhibit the same distribution. The denoising auto-encoder [43] provides remedies for this deficiency. Denoising auto-encoders receive as input data that have been corrupted by some form of noise, and are trained to reconstruct the uncorrupted data as their output. After denoising training, the network is forced to learn more robust invariant features and obtain more effective representations of the input. This is very similar to a contractive auto-encoder in the sense that the noise is considered a series of small perturbations to the input. The difference is that contractive auto-encoders make the feature extraction function resist small perturbations of the input, while denoising auto-encoders make the reconstruction function resist them [44]. The initial input can be corrupted by adding Gaussian noise or stochastically discarding certain features. The denoising auto-encoder employed in this paper is constructed with the same architecture as the stacked auto-encoder. The only difference is that the input is the corrupted data $\tilde{\boldsymbol{x}}$, as shown in **Figure 5**, and given by

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} + \eta \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \tag{8}$$

where $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ represents a multivariate standard normal distribution with a diagonal covariance structure, and η denotes noise intensity. The loss function for the denoising auto-encoder still computes the two-norm difference between the output vector $\boldsymbol{x}'$, and the original data $\boldsymbol{x}$.

The decoder networks built by the auto-encoders we have introduced above output a single value to describe each latent attribute. However, sometimes we hope to learn a probability distribution for each latent attribute to produce a better generalization and ensure that the latent space has properties that enable the generative process. This goal can be achieved by applying a well-known generative model, the variational auto-encoder [45, 46]. The special structure of the variational auto-encoder designed for the purpose of this paper is shown in **Figure 6**. Its encoder outputs parameters describing a distribution for each dimension in the latent space. Here we assume that the prior distribution $p(\boldsymbol{h})$ of the latent representation obeys a standard normal distribution, and the encoder therefore outputs two vectors describing the mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$ of the latent state distribution. The decoder will then generate a latent vector $\boldsymbol{h}$ by sampling from a multivariate Gaussian model with a diagonal covariance matrix and reconstruct the original input. It is worth noting that a simple trick, reparametrization, is used when sampling. It can be expressed as

$$\boldsymbol{h} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \tag{9}$$

This allows us to sample from a unit Gaussian $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ rather than sampling from the distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$, so as to ensure that the results of sampling are derivable and the error can be backpropagated through the network. The loss function for the variational auto-encoder is defined as

$$Loss = \mathcal{L}(\boldsymbol{x}, \boldsymbol{x}') + \lambda \sum_j D_{\mathrm{KL}}(q_j(h_j|\boldsymbol{x}) \| p(h_j)), \tag{10}$$

where

$$\begin{aligned} D_{\mathrm{KL}}(q_j(h_j|\boldsymbol{x}) \| p(h_j)) &= D_{\mathrm{KL}}(\mathcal{N}(\mu_j, \sigma_j^2) \| \mathcal{N}(0,1)) \\ &= \frac{1}{2}(-\log \sigma_j^2 + \mu_j^2 + \sigma_j^2 - 1). \end{aligned} \tag{11}$$

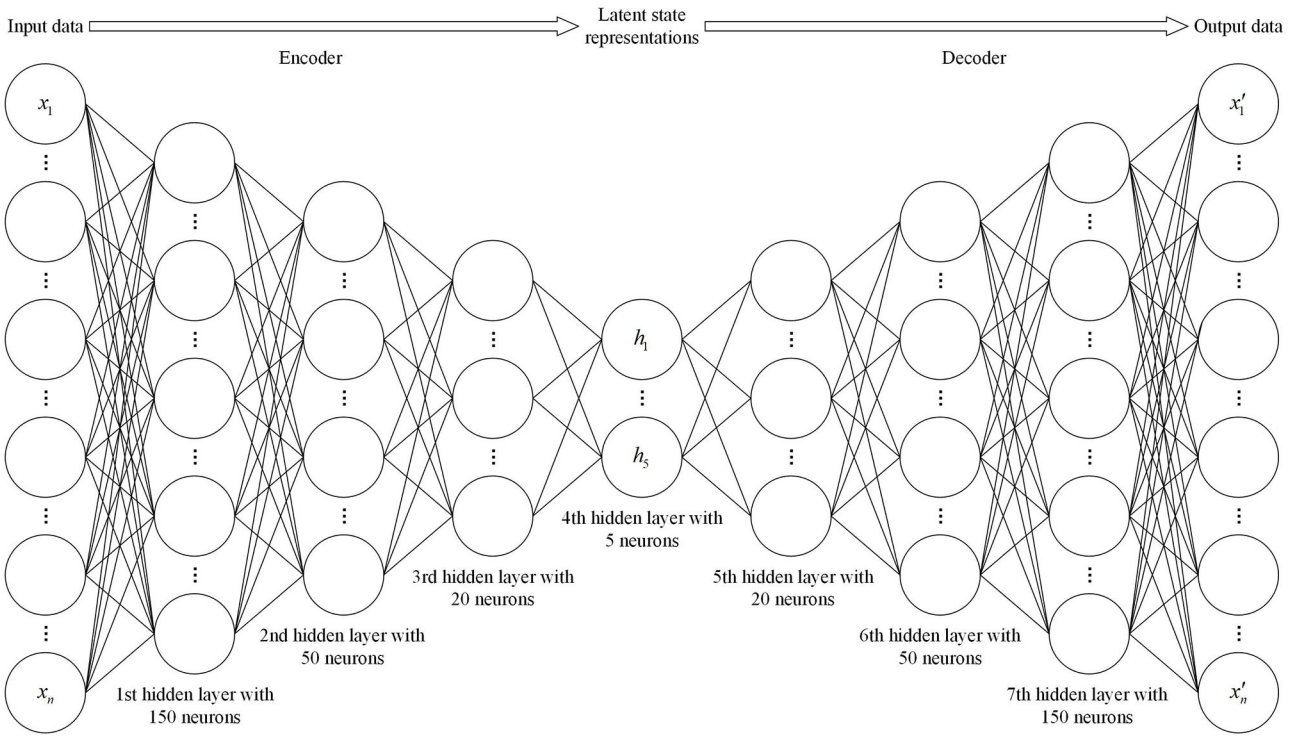


FIGURE 4 | Architecture of the stacked auto-encoder.

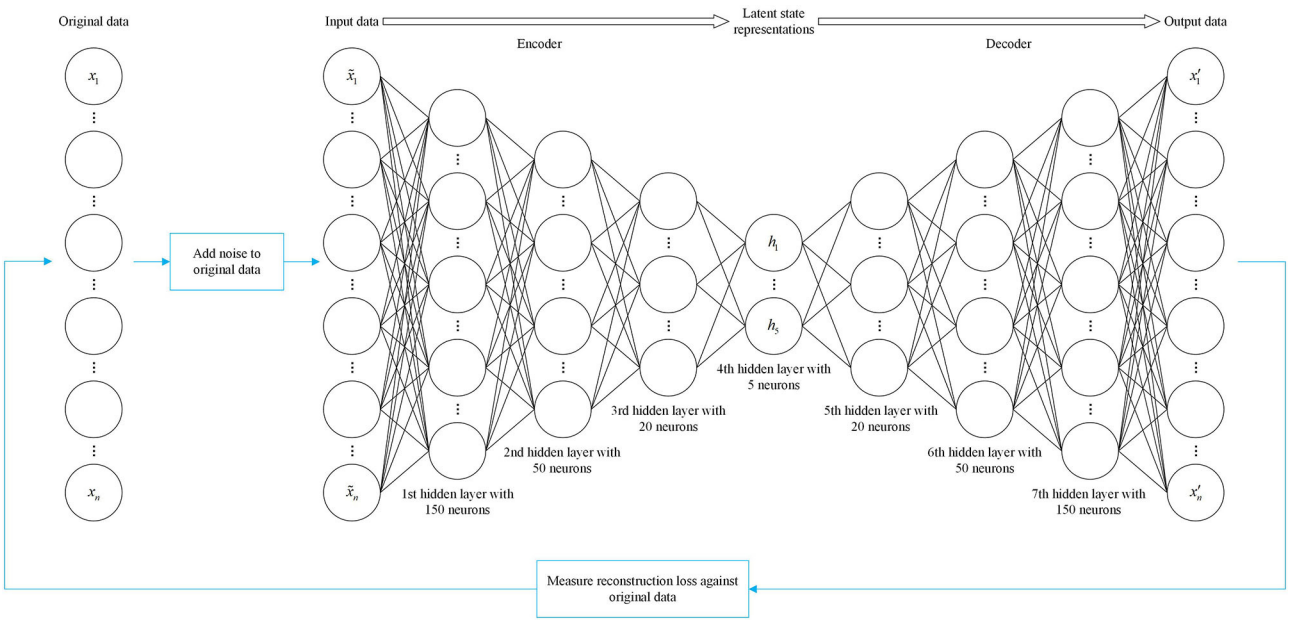


FIGURE 5 | Architecture of the denoising auto-encoder.

The first term in Equation (10) penalizes reconstruction errors (a feature also found in other auto-encoders). The second term encourages the learned latent-state distribution $q(\boldsymbol{h}|\boldsymbol{x})$ to be similar to the prior distribution $p(\boldsymbol{h})$, which minimizes the KL divergence between these two distributions. The relative weights of these two items are controlled by a hyperparameter λ.

After the auto-encoders have been trained, their encoders output an n-dimensional vector that contains n different latent

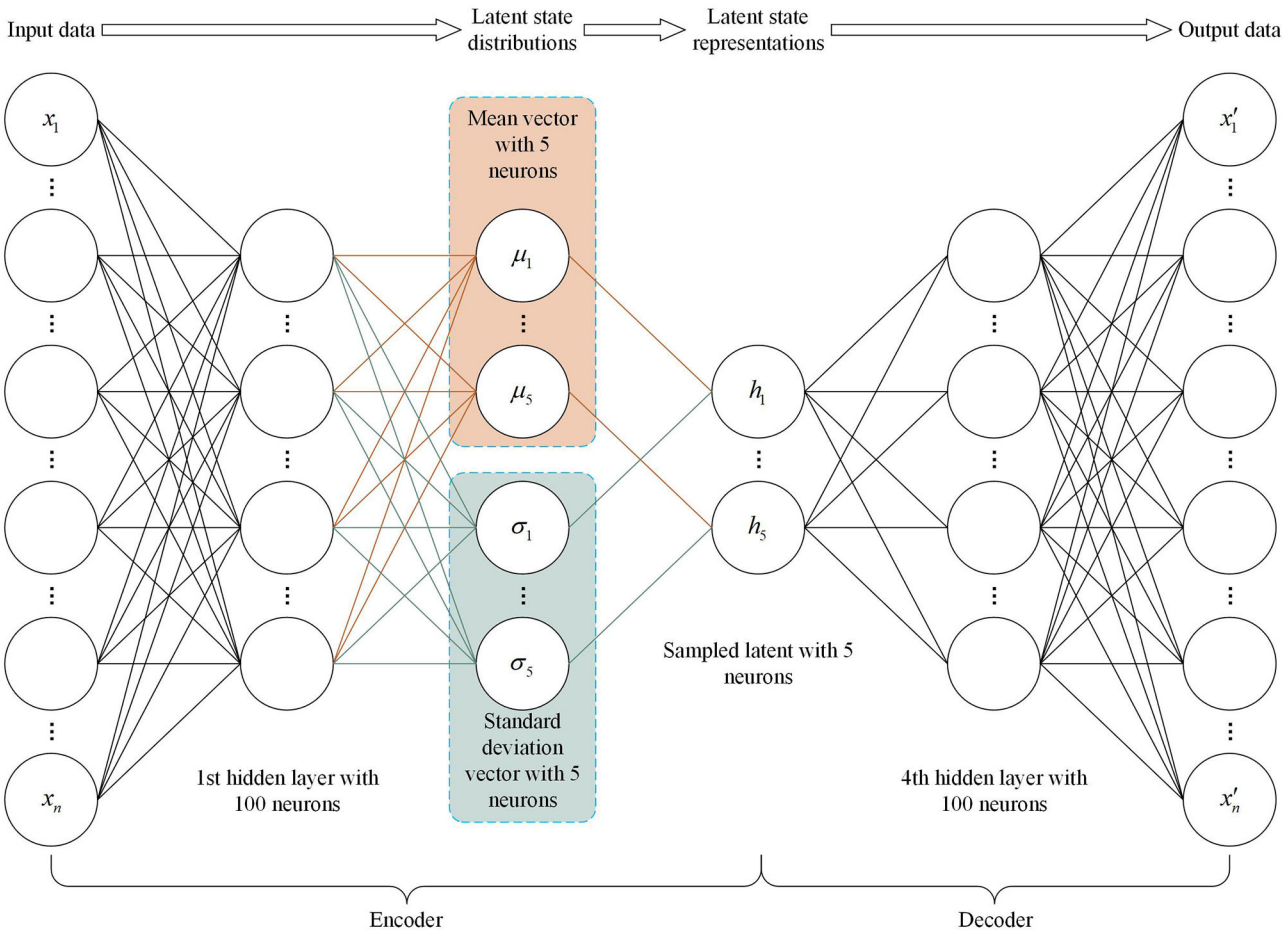


FIGURE 6 | Architecture of the variational auto-encoder.

factors. These latent factors are obtained by the process of dimensionality reduction or compression and can be used to represent n independent implied abstract features of the stock index market. This technique is of great significance in finance. Traditional financial pricing models with shallow architectures (at most two layers) typically describe market information based on linear portfolios. For example, the capital asset pricing model (CAPM) proposed by Sharpe [47] assumes that the market return is expressed by a linear combination of asset returns. In the arbitrage pricing theory (APT) proposed by Rosenberg and McKibbon [48] and Ross [49], a layer of linear factors is used to perform pricing. These traditional financial theories also apply the idea of dimensionality reduction, as they reduce a dataset of n observations (returns or factors) to one parameter. However, while the implied market prices capture linear features of the input asset returns or factors, they ignore a large amount of latent information and the non-linear relationship between the assets in a complex system with fractality properties. For this reason, we use the auto-encoder model with a hierarchical structure of univariate activation functions of portfolios to make up for the shortcomings of traditional financial models.

The decoders then proceed to reconstruct the input individual stock-returns from the latent representations of the stock index market. However, this process involves compression encoding, and therefore will inevitably bring information loss. Following Heaton et al. [25], we calculate the information loss of each stock during the encoding-decoding process by using Equation (12) below to measure the similarity of the j-th stock with the stock index market (i.e., the total two-norm difference between every original stock return and the corresponding reconstructed one on the training batch):

$$\mathcal{L}_j = \sum_{i=1}^{m} \left\| x_j^{(i)} - x_j'^{(i)} \right\|_2 . \tag{12}$$

The smaller $\mathcal{L}_j$ is, the less information the j-th stock loses, and therefore the more similar it is to the stock index market. We rank the stocks by their communal information content, i.e., the amount of information that they share with the stock index market. Since it is not beneficial for improving index tracking performance to include too many stocks contributing the same information, we select a fixed number of the most-communal

stocks plus a variable number of the least-communal stocks to construct a tracking portfolio.

In addition, in order to investigate the superiorities of auto-encoder-based stock selection strategies, we also adopt for comparison two conventional index-tracking stock-selection strategies: weight ranking and market-value ranking. We evaluate the tracking performance of these strategies under the same conditions.

Index Tracking Model

After selecting the representative stocks by the strategies above, we use an index tracking model to determine the investment weight allocated to each stock in the tracking portfolio, with the objective of minimizing tracking error and other constraints. The index tracking model established in this paper can be expressed as the following quadric programming problem:

$$\begin{aligned} \boldsymbol{w}^* &= \arg\min_{w} \|\mathbf{R}_I - \mathbf{R}_x \boldsymbol{w}\|_2^2 + \lambda \|\boldsymbol{w}\|_2^2 \\ \text{s.t.} \sum_{i=1}^{n} w_i &= 1, \\ w_i &\geq 0,\ i = 1, 2, \ldots, n, \end{aligned} \tag{13}$$

where $\boldsymbol{R}_I \in \mathbb{R}^m$ is a vector of the index return time series; $\boldsymbol{R}_x = [\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_n] \in \mathbb{R}^{m \times n}$ denotes the return matrix of the selected stocks; and $\boldsymbol{w} = [w_1, w_2, \ldots, w_n]^\top \in \mathbb{R}^n$ is a vector of stock weights. The objective function is complemented with a regularization term, $\lambda \|\boldsymbol{w}\|_2^2$, to avoid overfitting. In addition, the stock weights are kept non-negative, considering the short-selling restrictions in China's stock market.

EMPIRICAL ANALYSIS

Data Description and Processing

We investigate partial replication of the CSI 300 Index with the index tracking strategies we have proposed. The CSI 300 Index is a barometer of China's stock market. Its main income accounts for more than seventy percent of the Chinese market, and it well-represents emerging markets throughout the world. We use the daily closing prices of the CSI 300 Index and its constituent stocks from the sample period January 1, 2010 through December 31, 2018 (comprising 2,187 trading days). Because the constituents of the CSI 300 Index are adjusted semi-annually, generally in early January and early July, we obtain the daily closing prices of all the stocks that have been included in the constituents during the sample period. We also record the mid-year and end-year market values of the constituents and their weights from 2010 to 2018, for use in weight ranking and market-value ranking.

To ensure the analysis results are accurate and reliable, we first clean the original pricing data by the following steps:

(i) Exclude the stocks if more than 20% of the pricing data is missing in the training set (defined in the next sub-section).
(ii) Exclude the stocks if all pricing data for the first 5 days and the last 5 days is missing in the training set.
(iii) Exclude the stocks if they have been ejected from the constituents of the CSI 300 Index during the training set and the following testing set (defined in the next sub-section).
(iv) Perform linear interpolation to fill the missing prices of the retained stocks.

We obtain the daily return time series $r_{i,t}$ for each stock or the index by calculating $r_{i,t} = (P_{i,t} - P_{i,t-1})/P_{i,t-1}$, where $P_{i,t}$ denotes the daily closing price of stock (index) i on day t. Then all daily returns are standardized using z-score normalization as follows:

$$x_{i,t} = \frac{r_{i,t} - \bar{r}_i}{\sigma_i}, \tag{14}$$

where $\bar{r}_i$ and σ_i denote the mean and standard deviation of $r_{i,t}$, respectively.

Design of Tracking Strategy

In order to construct a dynamically adjusted out-of-sample portfolio to track the index, the data sample is divided into training and testing sets by the rolling-window approach [50]. The rolling-window approach keeps the features of time series in the data, making it match the investment decision-making process in practice. The training set is used to train the stock selection model to select a subset of constituents. The index tracking model which takes the returns of the selected stocks as input is then also trained on the training set to obtain the

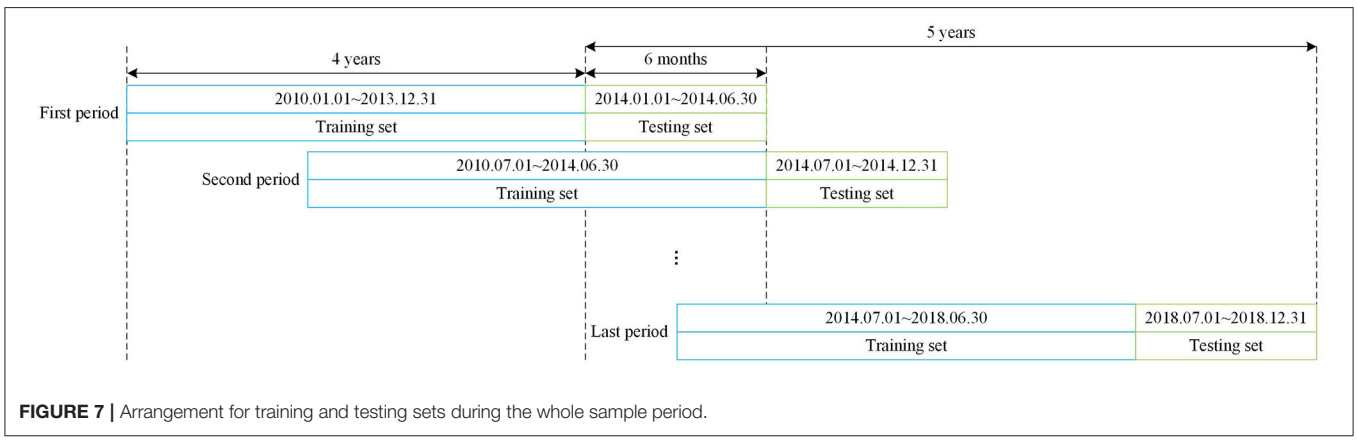


FIGURE 7 | Arrangement for training and testing sets during the whole sample period.

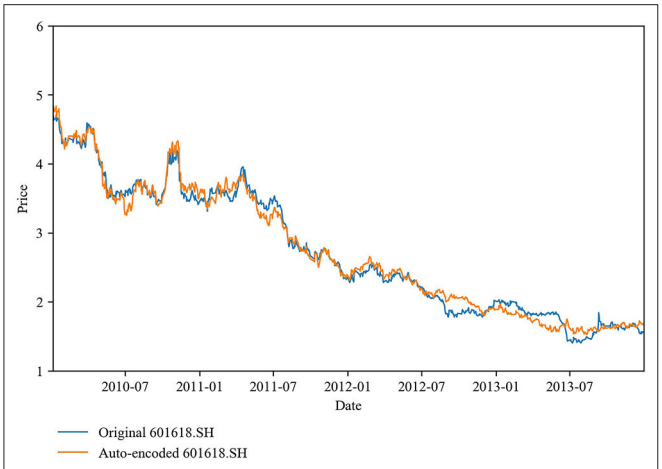


FIGURE 8 | Stock (601618.SH) with the most communal information: the prices of the actual and auto-encoded stocks are nearly the same across the period shown.

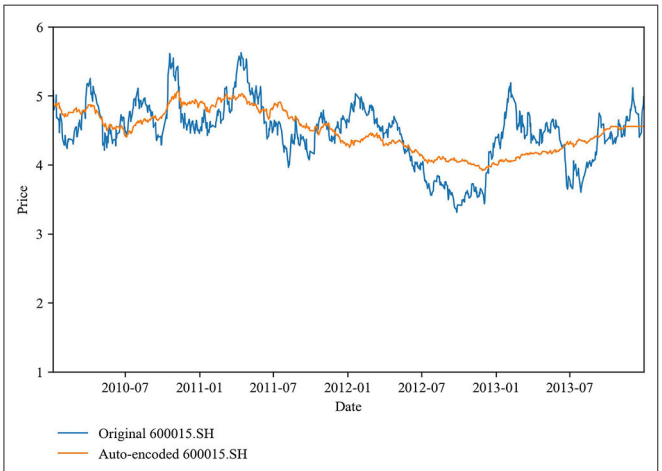


FIGURE 9 | Stock (600015.SH) with the least communal information: the prices of the actual and auto-encoded stocks differ significantly across the period shown.

stock weights. Afterwards, we construct a tracking portfolio with the selected stocks and corresponding weights obtained from the training set, and compute its portfolio return as well as the index tracking error on the testing set. We use the past four years' data as a training set. The dataset for the following 6 months is regarded as a testing set, in line with the adjustment frequency of the index constituents. This process continues for 5 years on each half-year from Jan. 2014 to Dec. 2018. For each stock selection model, there are in all 10 periods and 5 yearly index tracking results. The tracking procedure is illustrated in **Figure 7**.

Performance Measurement

We select stocks for each training set by employing eight selection approaches: six auto-encoder-based models, weight ranking, and market-value ranking. The auto-encoders are used to measure the degree of communal information between the stock index market and the constituent stocks. We then sort the constituents accordingly and select a subset of constituents that satisfy our requirements. As an example, **Figures 8**, **9** illustrate the stock 601618.SH, which shares the most communal information with the stock index market in the first period of the training sets (adopting the signal-hidden-layer undercomplete auto-encoder), and the stock 600015.SH, which shares the least. Obviously, stock 600015.SH loses much more information than stock 601618.SH during the encoding-decoding process. We already know that it is not necessary to add too much communal information to a portfolio. Following Heaton et al. [25], we select the 10 most communal stocks plus the $n - 10$ least communal stocks to construct a tracking portfolio, where n increases from 15 to 80 in steps of five. The weight (market value) ranking method is to select the n stocks with the largest half-yearly average weights (market values) for inclusion in a tracking portfolio.

After determining the stocks required for inclusion in the tracking portfolio, we apply the index tracking model introduced in section Index Tracking Model to determine the stock weights and construct a tracking portfolio to partially replicate the CSI 300 Index. We evaluate the tracking errors on portfolios with the same number of stocks selected by different strategies. A smaller tracking error indicates better tracking performance of the stock selection strategy. The equation for calculating the average tracking error ATE is

$$ATE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(R_{It} - R_{pt})^2}, \tag{15}$$

where the T represents the total number of out-of-sample trading days (which spans from January 1, 2014 to December 31, 2018 and covers 10 adjustment periods as the tracking portfolio is adjusted every half-year); R_{It} and R_{pt} are the returns of the index and of the tracking portfolio at time t.

Table 1 shows the out-of-sample tracking error values for the CSI 300 Index. **Figure 10** plots how the tracking error values change as a function of the number of stocks for all stock selection strategies. The tracking errors of all strategies decrease as the number of stocks in the tracking portfolio increases. In particular, the tracking error falls quickly when < 40 stocks are included in the portfolio. Furthermore, when the number of stocks included is < 30, the tracking errors of the six auto-encoder-based strategies are significantly smaller than those of the weight ranking and market-value ranking strategies. However, the falling rate of the tracking error slows down when over 40 stocks are required for inclusion. Moreover, the tracking errors of the six auto-encoder-based strategies exceed those of the weight ranking and the market-value ranking strategies when over 40 and 55 stocks are required for inclusion, respectively. We suggest the following explanations for the above results. When the tracking portfolio is constructed with many stocks selected by the weight ranking and market-value ranking strategies, the cumulative origin weight in the index of the selected stocks is larger, making the performance of the tracking portfolio closer to that of the index. While as the number of

TABLE 1 | Out-of-sample tracking error values ($\times 10^{-3}$) for all strategies covered by this study.

Number of stocks for inclusion	Stock selection strategy							
	Weight ranking	Market-value ranking	Single-hidden-layer undercomplete AE	Sparse AE	Contractive AE	Stacked AE	Denoising AE	Variational AE
15	6.554	8.590	5.285	4.976	5.294	5.329	5.115	5.814
20	5.982	7.128	4.770	4.246	4.648	4.491	4.396	4.874
25	5.224	5.940	4.134	3.971	3.993	4.058	3.940	4.403
30	4.351	5.243	3.977	3.737	3.688	3.910	3.659	4.232
35	3.728	4.645	3.812	3.681	3.606	3.766	3.617	4.126
40	3.464	4.077	3.656	3.636	3.517	3.696	3.507	4.052
45	3.208	3.793	3.697	3.492	3.498	3.641	3.309	3.838
50	3.034	3.561	3.535	3.338	3.479	3.426	3.241	3.830
55	2.893	3.155	3.507	3.259	3.353	3.345	3.211	3.820
60	2.736	2.947	3.366	3.151	3.261	3.296	3.195	3.700
65	2.661	2.823	3.320	3.072	3.233	3.268	3.104	3.604
70	2.563	2.722	3.273	3.046	3.248	3.236	3.084	3.569
75	2.489	2.598	3.254	2.965	3.195	3.185	3.053	3.479
80	2.379	2.504	3.222	2.976	3.100	3.171	3.031	3.374

"AE" is short for "auto-encoder".

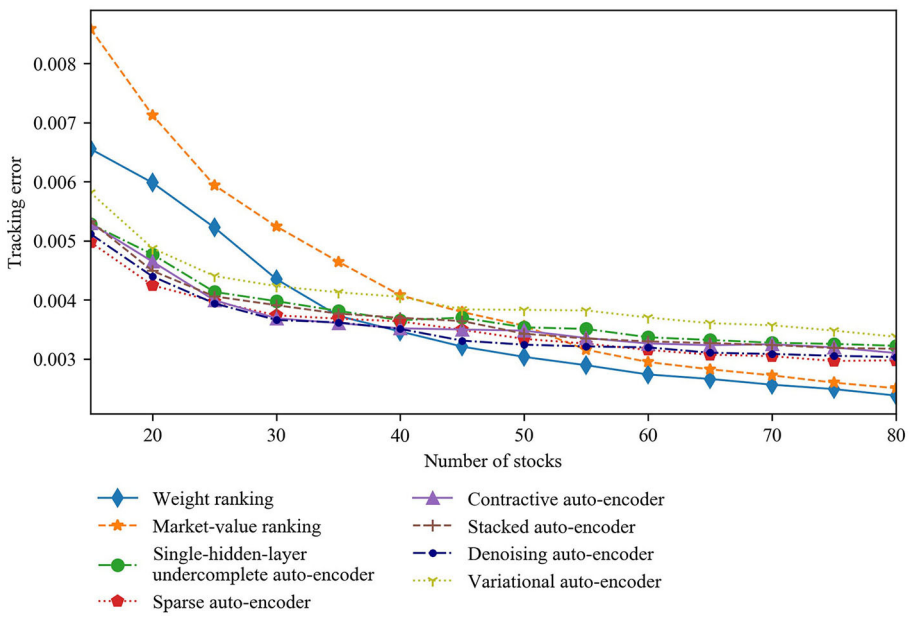


FIGURE 10 | Tracking error curves for all strategies covered by this study.

stocks in the tracking portfolio increases, the auto-encoder-based strategies append more stocks with medium communal information to the portfolio. The portfolios containing the most- and least-communal stocks are well able to reflect the market information. Thus, there is no benefit in having more medium-communal stocks.

Comparing the auto-encoder-based strategies to one another, the tracking errors of the strategies based on sparse, contractive, stacked, and denoising auto-encoders are almost always < that of the strategy based on single-hidden-layer undercomplete auto-encoder regardless of the number of stocks, although the difference is not sizeable. The explanation is that some of these four types of auto-encoders have a deeper structure that can learn more complex coding and deeper market information, while others are regularized to encourage the model to learn other features (except copying the input to the output) without limiting the model capacity by keeping the encoder and decoder shallow and the code size small. In either case, these auto-encoders can create more information-efficient representations of the market than the

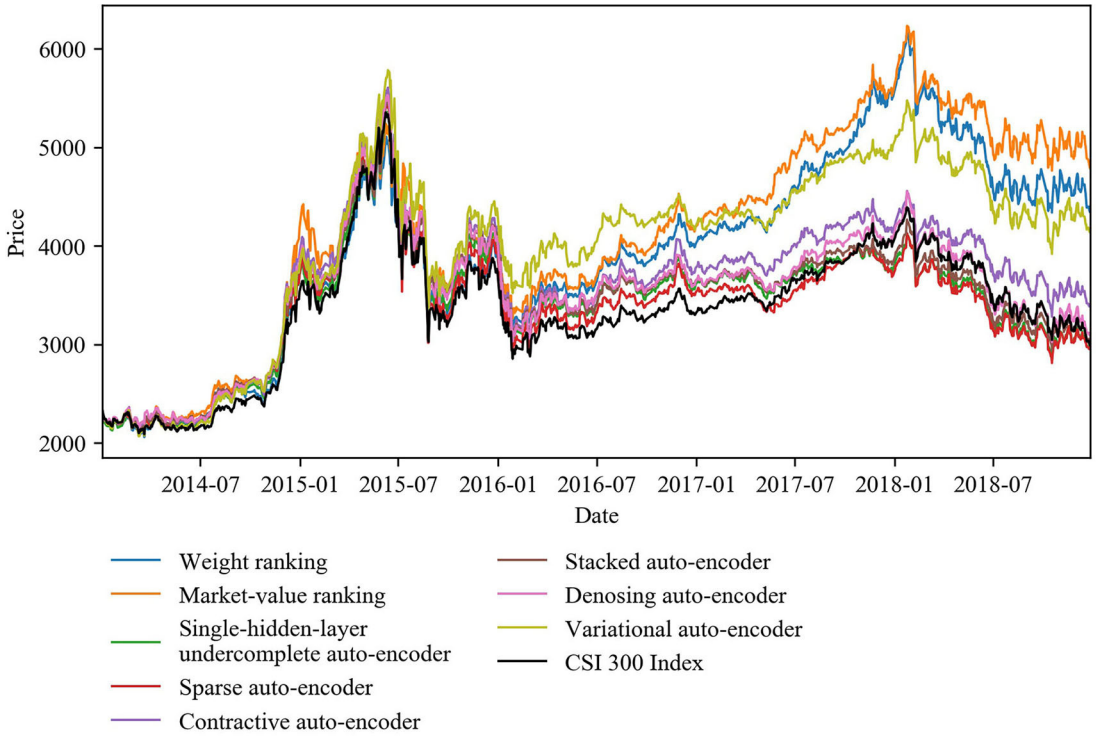


FIGURE 11 | Out-of-sample cumulative return curves of the CSI 300 Index and the 25-stock tracking portfolios constructed by all strategies covered by this study.

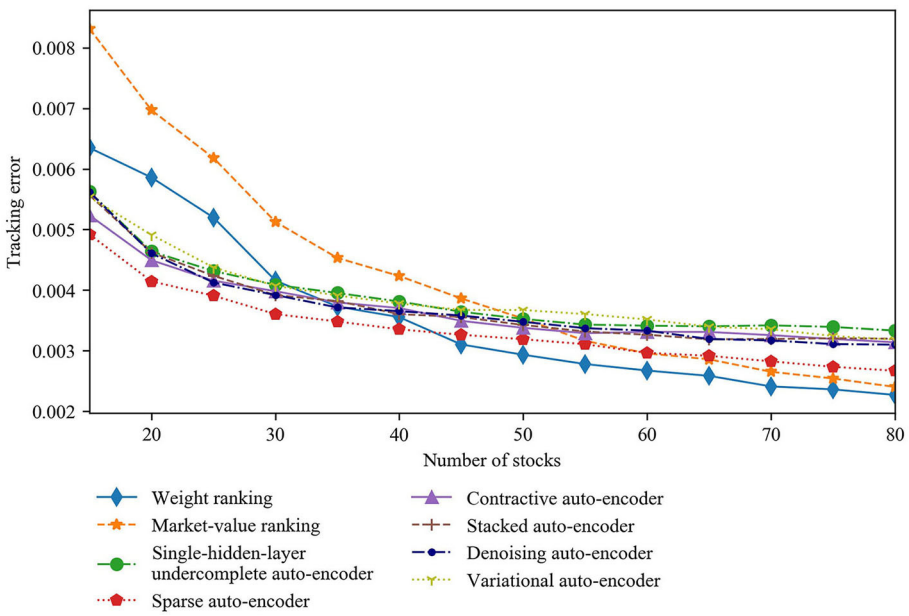


FIGURE 12 | Tracking error curves for all strategies with 3-year training sets.

single-hidden-layer undercomplete auto-encoder, so that the stocks selected by the strategies based on them better represent the entire market.

However, the strategy based on the variational auto-encoder does not perform better than that based on single-hidden-layer undercomplete auto-encoder. This can also be explained.

The purpose of an auto-encoder in the present work is to replicate the original input stock information from the latent space representing the compressed market information. However, a variational auto-encoder (mentioned in section Stock Selection Using Auto-Encoders, and normally used as a generative model) is meant to generate variations on an input vector from a continuous latent space: that is, its encoder only outputs a range of possible representations of the market, and these do not necessarily describe the market's current state. Therefore, the output reconstructed by the decoder is far from being a copy of the original input stock information. From this perspective, it is not surprising that the strategy based on the variational auto-encoder does not yield the desired result.

Although increasing the number of stocks in the tracking portfolio will reduce the tracking error, it will not significantly improve the tracking performance, while it will create additional transaction cost when the number of stocks included reaches a certain value. According to the previous analysis, the tracking error decreases rapidly as the number of stocks increases and the corresponding transaction cost is acceptable if a tracking portfolio is constructed with < 40 stocks. Therefore, the number of stocks in the tracking portfolio should be kept under 40 when balancing the tracking error and the transaction cost.

Considering the absolute tracking error values and the slope of the curves for the auto-encoder-based strategies in **Figure 10**, we find the tracking performance of auto-encoder-based strategies greatly surpasses that of conventional strategies for a 25-stock tracking portfolio. **Figure 11** shows the out-of-sample cumulative return curves of the CSI 300 Index and the 25-stock tracking portfolios constructed by our proposed strategies. The relative advantages of auto-encoder-based stock selection strategies can be seen clearly. In particular, the tracking error of the market-value ranking strategy is 5.940×10^{-3}, and that of the weight ranking strategy is 5.224×10^{-3}. Among the six auto-encoder-based strategies, the tracking error of the denoising auto-encoder-based strategy is the smallest at 3.940×10^{-3}, which is 33.67% lower than that of market-value ranking and 24.58% lower than that of weight ranking. The other five auto-encoder-based strategies also track better than the conventional strategies to varying degrees. Even the worst-performing auto-encoder-based strategy (the variational auto-encoder) has reductions of 25.88 and 15.72% compared to market-value ranking and weight ranking strategies, respectively. In conclusion, auto-encoder-based strategies outperform conventional strategies, provided that only a small number of stocks are required for inclusion in a tracking portfolio.

Robust Test

To evaluate the sensitivity of these empirical results to changes in the data sample, we perform various robustness checks.

First, variations in length of the training set may have an impact on the results. As a robustness check, we analyze the tracking performance when each training set length is changed to 3 or 5 years, respectively. Keeping each testing set length at 6 months, the length of the out-of-sample period accordingly changes to 6 and 4 years, respectively. **Figures 12**, **13** illustrate how the curves of the tracking error vary with the number of stocks when each training set has a length of 3 and 5 years, respectively. The results reveal that the

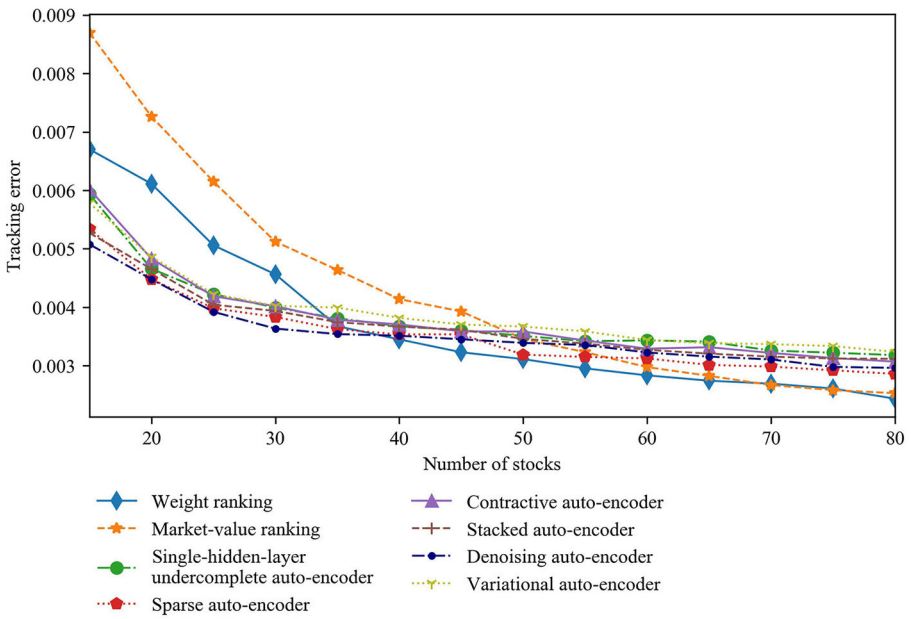


FIGURE 13 | Tracking error curves for all strategies with 5-year training sets.

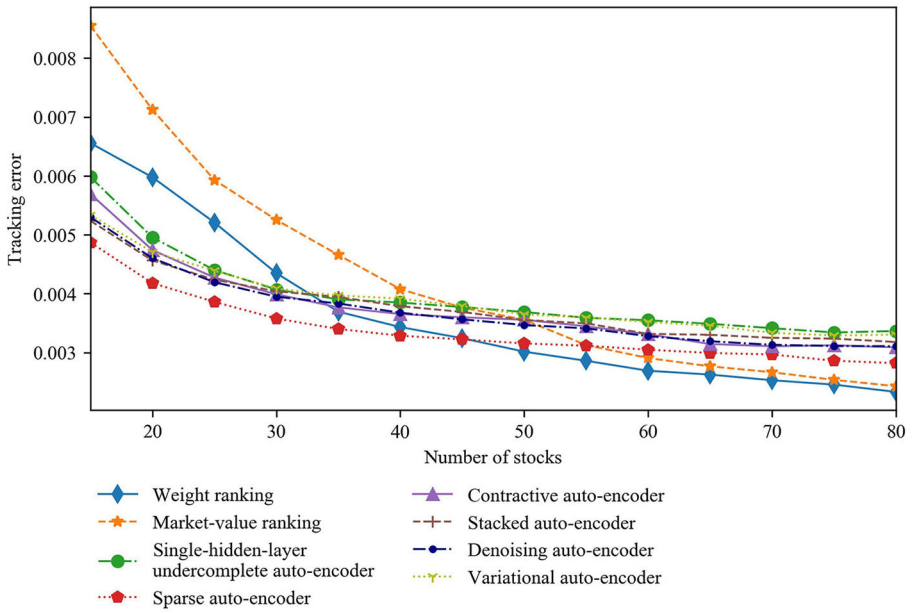


FIGURE 14 | Tracking error curves for all strategies with quarterly rebalancing.

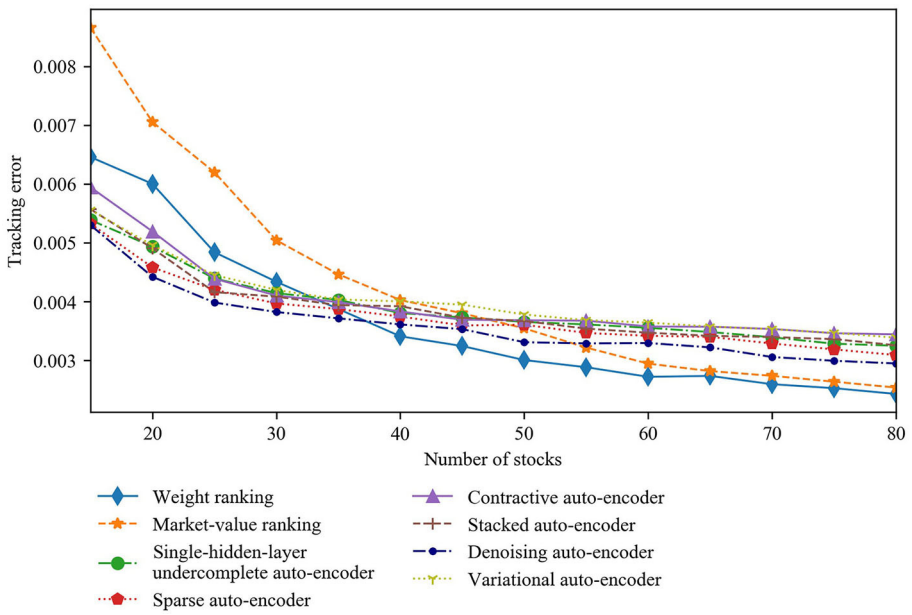


FIGURE 15 | Tracking error curves for all strategies with yearly rebalancing.

auto-encoder-based strategies tracks the index better than the conventional strategies when < 30 stocks are included in a tracking portfolio. In particular, the sparse auto-encoder-based strategy gets the lowest tracking error among all the auto-encoder-based strategies with 3-year training sets, whereas the denoising auto-encoder-based strategy performs best with 5-year training sets. In addition, the tracking error values change little in response to variations in the length of the training set. Thus, our base case results hold for these alternative training-set lengths.

Second, the rebalancing frequency, which is the reciprocal of the length of each testing set, will affect the performance of dynamic portfolio management. We compute quarterly and yearly rebalanced portfolios while keeping the training-set length unchanged to investigate the sensitivity of our results to alternative rebalancing frequencies. The results in **Figures 14**, **15**

verify that our base case results are robust to these changes. In particular, with quarterly rebalancing, the sparse auto-encoder-based strategy tracks the index best among all auto-encoder-based strategies and brings greater improvement on conventional strategies' tracking performance compared to the base case results. In the case of a 25-stock portfolio, the tracking error of the sparse auto-encoder-based strategy is 3.861×10^{-3}, which is 34.91% lower than the market-value ranking and 25.86% lower than the weight ranking. In contrast, with yearly rebalancing, the denoising auto-encoder-based strategy gets the best tracking performance. This proves that the sparse and denoising auto-encoder-based strategies are better at index tracking than the other four auto-encoder-based strategies.

CONCLUSIONS

We investigate the index tracking performance of deep learning-based tracking approaches. In particular, we use a variety of advanced auto-encoders: single-hidden-layer undercomplete, sparse, contractive, stacked, denoising, and variational auto-encoders to extract the complex non-linear relationship between stocks in a complex stock market system and construct dynamic tracking portfolios with subsets of stocks. Only one or two of these auto-encoders has previously been examined in the context of stock selection. Moreover, we evaluate for the first time whether auto-encoder-based strategies improve the tracking performance over the conventional strategies of weight ranking and market-value ranking.

In general, we find that whether auto-encoder-based strategies outperform conventional ones depends upon the number of stocks included in the tracking portfolio. When only a small number of stocks (probably < 30) are needed to construct a tracking portfolio, the auto-encoder-based strategies are generally superior to conventional strategies in terms of tracking performance. Furthermore, auto-encoders with particular architectures that can learn high-capacity, overcomplete encodings of the input, e.g., sparse and denoising auto-encoders, are better even than other auto-encoders at capturing complex latent representations of the market. The portfolios with stocks selected by these auto-encoders better replicate the index. However, if more than 40 stocks are required for inclusion, the conventional strategies still have the advantage.

Our findings suggest that deep learning algorithms are suitable for index tracking problems if the hierarchical architectures are explicitly designed. We expect these findings to be helpful in making asset-allocation decisions, especially, for indexing investment. Nonetheless, there are some limitations to the study: our analysis concerns a specific dataset; the impact of transaction costs on index tracking performance is not quantified; and hyper-parameter optimization is not well performed when constructing the models. Therefore, additional work with a more extensive dataset, optimized model settings, and greater practical realism would help to confirm our findings. This research can easily be extended to test other deep learning frameworks for index tracking in the future.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: the WIND database (http://www.wind.com.cn) provided by Shanghai Wind Information Co., Ltd.

AUTHOR CONTRIBUTIONS

CZ: conceptualization, methodology, and formal analysis. CZ and FL: writing-original draft and data curation. LF: writing-review, editing, supervision, and funding acquisition. SL: formal analysis, writing-review, editing, and data curation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (grant numbers 71771116, 71720107001, and U1811462), the Late-stage Subsidy Project for Philosophical and Social Sciences Research Foundation of China's Ministry of Education (grant number 18JHQ058), and the Fundamental Research Funds for the Central Universities (grant number 0118/14370107).

REFERENCES

1. Frino A, Gallagher DR. Is index performance achievable? An analysis of Australian equity index funds. *Abacus.* (2002) **38**:200–14. doi: 10.1111/1467-6281.00105
2. Masters SJ. The problem with emerging markets indexes. *J Portf Manag.* (1998) **24**:93. doi: 10.3905/jpm.24.2.93
3. Bogle JC. Selecting equity mutual funds. *J Portf Manag.* (1992) **18**:94. doi: 10.3905/jpm.1992.409402
4. Ouyang L, Wan L, Park C, Wang J, Ma Y. Ensemble RBF modeling technique for quality design. *J Manage Sci Eng.* (2019) **4**:105–18. doi: 10.1016/j.jmse.2019.05.005
5. Larsen GA, Resnick BG. Empirical insights on indexing: how capitalization, stratification and weighting can affect tracking error. *J Portf Manag.* (1998) **25**:51–60. doi: 10.3905/jpm.1998.409656
6. Shapcott J. *Index Tracking: Genetic Algorithms for Investment Portfolio Selection*. Edinburgh: Edinburgh Parallel Computing Centre (1992).
7. Pope PF, Yadav PK. Discovering errors in tracking error. *J Portf Manag.* (1994) **20**:27–32. doi: 10.3905/jpm.1994.409471
8. Rudolf M, Wolter HJ, Zimmermann H. A linear model for tracking error minimization. *J Bank Fin.* (1999) **23**:85–103. doi: 10.1016/S0378-4266(98)00076-4
9. Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics.* (1975) **31**:423–47. doi: 10.2307/2529430
10. Markowitz H. Portfolio selection. *J Fin.* (1952) **7**:77–91. doi: 10.1111/j.1540-6261.1952.tb01525.x
11. Roll R. A mean/variance analysis of tracking error. *J Portf Manag.* (1992) **18**:13–22. doi: 10.3905/jpm.1992.701922
12. Ammann M, Tobler J. Measurement and decomposition of tracking error variance. *Forschungsgemeinschaft für Nationalökonomie an der Universität St Gallen* St. Gallen (2000).

13. Dunis CL, Ho R. Cointegration portfolios of European equities for index tracking and market neutral strategies. *J Asset Manag*. (2005) **6**:33–52. doi: 10.1057/palgrave.jam.2240164
14. Chiam SC, Tan KC, Al Mamun A. Dynamic index tracking via multi-objective evolutionary algorithm. *Appl Comput*. (2013) **13**:3392–408. doi: 10.1016/j.asoc.2013.01.021
15. Filippi C, Guastaroba G, Speranza MG. A heuristic framework for the bi-objective enhanced index tracking problem. *Omega*. (2016) **65**:122–37. doi: 10.1016/j.omega.2016.01.004
16. Focardi SM, Fabozzi FJ. A methodology for index tracking based on time-series clustering. *Quant Fin*. (2004) **4**:417–25. doi: 10.1080/14697680400008668
17. Cai Z, Hong Y, Wang S. Econometric modeling and economic forecasting. *J Manage Sci Eng*. (2018) **3**:178–82. doi: 10.3724/SP.J.1383.304010
18. Jeurissen R, van den Berg J. Optimized index tracking using a hybrid genetic algorithm. In: 2008 *IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*. Hong Kong (2008).
19. Zorin A, Borisov A. Traditional and index tracking methods for portfolio construction by means of neural networks. In: *Network, in Scientific Proceedings of Riga Technical University*. Riga: Information Technology and Management Science (2002). p. 1–9.
20. Fernández A, Gómez S. Portfolio selection using neural networks. *Comput Operations Res*. (2007) **34**:1177–91. doi: 10.1016/j.cor.2005.06.017
21. Freitas FD, De Souza AF, de Almeida AR. Prediction-based portfolio optimization model using neural networks. *Neurocomputing*. (2009) **72**:2155–70. doi: 10.1016/j.neucom.2008.08.019
22. Chen Y, Yang B, Abraham A. Flexible neural trees ensemble for stock index modelling. *Neurocomputing*. (2007) **70**:697–703. doi: 10.1016/j.neucom.2006.10.005
23. Wu L, Yang Y, Liu H. Nonnegative-lasso and application in index tracking. *Comput Statis Data Anal*. (2014) **70**:116–26. doi: 10.1016/j.csda.2013.08.012
24. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) **521**:436–44. doi: 10.1038/nature14539
25. Heaton JB, Polson NG, Witte JH. Deep learning for finance: deep portfolios. *Appl Stochastic Models Bus Ind*. (2017) **33**:3–12. doi: 10.1002/asmb.2209
26. Ouyang H, Zhang X, Yan H. Index tracking based on deep neural network. *Cogn Syst Res*. (2019) **57**:107–14. doi: 10.1016/j.cogsys.2018.10.022
27. Kim S, Kim S. Index tracking through deep latent representation learning. *Quan Fin*. (2020) **20**:639–52. doi: 10.1080/14697688.2019.1683599
28. Hinton GE, Zemel RS. Autoencoders, minimum description length and helmholtz free energy. In: *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press (1993). p. 3–10.
29. Cybenko G. Approximation by superpositions of a sigmoidal function. *Math Control Signal Syst*. (1989) **2**:303–14. doi: 10.1007/BF02551274
30. Barron AR. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans Inform Theory*. (1993) **39**:930–45. doi: 10.1109/18.256500
31. Karlik B, Olgac AV. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Int J Artif Intell Exp Syst*. (2011) **1**:111—22. Available online at: http://www.cscjournals.org/library/manuscriptinfo.php?mc=IJAE-26
32. Jarrett K, Kavukcuoglu K, Ranzato MA, LeCun Y. What is the best multi-stage architecture for object recognition? In: *2009 IEEE 12th International Conference on Computer Vision*. Kyoto (2009). p. 2146–53. doi: 10.1109/ICCV.2009.5459469
33. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. Haifa (2010). p. 807–14.
34. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL (2011). p. 315–23.
35. Wold S, Esbensen K, Geladi P. Principal component analysis. Ch*emometr Intell Lab Syst*. (1987) **2**:37–52. doi: 10.1016/0169-7439(87)80084-9
36. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv [Preprint] arXiv:1412*. (2014). Available online at: https://arxiv.org/abs/1412.6980
37. Ranzato MA, Poultney C, Chopra S, Cun YL. Efficient learning of sparse representations with an energy-based model. In: *Advances in Neural Information Processing Systems*. Cambrdige, MA: MIT Press (2007). p. 1137–44.
38. Ranzato MA, Boureau YL, Cun YL. Sparse feature learning for deep belief networks. In: *Advances in Neural Information Processing Systems*. Cambrdige, MA: MIT Press (2008). p. 1185–92.
39. Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc Ser B*. (1996) **58**:267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
40. Kullback S, Leibler RA. On information and sufficiency. *Annals Math Stat*. (1951) **22**:79–86. doi: 10.1214/aoms/1177729694
41. Rifai S, Vincent P, Muller X, Glorot X, Bengio Y. Contractive auto-encoders: explicit invariance during feature extraction. In: *Proceedings of the 28th International Conference on Machine Learning*. Bellevue, WA (2011). p. 833–40.
42. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks, *Science*. (2006) **313**:504–7. doi: 10.1126/science.1127647
43. Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the 25th International Conference on Machine Learning*. Helsinki (2008). p. 1096–103. doi: 10.1145/1390156.1390294
44. Alain G, Bengio Y. What regularized auto-encoders learn from the data-generating distribution. *J Machine Learn Res*. (2014) **15**:3563–93. Available online at: https://jmlr.org/papers/v15/alain14a.html
45. Kingma DP, Welling M. Auto-encoding variational bayes. *arXiv [Preprint] arXiv*. (2013). Available online at: https://arxiv.org/abs/1312.6114
46. Kingma DP, Welling M. An introduction to variational autoencoders. *Found Trends Mach Learn*. (2019) **12**:307–92. doi: 10.1561/2200000056
47. Sharpe WF. Capital asset prices: a theory of market equilibrium under conditions of risk. *J Fin*. (1964) **19**:425–42. doi: 10.1111/j.1540-6261.1964.tb02865.x
48. Rosenberg B, McKibben W. The prediction of systematic and specific risk in common stocks. *J Fin Quant Anal*. (1973) **8**:317–33. doi: 10.2307/2330027
49. Ross SA. The arbitrage theory of capital asset pricing. *J Econ Theory*. (1976) **13**:341–60. doi: 10.1016/0022-0531(76)90046-6
50. DeMiguel V, Garlappi L, Uppal R. Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *Rev Fin Stud*. (2009) **22**:1915–53. doi: 10.1093/rfs/hhm075

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhang, Liang, Lyu and Fang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 02 October 2020
doi: 10.3389/fphy.2020.551501

Does Crude Oil Market Efficiency Improve After the Lift of the U.S. Export Ban? Evidence From Time-Varying Hurst Exponent

*Yinghui Shao**

School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, China

On 18 December 2015, the 40-years old U.S. crude oil export ban was repealed. Since then U.S. crude and oil producers were allowed to reach the global market. In this paper, we study if the crude oil market efficiency increases after the lift of the export ban via the Centered Detrending Moving average Analysis (CDMA) and the Detrended Fluctuation Analysis (DFA). We examine the time-varying market inefficiency from 2011 to 2020 with different rolling windows. The results indicate that WTI becomes inefficient after the lift in medium-term. Though in short and long-term, there is evidence for the improvement of the degree of market efficiency. Generally, the WTI market presents mixed efficiency behavior at different time horizons. In 1-year window, the degree of efficiency on Brent decreased while there isn't enough evidence to conclude that Brent market efficiency increased or decreased in medium and long term. In this sense, the lift of the ban might have significant impact on WTI but not on Brent.

Keywords: oil prices, spot market, Hurst exponent, market efficiency, WTI, Brent, detrended fluctuation analysis, detrending moving average analysis

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Ladislav Kristoufek,
Institute of Information Theory and Automation (ASCR), Czechia
Ken Kiyono,
Osaka University, Japan

***Correspondence:**
Yinghui Shao
yhshao@suibe.edu.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *13 April 2020*
Accepted: *19 August 2020*
Published: *02 October 2020*

Citation:
Shao Y (2020) Does Crude Oil Market Efficiency Improve After the Lift of the U.S. Export Ban? Evidence From Time-Varying Hurst Exponent. Front. Phys. 8:551501. doi: 10.3389/fphy.2020.551501

1. INTRODUCTION

Crude oil is one of the most important commodity in global economy. And crude oil market is increasingly connected to other financial markets. The market efficiency on crude oil market is crucial to price discovery, forecasting and investment management. There is a large number of literature on crude oil market weak-form efficiency.

The presence of long-range dependence in asset returns implies potential predictability, which contradicts the weak form of efficient market [1]. Hurst exponent is a popular measure of long range memory and has been widely applied to the test of weak form efficiency on crude oil market. Alvarez-Ramirez et al. [2] impose Rescaled Range (R/S) analysis on WTI from 1981 to 2002 and report the long-run memory. Serletis and Andreadis [3] get a similar conclusion with data from 1990 to 2001. Jiang et al. [4] verify the weak-form efficiency of the crude oil futures market from 1983 to 2012 with the Detrended Fluctuation Analysis (DFA) and Detrending Moving average Analysis (DMA).

One can conclude that the crude oil market is weak-form efficient in the long run, though over short time horizons the market might need a certain time to digest the information [4]. Tabak and Cajueiro [5] document that WTI and Brent become more efficient from 1983 to 2004. Via DFA, Alvarez-Ramirez et al. [6] find mixed scaling behavior in crude oil markets. The prices exhibit short-term predictability, while for large time horizons reflect non-correlated behavior. Wang and

Liu [7] report that short-term, medium-term and long-term behaviors of WTI crude oil market were generally turning into efficient behavior over time. Mensi et al. [8] find efficiency degree of WTI and Brent varies through time. Jiang et al. [4] uncover that the market is inefficient during turbulent events, including the oil price crash in 1985 and the Gulf war. Gu and Zhang [9] verify the non-linear relationship between multifractality and inefficiency in crude oil market. Kristoufek [10] reports that the WTI remains efficient except several periods, while Brent crude oil shows more inefficiency.

Notice that the crude oil market is affected by economic conditions [11], political events [12], policy changes [13, 14] and other factors, dynamics of the efficiency on crude oil market during events and crisis is of cardinal significance to investors as it indicates possible abnormal returns. This study focuses on the time-varying efficiency on crude oil market before and after the policy shift.

In 1970s the United States government enacted prohibitions on the export of crude oil to handle concerns on possible energy shortage [15]. The ban restricted most crude oil exports from U.S. to other countries. On 18 December 2015, President Obama signed a provision that lifted the 40-years old oil export ban. After the lift, the lead-lag relationship between WTI and Brent has changed [14] and the two benchmarks reconnected closely [13, 14]. However, the impact of this policy shift on crude oil market efficiency is still open to question.

In this paper we explore whether the crude oil market efficiency changes after the lift using WTI Brent daily price. We assess the time-varying market efficiency via the Centered Detrending Moving average Analysis (CDMA) and DFA. To compare the difference of market efficiency before and after the lift, we separate the whole sample into two time periods by the lift, then implement the Wilcoxon signed-rank test. We investigate the time-varying long range dependence over rolling windows. For the sake of consistency, we carry out the whole procedure with three rolling window sizes.

The paper is organized as follows. Section 2 presents the Data. Section 3 describes the Methodology. Section 4 shows the Results, and section 5 Concludes and Discusses.

2. DATA

We investigate the daily closing spot prices from the U.S. Energy Information Administration (EIA) website. The data spans from 14 October 2011 to 6 March 2020 with 2098 observations, which is plotted in **Figure 1**. Note that the vertical black line marks the date of 18 December 2015 when the lift was officially announced. Thereby the whole period is divided into two segments before

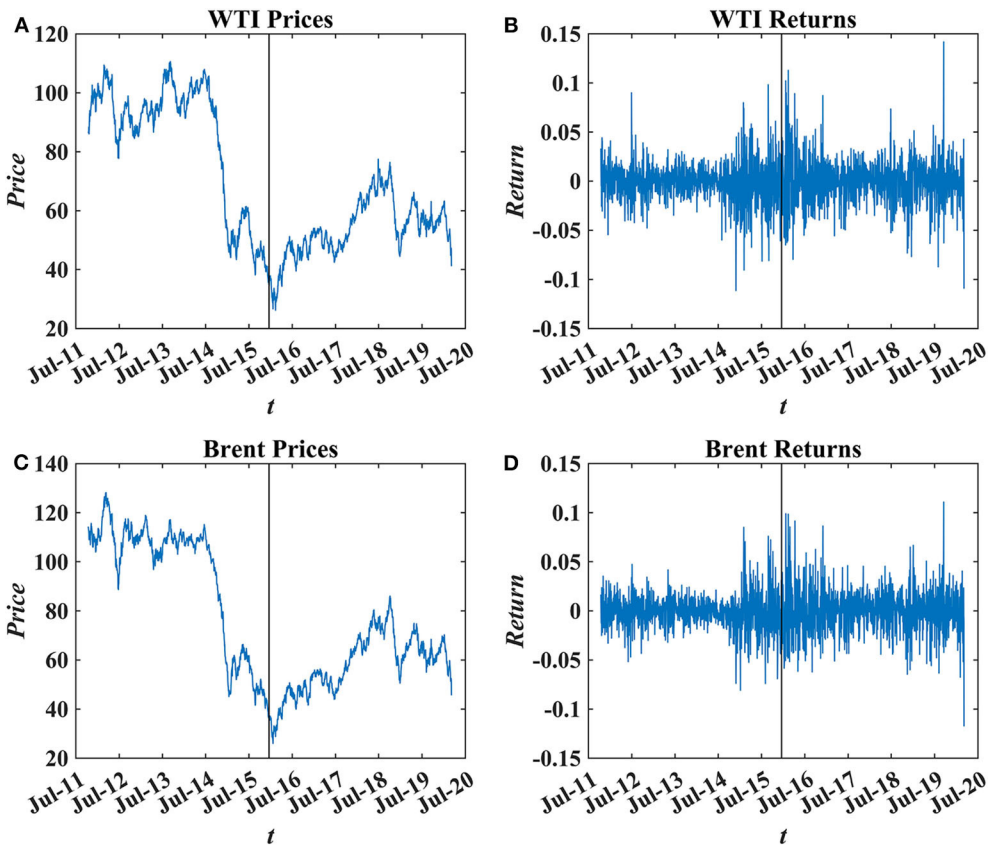


FIGURE 1 | The daily spot prices and returns for WTI and Brent. The vertical line marks the lift of the export ban. **(A)** WTI prices. **(B)** WTI returns. **(C)** Brent prices. **(D)** Brent returns.

and after the lift with the same 1049 sample sizes. The Brent and WTI price showed an impressive decrease of 60% from July 2014 to March 2015. Since 2016 until 2018 the price has steadily increased.

We employ daily returns for the analysis which are defined as the log difference in price:

$$r(t) = \ln p(t) - \ln p(t-1), \tag{1}$$

where $p(t)$ is the spot price at time t. The logarithm returns are illustrated in **Figure 1**. In 2014 and 2016 the process exhibits volatility clustering.

3. METHODOLOGY

3.1. Hurst Exponent Estimation Methods

We examine the market efficiency via the centered detrended moving average analysis (CDMA) and DFA. The two methodologies show comparable performance for the detection of long range correlation in time series [16–18], and have been widely used to determine the Hurst exponent in financial markets [19, 20].

3.1.1. The Detrended Fluctuation Analysis

The Detrended Fluctuation Analysis is a popular method to determine the long-range dependence in time series [21].

For a given time series $x(t)$, $t = 1, 2, \cdots, N$, compute the cumulative sum sequence

$$X(t) = \sum_{i=1}^{t} x(i), t = 1, 2, \cdots, N. \tag{2}$$

Detrend the integrated profile $X(t)$ by subtracting the local trend $\widetilde{X}(t)$ in each box of length s and get the residual series $\epsilon(t)$:

$$\epsilon(t) = X(t) - \widetilde{X}(t). \tag{3}$$

where $\widetilde{X}(t)$ is the polynomial fit of $X(t)$. Then separate the residual series $\epsilon(t)$ into N_s non-overlapped subseries with size s, where $N_s = \lfloor N/s - 1 \rfloor$. Obtain the fluctuation function via

$$F^2(s) = \frac{1}{N} \sum_{t=1}^{N} \left[\epsilon(t)\right]^2. \tag{4}$$

Repeat the procedure above with different box size s. Thus, one can determine the power-law relationship between $F(s)$ and box size s for long-range correlated time series

$$F(s) \sim s^H. \tag{5}$$

where H is an estimation of Hurst exponent. If $H > 0.5$, the time series $x(t)$ displays long-range dependent structure. When $H < 0.5$, $x(t)$ has long-range dependent structure. While $H = 0.5$ implies non-correlated behavior, which corresponds to weak-form efficiency in financial market.

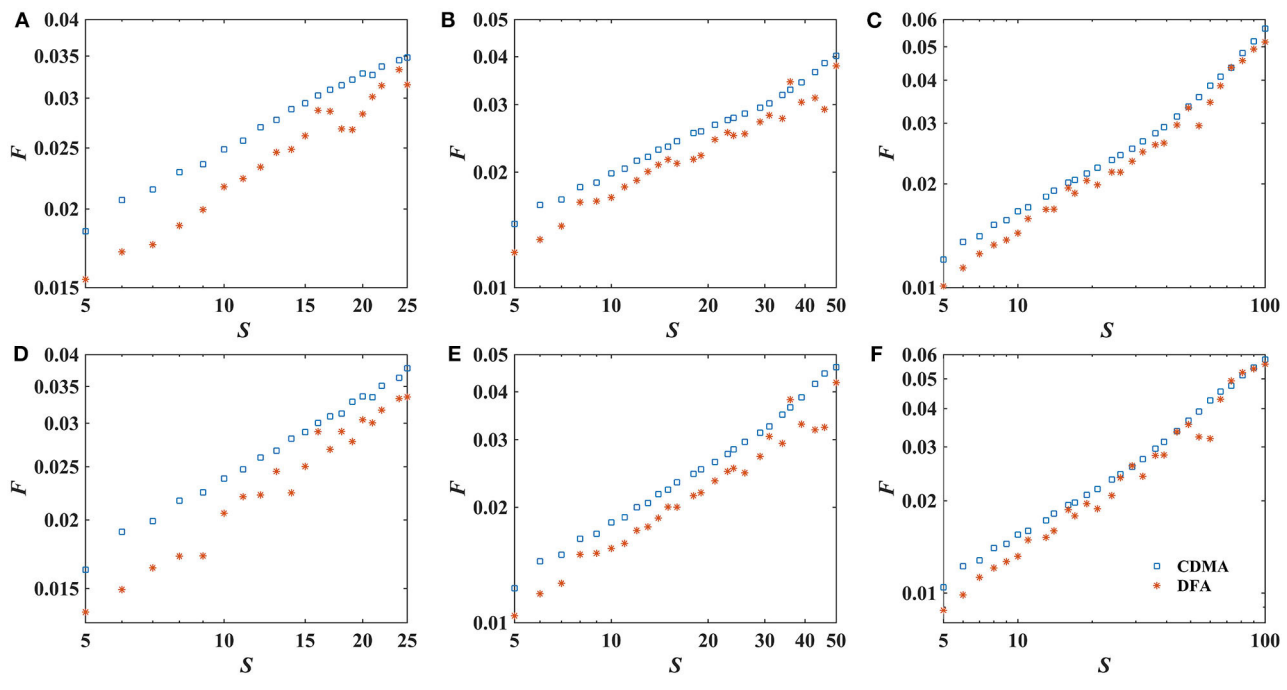


FIGURE 2 | Fluctuation function F with different rolling windows on 18 December 2015. **(A–C)** Depict the fluctuation function of WTI. **(D–F)** Illustrate the fluctuation function of Brent. From left to right, each column corresponds to 1, 2, and 4 years window.

3.1.2. The Detrending Moving Average Analysis

The DMA algorithm [22] is similar to DFA. Recently the understanding of DMA has been deepened and improved [23–27]. With DMA, one constructs the moving average function $\widetilde{X}(t)$ within a moving window with size s [28],

$$\widetilde{X}(t) = \frac{1}{s}\sum_{k=-s_2}^{s_1} X(t-k). \tag{6}$$

where $s = s_1 + s_2 + 1$, $s_1 = \lceil (s-1)(1-\theta) \rceil$, $s_2 = \lfloor (s-1)(1-\theta) \rfloor$ and θ is the position parameter with range of 0 to 1 (forward, centered, and backward moving average analysis for $\theta = 0$, $\theta = 0.5$, and $\theta = 1$, respectively).

We apply $\theta = 0.5$ (CDMA) in this study.

3.2. The Rolling Window Technique

The rolling window technique has been successfully used to assess time dependent efficiency in financial markets [7, 8, 29–33], predict crash in stock market [34], and evaluate the level of stability of financial firms [35].

The length of the subsample analyzed at each time is defined as the rolling window size. This technique works as follows [34]: (i) compute the Hurst exponent of the first subsample with a certain rolling window size. Thus, one can probe the local long range correlation at time $t = 1$. (ii) Move the time window with a specific number of observations, namely the step size. (iii) Repeat the process until end of whole sample. Thereby one can calculate the local Hurst exponent. Time series before a given time t contributes to the local Hurst exponent at time t [34].

4. RESULTS

We use CDMA and DFA to observe the long memory feature and time-varying inefficiency in WTI and Brent spot markets with rolling window technique. Rolling window size of DMA/DFA methods should be large enough to guarantee statistical significance and small enough to maintain the sensitivity to local changes [34–36]. Rolling window size under one trading year is recommended in case of possible seasonal periodicity and business cycles [34]. Another regular choice is 4-years window size [5, 7, 8, 10, 29], which corresponds to the political cycles in most countries [29]. Some studies apply an in-between size (2 years) [37, 38].

For the sake of consistency, we study market efficiency at different time horizons. We estimate the local Hurst exponent using three window sizes with a step size of one point: 250 datapoints (almost within 1 year), 500 datapoints (almost within 2 years), and 1,000 datapoints (almost within 4 years).

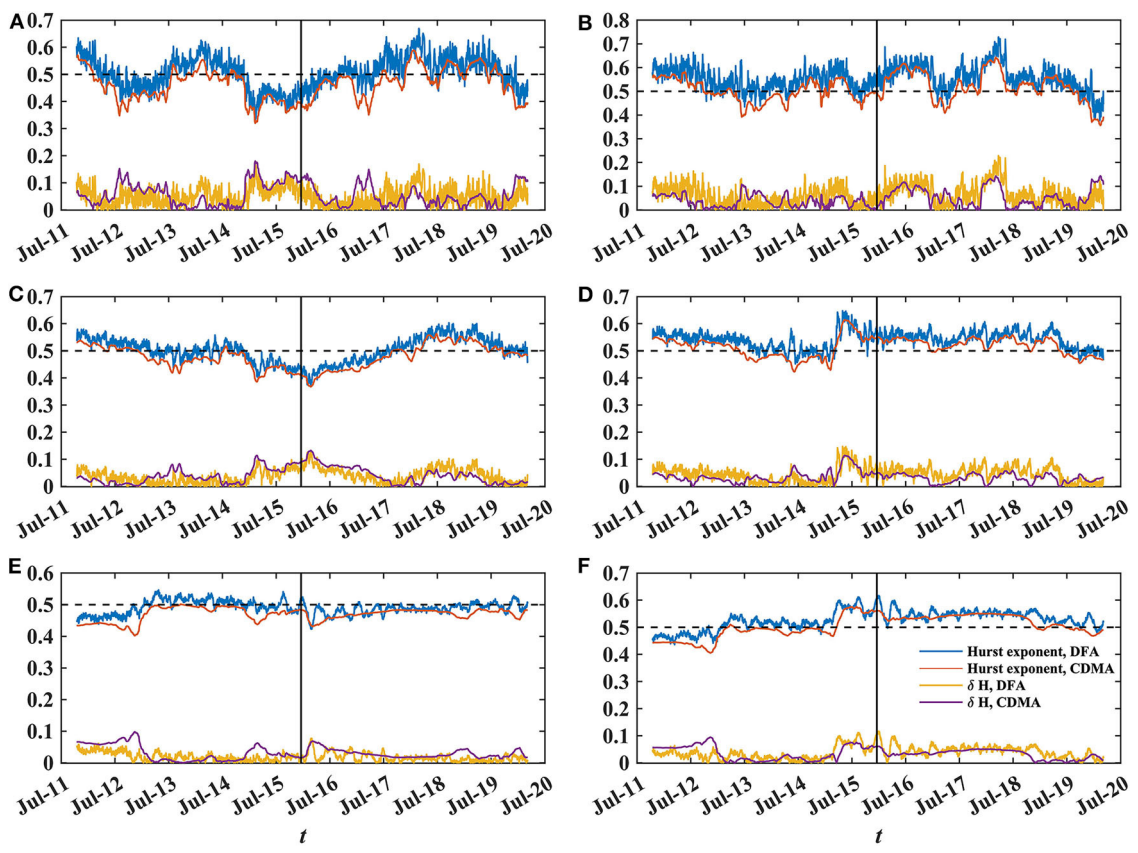


FIGURE 3 | Time-varying Hurst exponent and δH of WTI (left panel) and Brent (right panel). The black line marks the lift of the ban. **(A,B)** 1 year rolling window. **(C,D)** 2 years rolling window. **(E,F)** 4 years rolling window.

The choice of box size s is also essential for the detrending procedure. A too small or too big box size s can result in inaccurate estimate of local trend and curvature of the fluctuation function [16]. While box size $s \geqq 5$, the slope of log-log fluctuation plot can be almost approximated by a linear curve [25, 34]. Accordingly we set the range as $[5, L/10]$. **Figure 2** shows the fluctuation function F vs. s on 18 December 2015 with different rolling window sizes. Generally the CDMA curves seems quite straight. The DFA curves exhibit mild curvature.

There are various methods used to study market efficiency. Traditionally variance ratio test, unit root test as well as long memory estimation by the Hurst exponent are popular procedures [39]. Except those methods, researchers have developed different measurements based on fractality measures [40]. Kristoufek and Vosvrda introduced Efficiency Index to measure capital market efficiency [41], then extended and applied to commodity futures [42], gold, currencies [43], and cryptocurrencies [44]. Gu et al. [45] use $\delta H = |H - 0.5|$ to quantify the degree of market inefficiency, while H is the Hurst exponent. Smaller δH implies a more efficient market. In this paper, we apply δH to estimate market inefficiency, which is a simple and effective measurement.

The dynamics of local Hurst exponents and market inefficiency δH are shown in **Figure 3**. The black vertical line represents the lift of the U.S. crude oil export ban on 18 December 2015. The corresponding histograms are illustrated in **Figure 4**. Then we report the statistics of Hurst exponent and δH **Table 1**.

The local Hurst exponent moves around 0.5 and δH is generally <0.15. Both the mean and median of local δH is extremely close to 0. All these indicate that the crude oil markets are almost efficient with different degrees [2–5, 10].

4.1. Long Memory of WTI and Brent Within Rolling Windows

Here we give a quick look at the time-varying market efficiency of WTI and Brent. With rolling window of 1 year, the market displays higher fluctuation compared with longer time horizons (see in **Figures 3A,B**). The WTI returns shows short range correlation between 2012 and 2013 as well as 2014 and 2017. The Hurst exponent of Brent moves above 0.5. Both WTI and Brent display mixed behavior which reveals inefficiency of market at short time horizon.

In the case of 2-years rolling window, WTI demonstrates a mixed pattern of long and short-term correlation (see in **Figure 3C**). Local Hurst exponent is in the range of 0.35 to 0.55.

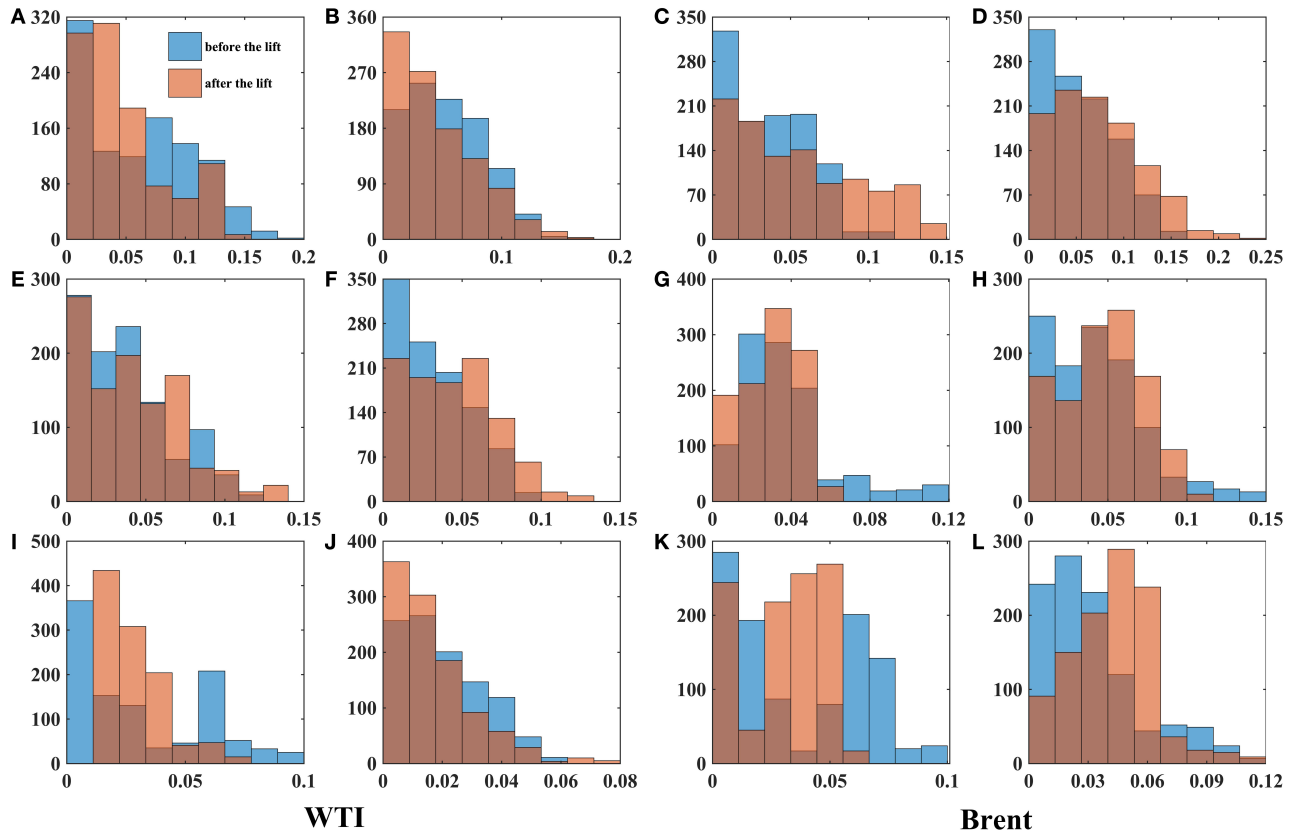


FIGURE 4 | Histograms of δH. The first and second columns correspond to results of WTI via CDMA and DFA respectively. The third and fourth columns correspond to the results of Brent via CDMA and DFA respectively. **(A–D)** Corresponds to a rolling window size of 1 year. **(E–H)** 2 years rolling window. **(I–L)** 4 years rolling windows.

TABLE 1 | Statistics for Hurst exponent and δH before and after the lift.

			1 year		2 years		4 years	
			Before	**After**	**Before**	**After**	**Before**	**After**
PANEL A: WTI								
CDMA	H	Mean	0.4500	0.4770	0.4710	0.4818	0.4681	0.4713
		Median	0.4444	0.4814	0.4679	0.4904	0.4774	0.4764
	δH	Mean	0.0617	0.0477	0.0379	0.0422	0.0319	0.0287
		Median	0.0605	0.0391	0.0330	0.0380	0.0226	0.0236
		p	0		0.9923		0.0804	
DFA	H	Mean	0.4926	0.5196	0.4966	0.5035	0.4950	0.4864
		Median	0.4945	0.5176	0.5003	0.5074	0.4986	0.4884
	δH	Mean	0.0535	0.0444	0.0317	0.0433	0.0201	0.0165
		Median	0.0511	0.0370	0.0280	0.0426	0.0178	0.0132
		p	0		1		0	
PANEL B: BRENT								
CDMA	H	Mean	0.5001	0.5228	0.5104	0.5205	0.4856	0.5261
		Median	0.4932	0.5256	0.5182	0.5288	0.4854	0.5340
	δH	Mean	0.0365	0.0550	0.0365	0.0299	0.0362	0.0314
		Median	0.0339	0.0471	0.0313	0.0314	0.0277	0.0340
		p	1		0		0.0002	
DFA	H	Mean	0.5480	0.5618	0.5371	0.5451	0.5131	0.5423
		Median	0.5473	0.5620	0.5403	0.5490	0.5129	0.5454
	δH	Mean	0.0530	0.0714	0.0428	0.0473	0.0324	0.0429
		Median	0.0480	0.0664	0.0406	0.0490	0.0267	0.0454
		p	1		1		1	

It decrease substantially from 2014 and gravitates around 0.55 since 2018. Obviously Brent returns display long-range memory (see in **Figure 3D**). The market is not weak-form efficient since it possesses long memory features.

With rolling window of 4 years, WTI exhibits extremely weak short range correlation until 2012 (see in **Figure 3E**), which is similar to Kristoufek's [10] work with longer step of 5 days. Since then our findings are slightly distinct. The cause might be the different step size. Basically the local Hurst exponent of WTI oscillates around 0.5, while Brent presents apparent long-range memory property (see in **Figure 3F**).

As is shown in **Table 1** and **Figure 4**, value of Hurst exponent and δH is more and more concentrated with window size enlarging. The inefficient behavior on crude oil markets at short time horizon are not surprising. The oil market is extremely fragile to economic and political factors as well as policy shift, which might cause the short time inefficiency. The smoother curve for the larger rolling window sizes implies the existence of weak-form efficiency in the crude oil market at longer time horizon. Over a long enough time, market participants could absorb the new information, resulting a weak-efficient market. Our findings is in line with work of Jiang et al. [4], Tabak and Cajueiro [5], Alvarez-Ramirez et al. [6], Wang and Liu [7], Gu and Zhang [9], and Kristoufek [10].

4.2. The Effects of the Lift

To investigate if the market efficiency dynamics change after the policy shift, we compare market inefficiency δH before and after the lift of export ban. Results in **Table 1** and **Figure 4** show a mild difference of δH between the two sub-periods.

Prior to hypothesis testing, we examine the distribution of local δH. Noticed that the histograms (see in **Figure 4**) show that the data are not normally distributed. Conducting the Lilliefors test for normality, we also reject the null hypothesis of normality at the 1% significance level. Hence commonly used paired t test isn't appropriate for the analysis of the data.

Here we implement the right-tailed Wilcoxon signed-rank test on δH to check if the market efficiency increased after the lift. The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test to analyze differences between two samples. Different from the paired t test, it does not require normal distribution of the data. The null hypothesis and alternative hypothesis are formulated as

H_0: The market inefficiency δH median is no greater than that after the lift.
H_1: δH median before the lift is greater than that after the lift.

If H_0 is rejected, we conclude that the market efficiency increased after the lift.

As is shown in **Table 1**, in the case of WTI for 1 year window the p value $= 0$ indicates a rejection of the null hypothesis at 5% significance level. There is enough statistical evidence to conclude that for 1-year rolling window size, the median market inefficiency δH before the lift is greater than that after the lift, which implies the increase of market efficiency. For 2-years window the result suggests that the null hypothesis cannot be rejected at 5% significance level, which means that the market efficiency of WTI didn't increase after the lift. However, at large time window the result is interesting. The result of DFA provides evidence that the market efficiency improved. For the result of CDMA, the $p-$value $=$ 0.0804 indicates that the null hypothesis cannot be rejected at 5% significance level though 10% significance level leads to opposite conclusion.

For Brent at the smallest time window, both CDMA and DFA results show that the null hypothesis cannot be rejected, which means that the market efficiency of Brent didn't improve after the lift. At larger time windows, via CDMA one concludes that the market efficiency increased after the policy shift while via DFA one obtains totally different results.

5. CONCLUSION

We studied whether or not the crude oil market moved toward efficiency after the lift of the U.S. export ban. We comprehensively examine the time-varying long-range dependence of WTI and Brent crude oil spot prices from 2011 to 2020 based on CDMA and DFA methods. Then we split the whole sample into two sub periods with the same sample sizes and carry out Wilcoxon signed rank test. To increase the reliability of the result, we estimate the time-varying Hurst exponents and market inefficiency with different rolling windows.

Our findings indicate that over long horizons the crude oil markets are close to weak form efficient with time-varying short-time inefficiency, which is consistent with other earlier studies [5–8, 10]. We also find a slightly lower degree of efficiency on Brent, which means a higher level of predictability.

The result of Wilcoxon signed rank test suggests that after the lift, degree of market efficiency of WTI increased in short term, while Brent market efficiency decreased. For the result of medium term, WTI efficiency declined after the lift. And within largest time windows, degree of efficiency of WTI is highly possible to improve after the lift. Consider that, investors with different or multiple time horizons (asset holding period) are suggested to adjust investment strategies accordingly. For arbitragers, there might exist arbitrage opportunity after the lift in medium time horizon.

However, CDMA and DFA get totally different results on Brent crude oil market in medium and long term. This might be caused by the difference between the two methods. Another possible explanation is that the change of Brent market efficiency after the lift is too small to be detected. From results of the two methods, we can hardly determine whether the market efficiency on Brent increased or not. The lift of the ban might only affect WTI in medium and long term.

In addition to the U.S. energy policy, the supply and demand situation and infrastructure issues also count for the WTI crude oil price and market efficiency [15, 46, 47]. With the hydraulic fracturing and horizontal drilling techniques, oil production from shale region in U.S. rose dramatically and led to the shale boom, while inadequate transportation infrastructures restricted the U.S. crude supply to the refining centers. Since mid-2013, new pipelines and rail transport eased this issue [48]. The effects of export ban, shale boom and transportation bottleneck on WTI market efficiency should be fully explored in future studies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.eia.gov.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 11805119).

ACKNOWLEDGMENTS

The author would like to thank the reviewers for their constructive comments.

REFERENCES

1. Fama EF. Efficient capital markets: a review of theory and empirical work. *J Finance.* (1970) **25**:383–417. doi: 10.2307/2325486
2. Alvarez-Ramirez J, Cisneros M, Ibarra-Valdez C, Soriano A. Multifractal Hurst analysis of crude oil prices. *Phys A Stat Mech Appl.* (2002) **313**:651–70. doi: 10.1016/S0378-4371(02)00985-8
3. Serletis A, Andreadis I. Random fractal structures in North American energy markets. *Energy Econ.* (2004) **26**:389–99. doi: 10.1016/j.eneco.2004.04.009
4. Jiang ZQ, Xie WJ, Zhou WX. Testing the weak-form efficiency of the WTI crude oil futures market. *Phys A Stat Mech Appl.* (2014) **405**:235–44. doi: 10.1016/j.physa.2014.02.042
5. Tabak BM, Cajueiro DO. Are the crude oil markets becoming weakly efficient over time? A test for time-varying long-range dependence in prices and volatility. *Energy Econ.* (2007) **29**:28–36. doi: 10.1016/j.eneco.2006.06.007
6. Alvarez-Ramirez J, Alvarez J, Rodriguez E. Short-term predictability of crude oil markets: a detrended fluctuation analysis approach. *Energy Econ.* (2008) **30**:2645–56. doi: 10.1016/j.eneco.2008.05.006
7. Wang Y, Liu L. Is WTI crude oil market becoming weakly efficient over time? New evidence from multiscale analysis based on detrended fluctuation analysis. *Energy Econ.* (2010) **32**:987–92. doi: 10.1016/j.eneco.2009.12.001
8. Mensi W, Beljid M, Managi S. Structural breaks and the time-varying levels of weak-form efficiency in crude oil markets: evidence from the Hurst exponent and Shannon entropy methods. *Int Econ.* (2014) **140**:89–106. doi: 10.1016/j.inteco.2014.10.001

9. Gu R, Zhang B. Is efficiency of crude oil market affected by multifractality? Evidence from the WTI crude oil market. *Energy Econ*. (2016) **53**:151–8. doi: 10.1016/j.eneco.2014.10.014
10. Kristoufek L. Are the crude oil markets really becoming more efficient over time? Some new evidence. *Energy Econ.* (2019) **82**:253–63. doi: 10.1016/j.eneco.2018.03.019
11. Basistha A, Kurov A. The impact of monetary policy surprises on energy prices. *J Futures Markets*. (2015) **35**:87–103. doi: 10.1002/fut.21639
12. Zhang X, Yu L, Wang S, Lai KK. Estimating the impact of extreme events on crude oil price: An EMD-based event analysis method. *Energy Econ*. (2009) **31**:768–78. doi: 10.1016/j.eneco.2009.04.003
13. Luong PV, Mizrach B, Otsubo Y. Location basis differentials in crude oil prices. *Energy J*. (2019) **40**:41–58. doi: 10.5547/01956574.40.SI2.pluo
14. Yang Y, Shao Y, Shao H, Song X. The time-dependent lead-lag relationship between WTI and Brent crude oil spot markets. *Front Phys*. (2020) **8**:132. doi: 10.3389/fphy.2020.00132
15. Colgan JD, Van de Graaf T. A crude reversal: the political economy of the United States crude oil export policy. *Energy Res Social Sci.* (2017) **24**:30–5. doi: 10.1016/j.erss.2016.12.012
16. Shao YH, Gu GF, Jiang ZQ, Zhou WX, Sornette D. Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series. *Sci Rep*. (2012) **2**:835. doi: 10.1038/srep00835
17. Shao YH, Gu GF, Jiang ZQ, Zhou WX. Effects of polynomial trends on detrending moving average analysis. *Fractals*. (2015) **23**:1550034. doi: 10.1142/S0218348X15500346
18. Kiyono K, Tsujimoto Y. Nonlinear filtering properties of detrended fluctuation analysis. *Phys A Stat Mech Appl*. (2016) **462**:807–15. doi: 10.1016/j.physa.2016.06.129
19. Carbone A, Castelli G, Stanley HE. Time-dependent Hurst exponent in financial time series. *Phys A Stat Mech Appl*. (2004) **344**:267–71. doi: 10.1016/j.physa.2004.06.130
20. Yang YH, Shao YH, Shao HL, Stanley HE. Revisiting the weak-form efficiency of the EUR/CHF exchange rate market: evidence from episodes of different Swiss franc regimes. *Phys A Stat Mech Appl*. (2019) **523**:734–46. doi: 10.1016/j.physa.2019.02.056
21. Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL. Mosaic organization of DNA nucleotides. *Phys Rev E*. (1994) **49**:1685. doi: 10.1103/PhysRevE.49.1685
22. Alessio E, Carbone A, Castelli G, Frappietro V. Second-order moving average and scaling of stochastic time series. *Eur Phys J B*. (2002) **27**:197–200. doi: 10.1140/epjb/e20020150
23. Carbone A, Kiyono K. Detrending moving average algorithm: frequency response and scaling performances. *Phys Rev E*. (2016) **93**:063309. doi: 10.1103/PhysRevE.93.063309
24. Kiyono K, Tsujimoto Y. Time and frequency domain characteristics of detrending-operation-based scaling analysis: exact DFA and DMA frequency responses. *Phys Rev E*. (2016) **94**:012111. doi: 10.1103/PhysRevE.94.012111
25. Kiyono K. Theory and applications of detrending-operation-based fractal-scaling analysis. In: *2017 International Conference on Noise and Fluctuations (ICNF)*. Vilnius: IEEE (2017). p. 1–4. doi: 10.1109/ICNF.2017.7985951
26. Höll M, Kiyono K, Kantz H. Theoretical foundation of detrending methods for fluctuation analysis such as detrended fluctuation analysis and detrending moving average. *Phys Rev E*. (2019) **99**:033305. doi: 10.1103/PhysRevE.99.033305
27. Nakata A, Kaneko M, Shigematsu T, Nakae S, Evans N, Taki C, et al. Generalized theory for detrending moving-average cross-correlation analysis: a practical guide. *Chaos Solit Fract*. (2019) **3**:100022. doi: 10.1016/j.csfx.2020.100022
28. Arianos S, Carbone A. Detrending moving average algorithm: a closed-form approximation of the scaling law. *Phys A Stat Mech Appl*. (2007) **382**:9–15. doi: 10.1016/j.physa.2007.02.074
29. Cajueiro DO, Tabak BM. The Hurst exponent over time: testing the assertion that emerging markets are becoming more efficient. *Phys A Stat Mech Appl*. (2004) **336**:521–37. doi: 10.1016/j.physa.2003.12.031
30. Alvarez-Ramirez J, Alvarez J, Rodriguez E, Fernandez-Anaya G. Time-varying Hurst exponent for US stock markets. *Phys A Stat Mech Appl*. (2008) **387**:6159–69. doi: 10.1016/j.physa.2008.06.056
31. Anagnostidis P, Varsakelis C, Emmanouilides CJ. Has the 2008 financial crisis affected stock market efficiency? The case of Eurozone. *Phys A Stat Mech Appl*. (2016) **447**:116–28. doi: 10.1016/j.physa.2015.12.017
32. Bariviera AF. The inefficiency of Bitcoin revisited: a dynamic approach. *Econ Lett*. (2017) **161**: 1–4. doi: 10.1016/j.econlet.2017.09.013
33. Alvarez-Ramirez J, Rodriguez E, Ibarra-Valdez C. Long-range correlations and asymmetry in the bitcoin market. *Phys A Stat Mech Appl*. (2018) **492**:948–55. doi: 10.1016/j.physa.2017.11.025
34. Grech D, Mazur Z. Can one make any crash prediction in finance using the local Hurst exponent idea? *Phys A Stat Mech Appl*. (2004) **336**:133–45. doi: 10.1016/j.physa.2004.01.018
35. Morales R, Di Matteo T, Gramatica R, Aste T. Dynamical generalized Hurst exponent as a tool to monitor unstable periods in financial time series. *Phys A Stat Mech Appl*. (2012) **391**:3180–9. doi: 10.1016/j.physa.2012.01.004
36. Sensoy A. Generalized Hurst exponent approach to efficiency in MENA markets. *Phys A Stat Mech Appl*. (2013) **392**:5019–26. doi: 10.1016/j.physa.2013.06.041
37. Wang Y, Wu C. Long memory in energy futures markets: further evidence. *Resour Policy*. (2012) **37**:261–72. doi: 10.1016/j.resourpol.2012.05.002
38. Jin X. Time-varying return-volatility relation in international stock markets. *Int Rev Econ Finance*. (2017) **51**:157–73. doi: 10.1016/j.iref.2017.05.015
39. Lim KP, Brooks R. The evolution of stock market efficiency over time: a survey of the empirical literature. *J Econ Surv*. (2011) **25**:69–108. doi: 10.1111/j.1467-6419.2009.00611.x
40. Jiang ZQ, Xie WJ, Zhou WX, Sornette D. Multifractal analysis of financial markets: a review. *Rep Prog Phys*. (2019) **82**:125901. doi: 10.1088/1361-6633/ab42fb
41. Kristoufek L, Vosvrda M. Measuring capital market efficiency: global and local correlations structure. *Phys A Stat Mech Appl*. (2013) **392**:184–93. doi: 10.1016/j.physa.2012.08.003
42. Kristoufek L, Vosvrda M. Commodity futures and market efficiency. *Energy Econ*. (2014) **42**:50–7. doi: 10.1016/j.eneco.2013.12.001
43. Kristoufek L, Vosvrda M. Gold, currencies and market efficiency. *Phys A Stat Mech Appl*. (2016) **449**:27–34. doi: 10.1016/j.physa.2015.12.075
44. Kristoufek L, Vosvrda M. Cryptocurrencies market efficiency ranking: not so straightforward. *Phys A Stat Mech Appl*. (2019) **531**:120853. doi: 10.1016/j.physa.2019.04.089
45. Gu R, Shao Y, Wang Q. Is the efficiency of stock market correlated with multifractality? An evidence from the Shanghai stock market. *Phys A Stat Mech Appl*. (2013) **392**:361–70. doi: 10.1016/j.physa.2012.09.008
46. Zhang B. Are the crude oil markets becoming more efficient over time? New evidence from a generalized spectral test. *Energy Econ*. (2013) **40**:875–81. doi: 10.1016/j.eneco.2013.10.012
47. Scheitrum DP, Carter CA, Revoredo-Giha C. WTI and Brent futures pricing structure. *Energy Econ*. (2018) **72:**462–9. doi: 10.1016/j.eneco.2018.04.039
48. Kilian L. The impact of the shale oil revolution on US oil and gasoline prices. *Rev Environ Econ Policy*. (2016) **10**:185–205. doi: 10.1093/reep/rew001

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Shao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 15 October 2020
doi: 10.3389/fphy.2020.00302

Initial Value Dependence of Urban Population's Growth-Rate Distribution and the Long-Term Growth

Atushi Ishikawa[1]*, *Shouji Fujimoto*[1], *Arturo Ramos*[2] and *Takayuki Mizuno*[3,4,5]

[1] Department of Economic Informatics, Kanazawa Gakuin University, Kanazawa, Japan , [2] Department of Economic Analysis, Universidad de Zaragoza, Zaragoza, Spain, [3] National Institute of Informatics, Tokyo, Japan, [4] The Graduate University for Advanced Studies [SOKENDAI], Tokyo, Japan, [5] Center for Advanced Research in Finance, The University of Tokyo, Tokyo, Japan

This paper uses census municipal population data for the United States, Italy, and Spain to analyze the statistical properties of their 10-year growth (short-term property). As a result, it was confirmed that the smaller the initial urban population is, the greater the probability that the urban population will decrease and that the probability that the urban population will increase does not depend on the initial urban population. We also observed the statistical properties of long-term growth of urban populations in each country over 100 years. Specifically, we identified the following properties by observing the geometric mean of logarithmically equal sized bins of the oldest urban population in the data used in the analysis. (1) The average urban population increases or decreases exponentially with time. (2) The smaller the initial average urban population, the smaller the exponent, which can be negative in Italy and Spain. (3) When the average urban population is large, exponential growth may stop. We showed that these long-term properties are derived from the short-term property by random sampling simulations from real data.

Keywords: urban population, city size distribution, growth-rate distribution, Gibrat's law, non-Gibrat's property, short-term growth, long-term growth

OPEN ACCESS

Edited by:
Hui-Jia Li,
Beijing University of Posts and Telecommunications (BUPT), China

Reviewed by:
Gui-Quan Sun,
North University of China, China
Keke Huang,
Central South University, China
Chengyi Xia,
Tianjin University of Technology, China

***Correspondence:**
Atushi Ishikawa
ishikawa@kanazawa-gu.ac.jp

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *12 April 2020*
Accepted: *30 June 2020*
Published: *15 October 2020*

Citation:
Ishikawa A, Fujimoto S, Ramos A and Mizuno T (2020) Initial Value Dependence of Urban Population's Growth-Rate Distribution and the Long-Term Growth. Front. Phys. 8:302. doi: 10.3389/fphy.2020.00302

1. INTRODUCTION

There are various universal structures in nature. In physics, various universal structures have been extracted from nature and described logically and mathematically, such as Newtonian mechanics in the seventeenth century, electromagnetism, relativity, and thermodynamics in the nineteenth century, and quantum mechanics in the twentieth century. We have deepened our understanding of nature by combining these. Interestingly, society also has a universal structure. Economics, of which we were able to collect data at a relatively early stage, paved the way for research in this area. The study of economics using methodology in physics (now called econophysics [1]) started with Pareto's discovery that the income distribution in the UK follows a power law [2–4]. The power-law distribution is also called Pareto distribution (so named after the discoverer). Later, as the accumulation of data progressed, it was found that power laws were observed for various economic sizes [5, 6]. Typical examples are sales, profits, the number of employees (they are referred as firm sizes), and the size of cities. Here, the city size mainly refers to the population of the city. In addition, the word city means all units, including municipalities.

A number of studies have shown that the size of these firms and cities also has a power law and that the distribution is restricted to the large-scale range. It is commonly recognized that the mid-scale range of firms and cities follows the lognormal (LN) distribution. Furthermore, it has been reported that a small-scale range of firm size distribution is described by Weibull distribution [7, 8]. For urban size distribution, various functions have been studied using statistical indicators such as LN, the double Pareto-lognormal distribution (dPLN), log-logistic, the threshold double Pareto Singh-Maddala, lognormal-upper Pareto, Pareto tails - lognormal, three log-normal (3LN), Pareto tails log-normal (PTLN), and threshold double Pareto Generalized Beta of second kind (tdPGB2). Relevant references to this research includes [9–24].

At the same time, studies on the growth rate of firms and cities have been carried out, and various data have confirmed that Gibrat's law is valid in large-scale ranges. Gibrat's law states that the size growth rate for two consecutive years is independent of the initial value [25, 26]. In [27], a simple unified model is proposed to explain the growth dynamics of cities and scaling laws, where the model predicts that the size of cities grows linearly regardless of its current size.

In a previous study, Ishikawa et al. found that the size growth-rate distribution of firms in a mid-scale range changes regularly depending on the initial value, and they called this non-Gibrat's property [28]. Specifically, in the case of firm sales, for example, the negative growth rate does not depend on the initial value, as in the case of Gibrat's law, but the positive growth rate increases as the initial value decreases [29]. Our previous studies have also shown that short-term properties of firm size lead to the long-term properties. In particular, we show that the size of newly established firms grows rapidly over time, according to the non-Gibrat property, and then shifts to a gradual exponential growth according to the Gibrat law using numerical simulations. Furthermore, this was confirmed by the observation of geometric mean of firm sales and number of employees [30, 31].

As mentioned earlier, there are many similarities between the study of firm size distribution and the study of urban size distribution. In this paper, we discuss the relationship between the short- and long-term properties of urban size based on our previous research on the short- and long-term properties of firm size.

The structure of this paper is as follows. First, in section 2, we describe the urban population data of the United States, Italy, and Spain that we analyzed in this paper. We also describe the firms' data that we used to review our previous work in section 3. In section 3, we briefly review the properties that we previously found regarding the initial dependence of a firm's sales growth rate and the long-term growth properties derived from them. In section 4, employing the data described in section 2, we describe the initial dependence of growth rate distributions on urban populations in the United States, Italy, and Spain and the properties observed in long-term growth. In section 5, the process of growth of cities with different population sizes is simulated using the growth rate distribution of urban populations sampled from real data, and it is confirmed that the properties observed in the long-term growth observed in section 4 are reproduced. Finally, section 6 summarizes this study through the interpretation of the simulation results in section 5 and describes the future prospects.

2. DATA

This paper uses census data for the United States, Italy, and Spain. Data for the United States are population data for cities, towns, villages, and Census-designed Places (CDP) from 1900 to 2010 at 10-year intervals. The number of cities, the number of people living in them, and the proportion of the population included in the data to the total population of the United States are shown in **Table 1**. In this paper, we will collectively refer to municipalities and the CDP as cities. The data for Italy are municipal population data from 1901 to 2011 at intervals of approximately 10 years (**Table 2**). The data for Spain are population data of municipalities from 1900 to 2010 at intervals of approximately 10 years (**Table 3**). Census data for Italy and Spain include nearly all citizens. Census data in the United States, on the other hand, have gradually increased in coverage from 47.0% in 1900 to 71.6% in 2010. In addition, the population of Italy and Spain is classified into one of the administrative divisions, while that of the United States is classified into CDP in addition to the municipality. There is therefore a significant difference in the completeness of data aggregation between Italy/Spain and the United States, which may be attributed to the fact that the United States is a relatively young country.

In addition, in order to compare the initial value dependence of the growth rate distribution of the urban population, which is the main focus of this paper, we introduce the initial value dependence of the growth rate distribution of firms' sales, which is our previous study, in section 3. The analysis will use data from the Orbis 2016 edition, the world's largest corporate information database, provided by Bureau van Dijk. Specifically, we use the sales and establishment year data of 944,116 Japanese firms for

TABLE 1 | Summary of U.S. population data.

Year	Number of cities	Urban population	Rate
1900	10,597	35,811,876	47.0%
1910	14,130	50,631,562	54.9%
1920	15,530	62,153,225	58.6%
1930	16,721	77,233,423	62.3%
1940	17,313	84,252,429	63.8%
1950	18,921	101,062,173	66.8%
1960	20,002	125,867,952	70.2%
1970	20,948	144,768,910	71.2%
1980	22,262	164,029,093	72.4%
1990	23,434	182,532,293	73.4%
2000	25,375	208,735,266	74.2%
2010	24,685	221,886,645	71.6%

Cities includes municipalities and Census-designed Place (CDP). Number of Cities is the number. Urban Population is the number of people living there. Rate is the ratio of the population in the data to the total U.S. population at the time.

which both 2013 and 2014 sales data exist. The most recent data in the 2016 edition of Orbis is from 2015, but since the database was in the process of being collected at the time it was available, and so was scarce (343, 473 firms), we used the largest data pair, 2014 (1, 029, 179 firms) and 2013 (1, 133, 993 firms), for our analysis. There were approximately 3.82 million firms in Japan in 2014, of which approximately 11, 000 were reported to be large firms, approximately 557, 000 to be medium-sized firms, and approximately 3, 252, 000 to be small-sized firms, according to the Small and Medium Business Administration of Japan. In Japan, small firms are defined by the number of employees under the Small and Medium-sized Enterprise Basic Law, which refers to firms with five employees or fewer in the retail, service, and wholesale industries and 20 or fewer in the manufacturing and other industries. Orbis is a comprehensive database of large and medium-sized Japanese firms, including small-sized firms whose sales are identifiable.

TABLE 2 | Summary of population data for Italy.

Year	Number of cities	Urban population	Rate
1901	8,100	32,963,316	100.0%
1911	8,100	35,841,563	100.0%
1921	8,100	39,396,757	100.0%
1931	8,100	41,043,489	100.0%
1936	8,100	42,398,489	100.0%
1951	8,100	47,515,537	100.0%
1961	8,100	50,623,569	100.0%
1971	8,100	54,136,547	100.0%
1981	8,100	56,556,911	100.0%
1991	8,100	56,778,031	100.0%
2001	8,100	56,871,757	100.0%
2011	8,081	60,429,103	100.0%

Cities includes municipalities. Number of Cities is the number. Urban Population is the number of people living there. Rate is the ratio of the population in the data to the total population of Italy at the time.

TABLE 3 | Summary of population data for Spain.

Year	Number of cities	Urban population	Rate
1900	8,077	17,802,721	95.7%
1910	8,077	19,140,404	96.1%
1920	8,077	20,482,448	96.1%
1930	8,077	22,775,879	96.6%
1940	8,077	25,114,397	97.0%
1950	8,077	27,494,367	97.9%
1960	8,077	30,071,527	98.9%
1970	8,077	33,741,276	99.8%
1981	8,077	37,771,008	99.4%
1991	8,077	39,434,102	100.0%
2001	8,077	40,703,018	99.8%
2010	8,074	46,853,613	100.0%

Cities includes municipalities. Number of Cities is the number. Urban Population is the number of people living there. Rate is the ratio of the population in the data to the total population of Spain at the time.

3. SHORT- AND LONG-TERM PROPERTIES OF FIRMS

In this section, we briefly review our previous study of the initial dependence of the firms' sales growth rate distribution [29] and its long-term growth properties [30, 31] using new data we have. As mentioned in the previous section, our database contains sales data for 2013 and 2014 for 944, 116 Japanese firms. The minimum sales for 2013 was 1, 000 USD, and the maximum was 249, 799, 825 USD. To observe the initial value dependence of the growth rate ($R = x_{2014}/x_{2013}$) from 2013 sales (x_{2013}) to 2014 sales (x_{2014}), the initial values are divided into logarithmically equally sized bins as follows $x_{2013} \in [10^{0.5(n-1)}, 10^{0.5n})$ ($n = 1, 2, \cdots, 10$). The number of firms in these 10 bins ($n = 1, 2, \cdots, 10$) is 4, 355, 6, 426, 18, 354, 65, 866, 163, 052, 266, 023, 221, 607, 114, 967, 50, 459, and 20, 498, respectively. The number of firms greater than $n = 10$ is 12, 509. Although there are arbitrary ways to divide these bins, if a somewhat smooth growth-rate distribution is observed, the data are divided as finely as possible. When the bin is divided more finely than this way, a smooth growth-rate distribution is not observed, and when the bin is divided more roughly, the properties described below are hardly observed. However, it is confirmed by considering the case where the division of the bins is slightly changed from that described above, that the properties described below do not depend on the manner in which the bins are divided.

Figures 1, **2** are their conditional probability density functions. The horizontal axis represents the logarithmic growth rate $r = \log_{10} R$, and the vertical axis represents the probability density function (PDF) $q(r|n)$. **Figure 1** shows that the smaller the initial value x_{2013} (the smaller n) is, the larger

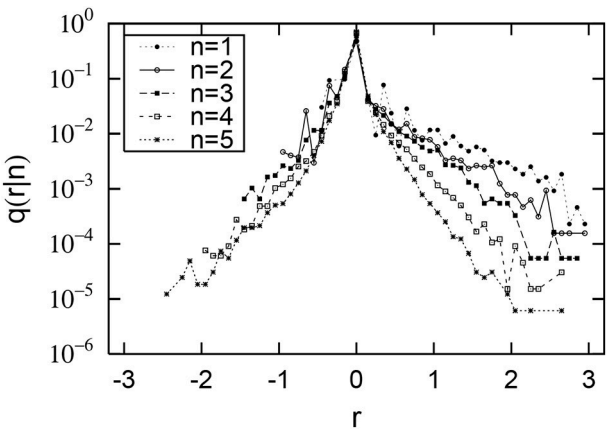


FIGURE 1 | Distributions of sales growth rate of Japanese firms. The horizontal axis shows the logarithmic growth rate: $r = \log_{10} \frac{x_{2014}}{x_{2013}}$. The vertical axis shows its conditional probability density function: $q(r|n)$. The initial sales value of x_{2013} for 2013 is divided into logarithmically equal sized bins: $x_{2013} \in [10^{0.5(n-1)}, 10^{0.5n})$ ($n = 1, 2, \cdots, 5$).

the positive growth rate, and that the negative growth rate is almost independent of the initial value. **Figure 2** also shows that when the initial value x_{2013} is $10^{2.5}$ thousand dollars or more (n is greater than or equal to 6), both the positive and negative growth-rate distributions are almost independent of the initial value. This property is called Gibrat's law [25, 26] and is observed in the large-scale range (Now x_{2013} is over $10^{2.5}$ thousand dollars). On the other hand, our previous studies have shown that the non-Gibrat's property in **Figure 1** is observed in the mid-scale range [28, 32, 33].

In our previous study, we also reported that there is a long-term growth property in the dependence of the geometric mean of firms' sales regarding their age [30, 31]. **Figure 3** depicts the firm age (t) dependency of the geometric mean sales ($\langle x_t \rangle$) in 2014. Here, the firm age is defined as the year of establishment of the firm, e.g., t =1. **Figure 3**, drawn on a semilogarithmic axis, shows that for the first 10 years, $\langle x_t \rangle$ rapidly grows following a power-law function:

$$\langle x_t \rangle \propto t^{\alpha_{\text{sales}} t} \tag{1}$$

with $\alpha_{\text{sales}} = 0.53 \pm 0.02$ (the dotted line). Then, $\langle x_t \rangle$ gradually grows following an exponential function:

$$\langle x_t \rangle \propto e^{\beta_{\text{sales}} t} \tag{2}$$

with $\beta_{\text{sales}} = 0.014 \pm 0.001$ (the solid line). These values are measured by applying $t = 1, 2, \cdots, 10$ for power-law growth (1) and $t = 11, 12, \cdots, 60$ for exponential growth (2).

Similar properties were confirmed not only in sales of Japanese firms but also in the number of employees of Japanese firms and sales and number of employees of French firms in previous studies [31]. Furthermore, we show by numerical simulation that these long-term growth properties are derived from the short-term growth properties mentioned above. The initial fast growth following the power-law function (1) is derived from the non-Gibrat's property, and the slow growth following the exponential function (2) is derived from Gibrat's law, using a stochastic process with numerical samples from real data.

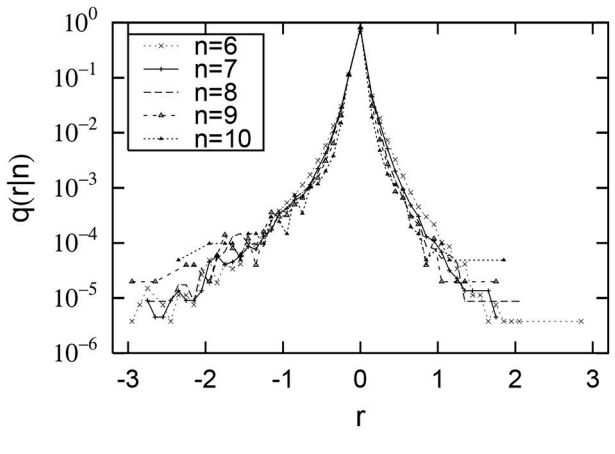


FIGURE 2 | Distributions of sales growth rate of Japanese firms. The horizontal axis shows the logarithmic growth rate: $r = \log_{10} \frac{x_{2014}}{x_{2013}}$. The vertical axis shows its conditional probability density function: $q(r|n)$. The initial sales value of x_{2013} for 2013 is divided into logarithmically equal sized bins: $x_{2013} \in [10^{0.5(n-1)}, 10^{0.5n})$ $(n = 6, 7, \cdots, 10)$.

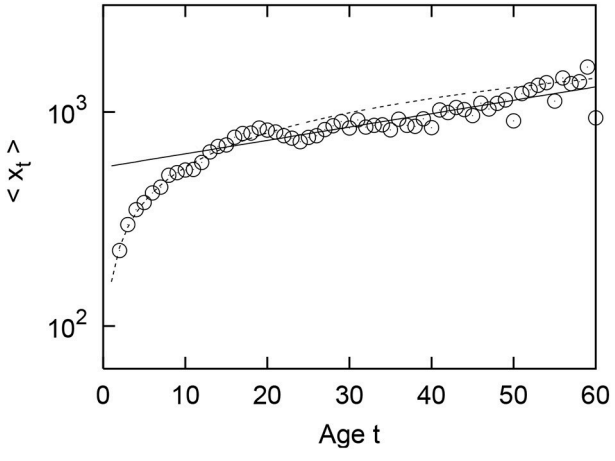


FIGURE 3 | Dependence of geometric mean of Japanese firms' sales $\langle x_t \rangle$ in 2014 on firm age t.

4. SHORT- AND LONG-TERM PROPERTIES OF URBAN POPULATION

The purpose of this study is to confirm whether the properties of short-term and long-term growth in firms' sales and the number of employees in our previous study are observed in an urban population. **Tables 1–3** show that the number of cities in the United States is around 10, 000 to 25, 000, and those of Italy and Spain number around 8, 000. As in the previous section, if these initial values are placed in logarithmically equal-sized bins and a somewhat smooth growth-rate distribution is to be observed, there will be only a few bins, and it will be difficult to observe the properties of the previous section. In the United States, we have increased the number of pairs of data by overlaying all 11 pairs of data, such as 1900 − 1910, 1910 − 1920, and $\cdots$, 2000 − 2010, and conducted the same analysis as in the previous section. This approach ignores changes in history over 110 years but has the advantage of being able to observe a macroscopic nature that is not influenced by the flow of history.

Using this method, pairs of 187,378 cities are created, each city's initial population is expressed as x_i, and the city's population 10 years later is expressed as x_{i+10}. The minimum value of x_i is 1 and the maximum value is 8, 008, 278. In the case of the United States, this small figure is recorded because it includes population data from Census-designed Places (CDP). To observe the dependence of the urban growth rate ($R = x_{i+10}/x_i$) on the initial value (x_i), we divide the initial value into logarithmically equidistant bins, as in the previous section, such as $x_i \in [10^{0.5(n-1)+2}, 10^{0.5n+2})$ $(n = 1, 2, \cdots, 7)$. The 7 bins contain the following cities: 34, 158 ($n = 1$), 56, 216 ($n = 2$), 41, 353 ($n = 3$), 22, 942 ($n = 4$), 10, 794 ($n = 5$), 3, 997 ($n = 6$), and 950 ($n = 7$). Number of cities less than $n = 1$ is 6, 278,

and number of cities greater than $n = 7$ is 361. The same discussion as in the previous section applies to the arbitrariness of division into bins. **Figure 4** shows the conditional PDF. The horizontal and vertical axes are the same as in the previous section. **Figure 4** shows that the smaller the initial value x_i is (The smaller n is), the larger the negative growth rate is, and the positive growth rate is almost independent of the initial value. Surprisingly, this property is symmetrical with that observed in **Figure 1**. This property is explicitly confirmed in **Figure 5**, which shows the n-dependence of positive standard deviation ($\sigma_+ = \sqrt{\Sigma_{r>0} r^2 q(r|n)}$) and negative one ($\sigma_- = \sqrt{\Sigma_{r<0} r^2 q(r|n)}$). **Figure 5** shows that σ_+ is almost n independent, but σ_- decreases depending on n, as approximated by

$$\sigma_- \propto e^{-\gamma^{\mathrm{US}} n} \tag{3}$$

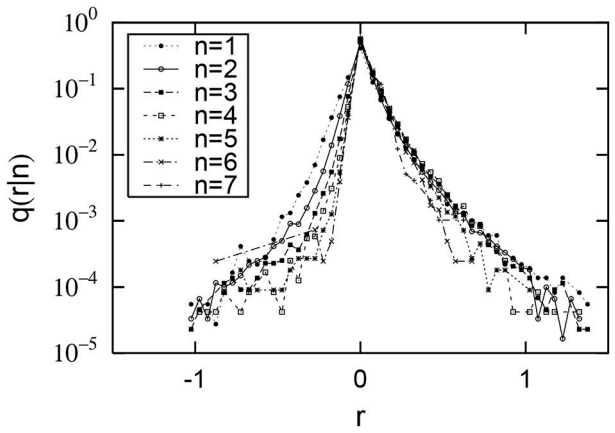


FIGURE 4 | Distributions of urban population growth rates in the United States. The horizontal axis shows the logarithmic growth rate: $r = \log_{10} \frac{x_{i+10}}{x_i}$. The vertical axis shows its conditional probability density function: $q(r|n)$. The initial population x_i is divided into logarithmically equal sized bins: $x_i \in [10^{0.5(n-1)+2}, 10^{0.5n+2})$ $(n = 1, 2, \cdots, 7)$.

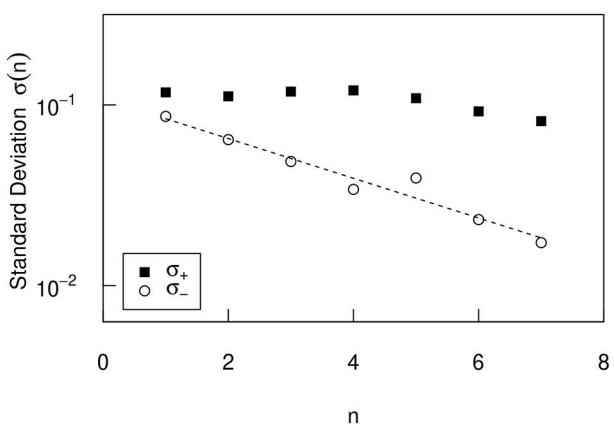


FIGURE 5 | The n dependence of the positive and negative standard deviations $\sigma_\pm$ of the conditional PDFs in the United States (**Figure 4**).

with $\gamma^{\mathrm{US}} = 0.25 \pm 0.03$.

In the previous section, we observed the long-term property of the growth of the geometric mean of firms' sales depending on the firm age. With a slightly different perspective on urban population data, we observe the long-term property of the growth of the geometric mean of the urban population over the observable years by size of the initial population. **Figure 6** shows the observed results for the urban population in the United States. The horizontal axis represents $1900, 1910, \cdots, 2010$ expressed as $T = 1, 2, \cdots, 12$, respectively, and the vertical axis represents the geometric mean of the urban population in each year ($\langle x_T \rangle$). In **Figure 6**, the original population of x_1 in 1900 ($T = 1$) is divided into logarithmically equally sized bins: $x_1 \in [10^{m+1}, 10^{m+2})$ $(m = 1, 2, 3)$, $x_1 \geq 10^5$ $(m = 4)$. The number of cities in each bin is 4,593 $(m = 1)$, 2,482 $(m = 2)$, 317 $(m = 3)$, and 28 $(m = 4)$. The number of cities less than $m = 1$ is 46. These totals are fewer than the 10,597 cities in 1900 because of the high frequency of urban renewal in the United States and the large number of cities that existed in 1900 but did not exist 110 years later. In **Figure 6**, which is the semilogarithmic axis, the optimum line is drawn by the least squares method for $T = 1, 2, 3, 4$. **Figure 6** shows that for $m = 1, 2$ the whole period is approximated by an exponential function:

$$\langle x_T \rangle \propto e^{\beta_{\mathrm{city}}(m) T}, \tag{4}$$

where $\beta_{\mathrm{city}}(m)$ is the index of urban growth corresponding to β_{sales} in Equation (2). Here, $\beta_{\mathrm{city}}(m)$ has a factor m because it varies depending on the bin containing the initial value x_1. On the other hand, at $m = 3, 4$, the exponential function is followed up to $T = 4$, after which the growth is negligible. Here, the exponents in U.S. $\beta^{\mathrm{US}}_{\mathrm{city}}(m)$ of the four straight lines (4) for $m = 1, 2, 3, 4$ in **Figure 6** are evaluated as $\beta^{\mathrm{US}}_{\mathrm{city}}(1) =$

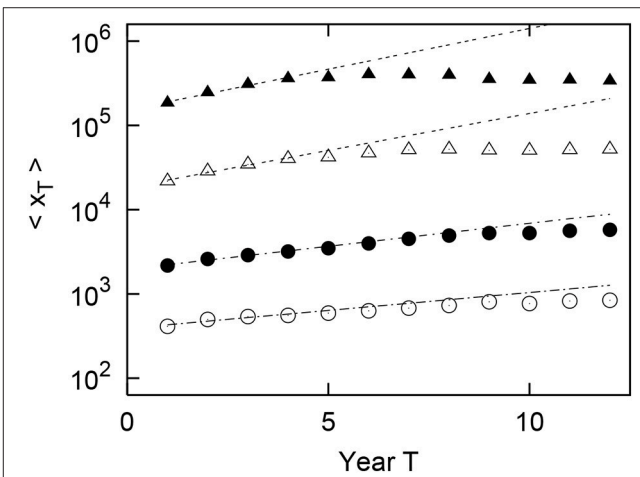


FIGURE 6 | Geometric mean growth of the urban population in the United States from 1900 to 2010 ($T = 1, 2, \cdots, 12$). The original population of x_1 in 1900 ($T = 1$) is divided into logarithmically equally sized bins: $x_1 \in [10^{m+1}, 10^{m+2})$ $(m = 1, 2, 3)$, $x_1 \geq 10^5$ $(m = 4)$. In the figure, ○, ●, △, and ▲each represent $m = 1, 2, 3$, and 4, respectively.

0.098 ± 0.026, $\beta^{\rm US}_{\rm city}(2) = 0.13 \pm 0.01$, $\beta^{\rm US}_{\rm city}(3) = 0.20 \pm 0.02$, and $\beta^{\rm US}_{\rm city}(4) = 0.22 \pm 0.02$, respectively. These values are measured by applying $T = 1, 2, 3, 4$ for exponential growth (4). These indicate that $\beta^{\rm US}_{\rm city}(m)$ is an increasing function of m. In other words, the geometric mean growth of the urban population will be faster as the initial urban population size increases.

Similar analyses were conducted on urban population data in Italy and Spain. As with the initial value dependence of the urban population growth-rate distribution in the United States, the negative growth rate increases as the initial value x_i decreases, and the positive growth rate hardly depends on the initial value. These properties are expressed as positive and negative standard deviations $\sigma_\pm$ in **Figures 7**, **8**. In **Figures 7**, **8**, as in the U.S., σ_+ is almost independent of n, and σ_- has the same n dependency as in Equation (3). The parameters for Italy and Spain are evaluated as $\gamma^{\rm IT} = 0.26 \pm 0.01$ and $\gamma^{\rm ES} = 0.30 \pm 0.06$, respectively. In Italy, the number of (x_i, x_{i+10}) pair is 88, 303, and the seven bins contain the following cities: 2, 008 ($n = 1$), 14, 721 ($n = 2$), 36, 913 ($n = 3$), 26, 002 ($n = 4$), 6, 781 ($n = 5$), 1, 453 ($n = 6$), and 241 ($n = 7$). Number of cities less than $n = 1$ is 948, number of cities greater than $n = 7$ is 0. In Spain, the number of (x_i, x_{i+10}) pair is 87, 141, and the seven bins contain the following cities: 13, 979 ($n = 1$), 30, 333 ($n = 2$), 23, 407 ($n = 3$), 11, 980 ($n = 4$), 3, 496 ($n = 5$), 720 ($n = 6$), and 244 ($n = 7$). Number of cities less than $n = 1$ is 2, 924, number of cities greater than $n = 7$ is 302.

As in the United States, **Figures 9**, **10** shows the long-term properties of the geometric mean of the urban population growing with the passage of observable years by size of the

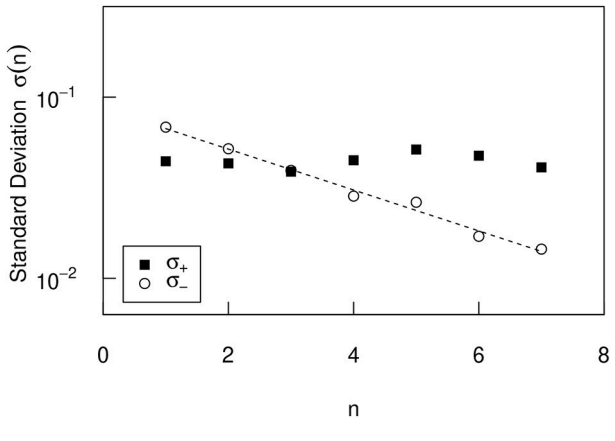


FIGURE 7 | The n dependence of the positive and negative standard deviations $\sigma_\pm$ of the conditional PDFs in Italy.

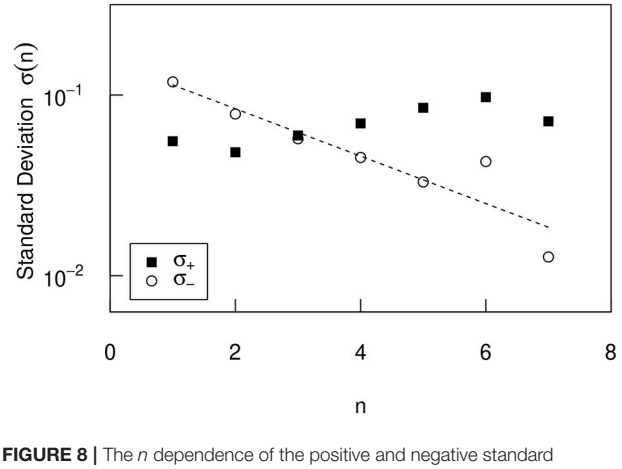


FIGURE 8 | The n dependence of the positive and negative standard deviations $\sigma_\pm$ of the conditional PDFs in Spain.

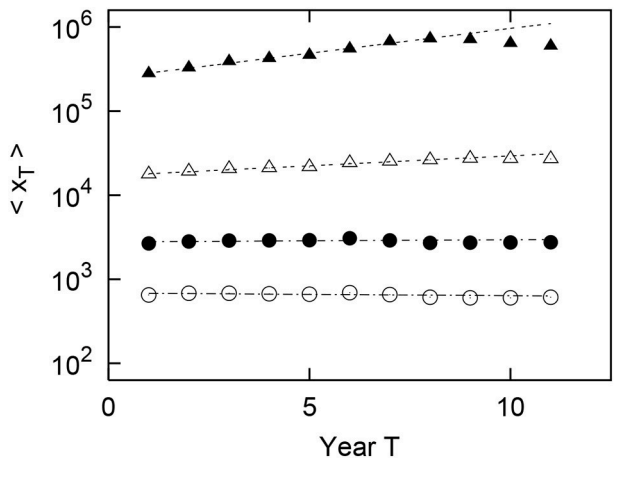


FIGURE 9 | Geometric mean growth of the urban population in Italy from 1901 to 2001 ($T = 1, 2, \cdots, 11$). The original population of x_1 in 1901 ($T = 1$) is divided into logarithmically equally sized bins: $x_1 \in [10^{m+1}, 10^{m+2})$ ($m = 1, 2, 3$), $x_1 \geq 10^5$ ($m = 4$). In the figure, ○, ●, △, and ▲each represent $m = 1, 2, 3$, and 4, respectively.

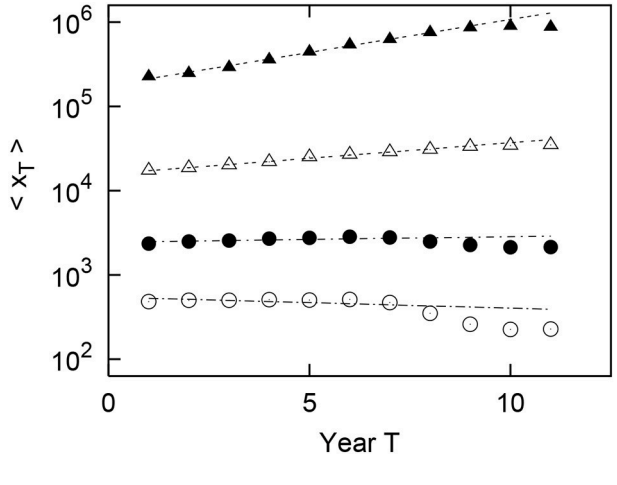


FIGURE 10 | Geometric mean growth of the urban population in Spain from 1900 to 2001 ($T = 1, 2, \cdots, 11$). The original population of x_1 in 1900 ($T = 1$) is divided into logarithmically equally sized bins: $x_1 \in [10^{m+1}, 10^{m+2})$ ($m = 1, 2, 3$), $x_1 \geq 10^5$ ($m = 4$). In the figure, ○, ●, △, and ▲each represent $m = 1, 2, 3$, and 4, respectively.

original population in Italy and Spain. For Italy, each bin contains 1,200 ($m = 1$), 6,031 ($m = 2$), 467 ($m = 3$), and 11 ($m = 4$) cities. For Spain, each bin contains 4,202 ($m = 1$), 3,375 ($m = 2$), 211 ($m = 3$), and 6 ($m = 4$) cities. The number of cities less than $m = 1$ is 2 in Italy and 6 in Spain. In **Figures 9**, **10** on the semilogarithmic axis, the optimum line is drawn by the least squares method for $T = 1, 2, \cdots, 8$. From these optimal straight lines, we confirm that the geometric mean of urban population grows approximately exponentially over almost all periods for the case of $m = 1, 2, 3$, and 4. Here, the exponents in Italy $\beta^{\mathrm{IT}}_{\mathrm{city}}(m)$ of the four straight lines (4) for $m = 1, 2, 3, 4$ in the Italian long-term growth **Figure 9** are evaluated as $\beta^{\mathrm{IT}}_{\mathrm{city}}(1) = -0.0071 \pm 0.0058$, $\beta^{\mathrm{IT}}_{\mathrm{city}}(2) = 0.0055 \pm 0.0071$, $\beta^{\mathrm{IT}}_{\mathrm{city}}(3) = 0.055 \pm 0.003$, and $\beta^{\mathrm{IT}}_{\mathrm{city}}(4) = 0.14 \pm 0.01$, respectively. The exponents in Spain $\beta^{\mathrm{ES}}_{\mathrm{city}}(m)$ of the four straight lines (4) for $m = 1, 2, 3, 4$ in the Spanish long-term growth **Figure 10** are evaluated as $\beta^{\mathrm{ES}}_{\mathrm{city}}(1) = -0.030 \pm 0.017$, $\beta^{\mathrm{ES}}_{\mathrm{city}}(2) = 0.016 \pm 0.009$, $\beta^{\mathrm{ES}}_{\mathrm{city}}(3) = 0.085 \pm 0.003$, and $\beta^{\mathrm{ES}}_{\mathrm{city}}(4) = 0.18 \pm 0.01$, respectively. These values are measured by applying $T = 1, 2, \cdots, 8$ for exponential growth (4). As in the U.S., $\beta^{\mathrm{IT}}_{\mathrm{city}}(m)$ and $\beta^{\mathrm{ES}}_{\mathrm{city}}(m)$ are also increasing functions of m. Significantly different from the U.S. is that the index $\beta_{\mathrm{city}}(1)$ for $m = 1$ becomes negative, i.e., for $m = 1$, the geometric mean of the urban population $\langle x_T \rangle$ decreases depending on T in Italy and Spain.

5. SIMULATION OF LONG-TERM GROWTH PROPERTY

In the previous section, we observed the dependence of the short-term growth-rate distribution on the initial value and observed the properties of long-term growth. With respect to the initial value dependence of the short-term growth rate, it was observed that the probability of a decrease in the urban population increases as the initial population decreases, and that the probability of an increase in the urban population does not depend on the initial population in any of the countries in which the urban population data were analyzed. As for long-term growth, it was confirmed that the geometric mean of the urban population $\langle x_T \rangle$ grew exponentially $e^{\beta_{\mathrm{city}}(m)T}$ as in Equation (4), depending on the year $T = (1, 2, \cdots)$. In Italy and Spain, an exponential decline ($\beta^{\mathrm{IT}}_{\mathrm{city}}(1), \beta^{\mathrm{ES}}_{\mathrm{city}}(1) < 0$) was also observed for small ($m = 1$) original ($T = 1$) urban populations ($x_1 \in [10^2, 10^3]$). Collectively, we observed that the larger the original urban population x_1 (the larger the m), the larger the exponent of the exponential function $\beta_{\mathrm{city}}(m)$, which indicates the rate of growth. It was also observed that as large cities grew, their growth slowed and stopped.

This section uses simulations to show how these short- and long-term growth properties are related. Specifically, we use data sampled from the short-term growth-rate distribution to grow cities with different initial values and confirm whether the long-term growth has the properties observed in the real data.

Since it is important to adopt a non-Gibrat's property that the growth-rate distribution differs depending on the initial value, the simulation was designed as follows. First, we divided the initial city population x_T into eight bins: $x_T \in [1, 10^2)$, $x_T \in [10^{0.5(n-1)+2}, 10^{0.5n+2})$ ($n = 1, 2, \cdots, 6$), and $x_T \in [10^5, \infty)$. The first and last bins differ from those in the previous section for the initial dependence of the growth-rate distribution. The first bin is needed if the city population is smaller than 10^2 in the simulation. The last bin removes the upper limit $10^{5.5}$ in the empirical data analysis to increase the number of data items in the bin. A bin corresponding to an initial value x_T is selected from these eight bins, and a growth rate R is extracted at random from the bin, and each urban population is grown by $x_{T+1} = R\, x_T$ to grow each urban population. In the simulation, this step is repeated 10 times to obtain $x_1, x_2, \cdots, x_{11}$.

We confirm that this growth depends on the original urban population size x_1 as follows. From the 1900 urban population, for seven cases with different original values, we randomly extracted the population of 10,000 with repetition: $x_1 \in [10^{0.5(m-1)+2}, 10^{0.5m+2})$ ($m = 1, 2, \cdots, 7$). The 10-step growth of the geometric mean of the urban population classified by the seven different original values is depicted in **Figure 11** for the United States and **Figure 12** for Italy. The results for Spain are so similar to those for Italy that they have been omitted. In **Figures 11**, **12**, as in Equation (4), it is confirmed that the geometric mean of the urban population $\langle x_T \rangle$ grows as an exponential function of step T: $\langle x_T \rangle \propto e^{\beta_{\mathrm{sim}}(m)T}$. In the following, the exponent of exponential growth observed in the simulation is expressed as $\beta_{\mathrm{sim}}(m)$. The exponent $\beta_{\mathrm{sim}}(m)$ increases with increasing m initially but decreases with increasing m.

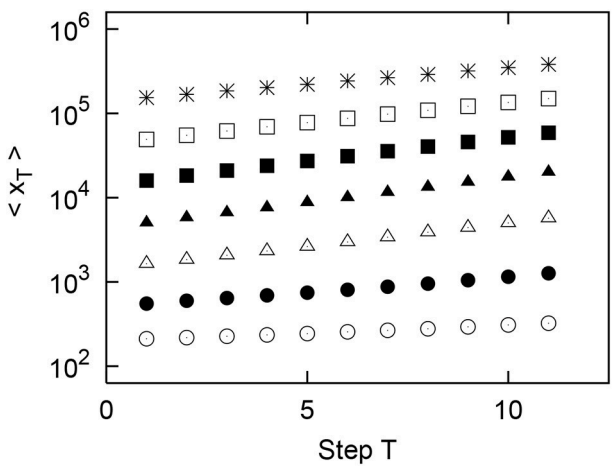


FIGURE 11 | Results of simulation of the growth of the geometric mean of the urban population using values randomly sampled from real data in the United States. Original ($T = 1$) populations x_1 are divided into logarithmically equally sized bins: $x_1 \in [10^{0.5(m-1)+2}, 10^{0.5m+2})$ ($m = 1, 2, \cdots, 7$). Points in the figure indicate $m = 1, 2, \cdots, 7$ in order from the bottom.

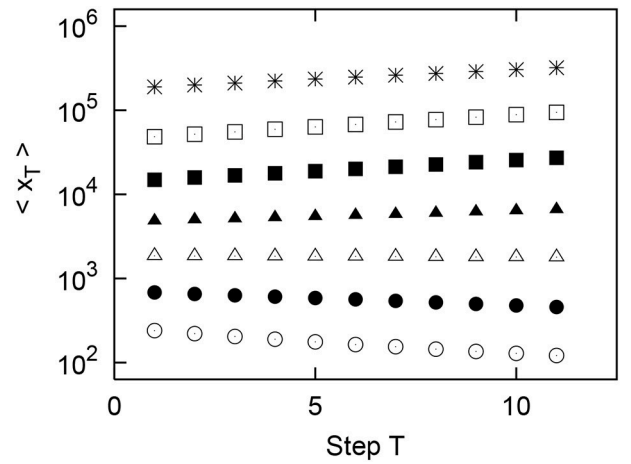


FIGURE 12 | Results of simulation of the growth of the geometric mean of the urban population using values randomly sampled from real data in Italy. Original ($T = 1$) populations x_1 are divided into logarithmically equally sized bins: $x_1 \in [10^{0.5(m-1)+2}, 10^{0.5m+2})$ ($m = 1, 2, \cdots, 7$). Points in the figure indicate $m = 1, 2, \cdots, 7$ in order from the bottom.

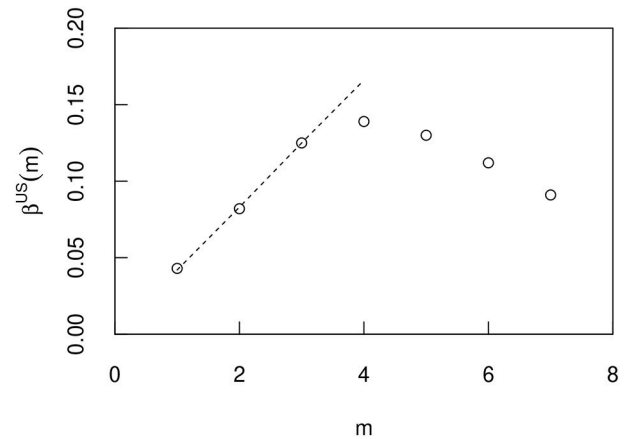


FIGURE 13 | The m dependence of the exponent $\beta^{US}(m)$ when the T dependence of the geometric mean $\langle x_T \rangle$ is approximated by an exponential function $e^{\beta^{US}(m)T}$ in a US simulation (**Figure 11**). The errors are so small that they do not appear in the figure, and they are therefore omitted.

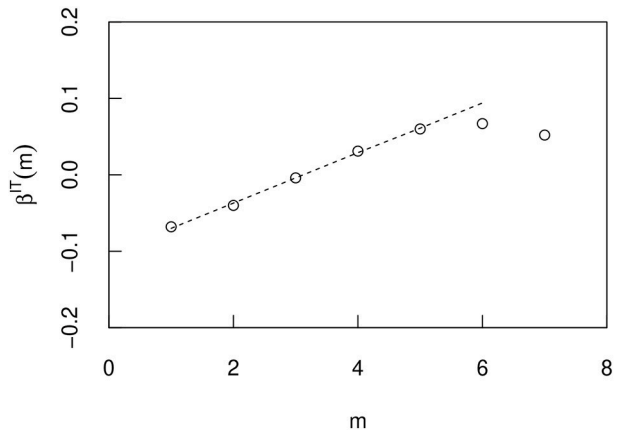


FIGURE 14 | The m dependence of the exponent $\beta^{IT}(m)$ when the T dependence of the geometric mean $\langle x_T \rangle$ is approximated by an exponential function $e^{\beta^{IT}(m)T}$ in a Italy simulation (**Figure 12**). The errors are so small that they do not appear in the figure, and they are therefore omitted.

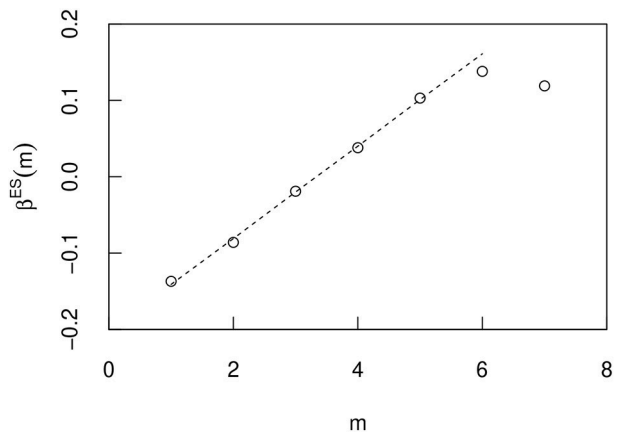


FIGURE 15 | The m dependence of the exponent $\beta^{ES}(m)$ when the T dependence of the geometric mean $\langle x_T \rangle$ is approximated by an exponential function $e^{\beta^{ES}(m)T}$ in a Spain simulation. The errors are so small that they do not appear in the figure, and they are therefore omitted.

Figures 13–15 show the correlation of this index $\beta_{\text{sim}}(m)$ and m in the U.S., Italy, and Spain, respectively. In these countries, $\beta_{\text{sim}}(m)$ initially increased linearly with m. The exponents are approximated by $\beta^{\text{US}}_{\text{sim}}(m) = 0.041m + 0.001$ in the U.S., $\beta^{\text{IT}}_{\text{sim}}(m) = 0.033m - 0.102$ in Italy, and $\beta^{\text{ES}}_{\text{sim}}(m) = 0.060m - 0.201$ in Spain. Here we consider the optimal line for the first 3 n in **Figure 13** and the first 5 n in **Figures 14**, **15**. In Italy and Spain, the intercept of these linear relationships is negative, so $\beta_{\text{sim}}(m)$ becomes negative when m is small. The value of the index β_{sim} obtained from the simulation results is closer to the value of $\beta_{\text{city}}(m)$ measured from the actual city growth. However, they are not exactly the same and do not have to be because they are simulated over different time periods. Importantly, the m dependency of β results from the non-Gibrat's property. Because, it can be confirmed that the m dependency of β_{sim} does not occur by simulation without the procedure of selecting the growth rate R from eight n bins divided by the initial value x_T.

6. CONCLUSION

This paper uses census municipal population data (these are collectively called urban populations in this paper) for the United States, Italy, and Spain to analyze the statistical properties of their 10-year growth (short-term growth). As a result, it was confirmed that the smaller the initial urban population is, the greater the probability that the urban population will decrease, and that the probability that the urban population will increase does not depend on the initial

urban population. We call this the non-Gibrat's property of the urban population. We also observed the statistical properties of long-term growth of urban populations in each country over 100 years. Specifically, we identified the following properties by observing the geometric mean of logarithmically equal sized bins of the oldest urban population in the data used in the analysis.

1. The average urban population increases or decreases exponentially with time.
2. The smaller the initial average urban population, the smaller the exponent, which can be negative in Italy and Spain.
3. When the average urban population is large, exponential growth may stop.

We conducted the following simulations to clarify the relationship between these short- and long-term properties. First, the original urban population was randomly extracted from the oldest urban population in the analysis, and using short-term growth data, they were grown by 10 steps. What is important here is that the growth rate varies according to the size of the initial urban population according to the non-Gibrat's property. As a result, it was confirmed that almost all of the above properties observed in real data for 100 years were reproduced. Specifically, the following long-term properties were confirmed.

4. The geometric mean of the urban population grows or declines exponentially over time.
5. The index increases with the size of the original urban population.
6. However, when the original urban population is very large, the index turns to decline.

The property 6. is considered to be the property 3. smoothed by simulation.

Finally, we consider how the short-term properties leads to the long-term properties 1. and 2. First, we assume that the definition of the growth rate approximately holds for the geometric mean. Furthermore, the growth rate does not change approximately over a period of around 100 years. In this case, it is easy to derive that the geometric mean of the urban population grows exponentially, and that the index is the geometric mean of the growth rate minus one. From the non-Gibrat's property observed in the short-term growth rate, it can be concluded that the smaller the urban population size, the smaller the geometric mean of the growth rate, and therefore the smaller the index of exponential growth. In this way, the non-Gibrat's property of short-term growth can be interpreted as being linked to the long-term growth property. In the case of firm sales, the non-Gibrat's property observed in the initial value dependence of the distribution of short-term growth rates produced the firms' initial rapid exponential growth. In the case of urban populations, on the other hand, the non-Gibrat's property controls the rate of long-term slow exponential growth through the mechanisms described above. It is very interesting that the different non-Gibrat's properties of firm sales and urban population lead to different long-term growths.

It was predictable that the distribution of the short-term growth rate in the mid-size range was dependent on the initial value, because the power-law distribution in the large-size range was collapsed in the mid-size range in both firm sales and urban population. In this study, we found that the urban population has a property that is completely opposite to the initial value dependence of the distribution of the short-term growth rate of firm sales, as described in 1. to 3. above. The decline in the urban population will be a policy issue. In the macro view of this paper, the solution is to merge cities with smaller populations. In Italy and Spain, cities with populations generally below tens of thousands tend to decline. This figure may serve as an indicator for policymaking. However, it is necessary to carefully examine the causal relationship between such figures and the results of the merger policy. This is a future issue.

In this paper, we derive these results from three data analyses, the U.S., Italy, and Spain Census. The remaining challenge is to clarify the relationship between the short-term growth parameter γ and the long-term growth parameter β. Since γ in the three countries matches within the margin of error, this may be a universal nature. It is also possible that γ is involved in the correlation between β and m. We are looking to solve this problem in the near future by analyzing the urban population data we are trying to obtain in France and Germany.

This paper discusses the macro-statistical properties of urban population. This discussion does not take into account the microscopic nature of individuals at all, and it is thus not possible to answer why the non-Gibrat's property occurs. In order to develop the results of this paper and better understand the nature of urban population, we need to take into account the microscopic perspective of human movement between cities. On an individual level, it is a likely scenario that people tend to move away from less populated cities because they are inconvenient and difficult to live in. It is also conceivable that the population of cities with too many people will not increase further because they are difficult to live in. In order to construct a theory incorporating such properties, a microscopic view of the network structure will be important [34, 35]. This is an important issue that should be addressed in the future.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: the census information of the municipal population data of the United States, Italy, and Spain.

AUTHOR CONTRIBUTIONS

AI and SF conceived of the presented idea. AI developed the theory and performed the computations. AR verified the analytical methods. TM encouraged AI to investigate non-Gibrat's property and supervised the findings of this work.

All authors discussed the results and contributed to the final manuscript.

FUNDING

This study was supported by JSPS KAKENHI Grant Numbers 17K01277, 16H05904, 18H05217, ROIS NII Open Collaborative Research 2019-FS01, the Obayashi Foundation, The Okawa Foundation for Information and Telecommunications, and ECO2017-82246-P of the Spanish Ministerio de Economía y Competitividad and ADETRE Reference Group of Gobierno de Aragón.

ACKNOWLEDGMENTS

The authors thank Dr. Tsutomu Watanabe with whom much of this work was discussed and Drs. Rafael González-Val, Luis Lanaspa, and Fernando Sanz-Gracia for the databases that we have used in this paper.

REFERENCES

1. Mantegna RN, Stanley HE. *An Introduction to Econophysics: Correlation and Complexity in Finance*. New York, NY: Cambridge University Press (1999).
2. Pareto V. *Cours d'Économie Politique*. London: Macmillan (1897).
3. Newman M. Power laws, Pareto distributions and Zipf's law. *Contemp Phys*. (2005) **46**:323–51. doi: 10.1080/00107510500052444
4. Clauset A, Shalizi CR, Newman M. Power-law distributions in empirical data. *SIAM Rev*. (2009) **51**:661–703. doi: 10.1137/070710111
5. Saichev A, Malevergne Y, Sornette D. *Theory of Zipf's Law and Beyond*. New York, NY: Springer (2009).
6. Aoyama H, Fujiwara Y, Ikeda Y, Iyetomi H, Souma W. *Econophysics and Companies -Statistical Life and Death in Complex Business Networks*. New York, NY: Cambridge University Press (2010).
7. Anazawa M, Ishikawa A, Suzuki T, Tomoyose M. Fractal structure with a typical scale. *Phys A*. (2004) **335**:616–28. doi: 10.1016/j.physa.2003.12.006
8. Ishikawa A, Suzuki T. Relations between a typical scale and averages in the breaking of fractal distribution. *Phys A*. (2004) **343**:376–92. doi: 10.1016/j.physa.2004.06.060
9. Reed WJ. On the rank-size distribution for human settlements. *J Reg Sci*. (2002) **42**:1–17. doi: 10.1111/1467-9787.00247
10. Reed WJ. The Pareto law of incomes–an explanation and an extension. *Phys A Stat Mech Appl*. (2003) **319**:469–86. doi: 10.1016/S0378-4371(02)01507-8
11. Reed WJ, Jorgensen M. The double Pareto-lognormal distribution -a new parametric model for size distributions. *Commun Stat Theor Methods*. (2004) **33**:1733–53. doi: 10.1081/STA-120037438
12. Eeckhout J. Gibrat's law for (all) cities. *Am Econ Rev*. (2004) **94**:1429–51. doi: 10.1257/0002828043052303
13. González-Val R. The evolution of US city size distribution from a long-term perspective (1900-2000). *J Reg Sci*. (2010) **50**:952–72. doi: 10.1111/j.1467-9787.2010.00685.x
14. Giesen K, Zimmermann A, Suedekum J. The size distribution across all cities–double Pareto lognormal strikes. *J Urban Econ*. (2010) **68**:129–37. doi: 10.1016/j.jue.2010.03.007
15. Giesen K, Suedekum J. City age and city size. *Eur Econ Rev*. (2014) **71**:193–208. doi: 10.1016/j.euroecorev.2014.07.006
16. González-Val R, Lanaspa L, Sanz-Gracia F. New evidence on Gibrat's law for cities. *Urban Stud*. (2013) **51**:93–115. doi: 10.1177/0042098013484528
17. González-Val R, Lanaspa L, Sanz-Gracia F. Gibrat's law for cities, growth regressions and sample size. *Econ Lett*. (2013) **118**:367–9. doi: 10.1016/j.econlet.2012.11.020
18. González-Val R, Ramos A, Sanz-Gracia F, Vera-Cabello M. Size distributions for all cities: which one is best? *Pap Reg Sci*. (2015) **94**:177–96. doi: 10.1111/pirs.12037/full
19. Puente-Ajovín M, Ramos A. On the parametric description of the French, German, Italian and Spanish city size distributions. *Ann Reg Sci*. (2015) **54**:489–509. doi: 10.1007/s00168-015-0663-3
20. Luckstead J, Devadoss S. Pareto tails and lognormal body of US cities size distribution. *Phys A Stat Mech Appl*. (2017) **465**:573–8. doi: 10.1016/j.physa.2016.08.061
21. Luckstead J, Devadoss S, Danforth D. The size distributions of all Indian cites. *Phys A Stat Mech Appl*. (2017) **465**:573–8. doi: 10.1016/j.physa.2017.01.065
22. Kwong HS, Nadarajah S. A note on "Pareto tails and lognormal body of US cities size distribution". *Phys A Stat Mech Appl*. (2019) **513**:55–62. doi: 10.1016/j.physa.2018.08.073
23. Bâncescu I, Chivu L, Preda V, Puente-Ajovín M, Ramos A. Comparisons of log-normal mixture and Pareto tails, GB2 or log-normal body of Romania's all cities size distribution. *Phys A Stat Mech Appl*. (2019) **526**:121017. doi: 10.1016/j.physa.2019.04.253
24. Su HL. On the city size distribution: a finite mixture interpretation. *J Urban Econ*. (2020) **116**:103216. doi: 10.1016/j.jue.2019.103216
25. Gibrat R. *Les Inegalites Economiques*. Paris: Sirey (1932).
26. Sutton J. Gibrat's Legacy. *J Econ Liter*. (1997) **35**:40–59.
27. Li R, Dong L, Zhang J, Wang X, Wang WX, Di Z, et al. Simple spatial scaling rules behind complex cities. *Nat Commun*. (2017) **8**:1841. doi: 10.1038/s41467-017-01882-w
28. Ishikawa A. Derivation of the distribution from extended Gibrat's law. *Phys A*. (2006) **367**:425–34. doi: 10.1016/j.physa.2005.12.005
29. Ishikawa A, Fujimoto S, Mizuno T. Shape of growth-rate distribution determines the type of Non-Gibrat's Property. *Phys A*. (2011) **390**:4273–85. doi: 10.1016/j.physa.2011.06.043
30. Ishikawa A, Fujimoto S, Mizuno T, Watanabe T. Firm growth function and extended-Gibrat's property. *Adv Math Phys*. (2016) **2016**:9303480. doi: 10.1155/2016/9303480
31. Ishikawa A, Fujimoto S, Mizuno T, Watanabe T. Long-term firm growth properties derived from short-term laws of sales and number of employees in Japan and France. *Evol Inst Econ Rev*. (2016) **2016**:409–22. doi: 10.1007/s40844-016-0055-0
32. Ishikawa A. The uniqueness of firm size distribution function from tent-shaped growth rate distribution. *Phys A*. (2007) **383**:79–84. doi: 10.1016/j.physa.2007.04.089
33. Tomoyose M, Fujimoto S, Ishikawa A. Non-Gibrat's law in the middle scale region. *Prog Theor Phys Suppl*. (2009) **179**:114–22. doi: 10.1143/PTPS.179.114/1914937
34. Wang J, Li C, Xia C. Improved centrality indicators to characterize the nodal spreading capability in complex networks. *Appl Math Comput*. (2018) **334**:388–400. doi: 10.1016/j.amc.2018.04.028
35. Li C, Wang L, Sun S, Xia C. Identification of influential spreaders based on classified neighbors in real-world complex networks. *Appl Math Comput*. (2018) **320**:512–23. doi: 10.1016/j.amc.2017.10.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ishikawa, Fujimoto, Ramos and Mizuno. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 23 October 2020
doi: 10.3389/fphy.2020.00339

Measuring Complexity in Financial Data

Gaurang Singh Yadav[1], *Apratim Guha*[2,3] *and Anindya S. Chakrabarti*[4]*

[1] *Indian Institute of Management Ahmedabad, Ahmedabad, India,* [2] *Production, Operations and Decision Sciences Area, XLRI, Xavier School of Management, Jamshedpur, India,* [3] *Production and Quantitative Methods Area, Indian Institute of Management Ahmedabad, Ahmedabad, India,* [4] *Economics Area, Indian Institute of Management Ahmedabad, Ahmedabad, India*

The stock market is a canonical example of a complex system, in which a large number of interacting agents lead to joint evolution of stock returns and the collective market behavior exhibits emergent properties. However, quantifying complexity in stock market data is a challenging task. In this report, we explore four different measures for characterizing the intrinsic complexity by evaluating the structural relationships between stock returns. The first two measures are based on linear and non-linear co-movement structures (accounting for contemporaneous and Granger causal relationships), the third is based on algorithmic complexity, and the fourth is based on spectral analysis of interacting dynamical systems. Our analysis of a dataset comprising daily prices of a large number of stocks in the complete historical data of NASDAQ (1972–2018) shows that the third and fourth measures are able to identify the greatest global economic downturn in 2007–09 and associated spillovers substantially more accurately than the first two measures. We conclude this report with a discussion of the implications of such quantification methods for risk management in complex systems.

Keywords: complex systems, networks, spectral analysis, mutual information, interaction, Granger causality, algorithmic complexity

OPEN ACCESS

Edited by:
Siew Ann Cheong,
Nanyang Technological University, Singapore

Reviewed by:
Gholamreza Jafari,
Shahid Beheshti University, Iran
Parongama Sen,
University of Calcutta, India

***Correspondence:**
Anindya S. Chakrabarti
anindyac@iima.ac.in

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *02 April 2020*
Accepted: *20 July 2020*
Published: *23 October 2020*

Citation:
Yadav GS, Guha A and Chakrabarti AS (2020) Measuring Complexity in Financial Data. Front. Phys. 8:339. doi: 10.3389/fphy.2020.00339

1. INTRODUCTION

How can complexity in financial markets be measured? Although financial markets are routinely thought of as complex systems, exact characterization of their embedded complexity seems non-existent, as pointed out in Brunnermeier and Oehmke [1]. In various contexts, different characterizations and underlying mechanisms have been proposed; explanations include the emergence of macroscopic properties from microscopic interactions [2, 3], the presence of power laws and/or long memory in fluctuations [4], and scaling behavior in growth rates of economic and financial entities [5], to name a few.

In this brief research report we investigate the following question: Given the realized dynamical behavior of a system, can we find the degree of complexity embedded in the system? We note that in the case of financial markets, while interactions between economic agents can be non-linear in nature (due to heterogeneity in behavioral aspects, institutional properties, or information processing abilities), a complete enumeration of all such non-linearities is almost impossible. In this work we do not attempt to find a microfoundation of complexity based on traders' behavior; instead, we aim to quantify complexity in terms of a summary statistic inferred from observed behavior that potentially evolves over time.

We consider four main candidate measures of complexity in multivariate financial asset return data. The dataset we analyze is extracted from complete historical data between 1972 and 2018 of

NASDAQ (National Association of Securities Dealers Automated Quotations), which is one of the largest stock markets in the world in terms of trading volume. We divide the whole time evolution into overlapping windows of 4 years long. To fix ideas, let N be the number of stocks in the stock market and T the number of return data points within each time window, where $N \ll T$ (in the actual implementation for each window of data, $N = 300$ and $T \geq 1,000$, which corresponds to 4 years of trading in NASDAQ; we elaborate on the data structure and sample selection in section 2.1). The first measure is based on mutual information across stocks. Mutual information is an entropy-based measure that generalizes the linear co-movement structure to non-linear co-movements. The second measure is based on dispersion in systemic risk captured via Granger causal relationships across stocks; Granger causality captures lagged co-movement structure in the data. The third measure is based on algorithmic complexity evaluated on the projection of the N-dimensional data onto a two-dimensional space. The fourth measure is based on a vector autoregression estimation of N-dimensional data. This measure is motivated by the famous May-Wigner result that characterizes the instability of many-dimensional heterogeneous interacting systems. We compute each of these measures on 4-years windows and study how the measure evolves when we move the windows by 1 year (from 1972 to 2018 there are 44 such windows).

We assess the usefulness of the measures by seeing whether they can identify the only major financial crisis in the time period under consideration (1972–2018), which occurred during 2007–09 (for an overview of the economic and financial impacts and the implications of the crisis, readers can consult [6, 7] and references therein). During this crisis, the housing market meltdown in the USA led to an avalanche of collapse in the global financial market. Therefore, if any of the four measures of complexity show an increase in magnitude during this time period (or around it), we will take this as a sign of increased instability and hence embedded complexity.

To summarize the results, we find that the first two measures do not exhibit any unusual behavior during or around 2007–09. However, the third and fourth measures (based on algorithmic complexity and heterogeneity of interactions, respectively) do show a substantial increase in magnitude during the crisis period.

Our work is related to several strands of the existing literature. First, it is related to an early attempt of Bonanno et al. to characterize levels of complexity in financial data [8]. They graded complexity in three levels: the lowest level has time series properties (such as volatility clustering); level two contains cross-correlations; and level three is characterized by extreme movements in the collective dynamics, signifying the highest level of complexity. The present work is an attempt toward numerically quantifying the third level, i.e., the highest level of complexity as described in Bonanno et al. [8]. Second, we note that the goal of finding complexity measures for financial data based on techniques from physics, economics, evolutionary biology, etc. has often been discussed, for example in Johnson and Lux [9] and references therein; however, to the best of our knowledge, currently there is still no measure available (apart from sudden changes in volatility) that can accurately identify periods of large-scale financial distress from only asset return data. We note that this goal is different from that of seeking statistical precursors to financial crises (or even identifying mechanisms correlated with financial crises), toward which some work has already been done (see e.g., [10, 11]).

There is large volume of work on construction and inference of network structures from multivariate stock return data (see [12–14] and references therein). Our first measure is based on non-linear relationships between stock returns, for which we adopt an entropy-based measure of mutual information (previously used in the context of financial time series, such as in Fiedor [15]), and we compare the dynamics of the corresponding eigenspectrum with that obtained from linear correlation matrices [16]. We see that there is an overall increase in the degree of correlation over time between what can be inferred from non-linear and from linear relationships, along with a cyclic oscillation in explanatory power. This indicates that a non-linear relationship between assets does not necessarily convey more information than a linear relationship.

Next, we quantify the behavior of a directional Granger causal network over time. The spread in centralities of the nodes in the directional lagged co-movement network (captured by Granger causation) remains fairly stable over time. This analysis is motivated by two influential papers in which the systemic risks of assets were constructed from return data (see [17, 18]). There is a related literature on characterizing shock spillover in a multi-dimensional return network. However, here we do not consider those constructions, since they do not directly relate to instability of the financial system.

We then implement a non-parametric, information-theoretic measure of complexity that is based on algebraic complexity [19–24]. Zenil et al. [25] applied an algorithmic measure of complexity to financial data. We adapt their measure to many-dimensional data by transforming the data through multi-dimensional scaling. This dimension-reduction technique makes the method very generally applicable to time series data, and the measure is able to accurately identify times of crisis.

Finally, the fourth measure is based on the ecology-inspired dynamical systems theory proposed by Robert May [26]. In reference [27] an adaptation of the original May-Wigner result is proposed in the context of a discrete-time vector autoregression model and applied to a limited set of data from the New York Stock Exchange. We adopt the same approach and construct the implied heterogeneity index of stock interactions, which exhibits sharp transitions during the crisis and also in the post-crisis period, indicating lagged effects.

2. MATERIALS AND METHODS

In this section we describe the data and the methods. All background material and a step-by-step description of the computational procedure are given in the **Supplementary Material**.

2.1. Data Description

We collected daily NASDAQ stock return data over a period of 47 years, from 1972 to 2018 [obtained from the Center for

Research in Security Prices (CRSP) database, http://www.crsp.org/, accessed through Wharton Research Data Services (WRDS), https://wrds-www.wharton.upenn.edu/]. Let $\mathbb{T}$ denote the length of the entire data series in years, so $\mathbb{T} = 47$. We considered moving windows of width equal to 4 years, i.e., 1972–75, 1973–76, and so on until 2015–18, with the windows indexed by $k = 1, 2, \ldots, 44$[1]. For each window, we calculated the market capitalization of all stocks at the end of the period and selected the top $N = 300$, with the restriction that the data for chosen stocks cannot have more than 5% missing values within a window (which we fill with zeros). This dataset covers pre-crisis, crisis, and post-crisis periods (the crisis period was 2007–09).

We denote each price series by $S_i^k(t)$, where i is the stock, t is the time period within a window, and k is the window. A 4-years window has ∼1,000 days (each year has slightly more than 250 trading days) and is denoted by T_k. All our analyses were conducted on the log-return data, defined as

$$G_i^k(t) = \log S_i^k(t+1) - \log S_i^k(t). \tag{1}$$

Next, we normalize the log-return data as follows:

$$g_i^k(t) = \frac{G_i^k(t) - \langle G_i^k(t) \rangle}{\sigma_i^k}, \tag{2}$$

where $\langle \cdot \rangle$ denotes the sample average and σ_i^k is the sample standard deviation of G_i.

2.2. Quantification of Linear and Non-linear Relationships

In this subsection we compare the information content in linear and non-linear relationships.

2.2.1. Correlation Matrix

We construct the cross-correlation matrix C^k as

$$C_{ij}^k = \langle g_i^k(t) g_j^k(t) \rangle \tag{3}$$

for all $i, j \leq N$ for the kth window, where $k = 1, \ldots, \mathbb{T}$. For the $\{i, j\}$th pair we construct a distance measure in the form of (this form is widely used; see e.g., [13]).

$$D_{ij}^k = \sqrt{2(1 - C_{ij}^k)}. \tag{4}$$

2.2.2. Entropy and Mutual Information Matrix

First we need to define entropy. For the probability distribution $p(x)$ of a discrete variable X defined over a domain $[x_1, x_2, \ldots, x_N]$, the Shannon entropy is given by [23]

$$H(X) = -\sum_i p(x_i) \log_2 p(x_i). \tag{5}$$

For two discrete variables X and Y with probability distributions $p(x)$ and $p(y)$, the joint entropy is given by [23]

$$H(X, Y) = -\sum_i \sum_j p(x_i, y_j) \log_2 p(x_i, y_j), \tag{6}$$

where $p(x_i, y_j)$ denotes joint probability. Mutual information is an entropy-based measure that is defined for two variables X and Y having probability distributions $p(x)$ and $p(y)$ [23]:

$$I(X; Y) = \sum_{i,j} p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(y_j) p(x_i)}, \tag{7}$$

which is always guaranteed to be non-negative and symmetric. We construct the mutual information matrix M^k for each window k, where the element M_{ij}^k of the matrix is defined as

$$M_{ij}^k = I(g_i^k; g_j^k). \tag{8}$$

By construction, M has all non-negative elements and is symmetric. We have used the Freedman-Diaconis rule here [28] to discretize the data. Further details are available in the **Supplementary Material**.

2.2.3. Comparison of Linear and Non-linear Relationships

We conduct an eigendecomposition of both the distance and the mutual information matrices for every window $k = 1, 2, \ldots, \mathbb{T}$. First we carry out eigendecompositions of the distance matrix D (from Equation 4) and the mutual information matrix M (from Equation 8):

$$D = \sum_{i=1}^{N} \nu_i^D (\nu_i^D)' \lambda_i^D \quad \text{and} \quad M = \sum_{i=1}^{N} \nu_i^M (\nu_i^M)' \lambda_i^M, \tag{9}$$

where λ_i is the ith eigenvalue, ν_i is the corresponding eigenvector, and a prime represents transpose. Since the dominant eigenvector represents the contribution of each asset to the aggregate interaction matrix, we extract the dominant eigenvectors from both the distance and the mutual information matrices for every window and regress the eigenvector obtained from the kth mutual information matrix ($\nu^{\mathrm{mi},k}$) on that obtained from the corresponding kth distance matrix ($\nu^{D,k}$):

$$\nu_j^{\mathrm{mi},k} = \alpha + \beta \nu_j^{D,k} + \epsilon_j \quad \text{for } j = 1, \ldots, N, \tag{10}$$

where α and β are constants and ϵ_j is an error term. The explained variation (i.e., the R^2 of the regression) over 47 windows is plotted in **Figure 1**. High explanatory power would indicate that the information content is similar in the two measures. Two features stand out from the results. First, there is substantial time variation and an almost cyclic oscillation in the explanatory power. Second, there seems to be a general increase over time in the degree of relationship, indicating that the information content is becoming more similar, at least for pairwise relationships. The mutual information estimates were computed by discretizing the data, with each series converted into an ordinal categorical series with b classes, where $b = 8$, 12, and 16, utilizing the useful property that mutual information is a probability-based measure. Upon varying number of bins, the results are similar in all cases [29]. Therefore, the information content seems to be captured well by linear correlation matrices, which are much less computationally intensive.

[1] Because of missing data, the first window contains 124 stocks.

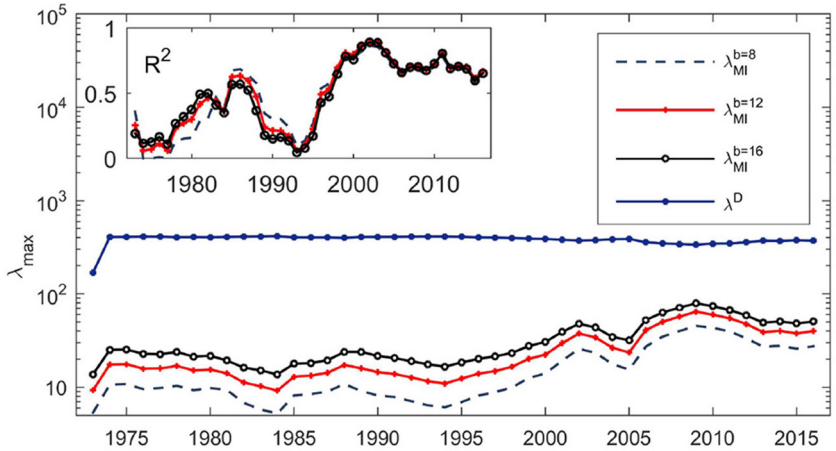


FIGURE 1 | Evolution of the dominant eigenvalues of the distance matrices (λ^D) and the mutual information matrices (λ^b_{MI} with $b = 8, 12, 16$) over the time period 1972–2018. The dominant eigenvalues of the mutual information matrices (for three bin choices) show variation over time in the semi-log plot. Due to scaling, the variation in the dominant eigenvalue of the distance matrix is subdued. Inset: Time series of R^2 obtained from regressing the dominant eigenvector of the mutual information matrix on the dominant eigenvector of the distance matrix over the period 1972–2018 comprising 44 time windows. The choice of the number of bins b seems to have a negligible effect (results are shown for $b = 8$, 12, and 16). The dominant eigenvectors representing market modes in the distance matrices D and the mutual information matrices M became strongly correlated after the year 2000.

2.3. Complexity Through Dispersion in Systemic Risk

The spread in systemic risk across different stocks may indicate the degree of complexity. A high spread would imply that some assets are extremely risky while other assets are safe; a low spread would indicate a similar risk profile for all stocks. We quantify systemic risk following the method proposed in Yun et al. [17], which uses the Granger causal network as the fundamental building block.

We construct the Granger causal network (GCN) for each window of data (excluding the first window as its network size was not comparable with that of the rest). Each network is constructed as follows. If the jth asset's return *Granger causes* the ith asset's return, then there exists an edge from j to i:

$$r_{it} = \alpha + \beta_{ii} r_{i,t-1} + \beta_{ij} r_{j,t-1} + \epsilon_{it} \quad \text{for } i, j = 1, \ldots, N, \qquad (11)$$

where α is a constant, β_{ij} is the parameter of interest, and ϵ_{it} is an error term. In the estimated model, if β_{ij} is significantly different from zero (evaluated at the standard 5% level of significance, with estimation done using the `lmtest` package in R), we connect i and j. We do the same for all $i, j = 1, \ldots, N$ and create a full Granger causal matrix $G_{N\times N}$. A visual example is shown in **Figure 2A**. High dispersion in the degree connectivity is evident.

Once the network is created, we find the PageRank [30] of the matrix as a measure of the systemic risk [17, 18]. The interpretation is that a high PageRank would imply a higher propensity of lagged movement with respect to other assets and, therefore, higher risk of spillover from other stocks (see the **Supplementary Material**).

We study the evolution of the influence of assets in the GCN by calculating the dispersion in PageRank. High dispersion would indicate high inequality in influence. We present the evolution of the standard deviation and the differential entropy, two well-known measures of dispersion, in **Figure 2B**. Both series seem to be quite stable, indicating low spread in the influence of assets in the predictive GCN, except for the high inequality around 2013 shown by the first series (the estimate for 2013 represents data from the window 2012–15).

2.4. Algorithmic Measures Based on Information Theory

In this subsection we treat the problem of defining complexity in the financial network from the point of view of replicability of the emergent pattern. Although the present approach is different, we note that in Zenil and Delahaye [25] it was proposed to apply an algorithmic complexity measure to financial price data. The authors analyzed deviation of financial markets from log-normal behavior in a parametric setup under distributional assumptions. Here we use a non-parametric formulation and study the time series behavior of the implied complexity measure.

Our main idea is as follows. Given financial time series data for a certain window, we first create a correlation matrix (as in Equation 3), and from that we construct a distance matrix in the form of an identity matrix minus the correlation matrix. Then, based on a clustering technique (multi-dimensional scaling, a non-linear dimension-reduction technique for information visualization that creates a pattern of the relative positions of a number of objects in a dataset; we employed Euclidean distance for the present implementation [31] using the `sklearn.manifold` package in Python), we project the distance matrix onto a two-dimensional plane. This step generates a data cloud on the two-dimensional plane.

By defining a fine grid on the plane, we convert the data cloud into a binary matrix, where each cell is evaluated according to whether or not it contains an asset's projection. Thus, we get

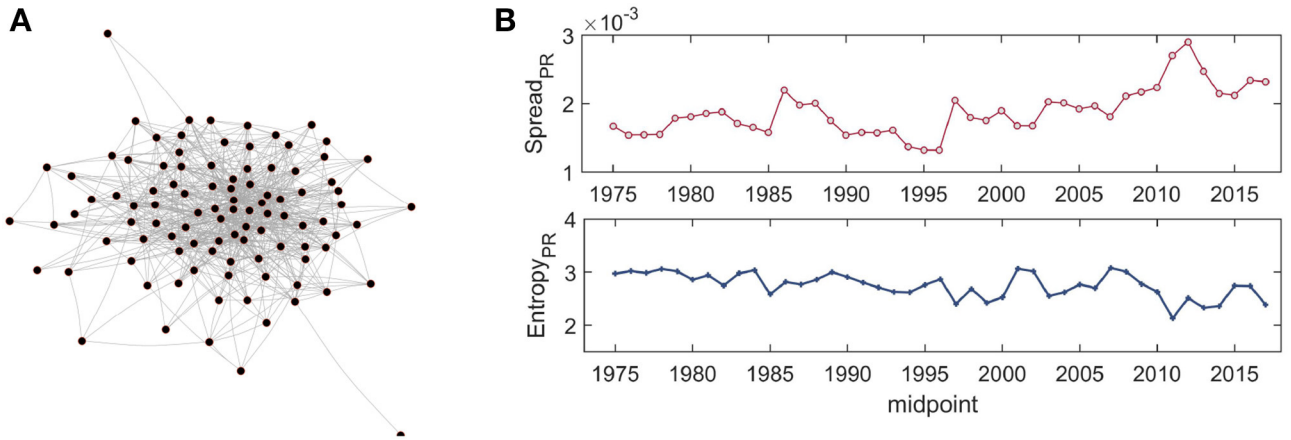


FIGURE 2 | (A) Illustrative example of a Granger causal network of 100 firms (nodes) with greatest market capitalization in the last window of data (2015–2018) from NASDAQ; an edge represents Granger causality between a pair of return time series. **(B)** Evolution of the spread in systemic risk as measured by PageRank computed on Granger causality: the top panel shows the evolution of the standard deviation of PageRank for the nodes; the bottom panel shows the evolution of the differential entropy (Jaynes' method). The spread in centrality of assets remained stable over time, with a large spike occurring around 2013 in the standard deviation, while the entropy remains quite stable.

a pattern on a two-dimensional grid. Given this binary matrix pattern, we can construct a complexity measure [21, 22, 32] based on how complex the pattern is that emerges on the matrix. Our main object of study is the evolution of this complexity measure (see the **Supplementary Material**).

Given this binary representation, one way to construct a complexity measure would be to employ a lossless compression method that captures statistical regularities related to information-theoretic measures, such as Shannon entropy, instead of algorithmic measures [22, 33]. A key limitation of such approaches is that they are not invariant with respect to different descriptions of the same object, while methods in algorithmic complexity, such as the "invariance theorem," can overcome this difficulty [32]. In the following we adopt an algorithm (the block decomposition method, BDM for short) developed in Zenil et al. [21, 22, 32] to construct a complexity measure which in our view is a potential candidate.

The algorithmic complexity of a string can be defined in terms of the shortest algorithm that generates that string [34–36]. The algorithmic complexity $K(s)$ of a string s is the length of the shortest program p that generates s when executed on a universal Turing machine U (prefix-free; for details see [37, 38]), which can be formally expressed as

$$K_U(s) = \min\{|p| : U(p) = s\}. \tag{12}$$

In the following we apply BDM estimation of the complexity of the projection of the data on a two-dimensional grid. For a complete discussion of the methodology of complexity calculations and the background, which is a vast literature in itself, one can consult [34, 35, 39, 40, 40–43].

2.5. Interactive Dynamics: Complexity Through Heterogeneity

Next, we explore an ecology-inspired [9] characterization of economic complexity in terms of the stability of interlinked dynamical systems [44], which comes from the work of Robert May. The result (which goes by the name of the May-Wigner result) is based on prior theoretical work done by E. Wigner on random matrices. The key idea is that as a first-order dynamical system defined on a vector of variables $X_{N\times 1}$ with random heterogeneous connections becomes larger (i.e., N increases), the system tends to be become unstable [26]. Formally, if Γ is an $N \times N$ interaction matrix with elements γ_{ij} such that $\text{Prob}(\gamma_{ij} = 0) = c$ and $\gamma_{ij} = f(0, \sigma^2)$ for all other elements, where f is some distribution with mean zero and variance σ^2, then in the limit $N \to \infty$, the probability that the linear system

$$\dot{X} = \Gamma X_t \tag{13}$$

is stable tends to 1 if $Nc\sigma^2 < 1$ and tends to 0 if $Nc\sigma^2 > 1$ [44]. Importantly, for us σ represents the heterogeneity in the strengths of connections of the interaction matrix Γ. In Rai et al. [27] this idea was applied to the stock market with a discrete-time formulation in the form of a vector autoregression ($X_t = \tilde{\Gamma} X_{t-1} + \epsilon_t$, where one allows for a constant vector c in the vector autoregression estimation; see the **Supplementary Material**). It is shown that during times of crisis the estimated heterogeneity parameter (σ obtained from the estimated $\tilde{\Gamma}$ matrix) increases substantially. However, the data considered in Rai et al. [27] was limited (spanning the 16 years 2002–17), the time windows were non-overlapping, and the analysis was done only on data from the New York Stock Exchange. In the present paper, we perform a complementary analysis with the same technique, using NASDAQ data from 1972 to 2018 with overlapping windows. We fit the vector autoregression model to the data and estimate the $\tilde{\Gamma}$ matrix for each window; then we compute

the standard deviation of the estimated parameters in the $\tilde{\Gamma}$ matrix, which represents the degree of heterogeneity in the interaction strengths.

2.6. Decomposability of Complexity Measures

We also explore whether a feature that we find at the level of raw data can be decomposed in terms of slices of data obtained via eigendecomposition. For this purpose, we consider singular value decomposition,

$$g = U\Sigma V^* \tag{14}$$

where the return matrix g (of size $T \times N$) is expressed as a product of three matrices, namely a $T \times T$ matrix U, an $N \times N$ orthonormal matrix V, such that $V = V^*$, and a $T \times N$ rectangular diagonal matrix Σ which contains non-negative numbers on the diagonal. In the present context, $T > N$.

After de-meaning the data matrix g, we consider the matrix Σ' which contains only a subset of entries on the diagonal while the rest of the entries are replaced by zeros. The original matrix Σ would have N entries on the diagonal. For Σ', we take the subsets to be the first to fourth singular values and the fifth to fifteenth singular values, implying that we can reconstruct the return series associated with the first four and the next eleven values by simply constructing

$$g' = U\Sigma' V^*. \tag{15}$$

We implement the complexity measures on these reconstructed data matrices as well, to see whether a complexity measure calculated from the whole data can be decomposed into complexity measures pertaining to the dominant eigenmodes in the data. We note that for the vector autoregression model estimation, the estimated interaction matrices would have only k non-trivial columns if we select k eigenmodes to construct the data slices.

3. RESULTS

In **Figure 3** we plot the complexity measure for rolling windows using the BDM for the whole data. The results were obtained by employing the multi-dimensional scaling method for fixed axes using the `scikit` library in Python [45]. Fixing the axes is required because the multi-dimensional scaling algorithm does not always compute the projection in the same way since the technique is invariant under rotation in the two-dimensional plane, whereas the complexity measure is not invariant under rotation. The BDM results were obtained using the Python module developed by the AlgoDyn Development team (publicly available at https://pybdm-docs.readthedocs.io/en/latest/index.html). For the present purposes, we used the 2D implementation and two symbols. For the heterogeneity estimates following the May-Wigner theory, in **Figure 4** we plot the evolution of the heterogeneity in the interaction matrix. The analysis was done using the VARS package in the R programming language. Both of the above analyses were complemented by computing the evolution of the same measures on the first four and the next eleven eigenmodes using the singular value decomposition (implemented using `quantmod` in R), as shown in the insets of **Figures 3**, **4**.

The main takeaway from these results is that both of the complexity measures correctly indicate the time of crisis. The BDM-based measure computes an analog of the dispersion in the clustering of data (even with normalized return data), whereas the vector autoregression-based measure captures the dispersion in terms of the strength of interactions. Interestingly, when we apply the same techniques to *slices* of data corresponding to different eigenmodes, similar features are absent. Therefore, these complexity measures, while reasonably correct at the aggregate level, do not seem to be decomposable.

4. SUMMARY AND DISCUSSION

The goal of this work was to extract statistical features from time-varying data that indicate evolution of complexity. Financial systems are thought of as canonical examples of complex systems in terms of interaction, emergence, evolution, and non-stationarity. Here we have analyzed historical financial data on a comprehensive set of stocks from NASDAQ, which is one of the three most followed indices of the US stock market and consists mostly of non-financial tech-oriented firms.

We have estimated four indices of complexity: a measure based on the information content of non-linear co-movements, a systemic risk-based measure constructed from Granger causal networks, an algorithmic complexity measure based on multi-dimensional scaling, and a heterogeneity-based measure motivated by dynamical systems theory. To summarize, the first two measures do not seem to indicate the crisis period (2007–09) clearly, whereas the third and fourth measures perform substantially better and are more accurate. However, neither of the latter two measures is decomposable, in the sense that for each of them the sum of the complexities of decomposed data is not the same as the complexity of the original data.

Some caveats and future directions for research are as follows. First, the results indicate that the information content of the mutual information matrix and that of the correlation matrix become quite similar after the year 2000, so a non-linear measure, such as mutual information is not very useful. There are, however, some new measures of association, with different asymptotic theory (e.g., [46]), that could be explored in future work. Second, an open problem relating to the construction of the Granger causal matrix from pairwise regression is that it does not test for joint significance and there can be type I error due to multiple testing, leading to false discovery of edges [47]. In future work we intend to explore this issue in more detail. Third, for the BDM-based measure of complexity, implementation with more symbols may yield better results, although this would be computationally quite costly. Fourth, following Rai et al. [27] we have shown that the heterogeneity of interaction strengths among the stocks significantly increases during the crisis period and attains an even higher level in the post-crisis period. Two major differences between our results and those of Rai et al. [27] are that (i) in the present work, the spike in heterogeneity has a much

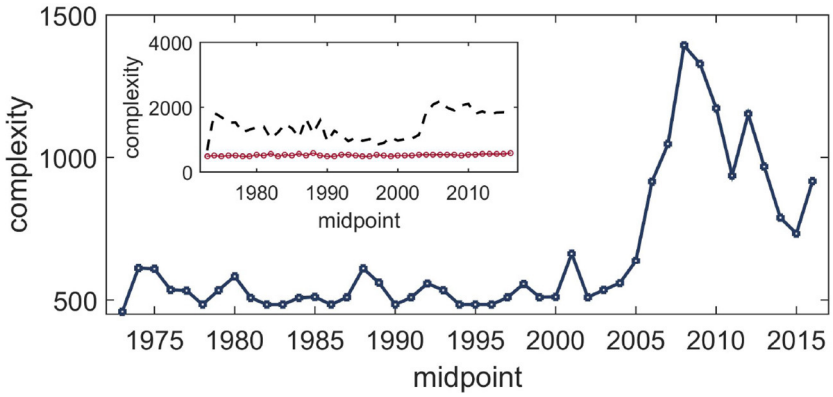


FIGURE 3 | Evolution of the complexity of financial linkages among the stocks over the period 1972–2018 obtained from the BDM. The dimension of the financial linkage data was reduced by mapping the dissimilarity matrices (constructed from the cross-correlation matrix $\rho_{N\times N}$ as $I_{N\times N} - \rho_{N\times N}$) onto two-dimensional grids using a multi-dimensional scaling technique, and the complexity measure was then evaluated on these. The measured value peaks around the crisis period. Inset: Result of the same procedure applied to data slices corresponding to the first four eigenmodes (black dashed line) and the next eleven eigenmodes (red circles).

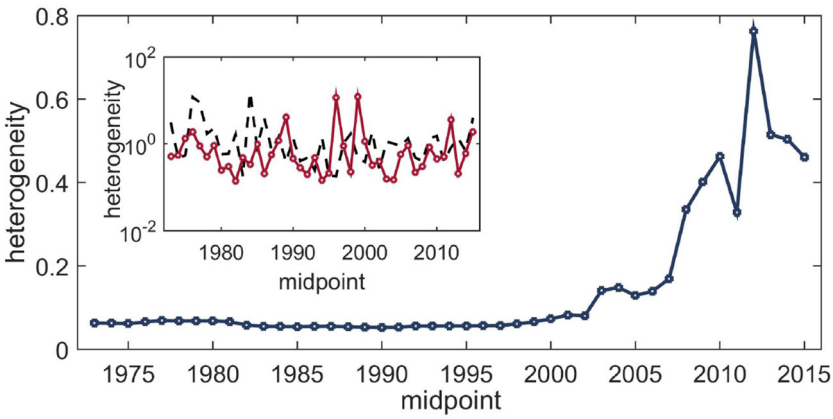


FIGURE 4 | Evolution of the heterogeneity in interaction strengths among the stocks over the period 1972–2018 obtained from the vector autoregression model. Each point estimate corresponds to a 4-years data slice. The *x*-axis plots the midpoints of the windows. Inset: Result of the same procedure applied to data slices corresponding to the first four eigenmodes (black dashed line) and the next eleven eigenmodes (red circles). No particular pattern emerges from the decomposition, but at the aggregate level heterogeneity increases substantially during the time of the crisis and rises further in the post-crisis period.

larger magnitude than that found in Rai et al. [27]; and (ii) in our results the greatest spike in heterogeneity occurs shortly after the crisis (rather than during the crisis as in the analysis of NYSE data in Rai et al. [27]) and seems to continue for a long time without tapering off.

Management of risk in complex systems, such as financial markets requires clear quantification of the complexity. The measures proposed in this paper complement the existing statistical finance literature on describing evolution of markets during crisis and non-crisis periods [11, 48–51]. In this work we have used the word *complexity* to mean emergent instability, similar to Kuyyamudi et al. [11]. It would be interesting to see whether similar ideas can be applied to other complex systems [12]. In the context of financial markets, such quantification of complexity brings us closer to answering the question of what factors (economic or financial) drive the evolution of complexity. A causal explanation of the mechanisms can inform policy-making with regard to complex financial systems.

DATA AVAILABILITY STATEMENT

The data can be accessed through WRDS (Wharton Research Data Services) by paying the required fees. Requests to access these datasets should be directed to the Center for Research in Security Prices: http://www.crsp.org/, WRDS: https://wrds-www.wharton.upenn.edu/.

AUTHOR CONTRIBUTIONS

AC designed the research. GY and AG carried out the numerical computations. AC and GY prepared the figures. All authors were involved in writing the paper.

FUNDING

This research was partially supported by an institute grant, IIM Ahmedabad.

ACKNOWLEDGMENTS

We thank the HPC lab at IIM Ahmedabad for computational support and the Vikram Sarabhai Library for providing the data used in this research. We would like to thank three reviewers, Guiseppe Brandi, Bikas K. Chakrabarti, Anirban Chakraborti, Diptesh Ghosh, Sitabhra Sinha, and Hector Zenil for useful discussions and for providing helpful references. We are grateful to Abinash Mishra for helping us with the collection and processing of the data. Jalshayin Bhachech and Abinash Mishra provided excellent research assistance. All remaining errors are ours.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00339/full#supplementary-material

REFERENCES

1. Brunnermeier M, Oehmke M. *Complexity in Financial Markets*. Working paper. Princeton, NJ: Princeton University (2009).
2. Di Guilmi C, Gallegati M, Landini S. *Interactive Macroeconomics: Stochastic Aggregate Dynamics With Heterogeneous and Interacting Agents*. Cambridge: Cambridge University Press (2017).
3. Gallegati M. *Complex Agent-Based Models*. Berlin: Springer (2018).
4. Lillo F, Mantegna RN. Power-law relaxation in a complex system: Omori law after a financial market crash. *Phys Rev E*. (2003) **68**:016119. doi: 10.1103/PhysRevE.68.016119
5. Stanley MH, Amaral LA, Buldyrev SV, Havlin S, Leschhorn H, Maass P, et al. Scaling behaviour in the growth of companies. *Nature*. (1996) **379**:804–6. doi: 10.1038/379804a0
6. Kashyap A, Zingales L. The 2007-8 financial crisis: Lessons from corporate finance. *J Financ Econ*. (2010) **97**:303–488. doi: 10.1016/j.jfineco.2010.05.010
7. Foster J, Magdoff F. *The Great Financial Crisis: Causes and Consequences*. New York, NY: NYU Press (2008).
8. Bonanno G, Lillo F, Mantegna RN. Levels of complexity in financial data. *Phys A*. (2001) **299**:16–27. doi: 10.1016/S0378-4371(01)00279-5
9. Johnson N, Lux T. Ecology and economics. *Nature*. (2011) **469**:302–3. doi: 10.1038/469302a
10. Sornette D, Zhou W. Predictability of large future changes in major financial indices. *Int J Forecast*. (2006) **22**:153–68. doi: 10.1016/j.ijforecast.2005.02.004
11. Kuyyamudi C, Chakrabarti AS, Sinha S. Emergence of frustration signals systemic risk. *Phys Rev E*. (2019) **99**:052306. doi: 10.1103/PhysRevE.99.052306
12. Newman ME. The structure and function of complex networks. *SIAM Rev*. (2003) **45**:167–256. doi: 10.1137/S003614450342480
13. Mantegna RN, Stanley HE. *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge: Cambridge University Press (1999).
14. Farmer JD, Lo AW. Frontiers of finance: evolution and efficient markets. *Proc Natl Acad Sci USA*. (1999) **96**:9991–2. doi: 10.1073/pnas.96.18.9991
15. Fiedor P. Networks in financial markets based on the mutual information rate. *Phys Rev E*. (2014) **89**:052801. doi: 10.1103/PhysRevE.89.052801
16. Sinha S, Chatterjee A, Chakraborti A, Chakrabarti BK. *Econophysics: An Introduction*. John Wiley & Sons (2010).
17. Yun T, Jeong D, Park S. "Too central to fail" systemic risk measure using PageRank algorithm. *J Econ Behav Organ*. (2019) **162**:251–72. doi: 10.1016/j.jebo.2018.12.021
18. Billio M, Getmansky M, Lo A, Pelizzon L. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J Financ Econ*. (2012) **104**:535–59. doi: 10.1016/j.jfineco.2011.12.010
19. Nirenberg S, Carcieri SM, Jacobs AL, Latham PE. Retinal ganglion cells act largely as independent encoders. *Nature*. (2001) **411**:698. doi: 10.1038/35079612
20. Williams PL, Beer RD. Generalized measures of information transfer. *arXiv*. (2011) 1102.1507.
21. Zenil H, Kiani NA, Zea AA, Tegnér J. Causal deconvolution by algorithmic generative models. *Nat Mach Intell*. (2019) **1**:58. doi: 10.1038/s42256-018-0005-0
22. Zenil H, Kiani NA, Tegnér J. Low-algorithmic-complexity entropy-deceiving graphs. *Phys Rev E*. (2017) **96**:012308. doi: 10.1103/PhysRevE.96.012308
23. Cover TM, Thomas JA. *Elements of Information Theory*. John Wiley & Sons (2012).
24. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. (1948) **27**:379–423.
25. Zenil H, Delahaye JP. An algorithmic information theoretic approach to the behaviour of financial markets. *J Econ Surv*. (2011) **25**:431–63. doi: 10.1111/j.1467-6419.2010.00666.x
26. May RM. Will a large complex system be stable? *Nature*. (1972) **238**:413–4. doi: 10.1038/238413a0
27. Rai A, Bansal A, Chakrabarti AS. Statistical estimation of time-varying complexity in financial networks. *Eur Phys J B*. (2019) **92**:239. doi: 10.1140/epjb/e2019-100161-1
28. Freedman D, Diaconis P. On the histogram as a density estimator: L2 theory. *Z Wahrsch Verwandte Geb*. (1981) **57**:453–76. doi: 10.1007/BF01025868
29. Paiva AR, Park I, Príncipe JC. A comparison of binless spike train measures. *Neural Comput Appl*. (2010) **19**:405–19. doi: 10.1007/s00521-009-0307-6
30. Newman M. *Networks*. Oxford: Oxford University Press (2018).
31. Borg I, Groenen P. Modern multidimensional scaling: theory and applications. *J Educ Meas*. (2003) **40**:277–80. doi: 10.1111/j.1745-3984.2003.tb01108.x
32. Zenil H, Badillo L, Hernández-Orozco S, Hernández-Quiroz F. Coding-theorem like behaviour and emergence of the universal distribution from resource-bounded algorithmic probability. *Int J Parallel Emerg Distrib Syst*. (2019) **34**:161–80. doi: 10.1080/17445760.2018.1448932
33. Cilibrasi R, Vitányi PM. Clustering by compression. *IEEE Trans Inform Theory*. (2005) **51**:1523–45. doi: 10.1109/TIT.2005.844059
34. Levin LA. Laws of information conservation (nongrowth) and aspects of the foundation of probability theory. *Probl Pered Inform*. (1974) **10**:30–5.
35. Chaitin GJ. On the length of programs for computing finite binary sequences. *J ACM*. (1966) **13**:547–69. doi: 10.1145/321356.321363
36. Kolmogorov AN. Three approaches to the quantitative definition of information. *Probl Inform Transm*. (1965) **1**:1–7.
37. Li M, Vitányi P. *An Introduction to Kolmogorov Complexity and Its Applications*, Vol. 3. Springer (2008).
38. Calude CS, Salomaa K, Roblot TK. Finite state complexity. *Theor Comput Sci*. (2011) **412**:5668–77. doi: 10.1016/j.tcs.2011.06.021
39. Solomonoff RJ. A formal theory of inductive inference. Part I. *Inform Control*. (1964) **7**:1–22. doi: 10.1016/S0019-9958(64)90223-2
40. Soler-Toscano F, Zenil H, Delahaye J-P, Gauvrit N. Calculating kolmogorov complexity from the output frequency distributions of small turing machines. *PLoS ONE*. (2014) **9**:e96223. doi: 10.1371/journal.pone.0096223
41. Delahaye JP, Zenil H. Numerical evaluation of algorithmic complexity for short strings: a glance into the innermost structure of randomness. *Appl Math Comput*. (2012) **219**:63–77. doi: 10.1016/j.amc.2011.10.006
42. Rado T. On non-computable functions. *Bell Syst Tech J*. (1962) **41**:877–84. doi: 10.1002/j.1538-7305.1962.tb00480.x
43. Zenil H, Hernández-Orozco S, Kiani NA, Soler-Toscano F, Rueda-Toicen A, Tegnér J. A decomposition method for global evaluation of shannon entropy

and local estimations of algorithmic complexity. *Entropy*. (2018) **20**:605. doi: 10.3390/e20080605
44. Sinha S. Complexity vs. stability in small-world networks. *Phys A Stat Mech Appl*. (2005) **346**:147–53. doi: 10.1016/j.physa.2004.08.062
45. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. (2011) **12**:2825–30.
46. Chatterjee S. A new coefficient of correlation. *J Am Stat Assoc*. (2020) 1–21. doi: 10.1080/01621459.2020.1758115
47. Drton M, Perlman MD. Multiple testing and error control in gaussian graphical model selection. *Stat Sci*. (2007) **22**:430–49. doi: 10.1214/088342307000000113
48. Verma A, Buonocore RJ, Di Matteo T. A cluster driven log-volatility factor model: a deepening on the source of the volatility clustering. *Quant Finance*. (2019) **19**:981–96. doi: 10.1080/14697688.2018.1535183
49. Kenett DY, Raddant M, Lux T, Ben-Jacob E. Evolvement of uniformity and volatility in the stressed global financial village. *PLoS ONE*. (2012) **7**:e31144. doi: 10.1371/journal.pone.0031144
50. Namaki A, Shirazi AH, Jafari GR. Network analysis of a financial market based on genuine correlation and threshold method. *Phys A*. (2011) **390**:3835–41. doi: 10.1016/j.physa.2011.06.033
51. Namaki A, Lai ZK, Jafari GR, Raei R, Tehrani R. Comparing emerging and mature markets during times of crises: a non-extensive statistical approach. *Phys A*. (2011) **392**:3039–44. doi: 10.1016/j.physa.2013.02.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yadav, Guha and Chakrabarti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 12 November 2020
doi: 10.3389/fphy.2020.590623

Time Scales and Characteristics of Stock Markets in Different Investment Horizons

Ajit Mahata and Md. Nurujjaman **

Department of Physics, NIT Sikkim, South Sikkim, India

Investors adopt varied investment strategies depending on the time scales (τ) of short-term and long-term investment time horizons (ITH). The nature of the market is very different in various investment τ. Empirical mode decomposition *(EMD)* based Hurst exponents (*H*) and normalized variance (NV) techniques have been applied to identify the τ and characteristics of the market in different time horizons. The values of *H* and NV have been estimated for the decomposed intrinsic mode functions (IMF) of the stock price. We obtained $H_1 = 0.5 \pm 0.04$ and $H_1 \geq 0.75$ for the IMFs with τ ranging from a few days to 3 months and $\tau \geq$ 5 months, respectively. Based on the value of H_1, two time series have been reconstructed from the *IMFs*: a) short-term time series [$X_{ST}(t)$] with $H_1 = 0.5 \pm 0.04$ and τ from a few days to 3 months; b) long-term time series [$X_{LT}(t)$] with $H_1 \geq 0.75$ and $\tau \geq$ 5 months. The $X_{ST}(t)$ and $X_{LT}(t)$ show that market dynamics is random in short-term *ITH* and correlated in long-term *ITH*. We have also found that the *NV* is very small in the short-term ITH and gradually increases for long-term *ITH*. The results further show that the stock prices are correlated with the fundamental variables of the company in the long-term *ITH*. The finding may help the investors to design investment and trading strategies in both short-term and long-term investment horizons.

Keywords: empirical mode decomposition, Hurst exponent, short-term investment time horizon, long-term investment time horizon, time scale, normalized variance

OPEN ACCESS

Edited by:
Anirban Chakraborti,
Jawaharlal Nehru University, India

Reviewed by:
Suchetana Sadhukhan,
National Autonomous University of Mexico, Mexico
Sunil Kumar,
University of Delhi, India

***Correspondence:**
Md. Nurujjaman
jaman_nonlinear@yahoo.co.in

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *02 August 2020*
Accepted: *29 September 2020*
Published: *12 November 2020*

Citation:
Mahata A and Nurujjaman M (2020) Time Scales and Characteristics of Stock Markets in Different Investment Horizons. Front. Phys. 8:590623. doi: 10.3389/fphy.2020.590623

1. INTRODUCTION

The stock market is a complex dynamical system where the evolution of the dynamics depends on the participation of different types of investors or traders [1–3]. Investors/traders participate in the stock market to gain profit implementing different investment and trading strategies depending on investment time horizons (ITH) [4, 5]. The participation of diversified investors, reaction to the information, and short-term and long-term investment approaches play crucial roles in the movement of stock prices [4].

In the stock markets, there are mainly two types of investors: short-term investors who invest for short-term gain and long-term investors who invest for long-term gain [6, 7]. Studies show that the *ITH* for short-term investors ranges from a single day to a few months, and for long-term investors, it usually ranges from a few months to several years [8, 9]. The fund managers and foreign exchange dealers of various countries use technical analysis for the short-term *ITH* and fundamental analysis for the long-term *ITH* [8, 10]. The time scales (τ) of short-term and long-term *ITH* by the investors are generally chosen in an arbitrary manner based on the investment experience [8, 10]. So the

identification of τ from stock price time series using a well defined technique may be helpful for both the short-term and long-term investors.

As the market is mean reversing in short-term *ITH* [9], traders fail to generate significant returns using technical analysis [11]. On the other hand, in the long-term *ITH*, an investor can generate significant return or help in making decision whether to exit from that particular stock to avoid loss by determining the financial health of a company by using fundamental variables [12–14]. Fundamental analysis is an essential tool to find out the relation between stock price and fundamental variables such as book to market (B/M), sales to price, debt to equity, earnings to price, and cash flow of a stock [12–14]. The stock price is found to be positively correlated with the essential fundamental variables [14–20]. The study of the correlation in the short-term *ITH* and long-term *ITH* is essential to take a fruitful investment decision.

In the short-term *ITH*, the market is generally considered to be governed by psychological behavior of the investors. However, the fundamental variables are the main determining crucial factors in the long-term *ITH*. Usually, investors choose the τ of short-term and long-term investment horizon in an arbitrary manner [8, 9]. Recently, we used structural break study to show that the τ for the short-term is usually less than a few months [9]. The separation of the short- and long-term dynamics in terms of τ plays a vital role in the prediction of future price movement. Hence, detailed studies are required to find the correlation of the stock price with fundamental variables and to identify the τ of market dynamics in the short-term *ITH* and long-term *ITH*.

In this article, we estimated the τ of the stock price in the short-term and long-term *ITH* for twelve leading global stock indices and the stock price of some companies using empirical mode decomposition (*EMD*) based Hurst exponent (*H*) analysis. We have reconstructed short-term and long-term time series based on the *H*. Finally, we estimated the correlation coefficient between long-term time series and fundamental variables. Herein we establish that short-term *ITH* is normally less than 3 months and long-term *ITH* is more than 5 months. Correlation analysis shows that the long-term stock price is positively correlated with the fundamental variables.

The remaining part of this paper is organized as follows: In **Section 2**, we introduce the method of analysis, while **Section 3** presents the data analyzed. Results and discussion and conclusion are delineated in **Sections 4 and 5**, respectively.

2. METHOD OF ANALYSIS

A nonlinear two-step technique—EMD followed by Hilbert–Huang Transform (*HHT*)—has been applied to analyze the stock data as it is nonlinear and nonstationary. Nonlinearity in the stock market appears due to the presence of market frictions and transaction costs, existence of bid-ask spread, and short selling, whereas nonstationarity appears due to different time scales present in the stock market [21, 22]. This approach helps us to identify the characteristic τ and the important trends and components present in the data [3].

The *EMD* method decomposes the stock index and stock price into the intrinsic oscillatory modes of different τ by preserving the nonstationarity and nonlinearity of the data. These oscillatory modes are termed intrinsic mode functions (*IMF*). The *IMFs* can be both amplitude and frequency modulated as well as nonstationary [23, 24]. The τ of each *IMFs* has been identified by *HHT*. The *HHT* eliminates the spurious harmonic components generated due to the nonlinearity and nonstationarity of the data [23, 24].

The *IMFs* satisfy the following two conditions; *i*) the number of extrema and the number of zero crossing must be equal or differ by one; and *ii*) mean values of the envelope, defined by the local maxima and local minima, for each point are zero. The *IMF* is calculated in the following way [23, 25]:

a. Lower envelope $U(t)$ and upper envelope $V(t)$ are drawn by connecting minima and maxima of the data, respectively, using spline fitting.
b. Mean value of the envelope $m = [U(t) + V(t)]/2$ is subtracted from the original time series to get new data set $h = X(t) - m$.
c. Repeat the processes (a) and (b) by considering h as a new data set until the *IMF* conditions (*i* and *ii*) are satisfied.

Once the conditions are satisfied, the process terminates, and h is stored as the first *IMF*. The second *IMF* is calculated repeating the above steps (a)–(c) from the data set $d(t) = X(t) - IMF_1$. When the final residual is monotonic in nature, the steps (a)–(c) are terminated and the original time series can be written as a set of *IMFs* plus residue,

$$X(t) = \sum_{i=1}^{n} IMF_i + residue,$$

where IMF_i represents the *i*th *IMF*, and residue represents the trend of the stock data.

IMFs are the signal with different τ. The IMF_1 is a signal with the smallest τ, the IMF_2 is the signal with the second smallest τ, and so on. Hence, *EMD* technique is useful to extract different τ from the signal. The characteristic τ of each *IMF* can be estimated from the frequency (ω) by using Hilbert Transform, which is defined as

$$Y(t) = \frac{P}{\pi}\int_{-\infty}^{\infty} \frac{IMF(t)}{t - t_,}\, dt,$$

where P is the Cauchy principle value, and $\tau = \frac{1}{\omega}$ where $\omega = \frac{d\theta(t)}{dt}$, and $\theta(t) = \tan^{-1}\frac{Y(t)}{IMF(t)}$ [23]. Identification of important *IMFs* is essential to differentiate the market dynamics in short-term from long-term *ITH*, and the differentiation can be done by evaluating the *H*.

Rescaled-range (R/S) analysis is a technique to estimate the correlation present in a time series by calculating H [26–28]. Details of the R/S technique are described below. Let us consider a time series of length L and divided into p subseries of length l. Each subseries is denoted as $Y_{j,t}$, where $t = 1, 2, 3, \ldots, p$. Mean and standard deviation of the subseries $Y_{j,t}$ are defined as

$$D_t = \frac{1}{l}\sum_{j=1}^{l} Y_{j,t}$$

and

$$S_t = \sqrt{\frac{1}{l}\sum_{j=1}^{l}\left(Y_{j,t} - D_t\right)^2},$$

respectively. Mean adjusted series is calculated as

$$Z_{j,t} = Y_{j,t} - D_t$$

for $j = 1, 2, 3, \ldots, l$. Cumulative time series is given by

$$X_{j,t} = \sum_{i=1}^{j} Z_{i,t}$$

for $j = 1, 2, 3, \ldots, l$.

Range of the series has been calculated as

$$R_t = \max\left(X_{1,t}, \ldots, X_{l,t}\right) - \min\left(X_{1,t}, \ldots, X_{l,t}\right).$$

Individual subseries range can be rescaled or normalized by dividing the standard deviation. So, R/S is written as

$$(R/S)_l = \frac{1}{p}\sum_{t=1}^{p} R_t \Big/ S_t.$$

The ratio of each subseries of length l is expressed as $(R/S)_l \propto l^H$, where H is the Hurst exponent. H can be estimated from the slope of ln(R/S) vs. ln(l). For a random time series, H is around 0.5, and for correlated and anticorrelated time series, H is greater than 0.5 and less than 0.5, respectively.

Normalized variance (NV) is another important statistical tool to identify the important *IMFs* based on the energy of the signal. The higher the NV value is, more significant the signal is. The technique estimates the energy of the ith *IMFs* by calculating variance [29, 30], and NV of ith *IMF* is defined as

$$NV_i = \frac{\sqrt{\sum_t IMF_i^2(t)}}{\sum_{i=1}^{q}\sqrt{\sum_t IMF_i^2(t)}},$$

where q is the total number of *IMF*.

3. DATA ANALYZED

We have analyzed the stock indices and stock prices of a few companies of different countries from December 1995 to July 2018, namely, 1) S&P 500 (USA), 2) Nikkei 225 (Japan), 3) CAC 40 (France), 4) IBEX 35 (Spain) 5) HSI (Hong Kong), 6) SSE (China), 7) BSE SENSEX (India), 8) IBOVESPA (Brazil), 9) BEL 20 (Euro-Next Brussels), 10) IPC (Mexico), 11) Russell 2000 (USA), and 12) TA125 (Israel), and stock prices of the companies 1) IBM (USA), 2) Microsoft (USA), 3) Tata Motors (India), 4) Reliance Communication (RCOM) (India), 5) Apple Inc. (USA), and 6) Reliance Industries Limited (RIL) (India). Stock indices and price data were downloaded from yahoo finance, and the analysis was carried out using MATLAB software.

4. RESULTS AND DISCUSSIONS

The stock market shows different behavior in different investment horizon. *EMD* based H and NV techniques have been applied to analyze the market dynamics as discussed below.

Figures 1A–J show the IMF_1 to IMF_9 and the residue of the S&P 500 index calculated using *EMD* technique as described in detail in **Section 2**. IMF_1 in **Figure 1A** represents the mode with the lowest τ, and it gradually increases with the increase in *IMF* numbers. **Figure 1J** represents the residue of the signal, which indicates the overall trend of the index. Similarly, we have calculated all *IMFs* for all the indices and companies to analyze the market.

4.1. *EMD* Based *H* and *NV* Analysis

H has been calculated for all the *IMFs*. **Figure 2A** shows the typical plot of H versus τ of all the indices and companies. We obtained single H from $IMF_1 - IMF_5$ and it is indicated as H_1. Higher-order *IMF* shows two H, namely, H_1 and H_2. We obtained $H_1 \approx 0.5 \pm 0.04$ for IMF_1 to IMF_5 with τ ranging from a few days (D) to 3 months (M). The value of H_1 jumps to ≥ 0.75 for IMF_6 with $\tau \approx 5$ M. It gradually increases for IMF_7 to IMF_9 with a τ ranging from 1 year (Y) to 12 Y. $H_1 \approx 0.5 \pm 0.04$ for IMF_1 to IMF_5 indicates that the nature of the first five *IMFs* is random. IMF_6 to IMF_9 show a long-range correlation up to one period *lag*. τ of IMF_1, IMF_2, IMF_3, IMF_4, and IMF_5 of all the indices and companies stock data analyzed here are in the range of 3–4 D, 7–10 D, 15–18 D, 1–1.5 M, and 2.5–3 M, respectively.

To further validate the robustness of the proposed method, analysis of the decomposed time series has been carried out using NV technique. **Figures 2B–D** represent NV of all the *IMFs* of all the indices and companies, where plots have been arranged according to the order of higher NV of *IMFs*. **Figures 2B–D** show that the value of NV is very low for all the indices and companies up to IMF_5, and NV increases significantly for IMF_6 to IMF_9. Hence, based on the value of NV the time series can also be decomposed into two time series with two distinct time horizons: short-term time series by adding IMF_1 to IMF_5 and long-term time series by adding IMF_6 to IMF_9 plus residue as described in **Section 4.2**. Further, NV technique can be used to find a time series with important time scale in the form of *IMF*. **Figures 2B–D** show that the value of NV is higher for IMF_7 with $0.8\ Y \leq \tau \leq 1.9\ Y$, IMF_8 with $2.0\ Y \leq \tau \leq 4.4\ Y$, and IMF_9 with $4.5\ Y \leq \tau \leq 12\ Y$, respectively, for the companies mentioned in the plots. The decomposed time series with higher value of NV may play important role to predict long-term behavior of the market [29]. More such studies in detail can be pursued in future.

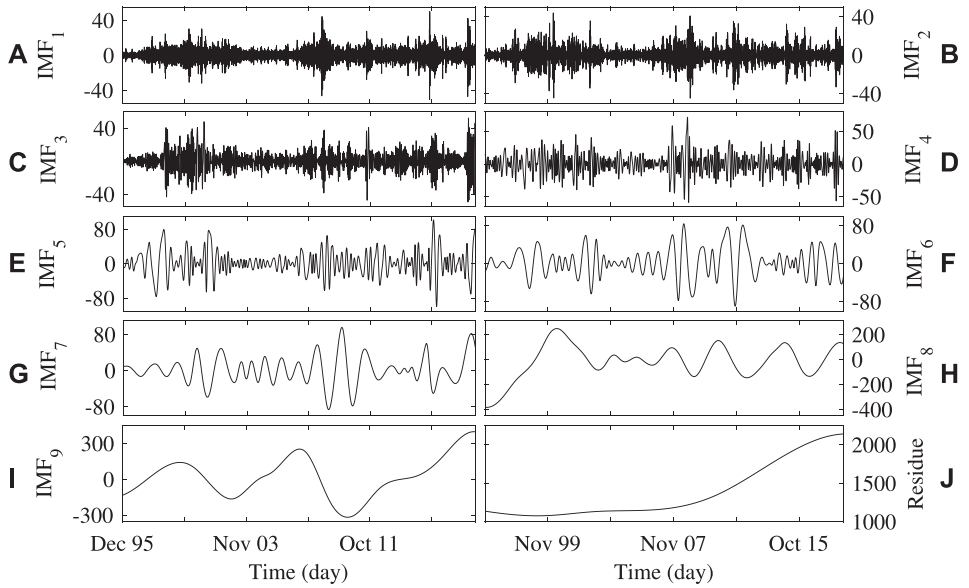


FIGURE 1 | The plots **(A)–(I)** represent the IMF_1 to IMF_9, respectively, and **(J)** represents residue of the S&P 500 index.

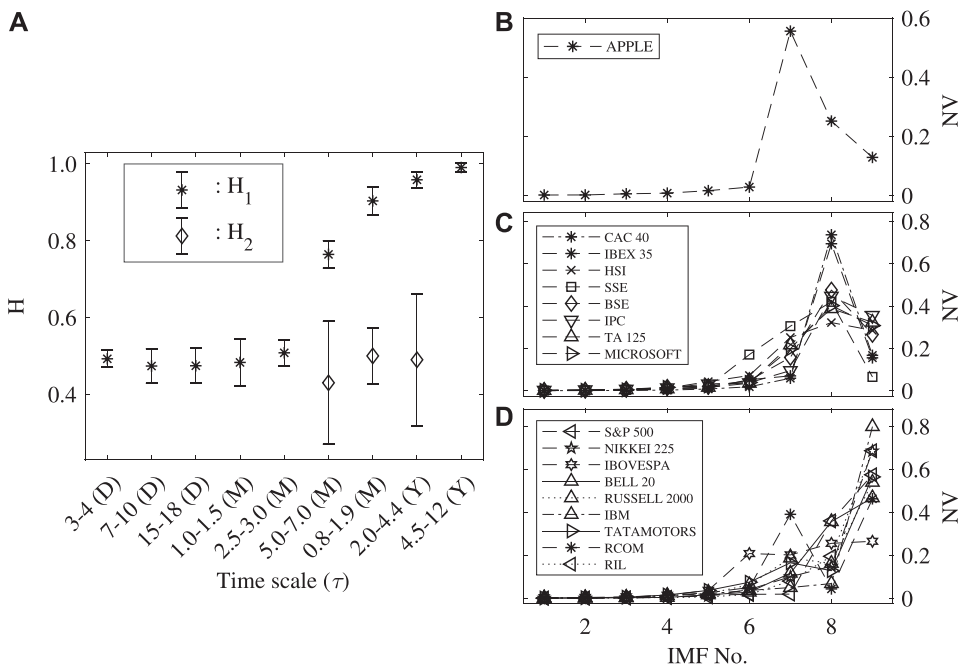


FIGURE 2 | (A) shows the Hurst exponents (H_1 and H_2) vs. τ of all the *IMFs* of all the indices and companies with 2σ error bar. The first point represents the average value of H_1 of all the first *IMFs* of all stock data, the second point represents the average value of H_1 of all the second *IMFs* of all stock data, and so on. For IMF_1 to IMF_5 of all indices and companies $H_1 = 0.5 \pm 0.04$ with a maximum τ of around 3 *M*. The value of H_1 jumps to 0.75 ± 0.04 for IMF_6 (with a $\tau \approx 5\,M$) and gradually increases for IMF_7 to IMF_9. H_1 value shows that nature of IMF_1 to IMF_5 is random and IMF_6 to IMF_9 have a long-range correlation. (*D*), (*M*), and (*Y*) in the *x*-axis represent the day, month, and year, respectively. **(B)–(D)** represent the normalized variance (*NV*) of *IMF*s of all the indices and company, respectively.

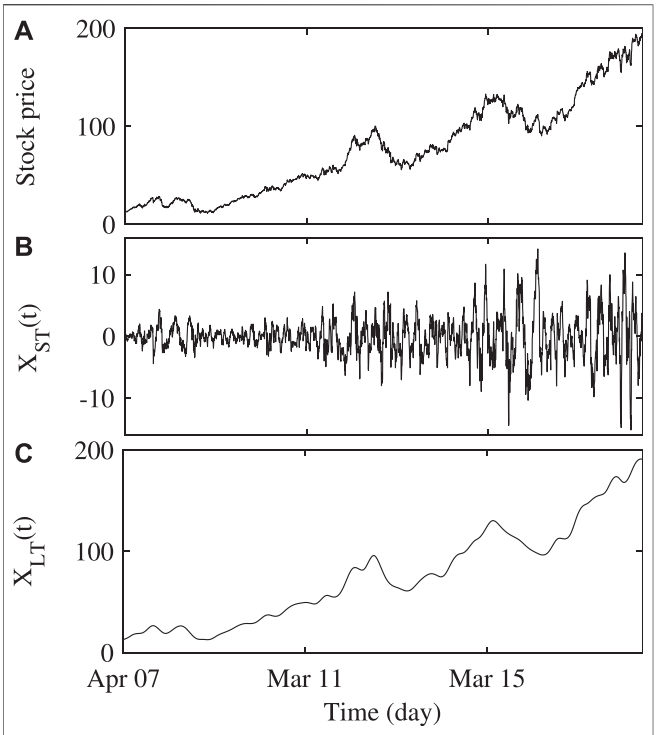


FIGURE 3 | (A) represents the daily price movement of Apple Inc. from April 2007 to March 2018. **(B)**,**(C)** represent the reconstructed short-term time series $[X_{ST}(t)]$ and long-term time series $[X_{LT}(t)]$, respectively.

4.2. Reconstruction of Short-Term and Long-Term Time Series

In order to analyze the market dynamics in short-term *ITH* and long-term *ITH*, we have reconstructed two time series from the decomposed *IMFs* as discussed below.

We have reconstructed a time series $[X_{ST}(t)]$ by adding the IMF_1 to IMF_5 whose $H_1 \approx 0.5 \pm 0.04$, that is, $X_{ST}(t) = \sum_{i=1}^{5} IMF_i$. The time scale of $X_{ST}(t)$ ranges in $3\,D \leq \tau \leq 3\,M$. **Figure 3B** shows the reconstructed time series $X_{ST}(t)$ obtained by decomposing the original time series of Apple Inc. which is shown in **Figure 3A**. $H_1 \approx 0.5 \pm 0.04$ shows that the stock market is random when τ ranging from a few days to 3 months. Hence, $X_{ST}(t)$ represents the short-term time series in $3\,D \leq \tau \leq 3\,M$. The above analysis shows that the market behavior is random in the short-term *ITH* when τ is in the range of a few days to 3 months.

Higher-order *IMF* shows two Hurst exponents (H_1 and H_2). We have reconstructed another time series $[X_{LT}(t)]$ by adding IMF_6 to IMF_9 whose $H_1 \geq 0.75$ and residue, that is, $X_{LT}(t) = \sum_{i=6}^{9} IMF_i +$ residue to understand the market dynamics in long-term *ITH*. The time scales of $X_{LT}(t)$ are $\geq 5\,M$. **Figure 3C** shows the reconstructed long-term time series $[X_{LT}(t)]$ obtained by decomposing the original time series of Apple Inc., which is shown in **Figure 3A**. The present analysis yielded a H_1 value of ≥ 0.75, which shows that the stock market has a long-range correlation with $\tau \geq 5\,M$. Hence, $X_{LT}(t)$ represents the long-term time series with $\tau \geq 5\,M$. From the above analysis, it can be concluded that the market has a long-range correlation in the long-term *ITH* with $\tau \geq 5\,M$ and hence may be used to predict a future price. Further, it has been observed that the future price of a stock is actually much more dependent on the fundamental variables of a company. In order to understand such dependence, we have studied the correlation between $X_{LT}(t)$ and the fundamental variables of the companies.

Table 1 shows that the correlation coefficient (*J*) between $X_{LT}(t)$ and three fundamental variables: sale, net profit (NP), and cash from operating activity (COA) for some Indian and American companies which are listed in NSE, NYSE, and NASDAQ, from March 2007 to March 2018 in the annual price level. Fundamental variables data have been downloaded from screener and macrotrends. We obtained a positive correlation between $X_{LT}(t)$ and sale, NP, and COA for all the years. It implies that stock price is correlated with the sale, NP, and COA. We have obtained a small *J* for a few stocks. These stocks show a small *J* for the following two possible reasons: a) the stock price of a company with strong growth prospect increases even though sale or NP decreases temporarily; b) the stock price of a company with weak growth prospect decreases even though sale or NP increases temporarily. Hence, for long-term investment, the fundamental variables are the most crucial variables for the prediction of the future price. In the future, we would like to study the correlation between stock price and other fundamental variables of companies.

5. CONCLUSION

In this paper, we have studied the stock market using the empirical mode decomposition (*EMD*) based Hurst exponent (*H*) analysis and normalized variance (*NV*) technique. *EMD*

TABLE 1 | Correlation coefficient (*J*) between reconstructed long-term time series $[X_{LT}(t)]$ and three fundamental variables of some Indian and American companies.

Company	Sale	NP	COA
NSE: ASIANPAINT	0.9930	1.0000	0.9441
NSE: BPCL	0.6923	0.8322	0.6853
NSE: CIPLA	0.8951	0.6713	0.6713
NSE: DRREDDY	0.9441	0.8671	0.9510
NSE: EICHERMOT	0.9441	0.9790	0.9860
NSE: GAIL	0.5664	0.5804	0.5804
NSE: GRASIM	0.8462	0.4615	0.4615
NSE: HCLTECH	0.9441	0.9231	0.8951
NSE: HEROMOTOCO	0.9650	0.9580	0.9021
NSE: HINDALCO	0.2587	0.4545	0.3776
NSE: HINDUNILVR	0.9720	0.9860	0.8252
NSE: TATAMOTORS	0.8462	0.7832	0.9091
NSE: RCOM	0.0490	0.9510	0.5035
NSE: RELIANCE	0.1119	0.7203	0.4476
NYSE: JNJ	0.8352	0.2747	0.8022
NASDAQ: AMZN	0.9890	0.4231	0.9890
NASDAQ: GOOGL	0.9231	0.8462	0.9231
NASDAQ: AAPL	0.9615	0.9396	0.9341
NASDAQ: MSFT	0.7253	0.0659	0.6484
NASDAQ: INTC	0.9011	0.5879	0.8681

First column: Sale, second column: NP, and third column: COA.

technique has been applied to decompose the time series in the form of *IMF*. *H* and *NV* have been calculated for all the *IMF* to understand the nature of the market dynamics.

The analysis yielded a value of $H_1 = 0.5 \pm 0.04$ for IMF_1 to IMF_5. A short-term time series $[X_{ST}(t)]$ is reconstructed by adding IMF_1 to IMF_5. The time scale of $X_{ST}(t)$ ranges from a few days to 3 months. The estimated value of $H_1 = 0.5 \pm 0.04$ which shows that the stock market is random in the short-term *ITH*. We have estimated the value of $H_1 = 0.75 \pm 0.04$, $H_1 = 0.90 \pm 0.04$, $H_1 = 0.96 \pm 0.02$, and $H_1 = 0.99 \pm 0.02$ for IMF_6, IMF_7, IMF_8, and IMF_9, respectively, for all the data. $H_1 > 0.5$ shows that the IMF_6 to IMF_9 have long-range correlation, and hence a long-term time series $[X_{LT}(t)]$ is reconstructed by adding IMF_6 to IMF_9 and residue. The time scale of $X_{LT}(t)$ is greater than 5 months. The results show that the market is random with $\tau \leq 3\,M$ and having a long-range correlation with $\tau \geq 5\,M$. The study of the correlation between $X_{LT}(t)$ and sale, net profit, and cash from operating activity of different companies shows that the market is positively correlated with the fundamental variables of a company in long-term *ITH*. Hence, the dynamics of the market may be predicted in long-term *ITH* using fundamental variables.

A detailed study of the market in the long-term *ITH* in terms of fundamental variables of a company is necessary to predict the future price. We believe that the outcome of the present study may help in making investment decisions in both short-term *ITH* and long-term *ITH*.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: https://in.finance.yahoo.com, https://www.screener.in, https://www.macrotrends.net

AUTHOR CONTRIBUTIONS

All the authors have equally contributed to preparing the manuscript.

ACKNOWLEDGMENTS

The authors acknowledge the help of Taraknath Kundu and suggestions of the anonymous reviewers in preparing and improving the manuscript.

REFERENCES

1. Mantegna RN, Stanley HE. *Introduction to econophysics: correlations and complexity in finance.* Cambridge, United Kingdom: Cambridge University Press (1999).
2. Bouchaud J-P, Potters M. *Theory of financial risk and derivative pricing: from statistical physics to risk management.* Cambridge, United Kingdom: Cambridge University Press (1999).
3. Huang NE, Wu M-L, Qu W, Long SR, Shen SS. Applications of Hilbert–Huang transform to non-stationary financial time series analysis. *Appl Stoch Model Bus Ind* (2003) 19:245–68. doi:10.1002/asmb.501
4. Peters EE, Peters ER, Peters D. *Fractal market analysis: applying chaos theory to investment and economics.*, Vol. 24. Hoboken, NJ: John Wiley & Sons (1994).
5. Jegadeesh N, Titman S. Returns to buying winners and selling losers: implications for stock market efficiency. *J Finance* (1993) 48:65–91. doi:10.1111/j.1540-6261.1993.tb04702.x
6. Campbell JY, Viceira LM, Viceira LM. *Strategic asset allocation: portfolio choice for long-term investors (Clarendon lectures in economic)* Oxford, United Kingdom: Oxford University Press (2002).
7. Cremers M, Pareek A, Sautner Z. Short-term investors, long-term investments, and firm value. *Recuperado el* (2017) 15:39.
8. Menkhoff L. The use of technical analysis by fund managers: international evidence. *J Bank Finance* (2010) 34:2573–86. doi:10.1016/j.jbankfin.2010.04.014
9. Mahata A, Bal DP, Nurujjaman M. Identification of short-term and long-term time scales in stock markets and effect of structural break. *Phys Stat Mech Appl* (2019) 12:123612. doi:10.1016/j.physa.2019.123612
10. Lui Y-H, Mole D. The use of fundamental and technical analyses by foreign exchange dealers: Hong Kong evidence. *J Int Money Finance* (1998) 17:535–45. doi:10.1016/s0261-5606(98)00011-4
11. Levy RA. The predictive significance of five-point chart patterns. *J Bus* (1971):316–23. doi:10.1086/295382
12. Samaras GD, Matsatsinis NF, Zopounidis C. A multicriteria dss for stock evaluation using fundamental analysis. *Eur J Oper Res* (2008) 187:1380–401. doi:10.1016/j.ejor.2006.09.020
13. Fama EF, French KR. Size and book-to-market factors in earnings and returns. *J Finance* (1995) 50:131–55. doi:10.2139/ssrn.498243
14. Lakonishok J, Shleifer A, Vishny RW. Contrarian investment, extrapolation, and risk. *J Finance* (1994) 49:1541–78. doi:10.3386/w4360
15. Basu S. Investment performance of common stocks in relation to their price-earnings ratios: a test of the efficient market hypothesis. *J Finance* (1977) 32, 663–82. doi:10.1111/j.1540-6261.1977.tb01979.x
16. Sloan RG. Do stock prices fully reflect information in accruals and cash flows about future earnings? *Acc Rev* (1996) 71:289–315.
17. Fama EF, French KR. The cross-section of expected stock returns. *J Finance* (1992) 47:427–65. doi:10.1108/03074350510770026
18. Barbee WC, Jr, Mukherji S, Raines GA. Do sales–price and debt–equity explain stock returns better than book–market and firm size? *Financ Anal J* (1996) 52:56–60. doi:10.2469/faj.v52.n2.1980
19. Mukherji S, Dhatt MS, Kim YH. A fundamental analysis of Korean stock returns. *Finance Anal J* (1997) 53:75–80. doi:10.2469/faj.v53.n3.2086
20. Bhandari LC. Debt/equity ratio and expected common stock returns: empirical evidence. *J Finance* (1988) 43:507–28. doi:10.1111/j.1540-6261.1988.tb03952.x
21. Hasanov M. Is South Korea's stock market efficient? Evidence from a nonlinear unit root test. *Appl Econ Lett* (2009) 16:163–7. doi:10.1080/13504850601018270
22. Hartman D, Hlinka J. Nonlinearity in stock networks. *Chaos Interdiscipl J Nonlinear Sci* (2018) 28:083127. doi:10.1063/1.5023309
23. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond A Math Phys Eng Sci* (1998) 454:903–95. doi:rspa.1998.0193
24. Huang W, Shen Z, Huang NE, Fung YC. Engineering analysis of biological variables: an example of blood pressure over 1 day. *Proc Natl Acad Sci USA* (1998) 95:4816–21. doi:10.1073/pnas.95.9.4816
25. Chowdhury S, Deb A, Nurujjaman M, Barman C. Identification of pre-seismic anomalies of soil radon-222 signal using Hilbert–Huang transform. *Nat Hazards* (2017) 87:1587–606. doi:10.1007/s11069-017-2835-1
26. Hurst HE. Long-term storage capacity of reservoirs. *Trans Am Soc Civ Eng* (1951) 116:770–99

27. Nurujjaman M, Iyengar AS. Realization of SOC behavior in a dc glow discharge plasma. *Phys Lett* (2007) 360:717–21. doi:10.1016/j.physleta.2006.09.005
28. Carreras B, Van Milligen BP, Pedrosa M, Balbın R, Hidalgo C, Newman D, et al. Self-similarity of the plasma edge fluctuations. *Phys Plasmas* (1998) 5: 3632–43. doi:10.1063/1.873081
29. Chatlani N, Soraghan JJ. EMD-based filtering (EMDF) of low-frequency noise for speech enhancement. *IEEE Trans Audio Speech Lang Process* (2012) 20: 1158–66. doi:10.1109/tasl.2011.2172428
30. Zao L, Coelho R, Flandrin P. Speech enhancement with emd and hurst-based mode selection. *IEEE/ACM Trans Audio Speech Lang Process* (2014) 22: 899–911. doi:10.1109/taslp.2014.2312541

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Nurujjaman and Mahata. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 24 November 2020
doi: 10.3389/fphy.2020.00366

Analyses of Daily Market Impact Using Execution and Order Book Information

Kenta Yamada[1]* *and Takayuki Mizuno*[2,3]

[1] Faculty of Global and Regional Studies, University of the Ryukyus, Nishihara, Japan, [2] Research Center for Financial Smart Data, National Institute of Informatics, Tokyo, Japan, [3] Department of Informatics, SOKENDAI (The Graduate University for Advanced Studies), Tokyo, Japan

We analyzed the Tokyo Stock Exchange (TSE) for a 29-month period from August 2014 to December 2016, including every transaction and order book snapshot, and confirmed through a simple statistical test that the market impact depends on each stock. Based on a correlation analysis, we found that the market impact slowly changes over time. From an order book analysis, negative correlations were found between the market impact and the averaged limit order volumes in the order book. We also clarified that one of the factors of market impact is the volume of limit orders in the order book.

Keywords: financial markets, stock price, market impact, market microstructure, order book analysis

OPEN ACCESS

Edited by:
Mahdi Jalili,
RMIT University, Australia

Reviewed by:
Gholamreza Jafari,
Shahid Beheshti University, Iran
Yuyao Wang,
Nanjing University of Science and Technology, China

***Correspondence:**
Kenta Yamada
k-yamada@grs.u-ryukyu.ac.jp

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *13 April 2020*
Accepted: *29 July 2020*
Published: *24 November 2020*

Citation:
Yamada K and Mizuno T (2020) Analyses of Daily Market Impact Using Execution and Order Book Information. Front. Phys. 8:366. doi: 10.3389/fphy.2020.00366

1. INTRODUCTION

Many previous studies have addressed market impact from theoretical [1–8] and empirical [9–13] viewpoints since it is related to market efficiency and trading costs, especially for trading large volumes.

The analyses of market impact using order book data has since the 2000s gradually become popular [14–18] thanks to the provision of datasets of the order book from stock exchange markets such as the London Stock Exchange (LSE), New York Stock Exchange (NYSE), and Tokyo Stock Exchange (TSE) and developments of computer technology that enable us to analyze huge size of order book data.

These studies mainly focused on short time scale market impact such as a single transaction, a few seconds, or a few minutes. It is useful to discuss how large volume order affects the market price over a long time scale. However long time scale of it, such as daily market impact, has not been studied enough compared to the short time scale. In order to estimate transaction costs, daily market impact analysis is important for asset management companies that manage their position daily.

In a previous study of a historical dataset of the Tokyo Stock Exchange (TSE) for a 17-month period from August 2014 to December 2015 [19], we identified the following two relationships: (i) a proportional relationship between the return of the market price and the order imbalance of the executed volume and (ii) an inverse proportional relationship between the market impact and the averaged order book volume. We also introduced a transaction cost model by calculating the spread cost and applying the relationship between the return and the order imbalance of execution.

We focused on the TOPIX Core 30 where the issues have very high market capitalization and liquidity, such as Toyota and Sony. However, in this study, we analyzed not only the TOPIX Core 30 but also such smaller stocks as the TOPIX Large 70 and Mid 400 and identified the following results about market impact: (i) it depends on issues; (ii) it slowly changes over time; and (iii) one of its factors is order book volume. These findings are useful to estimate transaction costs.

Our paper is organized as follows. In section 2, we present the empirical analysis results of the execution and order book data. Section 3 is our conclusion.

2. EMPIRICAL ANALYSIS

We analyzed the Tokyo Stock Exchange data for a 29-month period from August 2014 to December 2016, which recorded every transaction and every order book from the best price (0-th level) to the 7-th level for the TOPIX Core 30, the Large 70, and the Mid 400. The datasets can be purchased from Japan Exchange Group (https://www.jpx.co.jp/english/markets/paid-info-equities/historical/01.html).

In this section, the dataset is analyzed, and we show the market impact depends on each stock by a simple statistical test. We clarify the market impact slowly changes over time based on a correlation analysis of it. Additionally, dependencies of market impact on order book volume are investigated.

Figure 1 shows an example of Sony's market price (stock code: 6758) in the upper figure and the cumulative transaction volume of deal-ask (a market order hits the best-ask price) and deal-bid (a market order hits the best-bid price) in the lower figure from 1 second past 9 a.m. to 1 second before 3 p.m. on June 21, 2016. Execution of the opening and closing transactions usually tends to have very large volume. We focused on continuous sessions in this study, removed the opening and closing transactions, and defined the return of the market price for 1 day:

$$r = \frac{P_e - P_s}{P_s}, \tag{1}$$

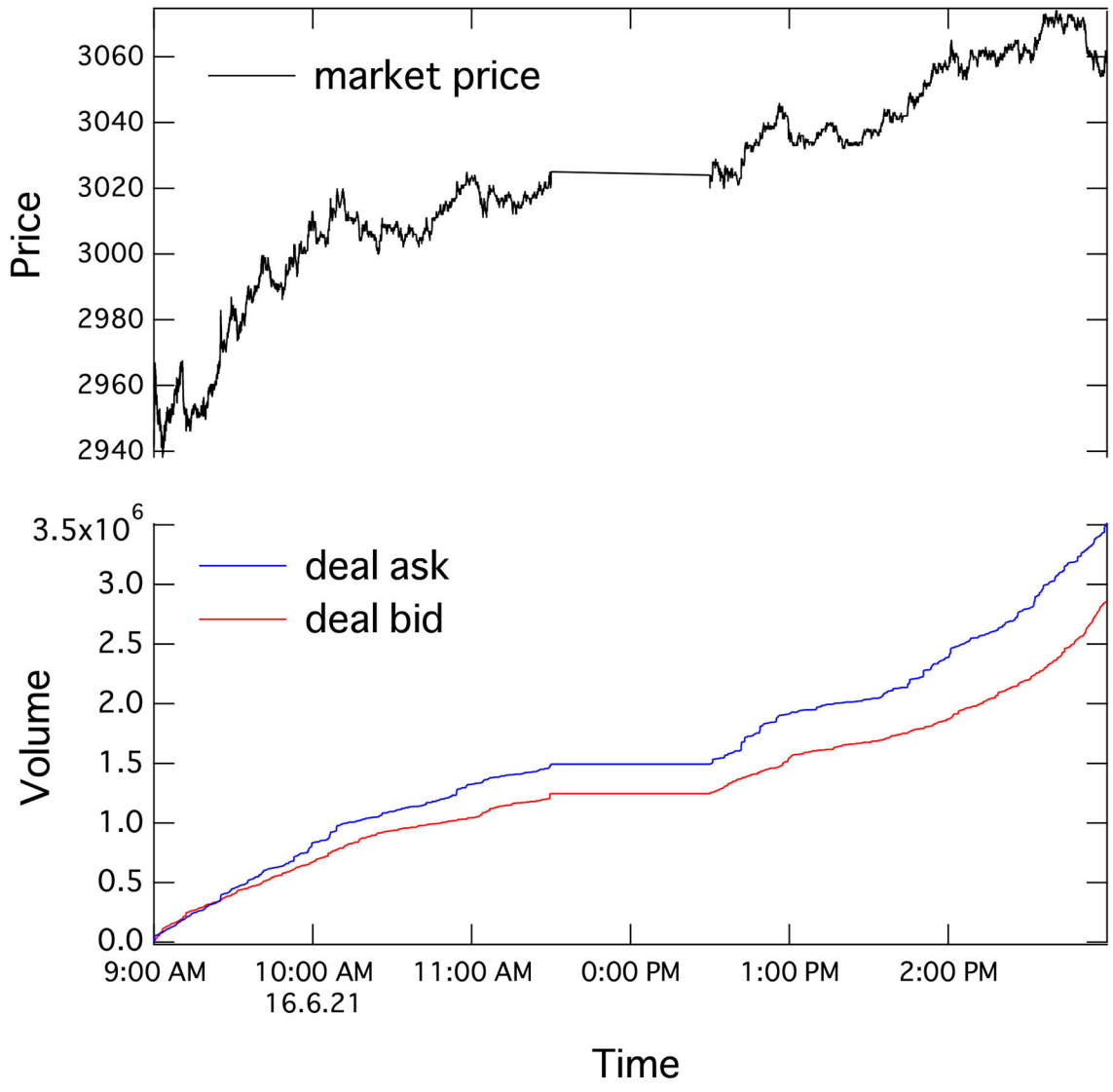


FIGURE 1 | Time series of market price and cumulative transaction volume of Sony (stock code: 6758) from 1 second past 9 a.m. to 1 second before 3 p.m. on June 21, 2016.

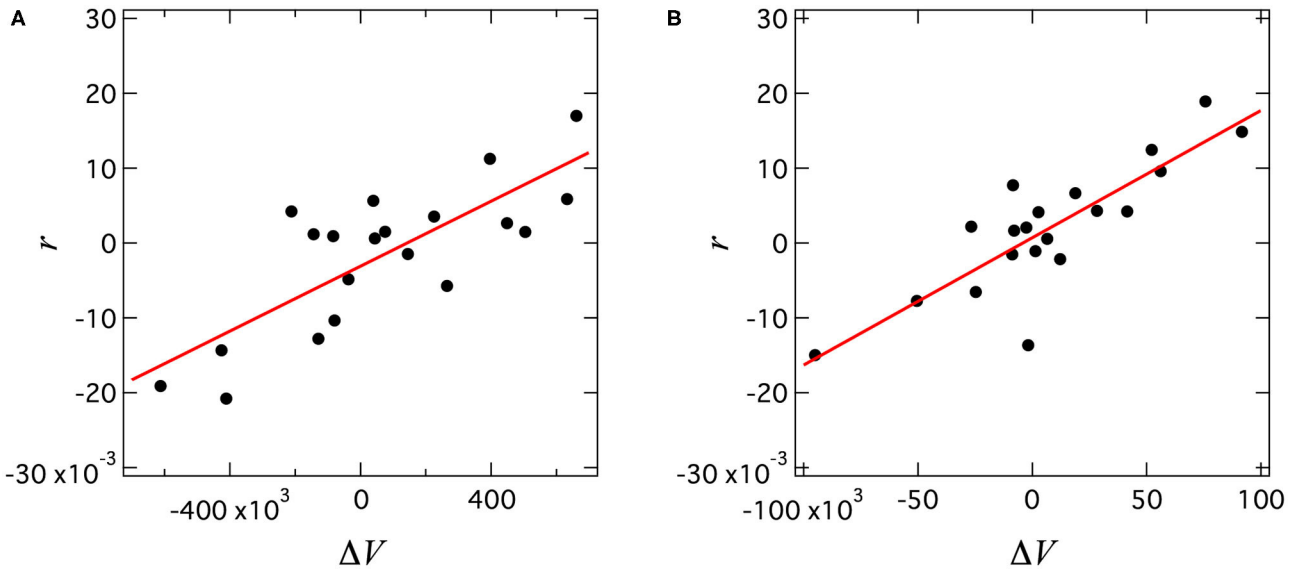


FIGURE 2 | Scatter plots between return (r) and executed order imbalance (ΔV) from October 1 to 31, 2016. **(A)** Sony corporation (stock code: 6758) in Core 30, **(B)** Nippon Paper Industries Co., Ltd. (stock code: 3863).

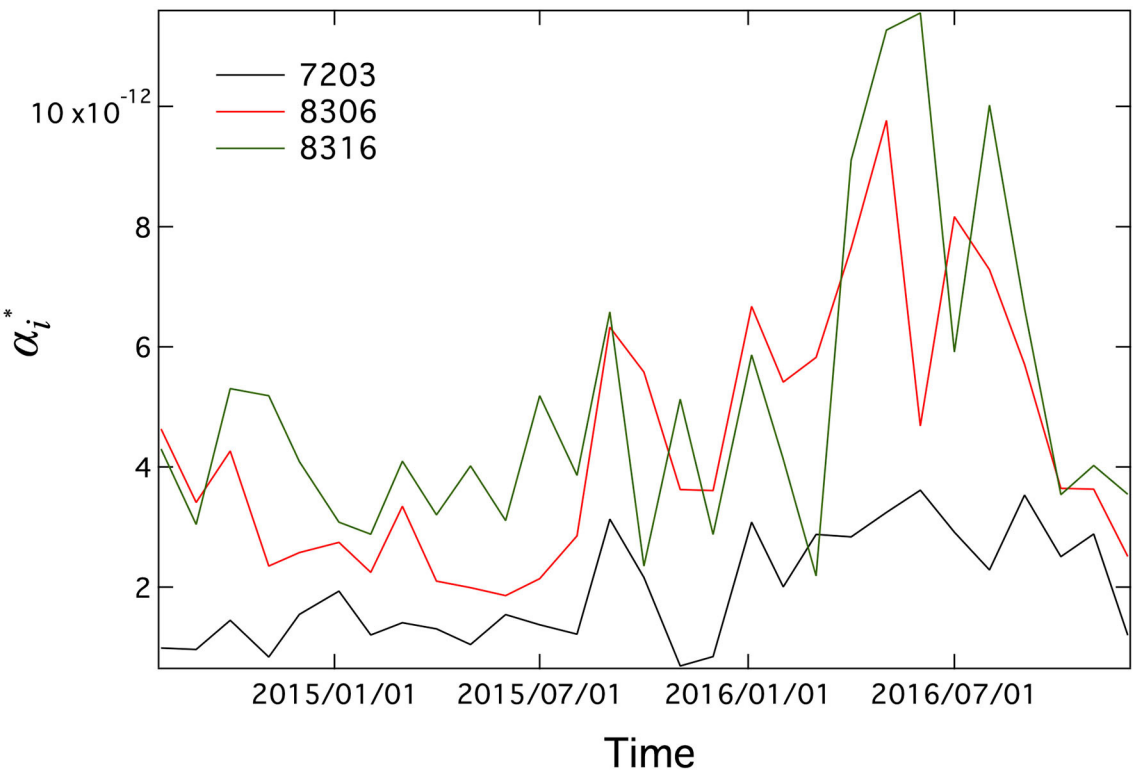


FIGURE 3 | Monthly time series of α_i^* during 29-month period for three companies: Toyota Motor Corp. (7203), Mitsubishi UFJ Financial Group, Inc. (8306), and Sumitomo Mitsui Financial Group, Inc. (8316).

where P_s and P_e are the prices at 9:00:01 and 14:59:59. The executed order imbalance (ΔV) is calculated as the cumulative deal-ask minus deal-bid 1 s before 3 p.m.

Figure 2 shows a scatter plot of the return (r) and the executed order imbalance (ΔV) of the Sony Corporation (stock code: 6758) in the Core 30 and Nippon Paper Industries Co., Ltd.

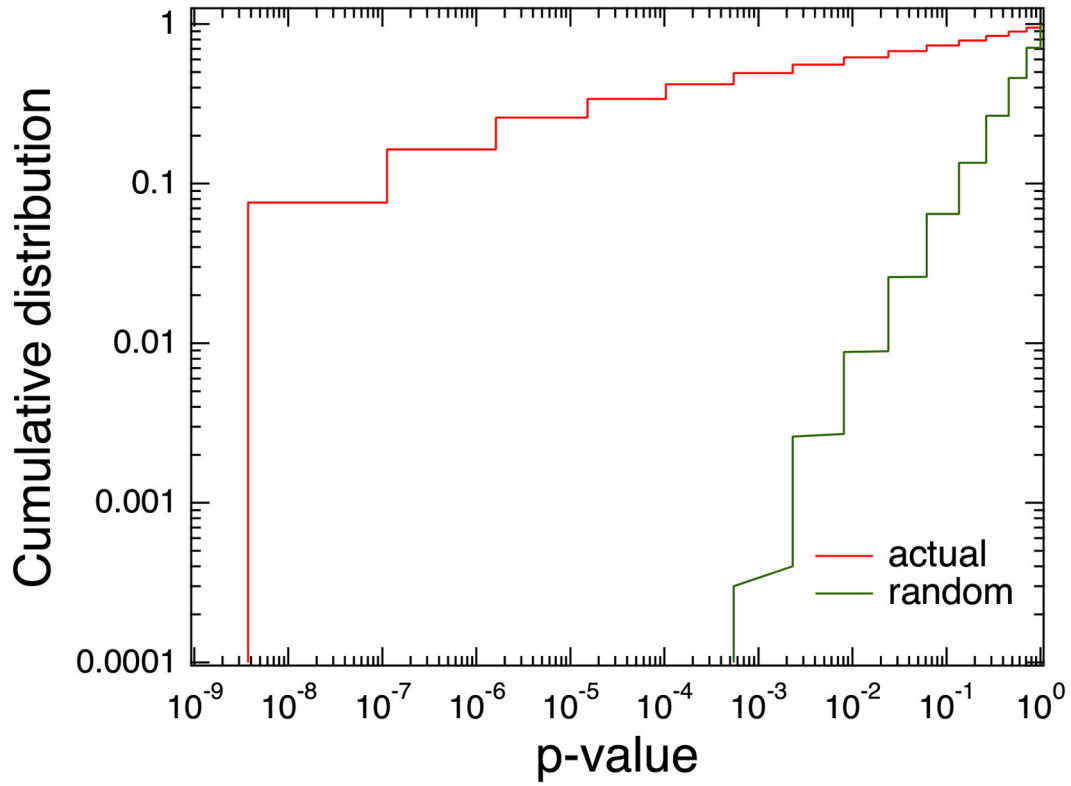


FIGURE 4 | Cumulative distributions of *p*-values for Core 30, Large 70, and Mid 400.

(stock code : 3863) in the Mid 400 from October 1 to 31, 2016. Each point is calculated by the data of a single day. We found a clear linear correlation between return (r) and executed order imbalance (ΔV):

$$r = \alpha_l \Delta V + \beta_l, \tag{2}$$

where gradient α is a measurement of market impact. In the case of Sony Corporation (stock code: 6758), $\alpha_l = 2.17 \times 10^{-8}$, standard error of α_l is 4.07×10^{-9}, and p-value for α_l is 4.49×10^{-5}. In the case of Nippon Paper Industries Co., Ltd. (stock code: 3863), $\alpha_l = 1.70 \times 10^{-7}$, standard error is 2.59×10^{-8}, and p-value is 3.78×10^{-6}.

To statistically check the existence of the differences of the market impact among stocks, we define the market impact for one yen as follows:

$$\alpha_i^* = \frac{\alpha_{l,i}}{P_i}, \tag{3}$$

where $\alpha_{l,i}$ describes the market impact for unit volume, P_i is the last price in a month, and i is the index of the stock. We show examples of monthly time series of α_i^* during 29-month period for three companies in **Figure 3**. We compare the monthly market impact, α_i^*, of companies A and B, and execute the a binomial test. A null hypothesis is the market impacts between company A and B is the same. In other words the probability where the market impact of company A is larger than B is half. If company A has a larger market impact n_A times than B, on the other hand company B has a larger market impact n_B times than A, a p-value is calculated:

$$p_{val} = \begin{cases} 2 \times \sum_{k=0}^{\min(n_A,n_B)} \binom{n}{k} \left(\frac{1}{2}\right)^n & (n_A \neq n_B) \\ 1 & (n_A = n_B) \end{cases} \tag{4}$$

where two times describes a two-tailed test and $\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$.

We applied a statistical test for all the combinations of Core 30, Large 70, and Mid 400. We are interested in small p-values and show a cumulative distribution of the p-values in a log-log scale (**Figure 4**) and found the actual data have a much bigger proportion of small p-values than a random case. For example, in the actual case, about 68% of the p-values are <0.05, implying the market impact depends on the issues.

To check the robustness of the market impact size, we show a scatter plot of the $\log_{10} \alpha^*$ at the M-th and $M+1$-th months in **Figure 5**. We added a median as well as 25 and 75% percentiles in each box and identified a clear correlation. The correlation coefficients have large values (0.67) in the Pearson product-moment correlation case and 0.71 in the Spearman rank-order correlation case.

We also checked the ΔM dependencies for the correlation shown in **Figure 6**. The correlation decreases until $\Delta M = 12$ (1 year) and saturates. For the Spearman rank-order correlation, it is

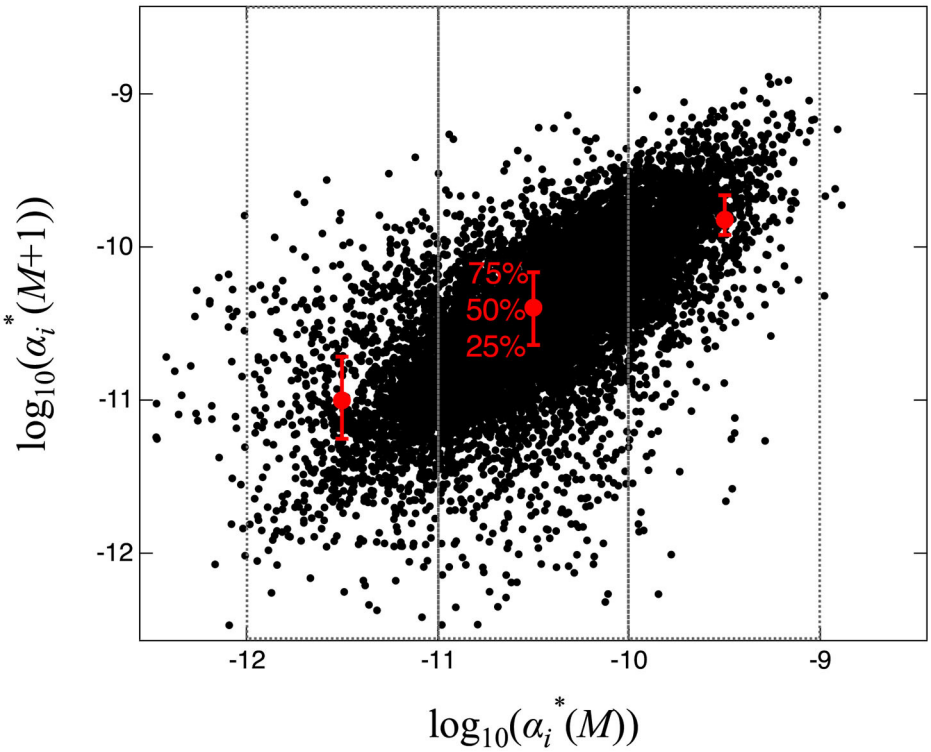


FIGURE 5 | Scatter plot of log α^* at M-th and $M+1$-th months.

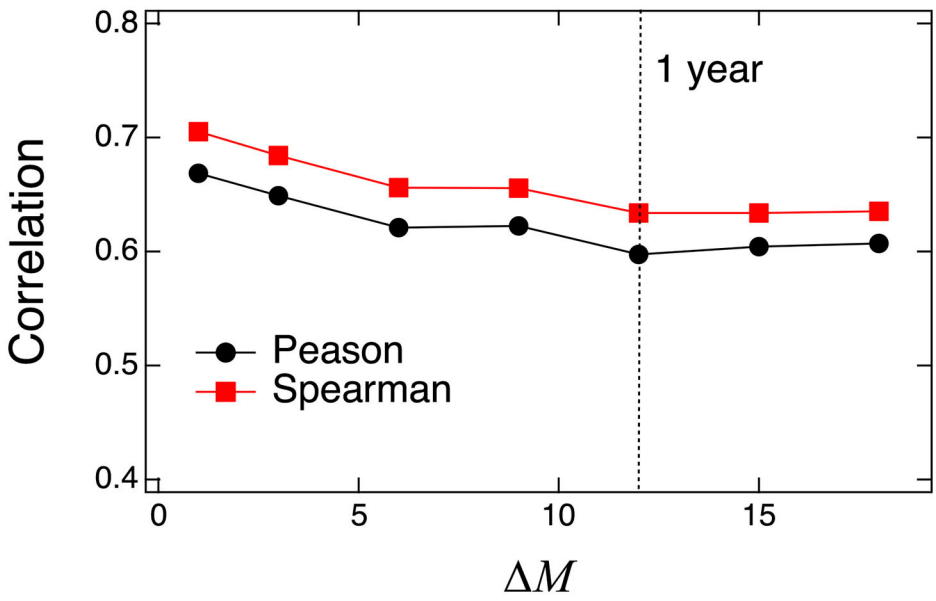


FIGURE 6 | Pearson product-moment correlation and Spearman rank-order correlation for log(α_i^*) at M and ΔM months.

approximately 0.64 when $\Delta M = 18$. The market impact, which does not change drastically, can be estimated using past market impact values.

Next we focused on the relationships between the market impact characterized by α_i^* and the order book volume. We defined the i-th best-ask and bid-side volumes at t as $V_{a,i}(t)$ and $V_{b,i}(t)$. For example, the 0th (best) ask-side volume at t is written as $V_{a,0}(t)$. and the average of both the bid- and ask-sides order book volumes from the 0-th to the 7-th levels at $t = t_s$ are given:

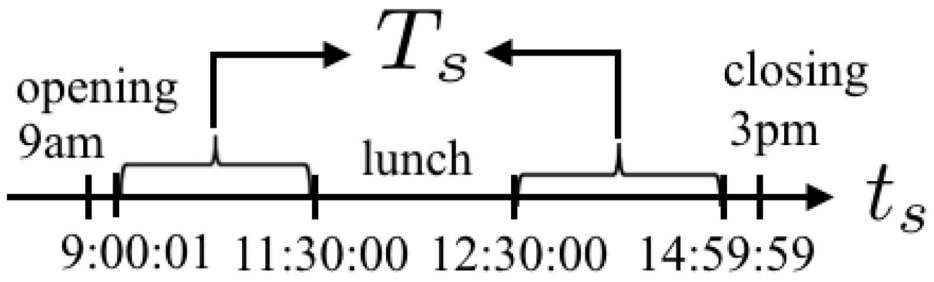


FIGURE 7 | Illustration of T_s.

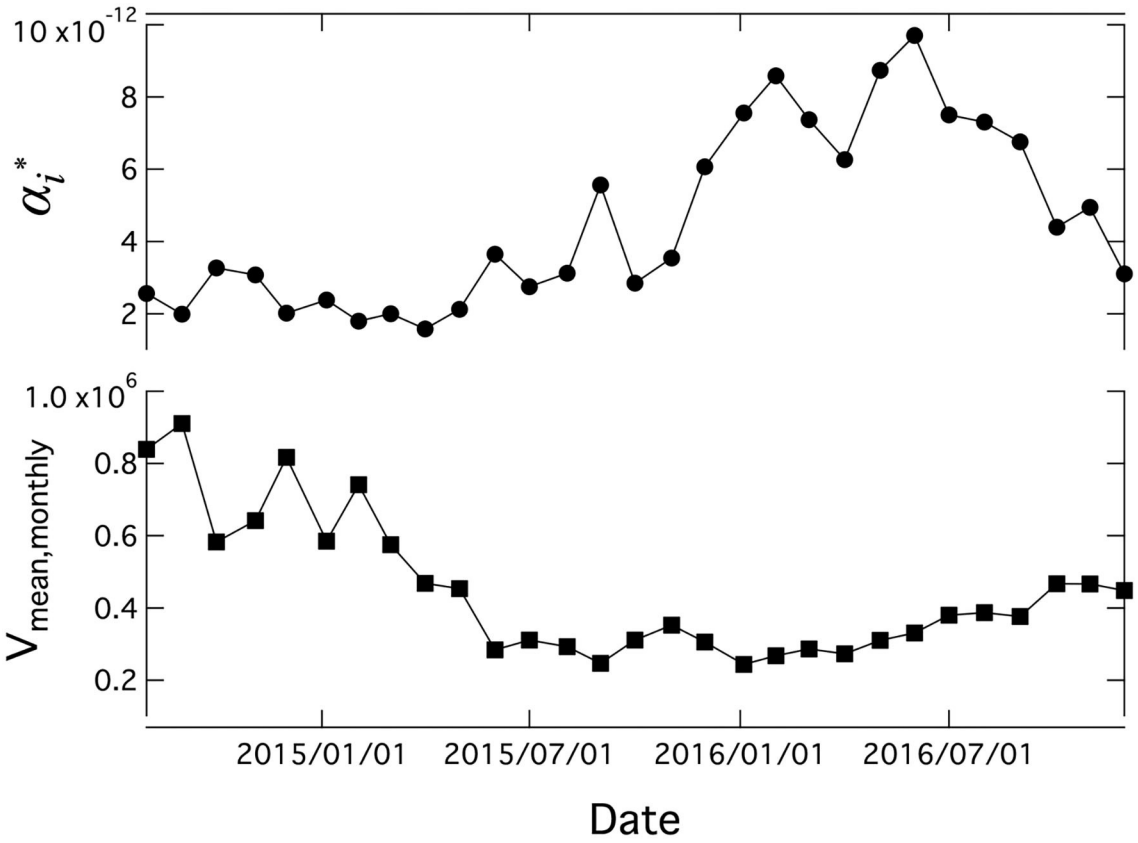


FIGURE 8 | Time series of monthly gradient (α_i^*) of Mizuho Financial Group, Inc. (stock code: 8411) and averaged order book volume ($V_{\text{mean, monthly}}$) for a 29-month period from August 2014 to December 2016.

$$V_m(t_s) = \frac{1}{2}\,\mathbb{E}_{i\in\{0,\cdots,7\}}\left(V_{b,i}(t_s) + V_{a,i}(t_s)\right), \tag{5}$$

where $\mathbb{E}$ represents the expectation value over the samples. We then defined the daily averaged order book volume:

$$V_{\text{mean,daily}}(t_d) = \mathbb{E}_{t_s\in T_s}\left(V_m(t_s)\right). \tag{6}$$

Here we ignore the opening (9 a.m.), closing (3 p.m.), and lunch (11:30 a.m.–0:30 p.m.) times (**Figure 7**) and finally calculate the monthly averaged order book volume:

$$V_{\text{mean,monthly}}(t_m) = \mathbb{E}_{t_d\in T_d}\left(V_m(t_d)\right), \tag{7}$$

where T_d represents trading days of TSE market in 1 month (excluding weekends and holidays).

The upper time series in **Figure 8** shows the monthly α_i^* of the Mizuho Financial Group, Inc. (stock code: 8411), and the lower time series shows the monthly averaged order book volume defined in Equation (7). We observed a negative correlation between these two values.

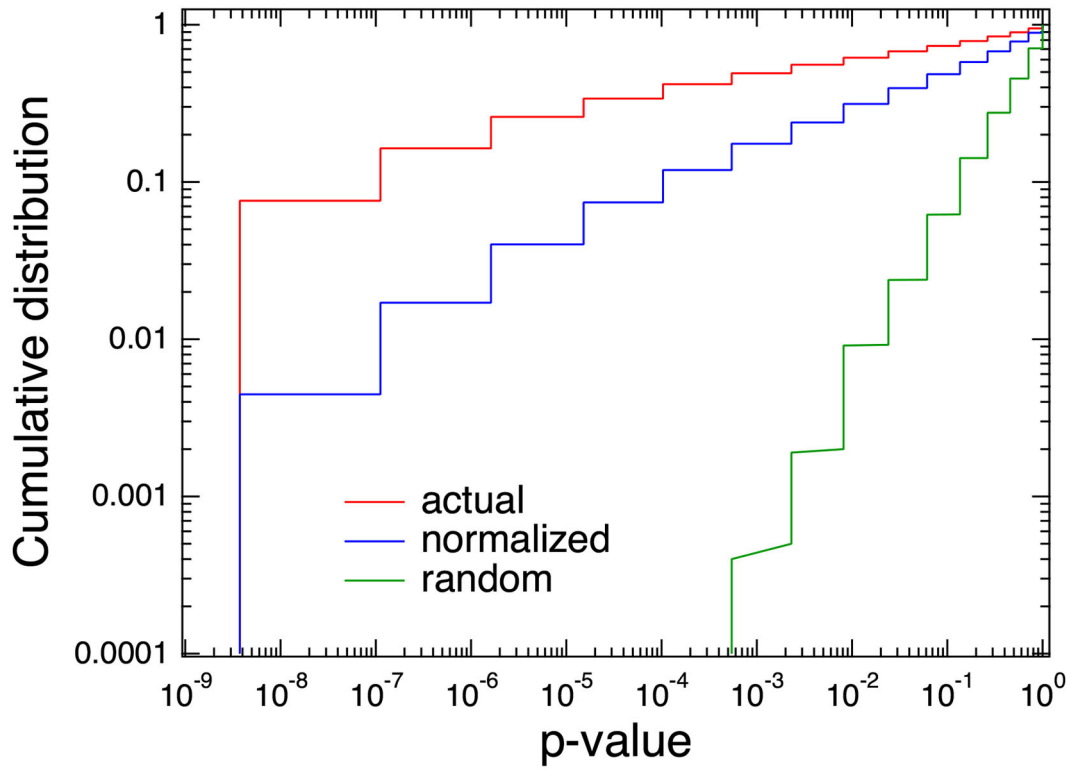


FIGURE 9 | Comparison cumulative distributions of *p*-values among actual, random, and normalized by order book volume.

One reason for the dependencies is the order book volume; if we normalized the market impact by order book volume, the cumulative distribution of the *p*-values shifted more to random cases than to unnormalized cases (**Figure 9**). Here, we focused on order book volumes; if we normalized the market impact by such other effects as order flow and the configuration of the stock holders, the *p*-value distribution would probably shift more to the random case.

3. CONCLUSION

We showed that the daily market impact depends on stocks by executing a binomial test for the Tokyo Stock Exchange (TSE) during a 29-month period from August 2014 to December 2016. We focused on a longer time scale (such as 1 day) than previous studies that used a single transaction or 5 min. We verified the market impact's robustness and found it does not change drastically and realized that we can estimate it using past market impact values. However, we need to pay attention to big news, such as the Lehman shock, that change market conditions drastically, causing sudden variation in terms of market impact. We also investigated the dependencies of the market impact on order book volume and concluded that if a stock has a larger order book volume, the market impact tends to be smaller. We expect our findings to be useful for discussions about controlling transaction costs.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The datasets analyzed for this study can be purchased from Japan Exchange Group (https://www.jpx.co.jp/english/markets/paid-info-equities/historical/01.html).

AUTHOR CONTRIBUTIONS

KY and TM carried out the conceptualization, methodology, investigation, and validation. KY carried out the formal analysis and prepared and wrote the original draft. TM carried out the project administration and funding acquisition. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was partly supported by JST (Japan Science and Technology Agency) Precursory Research for Embryonic Science and Technology, Grant Number JPMJPR14DA (KY).

REFERENCES

1. Demsetz H. The cost of transacting. *Q J Econ.* (1968) 82:33–53. doi: 10.2307/1882244
2. Scholes MS. The market for securities: substitution versus price pressure and the effects of information on share prices. *J Bus.* (1972) 45:179–211.
3. Kraus A, Stoll HR. Price impacts of block trading on the New York stock exchange. *J Finance.* (1972) 27:569–88. doi: 10.2307/2978262
4. Stoll HR. The supply of dealer services in securities markets. *J Finance.* (1978) 33:1133–51. doi: 10.2307/2326945
5. Kyle AS. Continuous auctions and insider trading. *Econometrica.* (1985) 53:1315–35. doi: 10.2307/1913210
6. Obizhaeva A, Wang J. *Optimal Trading Strategy and Supply/Demand Dynamics.* National Bureau of Economic Research (2005). Available online at: http://www.nber.org/papers/w11444.
7. Gatheral J. No-dynamic-arbitrage and market impact. *Quant Finance.* (2010) 10:749–59. doi: 10.1080/14697680903373692
8. Kyle AS, Obizhaeva AA. Market microstructure invariance: empirical hypotheses. *Econometrica.* (2016) 84:1345–404. doi: 10.3982/ECTA10486
9. Hasbrouck J. Measuring the information content of stock trades. *J Finance.* (1991) 46:179–207. doi: 10.2307/2328693
10. Kempf A, Korn O. Market depth and order size1. *J Financ Markets.* (1999) 2:29–48. doi: 10.1016/S1386-4181(98)00007-X
11. Evans MDD, Lyons RK. Order flow and exchange rate dynamics. *J Polit Econ.* (2002) 110:170–80. doi: 10.1086/324391
12. Breen WJ, Hodrick LS, Korajczyk RA. Predicting equity liquidity. *Manage Sci.* (2002) 48:470–83. doi: 10.1287/mnsc.48.4.470.210
13. Lillo F, Farmer JD, Mantegna RN. Econophysics: master curve for price-impact function. *Nature.* (2003) 421:129–30. doi: 10.1038/421129a
14. Weber P, Rosenow B. Order book approach to price impact. *Quant Finance.* (2005) 5:357–64. doi: 10.1080/14697680500244411
15. Cont R, Kukanov A, Stoikov S. The price impact of order book vents. *J Financ Econ.* (2014) 12:47–88. doi: 10.1093/jjfinec/nbt003
16. Eisler Z, Bouchaud JP, Kockelkoren J. The price impact of order book events: market orders, limit orders and cancellations. *Quant Finance.* (2012) 12:1395–419. doi: 10.1080/14697688.2010.528444
17. Taranto DE, Bormetti G, Bouchaud JP, Lillo F, Toth B. Linear models for the impact of order flow on prices. II. The Mixture Transition Distribution model. *Quant Finance.* (2018) 18:917–31. doi: 10.1080/14697688.2017.1397283
18. Taranto DE, Bormetti G, Bouchaud JP, Lillo F, Toth B. Linear models for the impact of order flow on prices. I. History dependent impact models. *Quant Finance.* (2018) 18:903–15. doi: 10.1080/14697688.2017.1395903
19. Yamada K, Mizuno T. Relationships between market impact characteristics and order book properties. In: *2017 IEEE International Conference on Big Data*, Boston: Big Data. (2017). p. 3156–61.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yamada and Mizuno. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 27 November 2020
doi: 10.3389/fphy.2020.565372

Model of Continuous Random Cascade Processes in Financial Markets

Jun-ichi Maskawa[1]* *and Koji Kuroda*[2]

[1]Department of Economics, Seijo University, Tokyo, Japan, [2]Graduate School of Integrated Basic Sciences, Nihon University, Tokyo, Japan

This article presents a continuous cascade model of volatility formulated as a stochastic differential equation. Two independent Brownian motions are introduced as random sources triggering the volatility cascade: one multiplicatively combines with volatility; the other does so additively. Assuming that the latter acts perturbatively on the system, the model parameters are estimated by the application to an actual stock price time series. Numerical calculation of the Fokker–Planck equation derived from the stochastic differential equation is conducted using the estimated values of parameters. The results reproduce the probability density function of the empirical volatility, the multifractality of the time series, and other empirical facts.

Keywords: Fokker–Planck equation, intermittency, Langevin equation, Markovian process, multifractal

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Haroldo V. Ribeiro,
State University of Maringá, Brazil
Joachim Peinke,
University of Oldenburg, Germany

***Correspondence:**
Jun-ichi Maskawa
maskawa@seijo.ac.jp

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *25 May 2020*
Accepted: *29 October 2020*
Published: *27 November 2020*

Citation:
Maskawa J and Kuroda K (2020) Model of Continuous Random Cascade Processes in Financial Markets. Front. Phys. 8:565372. doi: 10.3389/fphy.2020.565372

INTRODUCTION

In financial time series, past coarse-grained measures of volatility correlate better to future fine-scale volatility than the reverse process. Such a causal structure of financial time series was first reported by Müller et al. [1]. Since then, the causal structure between time scales, the flow of information from a long-term to a short-term scale, was investigated empirically in financial markets; it has been supported by multiple studies [2, 3] as a stylized fact of financial time series [4]. The asymmetric flow of information resembles an energy cascade found in conditions of turbulence. In a developed turbulent flow, the energy injected from the outside at macroscopic spatial scales is transferred to smaller scales and finally dissipated as heat at microscopic spatial scales [5–9]. Gashghaie et al. investigated details of the self-similar transformation rule of the probability density function of price fluctuations and the nonlinear scaling law of the structure function (nth moment of fluctuations), signifying the multifractality of the time series, in their study of the time series of foreign exchange. They pointed out the similarity of price changes in the financial time series to the velocity difference between two spatial points in turbulence [10, 11]. The intermittency in turbulence is a phenomenon characterized by the sudden temporal change of the statistical feature of fluctuations and the spatial coexistence of large and small fluctuations. Such intermittency, which is frequently encountered in heterogeneous complex systems, is well known in financial markets as volatility clustering [4, 12]. Intermittency at each time scale produces a characteristic hierarchical structure designated as multifractality [8, 9].

In the developed turbulence, the process by which mechanically generated vortices on a macroscale deform and destabilize according to the Navier–Stokes equation and then split into smaller vortices is regarded as an energy cascade. A similar idea of modeling multifractal time series by a recursive random multiplication process from a coarse-grained scale to a microscopic scale has offered an attractive means of describing financial time series [13, 14]. Chen et al. verified the

statistics of multiplier factors in the random multiplication process of turbulent flow by empirical studies using measured data and numerical experiments of Navier–Stokes equations [15]. Results show that the multiplier factors connecting two adjacent layers follow a Cauchy distribution in which all moments diverge and show that they are not independent. They show strongly negative correlation between the multiplier factors of adjacent layers. The authors verified the statistics of multipliers calculated backward from actual stock price fluctuations, finding a Cauchy distribution of multiplier factors and also the strongly negative correlation between the multiplier factors in financial markets. Results show that the discrete cascade model using the random multiplication process did not reproduce the statistical property of the multiplier factors. Therefore, as an alternative model, a discrete random multiplicative cascade process with additional additive stochastic processes [16–18], or a model formulated as the Fokker–Planck equation considering the cascade process as a continuous Markov process [19–23] was proposed. Those models have been applied to stock market or foreign exchange market data, yielding empirical results including the statistics of multipliers.

This study examines a continuous cascade model of volatility formulated as a stochastic differential equation including two independent modes of Brownian motion: one has multiplicative coupling with volatility; the other has additive coupling as in the discrete random multiplicative cascade process with additional additive stochastic processes described above. The model parameters are estimated by its application to the stock price time series. Numerical calculation of the Fokker–Planck equation derived from the stochastic differential equation is conducted using the estimated values of parameters resulting in successful reproduction of the pdf of the empirical volatility and the multifractality of the time series.

MATERIALS AND METHODS

Continuous Random Cascade Model

Stochastic Differential Equation

These analyses examine the following wavelet transform of the variation of the logarithmic stock price denoted by $Z(t) = \log S(t)/S(0), t \in [0, L]$:

$$W_\psi Z[u,s] = \int_{-\infty}^{+\infty} Z(t)\frac{1}{s}\psi^*\left(\frac{t-u}{s}\right)dt, \quad u \in [0,L], \tag{1}$$

where the function ψ is designated as the analyzing wavelet. When using the delta function $\psi(t) = \delta(t-1) - \delta(t)$ as the analyzing wavelet, the wavelet transform $W_\psi Z[u,s] = Z(u+s) - Z(u)$ is exactly the logarithmic return of the period s. Here, we use the second derivative of the Gaussian functions as

$$\psi(t) = \frac{d^2}{dt^2}\left(e^{-\frac{t^2}{2}}\right) = \left(t^2-1\right)e^{-\frac{t^2}{2}}. \tag{2}$$

In general, by using the nth derivative of the function having asymptotic fast decay as the analyzing wavelet, one can remove the local trend of mth order $(m \le n-1)$ because the function is orthogonal to mth-order polynomials. For the second derivative of the Gaussian functions, the linear trends of $Z(t)$ with scale s have been eliminated in the wavelet transform $W_\psi Z[u,s]$.

In actual financial market, the price fluctuation is nonstationary and the volatility is not observable. The quantity used herein is the absolute value of the wavelet transform $x(\lambda) = \left|W_\psi Z[t_0, s(\lambda)]\right|$ for arbitrary t_0 as a volatility proxy, where we use the variable $\lambda = \log L/s$. The quantity $x(\lambda)$ is thought to be a generalization of empirical volatility, whereas the wavelet transform $W_\psi Z[u,s]$ is exactly the absolute value of logarithmic return when we use $\psi(t) = \delta(t-1) - \delta(t)$.

The following stochastic equation is used to start.

$$x(\lambda + d\lambda) = x(\lambda)\cdot e^{\sigma dB(\lambda) + \mu d\lambda} \quad (d\lambda > 0) \tag{3}$$

In that equation, $B(\lambda)$ represents the Brownian motion. **Equation 3** expresses that the value of the quantity $x(\lambda + d\lambda)$ at scale $\lambda + d\lambda$ is obtained stochastically from $x(\lambda)$ at just a slightly larger scale λ by multiplying the stochastic variable $W(\lambda, \lambda + d\lambda) = e^{\sigma dB(\lambda) + \mu d\lambda}$. The stochastic multiplier $W(\lambda, \lambda + d\lambda)$ follows a logarithmic normal distribution $LN(\mu d\lambda, \sigma^2 d\lambda)$ because $dB(\lambda) \sim N(0, d\lambda)$. One can derive the following stochastic differential equation using $dB(\lambda)^2 = d\lambda$ as

$$\begin{aligned} dx(\lambda) &= x(\lambda + d\lambda) - x(\lambda) \\ &= x(\lambda)\cdot\left(e^{\sigma dB(\lambda) + \mu d\lambda} - 1\right) \\ &= x(\lambda)\cdot\left[\sigma dB(\lambda) + \left(\mu + \frac{1}{2}\sigma^2\right)d\lambda\right]. \end{aligned} \tag{4}$$

The solution is obtained easily using Ito's formula as [24].

$$x(\lambda) = x(0)\cdot e^{\sigma B(\lambda) + \mu\lambda}. \tag{5}$$

The power law behavior of the qth moment $E[x(\lambda)^q]$ (qth structure function) as a function of scale s is proved by the solution of **Eq. 5** as follows:

$$\begin{aligned} E[x(\lambda)^q] &= E[x(0)^q]\exp\left\{\mu\lambda q + \frac{1}{2}\sigma^2\lambda q^2\right\} \\ &= E[x(0)^q]\left(\frac{s}{L}\right)^{-\mu q - \frac{1}{2}\sigma^2 q^2} \end{aligned} \tag{6}$$

The multifractality of signal $Z(t)$ for which the wavelet transform follows the stochastic **Eq. 4** is verified because the scaling exponent $\tau(q) = -\mu q - \frac{1}{2}\sigma^2 q^2 - 1$ is a convex upward nonlinear function. However, in this model, the stochastic multiplier $W(\lambda_2, \lambda_1) = x(\lambda_2)/x(\lambda_1)$ $(\lambda_2 \le \lambda_1)$ linking two scales follows the logarithmic normal distribution $LN(\mu(\lambda_1 - \lambda_2), \sigma^2(\lambda_1 - \lambda_2))$. It is independent of the multiplier $W(\lambda_3, \lambda_2)$ $(\lambda_3 \le \lambda_2)$ linking two adjacent scales. That result is contrary to the empirical results described in *Introduction*.

We introduce an additional additive stochastic process as we have done in the discrete cascade model. We first consider the following stochastic differential equation.

$$dx(\lambda) = x(\lambda)\cdot\left(-\gamma_M d\lambda + \sigma_M dB_M(\lambda)\right) + a_A(\lambda)d\lambda + b_A(\lambda)dB_A(\lambda) \tag{7}$$

The equation is produced on the assumption that Brownian motions $dB_M(\lambda)$ and $dB_A(\lambda)$ are mutually independent. The

first two terms correspond to **Eq. 4**. The origin of those random sources triggering volatility cascade in financial markets remains unclear.

To solve the stochastic differential **Eq. 7**, we consider the following stochastic differential equation:

$$dw(\lambda) = w(\lambda) \cdot \left(-\gamma_M d\lambda + \sigma_M dB_M(\lambda) \right), \tag{8}$$

which is the same as **Eq. 4**. Using the solution of **Eq. 8**

$$w(\lambda) = w(0) \cdot \exp\left\{ -\left(\gamma_M + \frac{1}{2}\sigma_M^2 \right)\lambda + \sigma_M B_M(\lambda) \right\}, \tag{9}$$

the solution of **Eq. 7** is expressed as shown below:

$$x(\lambda) = w(\lambda) \cdot \left(\int_0^{\lambda} \frac{a_A(u)}{w(u)} du + \int_0^{\lambda} \frac{b_A(u)}{w(u)} dB_A(u) + \frac{x(0)}{w(0)} \right). \tag{10}$$

Statistics of Multipliers

We have mentioned the statistics of multipliers in *Introduction*:

(1) The stochastic multiplier $W(\lambda_2, \lambda_1) = x(\lambda_2)/x(\lambda_1)$ $(\lambda_2 \le \lambda_1)$ linking two different scales follows a Cauchy distribution.
(2) When considering the three scales $\lambda_1 < \lambda_2 < \lambda_3$ $(s_1 > s_2 > s_3)$, the adjacent multipliers $W(\lambda_2, \lambda_1) = x(\lambda_2)/x(\lambda_1)$ and $W(\lambda_3, \lambda_2) = x(\lambda_3)/x(\lambda_2)$ show strongly negative correlation.

Here, we show property (1) and infer the existence of correlation between adjacent multipliers under some reasonable approximations. The parameter σ_M is an important model parameter for the signal to have multifractality. As presented in a later section, in spite of the importance, the value of the parameter σ_M^2 is small, about 0.02 to -0.03 in stock markets, irrespective of the stock issue. To specifically examine the role of additional stochastic processes, we investigate the 0th-order approximation of small σ_M. When setting $\sigma_M = 0$, the solution of **Eq. 10** becomes

$$x(\lambda) = \left(\int_{\lambda_0}^{\lambda} a_A(u) e^{\gamma_M (u-\lambda_0)} du + \int_{\lambda_0}^{\lambda} b_A(u) e^{\gamma_M (u-\lambda_0)} dB_A(u) + x(\lambda_0) \right) \tag{11}$$

Therefore, the difference $\Delta x(\lambda, \lambda_0) = x(\lambda) - x(\lambda_0)$ follows a normal distribution.

$N\left(\int_{\lambda}^{\lambda_0} a_A(u) e^{\gamma_M (u-\lambda_0)} du, \int_{\lambda_0}^{\lambda} (b_A(u))^2 e^{2\gamma_M (u-\lambda_0)} du \right)$. If one simply assumes that $x(\lambda_0)$ follows a normal distribution, then the ratio $\Delta x(\lambda_0, \lambda_1)/x(\lambda_0)$ of two independent stochastic variables following normal distributions follows a Cauchy distribution. So, $x(\lambda_1)/x(\lambda_0)$ is the same.

By defining the differences $\Delta x(\lambda_2, \lambda_1) = x(\lambda_2) - x(\lambda_1)$ and $\Delta x(\lambda_3, \lambda_2) = x(\lambda_3) - x(\lambda_2)$ for the three scales $\lambda_1 < \lambda_2 < \lambda_3$, it is readily apparent that $W_1 = x(\lambda_2)/x(\lambda_1) = 1 + \Delta x(\lambda_2, \lambda_1)/x(\lambda_1)$ and $W_2 = x(\lambda_3)/x(\lambda_2) = 1 + \Delta x(\lambda_3, \lambda_2)/x(\lambda_2)$ show correlation. In this framework, it was difficult to show that they have strongly negative correlation. Those statistics of multipliers have also been considered in earlier work by Siefert and Peinke [22]. The same result can be shown using a Fokker–Planck equation under some approximations. In a later section, we show a similar Fokker–Planck equation derived from the stochastic differential **Eq. 7**.

Relation to Discrete Random Cascade Model

Assuming that $\Delta\lambda$ is sufficiently small, then when we use the following approximation of Ito's stochastic integration [24] as

$$\int_{\lambda}^{\lambda+\Delta\lambda} \frac{b_A(u)}{w(u)} dB_A(u) \sim \frac{b_A(\lambda)}{w(t)} \left(B_A(\lambda + \Delta\lambda) - B_A(\lambda) \right), \tag{12}$$

we obtain the discrete random cascade equation as

$$x(\lambda + \Delta\lambda) = W_M(\lambda, \lambda + \Delta\lambda) \cdot \left(x(\lambda) + a_A(\lambda)\Delta\lambda + b_A(\lambda)\left(B_A(\lambda + \Delta\lambda) - B_A(\lambda) \right) \right), \tag{13}$$

where $W_M(\lambda, \lambda + \Delta\lambda) = e^{-\left(\gamma_M + \frac{1}{2}\sigma_M^2\right)\Delta\lambda + \sigma_M \left(B_M(\lambda+\Delta\lambda) - B_M(\lambda) \right)}$. The conditional expectation value of the square of $x(\lambda + \Delta\lambda)$, as the function of $x^2(\lambda)$,

$$E\left(x^2(\lambda + \Delta\lambda) \middle| x(\lambda) \right) = e^{\left(2\mu_M + 2\sigma_M^2\right)\Delta\lambda} \left(x^2(\lambda) + \left(2a_A(\lambda) x(\lambda) + b_A^2(\lambda) \right), \Delta\lambda \right) \tag{14}$$

shows that deviation of the quadratic curve from the origin results from the parameter $b_A(\lambda)$, as demonstrated from an empirical study in [18].

Constraint Condition From the pdf of $x(\lambda)$.

A remarkable feature of the probability density function (pdf) of the quantity $x(\lambda) = \left| W_\psi Z[., s(\lambda)] \right|$ is the coincidence of the expected value $E\left(\left| W_\psi Z[., s(\lambda)] \right| \right)$ with standard deviation $V\left(\left| W_\psi Z[., s(\lambda)] \right| \right)^{1/2}$, as shown in **Figure 1** for the data examined in this study (see also **Figure 10** for the pdf of $x(\lambda)$). It indicates the constraint condition as

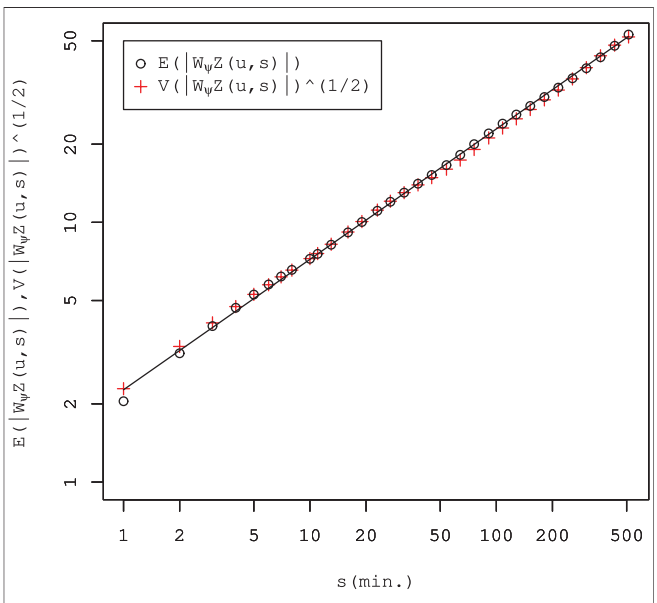


FIGURE 1 | Scaling properties of $E\left(\left| W_\psi Z[., s(\lambda)] \right| \right)$ and $V\left(\left| W_\psi Z[., s(\lambda)] \right| \right)^{1/2}$. The expected value almost perfectly coincides with the standard deviation at all scales. The solid line represents the least-squares fit to the power law function, $2.27 s^{0.5}$.

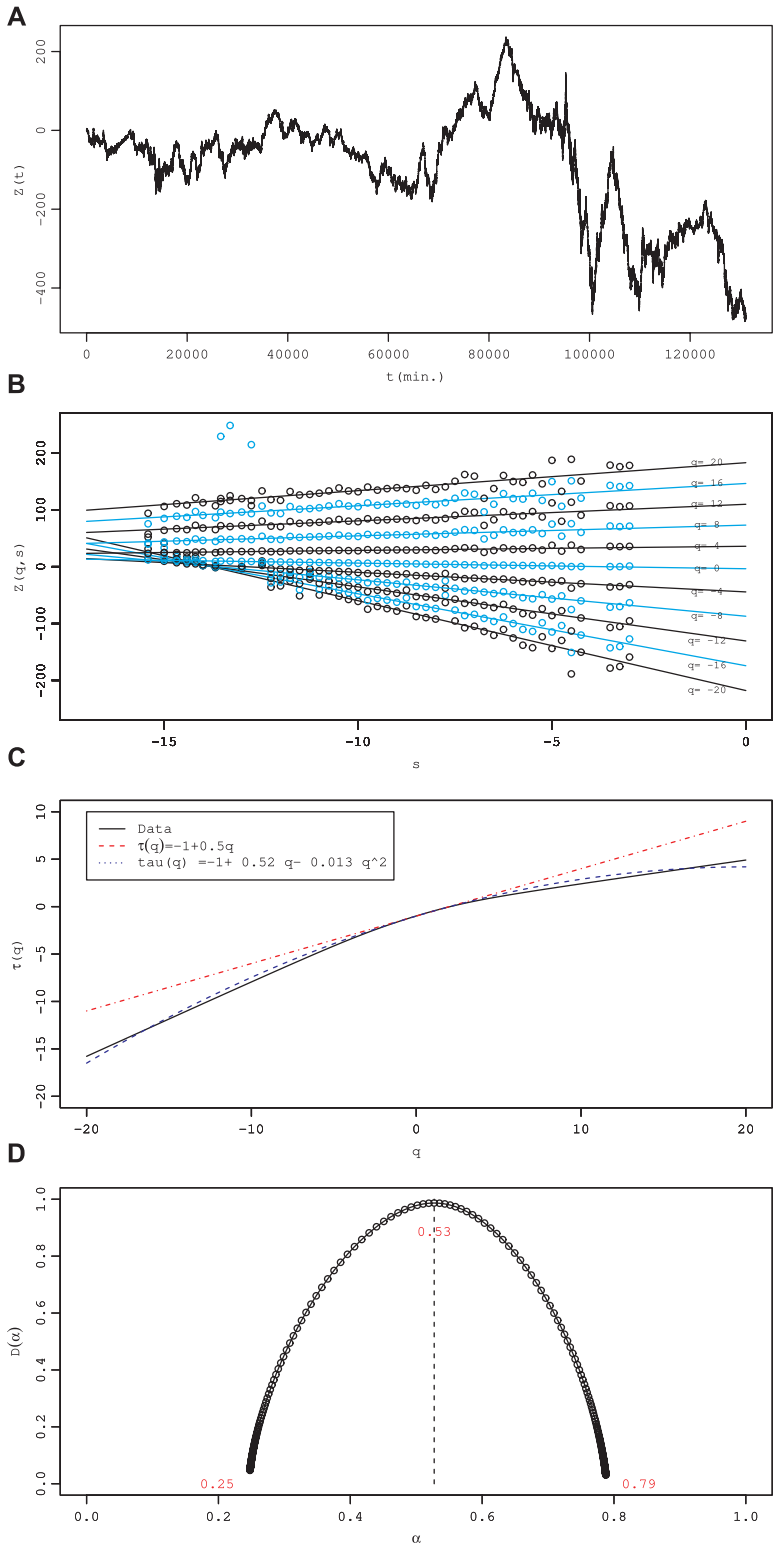


FIGURE 2 | Results of multifractal analysis. **(A)** $Z(t)$. **(B)** $Z(q,s)$ for $q = -20, \ldots, 20$ ($\circ$) and regression lines. **(C)** Scaling exponent $\tau(q)$ (solid line). The dashed blue line is the least-squares fit to the quadratic function $\tau(q) = -1 + 0.52q - 0.013q^2$. The dotted red line $\tau(q) = -1 + 0.5q$ corresponds to Brownian motion. **(D)** Singular spectrum $D(\alpha)$.

$$a_A(\lambda)E(x(\lambda)) = b_A^2(\lambda). \tag{15}$$

Derivation of the constraint condition **Eq. 15** is given in **Appendix 1**.

The additional additive stochastic process in model **Eq. 7** is expected to be a small perturbation to basic model **Eq. 4** to avoid violating multifractality. We also impose the following condition for all scales s:

$$\frac{a_A(s)}{E\left(\left|W_\psi Z[.,s]\right|\right)} << 1, \frac{b_A(s)}{E\left(\left|W_\psi Z[.,s]\right|\right)} << 1. \tag{16}$$

The power law scaling shown in **Figure 1**,

$$E\left(\left|W_\psi Z[,s]\right|\right) \sim s^{0.5}, \tag{17}$$

and condition **Eq. 16** show the following constraint condition:

$$a_A(s) \sim s^{0.5}, b_A(s) \sim s^{0.5}. \tag{18}$$

Inserting **Eq. 18** into **Eq. 15**, we also have the equation:

$$a_A(1)E\left(\left|W_\psi Z[,1]\right|\right) = b_A^2(1). \tag{19}$$

Fokker–Planck Equation

We can derive the Fokker–Planck equation for the stochastic process $\{x(\lambda)\}$ expressed by the stochastic differential **Eq. 7** as the following [24], which is the master equation that the density of the transition probability $p(x,\lambda|x_0,\lambda_0)$ follows:

$$\begin{aligned}&\frac{\partial}{\partial\lambda}p(x,\lambda|x_0,\lambda_0)\\ &\quad= \left[-\frac{\partial}{\partial x}D_1(\lambda,x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}D_2(\lambda,x)\right]p(x,\lambda|x_0,\lambda_0).\end{aligned} \tag{20}$$

Therein, the functions $D_1(\lambda,x)$ and $D_2(\lambda,x)$ are defined as

$$\begin{aligned}D_1(\lambda,x) &= a_A(\lambda) - \gamma_M x,\\ D_2(\lambda,x) &= b_A(\lambda)^2 + \sigma_M^2 x^2.\end{aligned} \tag{21}$$

The kth moment of the change $\delta x(\lambda) = x(\lambda+\delta\lambda) - x(\lambda)$ induced by the infinitesimal scale transformation $\delta\lambda$ is derived as shown below:

$$\begin{aligned}E\left(\delta x^k\middle|x(\lambda)=x\right) &= \int_{-\infty}^{+\infty}(y-x)^k p(y,\lambda+\delta\lambda|x,\lambda)dy\\ &= \int_{-\infty}^{+\infty}(y-x)^k\Bigg(p(y,\lambda|x,\lambda)\\ &\quad+\delta\lambda\frac{\partial}{\partial\lambda_\prime}p\left(y,\lambda_\prime\middle|x,\lambda\right)|_{\lambda_\prime=\lambda} + O(\delta\lambda^2)\Bigg)dy\\ &= \int_{-\infty}^{+\infty}(y-x)^k\Bigg(\delta\lambda\left[-\frac{\partial}{\partial y}D_1\left(\lambda_\prime,y\right)\right.\\ &\quad\left.+\frac{1}{2}\frac{\partial^2}{\partial y^2}D_2\left(\lambda_\prime,y\right)\right]p\left(y,\lambda_\prime\middle|x,\lambda\right)|_{\lambda_\prime=\lambda} + O(\delta\lambda^2)\Bigg)dy\\ &= \int_{-\infty}^{+\infty}\Bigg(\delta\lambda\Big[k(y-x)^{k-1}D_1(\lambda,y)\\ &\quad+\frac{1}{2}k\left(\frac{\partial}{\partial y}(y-x)^{k-1}\right)D_2(\lambda,y)\Big]\delta(y-x) + O(\delta\lambda^2)\Bigg)d\end{aligned} \tag{22}$$

Therein, we used the identity $p(y,\lambda|x,\lambda) = \delta(y-x)$. Coefficients $D_1(\lambda,x)$ and $D_2(\lambda,x)$ show a relation to the first and second moments of $\delta x(\lambda)$ in the following way:

$$\lim_{\delta\lambda\to 0}\frac{E\left(\delta x^k\middle|x(\lambda)=x\right)}{\delta\lambda} = \begin{cases}D_k(\lambda,x) & k=1,2\\ 0 & others\end{cases} \tag{23}$$

Coefficients D_k are designated as Kramers–Moyal coefficients [24, 25]. We use **Eq. 23** to estimate the function $a_A(\lambda)$ and $b_A(\lambda)$ and parameters γ_M and σ_M. To validate model **Eq. 7**, it is necessary to confirm vanishing of the kth moments for $3 \le k$ in the limit of $\delta\lambda \to 0$. Renner et al. proposed almost identical equations (**Eq. 20**) within the literature [20, 21], in which they deal with the price change itself as an analogy of the velocity difference in turbulence [19]. They derived a Fokker–Planck equation as a result of their empirical studies using Kramers–Moyal expansion of the Chapman–Kolmogorov equation, regarding the process as a Markovian process.

Empirical Study

Data

We analyze the normalized average of the logarithmic stock prices of the constituent issues of the FTSE 100 Index listed on the London Stock Exchange for November 2007 through January 2009, which includes the Lehman shock of September 15, 2008 and the market crash of October 8, 2008.

Data Processing

First, we calculate the average deseasonalized return of each issue $\delta Z_i(t) = \log(S_i(t)) - \log(S_i(t-\delta t))$, which describes the average change of the portfolio as follows:

$$\delta Z(k\delta t) = \frac{1}{N_F}\sum_{i=1}^{N_F}\frac{\delta Z_i(k\delta t) - \mu_i}{\sigma_i}, \tag{24}$$

where μ_i and σ_i, respectively, denote the average and the standard deviation of δZ_i and where N_F represents the number of constituent stock issues (stocks). The constituents of the FTSE 100 Index are updated frequently. We selected $N_F = 111$ stocks that remained listed on the London Stock Exchange throughout the period. Here, we set $\delta t = 1$ min and examine the 1-min log return. We excluded the overnight price change and specifically examine the intraday evolutions of returns. To remove the effect of intraday U-shaped patterns of market activity from the time series, the return was divided by the standard deviation of the corresponding time of the day for each issue i. Then, we cumulate $\delta Z(t)$ to obtain the path of process $Z(k\delta t)$ $(k = 1,\ldots,L)$ (**Figure 2A**) as follows:

$$Z(k\delta t) = \sum_{k_\prime=1}^{k}\delta Z\left(k_\prime\delta t\right). \tag{25}$$

The data size L is 2^{17}.

RESULTS

Multifractal Analysis

First, we analyze the multifractal properties of the path $Z(t)$ using an approach with wavelet-based multifractal formalism proposed

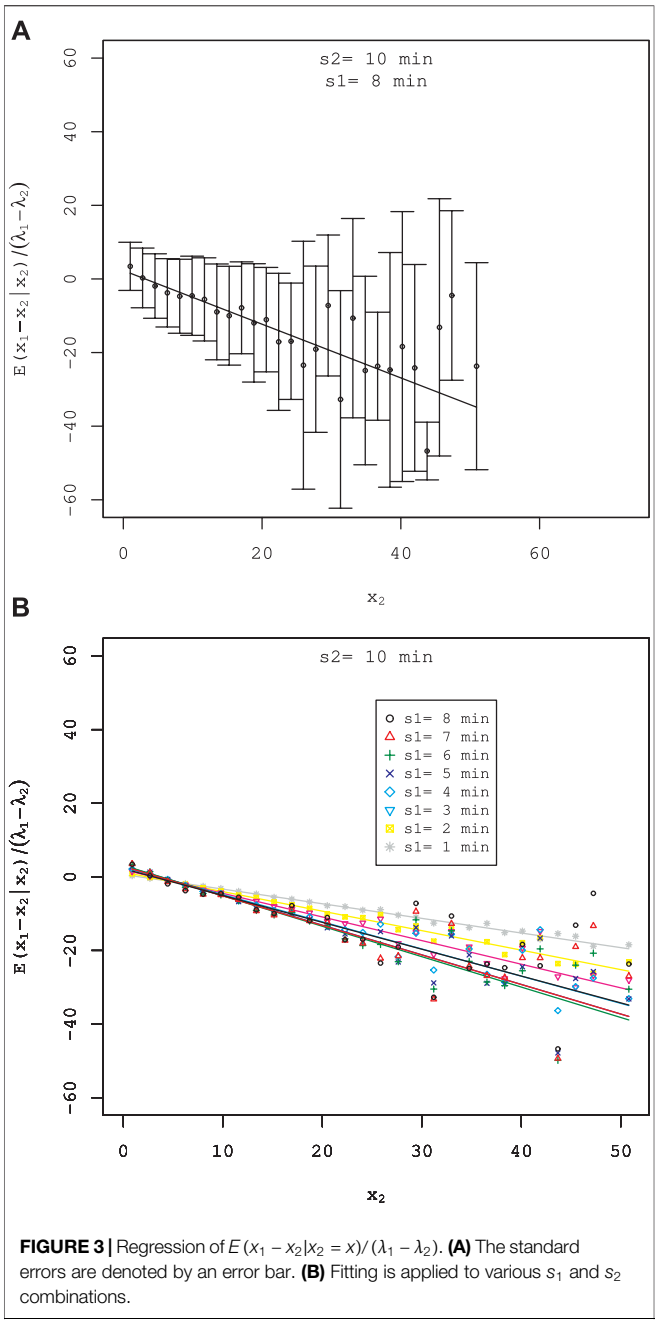


FIGURE 3 | Regression of $E(x_1 - x_2|x_2 = x)/(\lambda_1 - \lambda_2)$. **(A)** The standard errors are denoted by an error bar. **(B)** Fitting is applied to various s_1 and s_2 combinations.

by Muzy, Bacry, and Arneodo [26, 27]. Initially, we define two mathematical terms. The Hölder exponent $\alpha(x_0)$ of a function $f(x)$ at x_0 is defined as the largest exponent such that there exist an nth-order polynomial $P_n(x)$ and constant C that satisfy

$$\left|f(x) - P_n(x - x_0)\right| \leq C\left|x - x_0\right|^{\alpha}, \tag{26}$$

where x in a neighborhood of x_0, characterizing the regularity of the function $f(x)$ at x_0. The singular spectrum $D(\alpha)$ is the Hausdorff dimension of the set where the Hölder exponent is equal to α, as follows:

$$D(\alpha) = dim_H\{x|\alpha(x) = \alpha\}. \tag{27}$$

For multifractal paths, the Hölder exponent α is distributed in a range; for paths of the Brownian motion, which are fractal, $D(0.5) = 1$ and $D(\alpha) = 0$ for $\alpha \neq 0.5$.

Muzy, Bacry, and Arneodo proposed the wavelet transform modulus maxima (WTMM) method based on continuous wavelet transform of function to calculate the singular spectrum $D(\alpha)$. We briefly sketch the WTMM method in **Appendix 2**. We calculate the partition function $Z(q, s)$ of the qth moment of wavelet coefficients using **Eq. 37** for the path of our data. Results are presented in **Figure 2B**. The partition function $Z(q, s)$ for each order q shows power law behavior in the range of scales $s/L < 2^{-5}$. Exponents $\tau(q)$ are derived by **Eq. 38**. **Figure 2C** shows that it is a convex function of q. Those results underscore the multifractality of the data path. The singular spectrum $D(\alpha)$ derived as the Legendre transformation of the function $\tau(q)$ by **Eq. 39** is a convex function that has compact support $[0.25, 0.79]$ taking the peak at $\alpha = 0.53$, as shown in **Figure 3D**.

Parameter Estimations

a_A and γ_M

Parameters $a_A(\lambda)$ and γ_M are estimated by taking the limit $\lambda_1 - \lambda_2 \to 0$ of the first moment $E(x_1 - x_2|x_2 = x)/(\lambda_1 - \lambda_2)$ (LABEL : K − M). The first moment $E(x_1 - x_2|x_2 = x)/(\lambda_1 - \lambda_2)$ is fitted by a linear function as follows:

$$\frac{E(x_1 - x_2|x_2 = x)}{d\lambda} = a_A(\lambda_2, d\lambda) - \gamma_M(\lambda_2, d\lambda)x, \tag{28}$$

where $d\lambda = \lambda_1 - \lambda_2$. As shown in **Figure 3A**, the first moment is well fitted by a linear function. Fitting of this kind is applied to various $\lambda_1 = \log(L/s1)$ and $\lambda_2 = \log(L/s2)$ combinations (**Figure 3B**). Taking the limit $d\lambda \to 0$ ($ds/s = (s_2 - s_1)/s_2 \to 0$), one obtains, $a_A(\lambda_2) = \lim_{d\lambda \to 0} a_A(\lambda_2, d\lambda)$ ($a_A(s_2) = \lim_{ds/s \to 0} a_A(s_2, ds/s)$) and $\gamma_M = \lim_{d\lambda \to 0} \gamma_M(\lambda_2, d\lambda)$ ($\gamma_M = \lim_{ds/s \to 0} \gamma_M(s_2, ds/s)$). **Figure 4A** presents examples of $a_A(s_2, ds/s)$ and nonlinear fittings by the function $log(a_A(s, ds/s)) = a + b(ds/s) + c(ds/s)^2$. We estimate $a_A(s)$ by $a_A(s) = \exp(a)$ for each line. The result is presented in **Figure 4B**. The solid line is the least-squares fit $a_A(s)$ to a power law function as follows:

$$\log(a_A(s)) = -1.50(0.41) + 0.58(0.11)\log s, \tag{29}$$

where the standard errors are in parentheses. The estimated exponent $0.58(0.11)$ is consistent with the constraint condition (**Eq. 18**) within the standard error. By a similar extrapolation $\log(\gamma_M(s, ds/s)) = a + b(ds/s) + c(ds/s)^2$, we estimate $\gamma_M(s) = exp(a)$. **Figure 5A** presents examples of $\gamma_M(s_2, ds/s)$ and nonlinear fittings. We estimate $\gamma_M(s)$ by $\gamma_M(s) = \exp(a)$ for each line. The result is presented in **Figure 5B**. We estimate the parameter γ_M by the average value weighted by the reciprocals of the standard errors as follows:

$$\gamma_M = 0.64(0.21), \tag{30}$$

where the standard error is the value in the parenthesis.

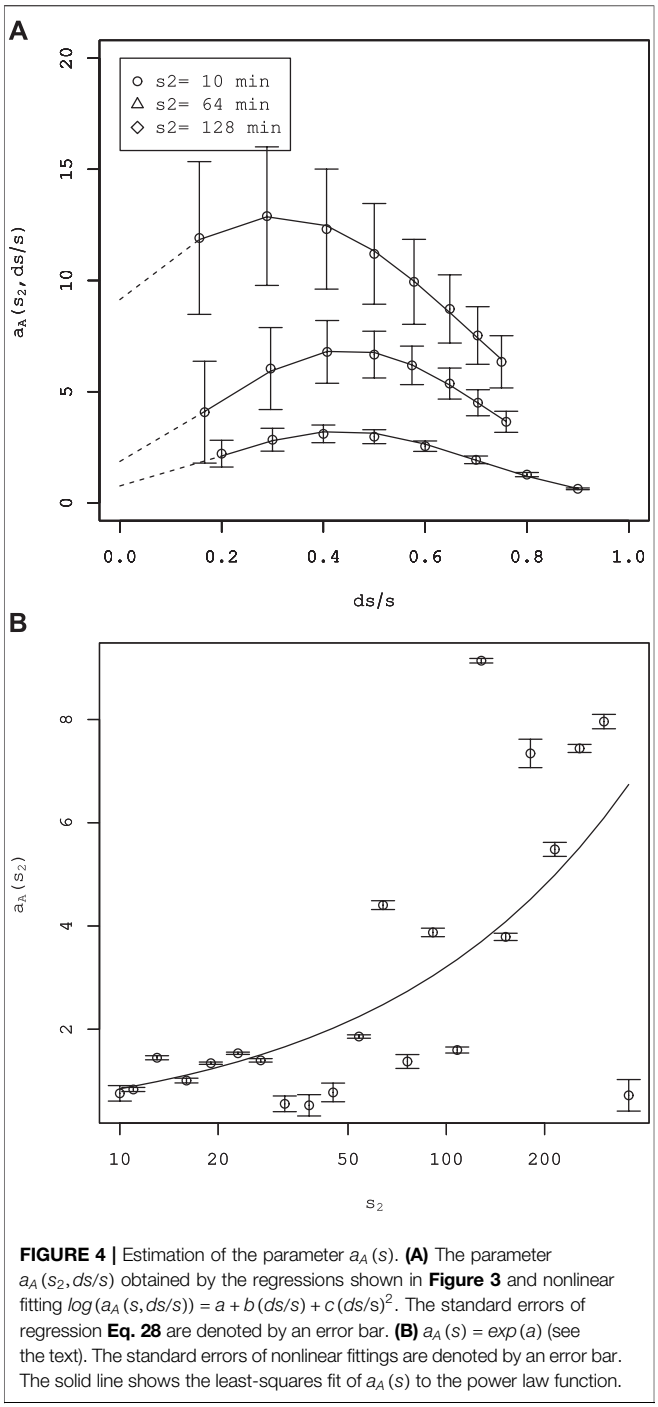


FIGURE 4 | Estimation of the parameter $a_A(s)$. **(A)** The parameter $a_A(s_2, ds/s)$ obtained by the regressions shown in **Figure 3** and nonlinear fitting $log(a_A(s, ds/s)) = a + b(ds/s) + c(ds/s)^2$. The standard errors of regression **Eq. 28** are denoted by an error bar. **(B)** $a_A(s) = exp(a)$ (see the text). The standard errors of nonlinear fittings are denoted by an error bar. The solid line shows the least-squares fit of $a_A(s)$ to the power law function.

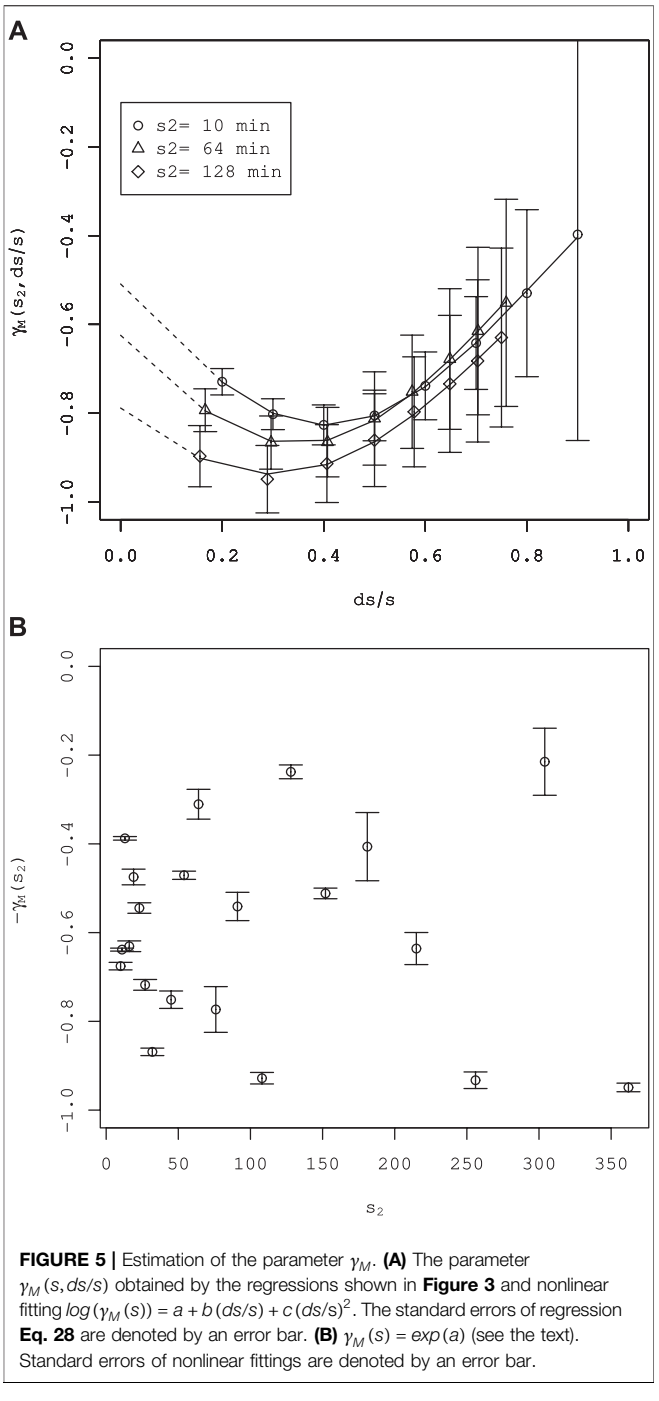


FIGURE 5 | Estimation of the parameter γ_M. **(A)** The parameter $\gamma_M(s, ds/s)$ obtained by the regressions shown in **Figure 3** and nonlinear fitting $log(\gamma_M(s)) = a + b(ds/s) + c(ds/s)^2$. The standard errors of regression **Eq. 28** are denoted by an error bar. **(B)** $\gamma_M(s) = exp(a)$ (see the text). Standard errors of nonlinear fittings are denoted by an error bar.

b_A and σ_M

Similarly, we estimate parameters b_A and σ_M by taking the limit $\lambda_1 - \lambda_2 \to 0$ of the second moment $E((x_1 - x_2)^2 | x_2 = x)/(\lambda_1 - \lambda_2)$(LABEL : K – M). The second moment $E((x_1 - x_2)^2 | x_2 = x)/(\lambda_1 - \lambda_2)$ is fitted by a quadratic function (a regression against x^2) as follows:

$$\frac{E((x_1 - x_2)^2 | x_2 = x)}{d\lambda} = b_A(\lambda_2, d\lambda) + \sigma_M(\lambda_2, d\lambda)x^2. \quad (31)$$

As shown in **Figure 6A**, the second moment is well fitted by a quadratic function. Fitting of this kind is applied to various λ_1 and λ_2 combinations (**Figure 6B**). Taking the limit $d\lambda \to 0$, one obtains $b_A^2(\lambda_2) = \lim_{d\lambda \to 0} b_A^2(\lambda_2, d\lambda)$ and $\sigma_M^2 = \lim_{d\lambda \to 0} \sigma_M^2(\lambda_2, d\lambda)$. **Figure 7A** presents examples of $b_A^2(s_2, ds/s)$ and nonlinear fitting by the function $log(b_A^2(s, ds/s)) = a + b(ds/s) + c(ds/s)^2$. We estimate $b_A^2(s)$ for each line by $b_A^2(s) = \exp(a)$. The result is presented in **Figure 7B**. The solid line is the least-squares fit $b_A^2(s)$ to a power law function as follows:

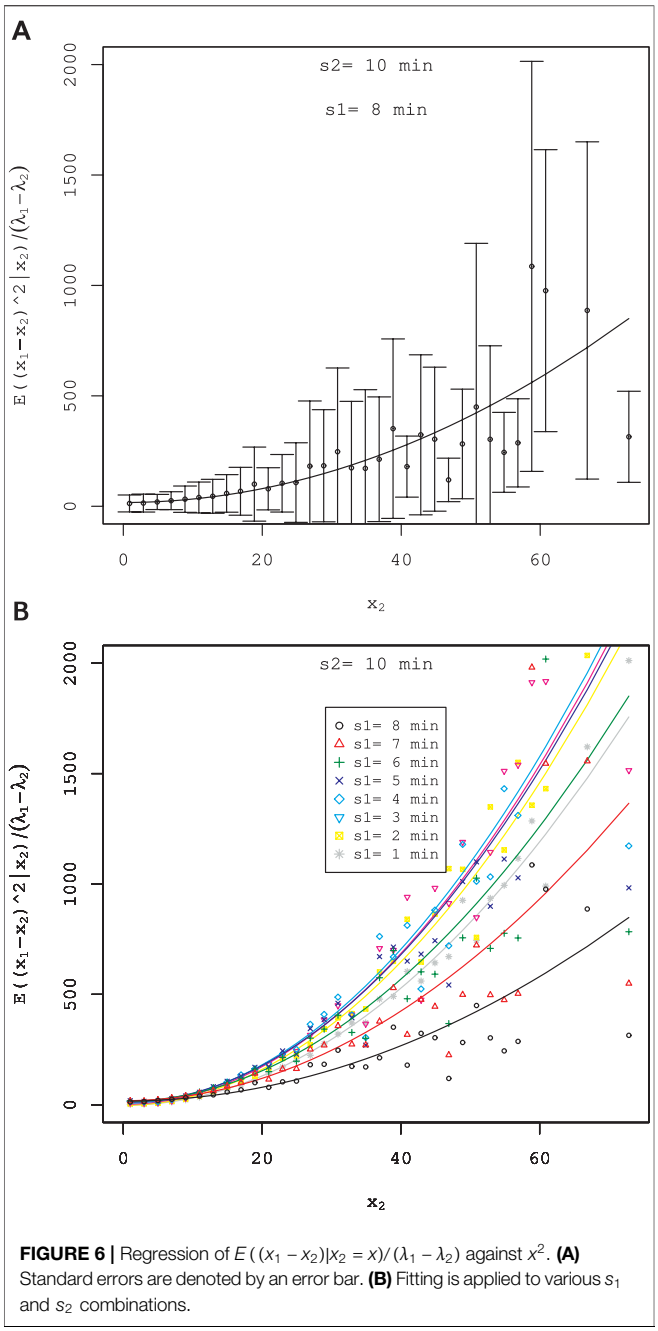


FIGURE 6 | Regression of $E((x_1 - x_2)|x_2 = x)/(\lambda_1 - \lambda_2)$ against x^2. **(A)** Standard errors are denoted by an error bar. **(B)** Fitting is applied to various s_1 and s_2 combinations.

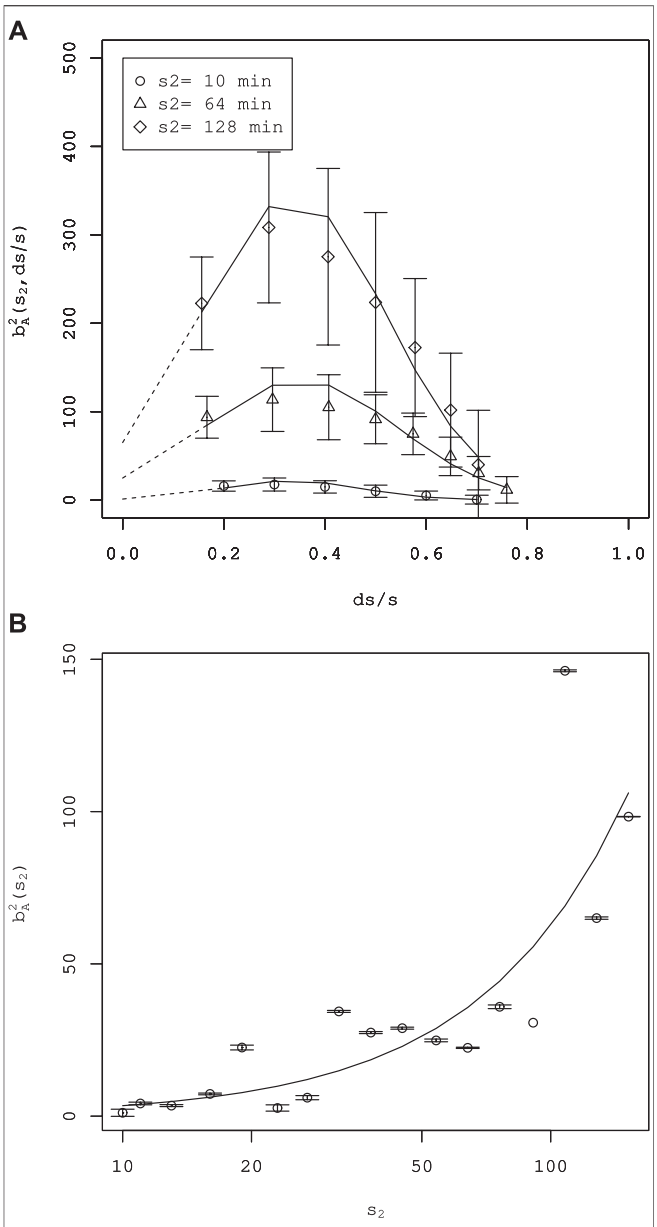


FIGURE 7 | Estimation of the parameter $b_A(s)$. **(A)** The parameter $b_A(s_2, ds/s)$ obtained by the regressions shown in **Figure 6** and nonlinear fitting $log(b_A^2(s, ds/s)) = a + b(ds/s) + c(ds/s)^2$. The standard errors of the regression **Eq. 31** are denoted by an error bar. **(B)** $b_A^2(s) = exp(a)$ (see the text). Standard errors of nonlinear fittings are denoted by an error bar. The solid line shows the least-squares fit of $b_A^2(s)$ to the power law function.

$$\log\left(b_A^2(s)\right) = -1.67\,(0.56) + 1.26\,(0.13)\log s, \tag{32}$$

where the standard errors are in parentheses. The estimated exponent 1.26 (0.13) is slightly higher than the constraint condition (**Eq 18**) $(b_A^2(s) \sim s)$. However, it is acceptable with accuracy. By a similar extrapolation $\log(\sigma_M^2(s, ds/s)) = a + b(ds/s) + c(ds/s)^2$, we estimate $\sigma_M^2(s) = exp(a)$. **Figure 8A** presents an example of $\sigma_M^2(s_2, ds/s)$. and an estimate $\sigma_M^2(s)$ by $\sigma_M^2(s) = \exp(a)$ for each line. The result is shown in **Figure 8B**. We estimate parameter σ_M^2 by the average value weighted by the reciprocals of the standard errors.

$$\sigma_M^2 = 0.05\,(0.03) \tag{33}$$

Therein, the standard error is in the parenthesis.

Higher Moments

Similarly, it is possible to show the kth $(3 \le k)$ moment $E((x_1 - x_2)^k | x_2 = x)/(\lambda_1 - \lambda_2)$ of the transition probability density $p(x_1, \lambda_1 | \lambda_2, x_2)$ vanishes in the limit $\lambda_1 - \lambda_2 \to 0$. As portrayed in **Figure 9A**, the fourth moment is well fitted by a quartic function. Applying the fitting to various λ_1 and λ_2 combinations (**Figure 9B**), we have convinced that the fourth

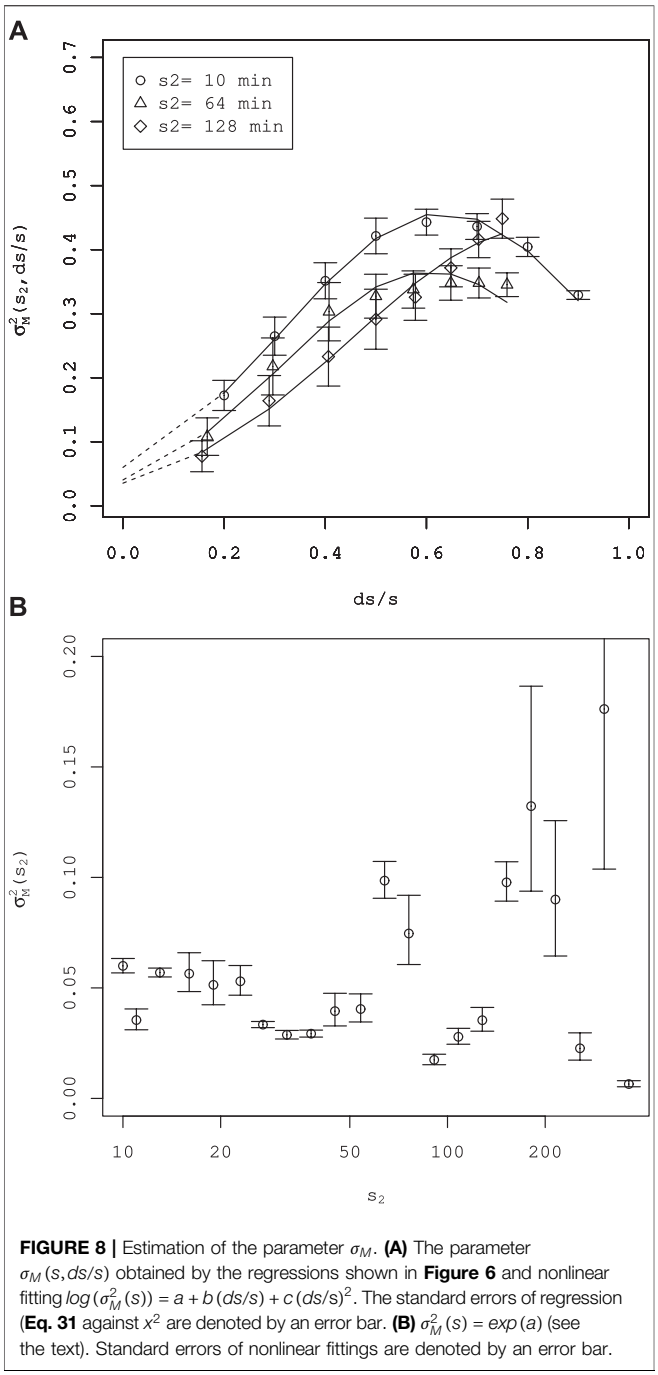


FIGURE 8 | Estimation of the parameter σ_M. **(A)** The parameter $\sigma_M(s, ds/s)$ obtained by the regressions shown in **Figure 6** and nonlinear fitting $log(\sigma_M^2(s)) = a + b(ds/s) + c(ds/s)^2$. The standard errors of regression (**Eq. 31** against x^2 are denoted by an error bar. **(B)** $\sigma_M^2(s) = exp(a)$ (see the text). Standard errors of nonlinear fittings are denoted by an error bar.

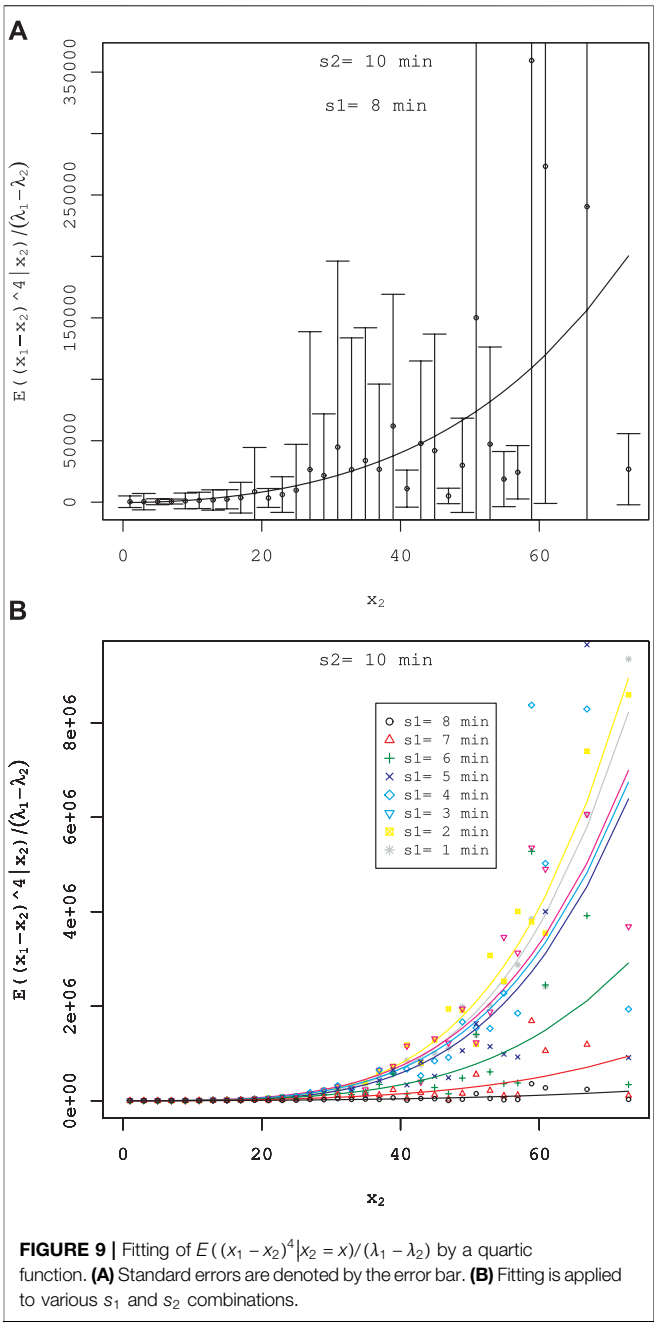


FIGURE 9 | Fitting of $E((x_1 - x_2)^4 | X_2 = x)/(\lambda_1 - \lambda_2)$ by a quartic function. **(A)** Standard errors are denoted by the error bar. **(B)** Fitting is applied to various s_1 and s_2 combinations.

moment vanishes in the limit $d\lambda \to 0$. The Pawula theorem states that all higher Kramers–Moyal coefficients D_k $(3 \leq k)$ vanish if D_4 vanishes [25]. Therefore, we verified **Eq 23**.

Numerical Calculation of the Fokker–Planck Equation

We confirmed that estimation of the parameter $a_A(\lambda)$ and $b_A(\lambda)$ by the $E((x_1 - x_2)^k | x_2 = x/(\lambda_1 - \lambda_2))(k = 1, 2)$ is consistent with the constraint condition (**Eq. 18**) with accuracy. If one imposes the other constraint (**Eq. 19**, then the parameters take the following functional form:

$$a_A(\lambda(s)) = \epsilon s^{0.5}$$
$$b_A^2(\lambda(s)) = 2.27\epsilon s, \tag{34}$$

where ϵ is a small parameter. The consistent range of ϵ found by estimation of **Eq 29** and **Eq 32** is $0.15 \leq \epsilon \leq 0.34$. To fix parameters γ_M and σ_M, we use the empirical value of the scaling exponent $\tau(q)$, which is fitted by the quadratic function $\tau(q) = -1 + 0.52q - 0.013q^2$ (see **Figure 2C**). One can derive $\tau(q) = -1 + (\gamma_M + \frac{1}{2}\sigma_M^2)q - \frac{1}{2}\sigma_M^2 q^2$ for the basic model (**Eq. 4** without additional stochastic processes. Again using the assumption of slight perturbation, then from the coefficients of the quadratic function, the parameters γ_M and σ_M are expected to exist

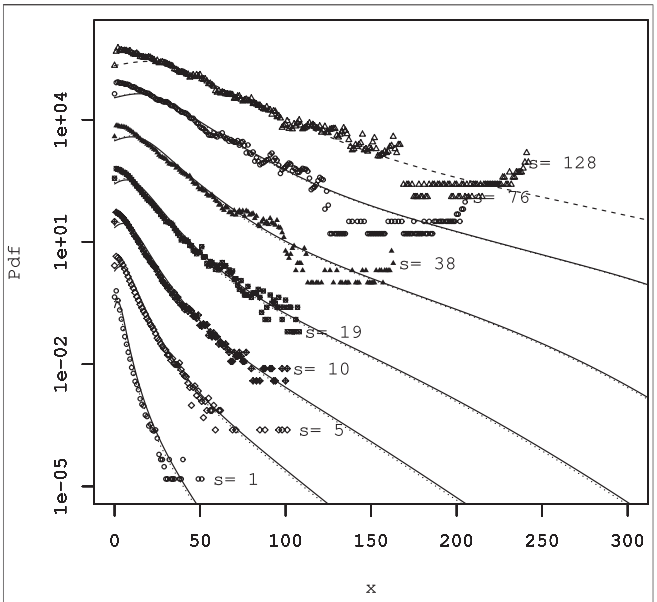


FIGURE 10 | Pdf of measured $x(\lambda)$ and numerical calculation of the Fokker–Planck equation. The result of numerical calculation is represented by the solid lines. Marks are measured values. The scale is attached to each line.

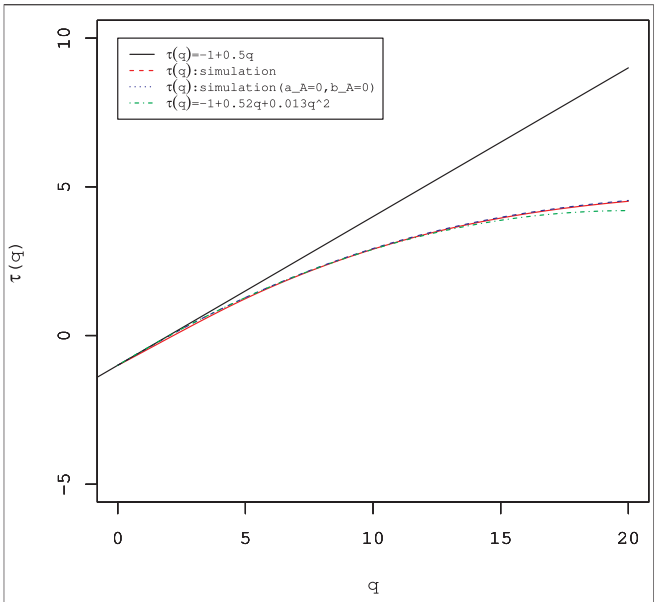


FIGURE 11 | Scaling exponent $\tau(q)$. A representation of each line is shown in the legend.

respectively in the neighborhood of $\gamma_M = 0.51$ and $\sigma_M = 0.026$. Next, we try the value of the parameters $\gamma_M = 0.51$, $\sigma_M = 0.026$, and $\epsilon = 0.16$ for numerical calculation of the Fokker–Planck equation. Results are presented in **Figure 10**. The initial pdf of the numerical calculation represented by the dashed line was based on the measured pdf on scale $s = 128\,(min)$. In the initial values, the fine fluctuation was smoothed using a spline function with the rationale that small fluctuations in the measured pdf are attributable to the finiteness of the number of observations. The tails are extrapolated using a power function with index −4.9 which is obtained empirically. For time evolution, the fourth-order explicit Runge–Kutta method was used. The solid line is the calculation result obtained using the estimated value of the parameters $\gamma_M = 0.51$, $\sigma_M = 0.026$, and $\epsilon = 0.16$. The dotted line is the result obtained when $\epsilon = 0$. The difference between the two was very small. The results closely matched the actual pdf. In the data and the numerical calculation, the probability density function does not converge to zero at the origin because of the finite size of the bin. Although the details around the origin $x = 0$ cannot be empirically discussed due to the finiteness of the observed data, the probability density function must converge to zero at the origin if the negative qth moment of the fluctuation is requested to converge.

Using results of the numerical calculation of the pdf obtained at each scale, we calculate the scaling exponent $\tau(q)$ as shown follows:

$$E\left(\left|W_\psi Z[u,s]\right|^q\right) \sim s^{\tau(q)}\,(0 \le q) \tag{35}$$

The result is presented in **Figure 11**. No difference exists between the two numerical calculation results. Both curves are convex upward, indicating multifractal properties. Comparison with measured values is also good. These results, when combined with consideration of the statistics of multipliers given in 2.1.2, underscore the effectiveness of the continuous cascade model **Eq. 7** with additive stochastic processes proposed.

DISCUSSION

The random cascade model has evolved as a model of developed turbulence. The original model, in which the stochastic process that connects each layer of the spatial scale is an independent random multiplication process, contradicts results obtained through empirical research. Therefore, an improved discrete random multiplicative cascade model with additional additive stochastic processes was proposed along with a model formulated as a Fokker–Planck equation by considering cascade processes as a continuous Markov process. Moreover, those models have been applied to data analysis of the stock market and the foreign exchange market, where they have been successful. Herein, we propose a continuous cascade model formulated as a stochastic differential equation of volatility including two independent modes of Brownian motion: one has multiplicative coupling with volatility; the other has additive coupling, as in an improved discrete cascade model for the stock market, with effectiveness clarified by results of earlier research [18]. The model parameters were estimated by application to a stock price time series. The Fokker–Planck equation was derived from the stochastic differential equation as a master equation with the transition probability density function of volatility. Furthermore, the model parameters were estimated by its application to the average stock price time series made from FTSE 100 constituents listed on the London Stock Exchange. At

that time, as an alternative variable of volatility, the wavelet transform coefficient with the second derivative of the Gaussian function as an analyzing wavelet was used. Numerical calculation of the Fokker–Planck equation was conducted using the estimated parameter values. The results reported herein faithfully reproduce the results of an earlier empirical study. This model includes information about neither the time axis nor the sign of the price fluctuation, which is necessary for a model of price fluctuations. The actual stock market exhibits well-known properties that break symmetry with respect to the time axis, such as the causal structure from long-term to short-term scale volatility described first in *Introduction* and price–volatility correlation (leverage effect) [4, 12]. Therefore, the extension of the random cascade model to encompass these phenomena remains as a subject for future work.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: Historic Price Service (HPS) provided by the London Stock Exchange (https://www.londonstockexchange.com/products-and-services/reference-data/hps/hps.htm).

AUTHOR CONTRIBUTIONS

JM conducted the empirical and numerical study of the model based on the dataset. KK theoretically analyzed the model. All authors agree to be accountable for the contents of the work.

FUNDING

This research was partially supported by a Grant-in-Aid for Scientific Research (C) No. 16K01259.

ACKNOWLEDGMENTS

One author, JM, expresses special appreciation for support by a Seijo University Special Research Grant.

REFERENCES

1. Müller UA, Dacorogna MM, Davé RD, Olsen RB, Pictet OV, von Weizsäcker JE. Volatilities of different time resolutions–analyzing the dynamics of market components. *J Empir Finance.* (1997). 4:213–39. doi:10.1016/s0927-5398(97)00007-8.
2. Arnéodo A, Muzy J-F, Sornette D. "Direct" causal cascade in the stock market. *Eur Phys J B.* (1998). 2:277–82. doi:10.1007/s100510050250.
3. Lynch PE, Zumbach GO. Market heterogeneities and the causal structure of volatility. *Quant Finance.* (2003). 3:320–31. doi:10.1088/1469-7688/3/4/308.
4. Cont R. Empirical properties of asset returns: stylized facts and statistical issues. *Quant Finance.* (2001). 1:223–36. doi:10.1080/713665670.
5. Richardson LF. *Weather prediction by numerical process.* Cambridge UK: Cambridge University Press (1922). p. 236.
6. Kolmogorov AN. The local structure of turbulence in incompressible viscous fluids for very large Reynolds numbers. *Dokl Akad Nauk SSSR.* (1941). 30:301–5.
7. Kolmogorov AN. A refinement of previous hypotheses related to the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *J Fluid Mech.* (1962). 13:301–5. doi:10.1017/s0022112062000518.
8. Mandelbrot BB. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *J Fluid Mech.* (1974). 62:331–58. doi:10.1017/s0022112074000711.
9. Frisch U. *Turbulence: the legacy of A. Kolmogorov.* Cambridge, UK: Cambridge University Press (1997). p. 312.
10. Ghashghaie S, Breymann W, Peinke J, Talkner P, Dodge Y. Turbulent cascades in foreign exchange markets. *Nature.* (1996). 381:767–70. doi:10.1038/381767a0.
11. Schmitt F, Schertzer D, Lovejoy S. Multifractal analysis of foreign exchange data. *Appl Stoch Model Data Anal.* (1999). 15:29–53. doi:10.1002/(sici)1099-0747(199903)15:1<29::aid-asm357>3.0.co;2-z.
12. Bouchaud J-P, Potter M. *Theory of financial risk and derivative pricing: from statistical physics to risk management.* 2nd ed. Cambridge, UK: Cambridge University Press (2003). p. 400.
13. Arneodo A, Bacry E, Muzy JF. Random cascades on wavelet dyadic trees. *J Math Phys.* (1998). 39:4142–64. doi:10.1063/1.532489.
14. Breymann W, Ghashghaie S, Talkner P. A stochastic cascade model for Fx dynamics. *Int J Theor Appl Finance.* (2000). 03:357–60. doi:10.1142/s021902490000019x.
15. Chen Q, Chen S, Eyink GL. Kolmogorov's third hypothesis and turbulent sign statistics. *Phys Rev Lett.* (2003). 90:254501. doi:10.1103/physrevlett.90.254501.
16. Jiménez J. Intermittency and cascades. *J Fluid Mech.* (2000). 409:99–120. doi:10.1017/s0022112099007739.
17. Jiménez J. Intermittency in turbulence. Proceedings of the 15th "Aha Huliko" a Winter Workshop, 2007 January 23–26; Honolulu,Publication Services, SOEST, University of Hawaii; (2007). p. 99–120.
18. Maskawa J-i., Kuroda K, Murai J. Multiplicative random cascades with additional stochastic process in financial markets. *Evol Inst Econ Rev.* (2018). 15:515–29. doi:10.1007/s40844-018-0112-y.
19. Friedrich R, Peinke J. Description of a turbulent cascade by a Fokker-Planck equation. *Phys Rev Lett.* (1997). 78:863–6. doi:10.1103/physrevlett.78.863.
20. Renner C, Peinke J, Friedrich R. Evidence of Markov properties of high frequency exchange rate data. *Phys Stat Mech Appl.* (2001). 298:499–520. doi:10.1016/s0378-4371(01)00269-2.
21. Renner C, Peinke J, Friedrich R. Experimental indications for Markov properties of small-scale turbulence. *J Fluid Mech.* (2001). 433:383–409. doi:10.1017/s0022112001003597.
22. Siefert M, Peinke J. Complete multiplier statistics explained by stochastic cascade processes. *Phys Lett.* (2007). 371:34–8. doi:10.1016/j.physleta.2007.05.111.
23. Reinke N, Fuchs A, Nickelsen D, Peinke J. On universal features of the turbulent cascade in terms of non-equilibrium thermodynamics. *J Fluid Mech.* (2018). 848:117–53. doi:10.1017/jfm.2018.360.
24. Peinke C. *Stochastic methods.* 4th ed. Berlin Heidelberg, Germany: Springer-Verlag (2009). p. 447.
25. Risken H. *The fokker–planck equation.* 2nd ed. Berlin Heidelberg, Germany: Springer-Verlag (1889). p. 472.
26. Muzy JF, Bacry E, Arneodo A. Multifractal formalism for fractal signals: the structure-function approach versus the wavelet-transform modulus-maxima method. *Phys Rev E.* (1993). 47:875–84. doi:10.1103/physreve.47.875.
27. Bacry E, Muzy JF, Arneodo A. Singularity spectrum of fractal signals from wavelet analysis: exact results. *J Stat Phys.* (2013). 70:635–74. doi:10.1007/BF01053588

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relation that could be construed as a potential conflict of interest.

Copyright © 2020 Maskawa and Kuroda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX 1

Derivation of Eq 15

We introduce some notation for simplification of the description:

$$E_1(\lambda) = E(x(\lambda)), E_2(\lambda) = E(x^2(\lambda)), \Delta B_\lambda = B_A(\lambda + d\lambda) - B_A(\lambda), \mu_M = -\left(\gamma_M + \frac{1}{2}\sigma_M^2\right).$$

From **Eq. 13**, we have

$$\begin{aligned} E_1(\lambda + d\lambda) &= E(W_M(\lambda, \lambda + d\lambda) \cdot (x(\lambda) + a_A(\lambda)d\lambda + b_A(\lambda)\Delta B_\lambda)) \\ &= E(W_M(\lambda, \lambda + d\lambda))E(x(\lambda) + a_A(\lambda)d\lambda + b_A(\lambda)\Delta B_\lambda)) \\ &= E(W_M(\lambda, \lambda + d\lambda))(E_1(\lambda) + a_A(\lambda)d\lambda) \\ &= e^{(\mu_M + \frac{1}{2}\sigma_M^2)d\lambda}(E_1(\lambda) + a_A(\lambda)d\lambda), \end{aligned}$$

$$\begin{aligned} E_2(\lambda + d\lambda) &= E(W_M^2(\lambda, \lambda + d\lambda) \cdot (x(\lambda)2 + (2a_A(\lambda)x(\lambda) + b_A^2(\lambda))d\lambda) \\ &= E(W_M^2(\lambda, \lambda + d\lambda))E(x(\lambda)^2 + (2a_A(\lambda)x(\lambda) + b_A^2(\lambda))d\lambda) \\ &= E(W_M^2(\lambda, \lambda + d\lambda))(E_2(\lambda) + (2a_A(\lambda)E_1(\lambda) + b_A^2(\lambda))d\lambda)) \\ &= e^{(2\mu_M + 2\sigma_M^2)d\lambda}(E_2(\lambda) + (2a_A(\lambda)E_1(\lambda) + b_A^2(\lambda))d\lambda)) \end{aligned}$$

We also have

$$E_1^2(\lambda + d\lambda) = e^{(2\mu_M + \sigma_M^2)d\lambda}(E_1^2(\lambda) + 2a_A(\lambda)E_1(\lambda)d\lambda).$$

Because of the coincidence of the expected value and the standard deviation, we have $E_2(\lambda) = 2E_1^2(\lambda)$ and $E_2(\lambda + d\lambda) = 2E_1^2(\lambda + d\lambda)$.

Inserting those equalities and using approximation $e^{\sigma_M^2 d\lambda} = 1$, we have the constraint condition **15**.

APPENDIX 2

WTMM Method

The WTMM method builds a partition function from the modulus maxima of the wavelet transform defined at each scale s as the local maxima of $|W_\psi[f](x, s)|$ regarded as a function of x. Those maxima mutually connect across scales and form ridge lines designated as maxima lines. The set $\mathcal{L}(s_0)$ is the set of all the maxima lines l which satisfy

$$(x, s) \in l \Rightarrow s \leq s_0, \forall s \leq s_0 \Rightarrow \exists (x, s) \in l. \tag{36}$$

The partition function is defined by the maxima lines as

$$Z(q, s) = \sum_{l \in \mathcal{L}(s)} \left(\sup_{(x,s') \in l} |W_\psi\left[x, s'\right]| \right)^q. \tag{37}$$

Assuming power law behavior of the partition function

$$Z(q, s) \sim s^{\tau(q)}, \tag{38}$$

one can define the exponents $\tau(q)$. The singular spectrum $D(\alpha)$ can be computed using the Legendre transform of $\tau(q)$:

$$D(\alpha) = \min_q (q\alpha - \tau(q)). \tag{39}$$

frontiers in Physics

ORIGINAL RESEARCH
published: 27 November 2020
doi: 10.3389/fphy.2020.539521

Analysis of Stock Price Motion Asymmetry via Visibility-Graph Algorithm

Ruiyun Liu and Yu Chen**

Department of Human and Engineered Environmental Studies, Graduate School of Frontier Science, The University of Tokyo, Kashiwa, Japan

This paper is the first to differentiate between concave and convex price motion trajectories by applying visibility-graph and invisibility-graph algorithms to the analyses of stock indices. Concave and convex indicators for price increase and decrease motions are introduced to characterize accelerated and decelerated stock index increases and decreases. Upon comparing the distributions of these indicators, it is found that asymmetry exists in price motion trajectories and that the degree of asymmetry, which is characterized by the Kullback-Leibler divergence between the distributions of rise and fall indictors, fluctuates after a change in time scope. Moreover, asymmetry in price motion speeds is demonstrated by comparing conditional expected rise and fall returns on the node degrees of visibility and invisibility graphs.

Keywords: asymmetry, stock index, price motions, Kullback-Leibler divergence, visibility graph

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Huijie Yang,
University of Shanghai for Science and Technology, China
Yong Zou,
East China Normal University, China

****Correspondence:***
Ruiyun Liu
ry.liu@scslab.k.u-tokyo.ac.jp
Yu Chen
chen@edu.k.u-tokyo.ac.jp

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *01 March 2020*
Accepted: *23 October 2020*
Published: *27 November 2020*

Citation:
Liu R and Chen Y (2020) Analysis of Stock Price Motion Asymmetry via Visibility-Graph Algorithm. Front. Phys. 8:539521. doi: 10.3389/fphy.2020.539521

1 INTRODUCTION

The use of network science to perform time series analysis has emerged in recent decades. Of the numerous approaches to rendering a time series into a complex network, three major categories of approaches have most attracted researchers' attention [1–11]. The first approach uses recurrence networks and was introduced by Donner and Zou et al. in 2009 [5–8]. This approach analyzes phase space recurrence of a time series from a geometric point of view by interpreting the recurrence matrix of a time series as the adjacency matrix of a complex network. Transition networks represent the second major approach to transform a time series into a complex network. These networks are constructed by partitioning the phase space of a dynamic system and were introduced by Nicolis et al. in 2005 [9]. Hence, a node in a transition network represents a certain discrete state or pattern that describes the dynamic system. Direct links are established if one of the nodes is followed by another with nonzero probability along the time series [10]. The third category is the algorithmic group of visibility graphs (VG) [11]. In 2008, Lacasa et al. proposed an effective method called the visibility-graph algorithm (VGA) for converting a time series into a graph network by analyzing the mutual visibility relationships between points and cutting points in a computational geometry landscape [12, 13]. This concept has attracted great interest and numerous extensions of the standard VGA have been proposed. Luque et al. [14] came up with a simplified VGA called a horizontal visibility graph (HVG) to transform a time series into a complex network. Specifically, two observations are connected in an HVG if and only if there are no obstacles in between [15]. Based on the concepts of the VG and HVG, parametric VGs introduce a viewing angle α and allow one to study the dependence of network structural measures on α [16]. Limited penetrable VG (LPVG) is a less restricted HVG in which two observations are connected if either one has a larger value than the obstacles in between [17, 18].

Analyses of financial time series via a VG approach have been studied intensively [19–29]. For example, Long Yu discovered small-world characteristics in visibility-graph networks converted from the time series of the price of gold and its returns [24]. Moreover, Yao et al. found that exchange-rate networks converted from the currency-rate time series of the US dollar, euro, yen, and sterling against the Chinese yuan share consistent topological characteristics with hierarchical structures and mixed small-world and scale-free properties. They also discovered that network communities are actually composed of large numbers of trending points and small numbers of discrete peaks and trough points [25]. Furthermore, a novel method that combined VGA with link prediction was proposed by Zhang et al. to forecast the time series. Using fuzzy logic, better predictability can be achieved by fusing the direct and indirect effects of historical data [26].

Asymmetry in financial time series has generally been explored via statistical analysis [30–42]. Typically, it is found that the distribution of time horizons over which a detrended stock index moves from an arbitrary initial return to a predetermined positive level deviates to a symmetrically predetermined negative level [37–40]. This property is known as the gain-loss asymmetry and has been regarded as a characteristic of financial time series [30, 34, 36, 37]. Another well-known asymmetry, which describes the negative correlation between volatility and the direction of price motion, is the leverage effect [41]. Recently, Jiang et al. investigated asymmetry in large-scale price fluctuations. Analyses reveal that dynamic relaxation before and after large fluctuations is characterized by a power law with exponents p_+ and p_-. On minute time scales, large-fluctuation dynamics are time-reversely symmetric with $p_+ = p_-$. On daily time scales, however, large price fluctuations that approach financial crashes are asymmetric with $p_- \neq p_+$ [42]. We shall point out that the results of these studies are rather generic in the sense that only the price increases and decreases are considered. In fact, price motions can be classified in more detail into accelerated or decelerated rise (AR; DR) and fall (AF; DF), depending upon the convexity or concavity of the price motion trajectories. Symmetry analysis should also be performed by taking these four types of price motion trajectories into account. However, since different price motion modes can form different convoluted temporal structures, it would be quite difficult to decompose these price movements via statistical approaches.

In this paper, we propose to study the financial time series asymmetry via visibility-graph networks based on the intuition that network approaches may be more effective in identifying different price motion trajectories. Hence the terminology of symmetry in this study specifically refers to the topological symmetry of the price motion trajectories. In particular, in a stock price series, we are concerned with whether concave and convex price motions can form a time-reversed symmetry. The research question reads as follows: whether those accelerated/decelerated price rises are statistically symmetric with those decelerated/accelerated falls. It is worth mentioning that conventional VGA analysis cannot be applied directly to investigation of the topologically asymmetric properties of financial series because the method is incapable of distinguishing different stock price rise and fall trajectories by mapping the time series as a whole onto an undirected network. To solve this problem, an idea from Yan et al. (2012) [43] is borrowed to address discrimination between price movements via visibility and invisibility-graph (IVG) networks. Using these graph networks, asymmetry in stock index motion can be measured using concave or convex indicator distributions or expected returns that are conditional on node degrees, instead of the conventional waiting-time statistics.

Recently, fruitful results have been achieved in the investigation of time series time reversibility using the HVG method [44–51]. In a study on how crises affect the motions of US stock prices, different market price behaviors are identified by examining the series irreversibility evolved over the time [50, 51]. Based on the notion that the reversed and original processes are statistically distinguishable if a stationary process is time-reversible, we may postulate that a topological symmetric time series must be a time-reversible one, and vice versa. A quantitative analysis of this postulation is also conducted.

The whole paper consists of four sections. Following the introduction of the study in **Section 1**, the methodology is detailed in **Section 2**. Asymmetry in price motion trajectories and speeds is analyzed using graph networks in **Section 3 and 4**, respectively. Finally, conclusions are described in **Section 5**.

2 METHODOLOGY

2.1 Basic Algorithms

Graph networks for AR and DF motions of a stock index can be constructed by mapping a time series of length L, $X(t_i)(i < L)$ onto a graph network using VGA. To start, two arbitrary data points (t_a, x_a) and (t_b, x_b), where $a < b$, are specified. Two vertical lines are drawn exactly at t_a and t_b, with heights equal to the values of x_a and x_b, respectively. Next, the endpoints of the two vertical lines are connected via a straight line whenever vertical lines from any other data points within the range $(t_a,\ t_b)$ do not cut off the connection. That is, if any intermediate data point (t_c, x_c) fulfills the condition

$$\boldsymbol{x_c < x_b + (x_a - x_b)\frac{t_b - t_c}{t_b - t_a},\ a < c < b}, \tag{1}$$

the two data points (t_a, x_a) and (t_b, x_b) are visible to each other.

The invisibility-graph algorithm (IVGA) [43] can be used to build up the networks that describe DR and AF motions of a stock index. In contrast to the VGA, here, the data points (t_a, x_a) and (t_b, x_b) are connected only if the point (t_c, x_c) intersects the connecting line. Hence, the relationship between these three data points is transformed into the following:

$$\boldsymbol{x_c > x_b + (x_a - x_b)\frac{t_b - t_c}{t_b - t_a},\ a < c < b}. \tag{2}$$

Based on Yan's study [43], three conditions are further applied to distinguish AR and DF in VG as well as DR and AF in IVG. These conditions are stated as follows:

1. Any data point in the time series can be linked only to data points located on its left-hand side.
2. The link between (t_a, x_a) and (t_b, x_b) $(t_a < t_b)$ is connected only if $x_b > x_a$ when constructing the graph network for price increases, and vice versa when constructing the graph network for price decreases.
3. A moving time scope S is used to construct a graph network throughout the entire time series; therefore, there is no link between (t_a, x_a) and (t_b, x_b) if $|a - b| > S$.

With these three additional conditions, VGA and IVGA can capture rise and fall trends by eliminating links converted from the short-term disturbing price motions. We demonstrate the capture of time series topological symmetry in **Figures 1** and **2**. The networks in **Figures 1** and **2** are built from an artificial time series that combines the original Hang Seng Index sample data with its mirror-symmetric counterpart. **Figure 1** shows that VGA can distinguish the concave trajectory of the price rise from other price motions. For those data points located on the concave upward trend (from node 11 to node 14 in the original HSI section and from node 24 to node 26 in the reversed HSI section), the degree numbers of AR nodes are higher. For other data points, in contrast, the degree numbers are lower. VGA can also distinguish concave decreasing price motions as well (from node 3 to node 5 in the original HSI section and from node 15 to node 19 in the reversed HSI section). In **Figure 2**, however, the convex trajectories of price motions are captured effectively by IVGA based on the degree distributions for price rise (from node 10 to node 12 in the original HSI section and from node 26 to node 28 in the reversed HSI section) and price fall (from node 1 to node 3 in the original HSI section and node 17 to node 19 in the reversed HSI section). Obviously, the artificial time series has perfectly symmetric price motion trajectories since the concave (convex) rise in the original HSI section has a corresponding convex (concave) fall in the reversed HSI section. The identical rising and falling degree distributions of VGs and IVGs in **Figures 1** and **2** reveal that this artificial time series does have a perfect topological symmetry. Given the definition of time reversibility, a topologically symmetrical time series must be a time-reversible one. The reason lies in that if a time series is topologically symmetrical, the reversed and original processes should be statistically indistinguishable with regard to the degree distributions of nodes.

2.2 Concave and Convex Motion Indicators

To formalize the algorithms shown above, we specify a node b in a time series of length L and set a time scope S $(S < L/2)$. Based on **Eq. 1**, node a, which has a link to node b in the visibility-graph network, should belong to the following set:

$$\boldsymbol{A} = \left[(t_a, x_a) | x_a > x_b + (x_c - x_b)\frac{t_b - t_a}{t_b - t_c}, a < c < b < L \right]. \quad (3)$$

Employing the three additional conditions described in **Section 2.1**, the subsets of nodes that connect to node b in the price rise and price fall trajectories are defined as follows:

$$\begin{aligned} \boldsymbol{A_R} &= \{(t_a, x_a) | (t_a, x_a) \in \boldsymbol{A}, x_b > x_a \quad \text{and} \quad |a - b| \le S\}, \\ \boldsymbol{A_F} &= \{(t_a, x_a) | (t_a, x_a) \in \boldsymbol{A}, x_b < x_a \quad \text{and} \quad |a - b| \le S\}. \end{aligned} \quad (4)$$

Hence, the degree of node b in the visibility-graph rise and fall networks reads as

$$\boldsymbol{D^X_{VG}(b) = n(A_X)}, \quad \text{with} \quad \boldsymbol{X = R, F}. \quad (5)$$

By definition, $D^X_{VG} \in [1, 2S]$, $X = R, F$. The concave motion indicator for a node i is proposed to be the following:

$$\boldsymbol{I^X_{CC}(i) = \frac{D^X_{VG}(i)}{S}}, \quad \text{with} \quad \boldsymbol{X = R, F}. \quad (6)$$

Note that, for an ideal concave trajectory consisting of L data points, the concave indicators along the time axis can be sketched as in **Figure 3**. For a realistic time series, the distribution of this indicator measures how perfectly a concave curve could fit the ideal AR or DF price motion trajectory.

Hence, the mean value of $I^{R,F}_{CC}$ can be a measure of the smoothness of AR or DF price motions. The higher the $\overline{I^{R,F}_{CC}}$, the less the zigzag price variation, and vice versa.

In the same manner, the mathematical set that describes the invisibility-graph network can be written as

$$\boldsymbol{A'} = \left\{ (t_a, x_a) | x_a < x_b + (x_c - x_b)\frac{t_b - t_a}{t_b - t_c}, a < c < b < L \right\}. \quad (7)$$

Subsets of the rise and fall trend read, respectively, as follows:

$$\begin{aligned} \boldsymbol{A'_R} &= \{(t_a, x_a) | (t_a, x_a) \in \boldsymbol{A'}, x_b > x_a \quad \text{and} \quad |a - b| \le S\}, \\ \boldsymbol{A'_F} &= \{(t_a, x_a) | (t_a, x_a) \in \boldsymbol{A'}, x_b < x_a \quad \text{and} \quad |a - b| \le S\}. \end{aligned} \quad (8)$$

The degree of node b in the invisibility rise and fall networks is

$$\boldsymbol{D^X_{IVG}(b) = n\left(A'_X\right)}, \quad \text{with} \quad \boldsymbol{X = R, F}. \quad (9)$$

Convex motion indicators can thus be defined via

$$\boldsymbol{I^X_{CV}(i) = \frac{D^X_{IVG}(i)}{S}}, \quad \text{with} \quad \boldsymbol{X = R, F}, \quad (10)$$

the distribution of which measures how perfectly a convex curve can fit an ideal DR or AF motion price trajectories. Likewise, $\overline{I^X_{CV}}$ measures the smoothness of the price motion trajectory.

A quantitative measure of the topological asymmetry in price motion trajectory can be done via the distinguishability between distributions of the concave and convex rise/fall indicators defined above. Specifically, denoting the distribution of rise indicators as $P(I^R_{CC,CV})$ and the distribution of fall indicators as $P(I^F_{CC,CV})$, a topologically symmetric time series should have $P(I^R_{CC,CV}) = P(I^F_{CC,CV})$.

On the other hand, the degree of topological asymmetry is measured by calculating the Kullback-Leibler divergence (KLD) of $P(I^R_{CC,CV})$ and $P(I^F_{CC,CV})$. Stemming from information theory, KLD is employed as a measure of the distance between two probability distributions [52, 53]. KLD of concave/convex rise and fall indicators distributions can be calculated as follows:

$$\boldsymbol{D(P\|Q) \underline{\underline{def}} \sum P(I^R_Y) \log \frac{P(I^R_Y)}{Q(I^F_Y)}}, \quad \text{with} \quad \boldsymbol{Y = CC, CV}, \quad (11)$$

which equals 0 if and only if $P(I^R_Y) = Q(I^F_Y)$ and exceeds 0 otherwise.

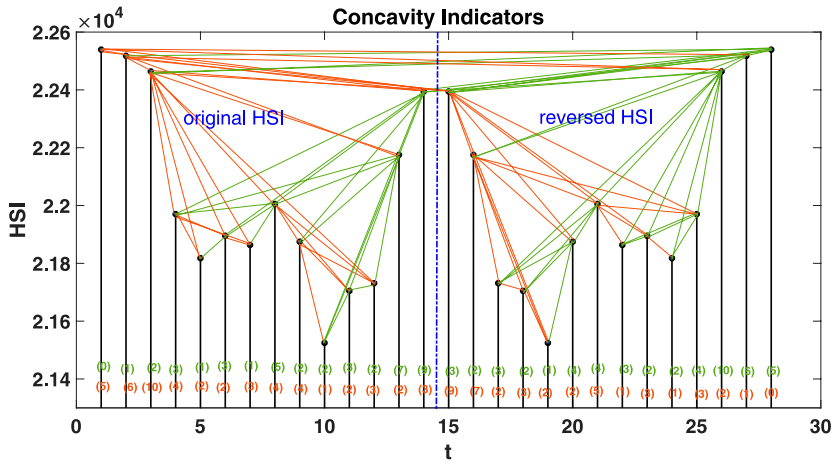


FIGURE 1 | Application of VGA to the daily Hang Seng Index from August 13, 2013, to September 2, 2013, and its time-reversely mirrored data set. Note that the perfect reflection symmetry between AR [nodes 10–14] and DF [nodes 1–5] is demonstrated via identically distributed degrees in the two networks.

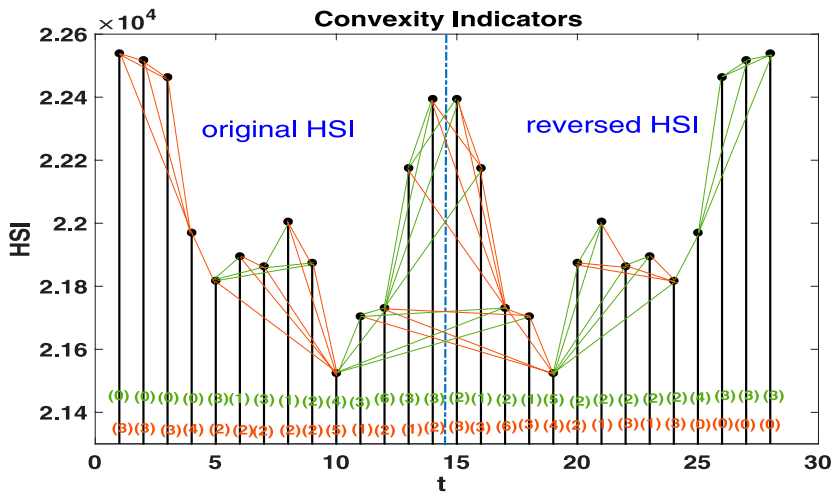


FIGURE 2 | Application of the modified IVGA to the daily Hang Seng Index from August 13, 2013, to September 2, 2013, and its time-reversely mirrored data. Note that perfect reflectional symmetry between DR [nodes 5–8 in **(A)**] and AF [nodes 6–10 in **(B)**] is demonstrated via identically distributed degrees in the two networks.

3 ASYMMETRY IN TRAJECTORIES OF PRICE MOTION

Analyses that include the application of VGA and IVGA to the stock indices of various countries and regions are presented in this section. Eight data sets from international stock market indices that span from June 28, 1999, to June 28, 2019, were selected. These include the Hong Kong Hang Seng, Dow Jones Industrial Average, Japanese Nikkei 225, London FTSE 100, German DAX, French CAC 40, Shanghai SSE Composite, and Indian BSE. Here, we set the time scope as S = 262 since the number can be interpreted as the trading days in a year.

Networks converted from the Hang Seng Index are used to demonstrate how the concave and convex motion indicators [**Eqs 6** and **10**] change along with the time evolution of price.

In the upper panel of **Figure 4**, the long-lasting bubble right before the subprime mortgage crisis in 2008 is characterized by clusters of large concave rise indicators. The large concave fall indicators in the lower panel of **Figure 4**, however, characterize the decelerated fall of the index over 3 years after it reaches 18,000 points on March 27, 2000. The large convex rise indicators in the upper panel of **Figure 5** show that the Hang Seng Index exhibits a decelerated rise from 2009 to 2011 after the subprime mortgage crisis. The notorious 2008–2009 crisis is represented by the extraordinarily large values of convex fall indicators in the lower panel of **Figure 5**. Both figures show that the concave rise indicator changes asynchronously with its fall counterpart, just as the convex fall indicator changes asynchronously with its rise counterpart.

To illustrate asymmetry in the price motion trajectories for these eight financial time series, distributions of $I_{CC}^{R,F}$ and $I_{CV}^{R,F}$ are

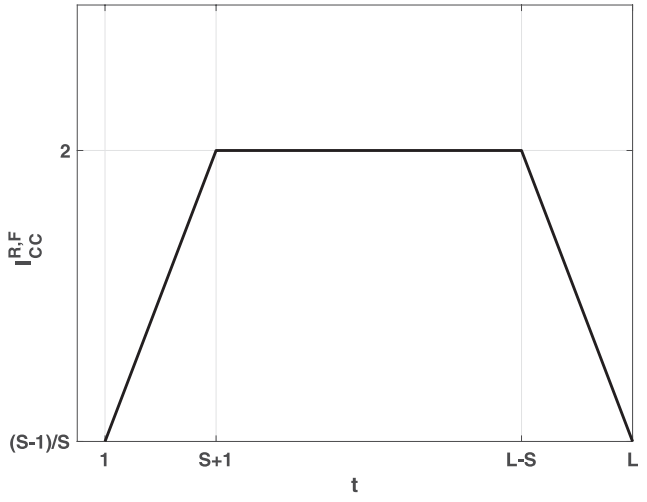


FIGURE 3 | The distribution of concave indicators $I^{R,F}_{CC}$ in an ideal concave trajectory of length L with time scope S.

obtained from the statistics of VG and IVG networks, respectively. Distributions of rise and fall I_{CC} for the eight stock indices are shown in **Figures 6A–H**. As in the small I_{CC} regime, the rise distributions are similar to the fall distributions in all cases. However, as I_{CC} becomes larger than 10^{-2}, the rise distributions start to deviate from the fall distributions. This suggests that the AR and DF motions of stock indices are essentially asymmetric. Similarly, distributions of rise and fall I_{CV} are displayed in **Figures 7A–H**. Again, deviation in rise and fall distributions can be found in the range $I_{CV} > 10^{-2}$, indicating that DR and AF motions of stock indices are essentially asymmetric, too.

Average values of rise and fall indicators are also calculated for the eight stock indices. The rise and fall $\overline{I_{CC}}$ and $\overline{I_{CV}}$ values are listed in **Tables 1** and **2**, respectively. The SSEC AR and DF motions are smoothest since its $\overline{I^R_{CC}}$ and $\overline{I^F_{CC}}$ rank first in magnitude among the others in the table. With regard to the DR motions of stock indices, the Indian market behaves in the smoothest manner because the $\overline{I^R_{CV}}$ of the BSESN is larger than those of any other indices. On the other hand, the zigzag AF appears less frequently in the Chinese stock market than in other markets, as suggested by the fact that the SSEC has the largest $\overline{I^F_{CV}}$ in **Table 2**. These observations are consistent with empirical evidence that emerging financial markets are less efficient than developed markets. The presence of fewer price oscillation in emerging markets implies that investors are more likely to form a herd.

As the time scale can be an important factor that influences the topological asymmetry in stock price motion, we measure the KLD between the rise and fall distributions of I_{CC} and I_{CV} with different time scopes as $S = 100, 130, 160, \ldots, 610$. Values of KLD between the rise and fall I_{CC} distributions are shown in **Figure 8** as functions of the time scope for the eight stock indices, as well as for a purely random time series. Compared with the KLD of the random series depicted in blue line dots, the DJIA and BSESN have an impressively higher value than any other stock indices. This indicates higher degrees of asymmetry between AR and DF price motions for these two indices. On the other hand, the dependence of KLD on the time scope is rather weak outside of the DJIA, SSEC, and BSESN.

The KL-divergence values between rise and fall I_{CV} distributions are shown as functions of time scopes in **Figure 9** for the eight stock indices, as well as for the purely random series. Except in the case of BSESN data, the overall degree of asymmetry between DR and AF price motions is weaker than that in **Figure 8**. However, the dependence of the KL divergence on the time scope strengthens in all cases except for the FCHI, FTSE, and Nikkei 225.

In **Figures 8** and **9**, we note that the KLDs for random series are close to 0 and vary little as the parameter S changes. This is in agreement with the postulation that time-reversible time series are topologically symmetrical. On the other hand, the bigger KLDs for stock indices shown in the same figures are consistent with the finding in the previous study on time irreversibility in stock indices via the HVG method [45], which states that a chaotic time series results in a bigger KLD between in- and out-distributions than a Gaussian time series does.

In addition, KLDs of the BSESN are found to be dramatically higher than those of other stock markets in **Figures 8** and **9**. In particular, the observation that the BSESN KLD follows an increasing trend in **Figure 9** indicates that the Indian stock market index exhibits a long-term, low-speed rise. This is in line with observations that the Indian stock market was in a bull market for over 20 years until the coronavirus outbreak. The KLD results in **Figure 8** also show that the DJIA has relatively large topological asymmetry between AR and DF price motions during the period from 1999 to 2019, which indicates that the price is pushed upwards mostly by AF motions in the USA bull markets. As Yan et al. [43] published, the AR price motion implies a superexponential growth typically caused by investors' herding behavior. Over the past decade, there have been several reports on herd buying behavior of AAPL and MSFT [54, 55]. A report published on December 4, 2019 [56], said "The Dow Jones Industrial Average owes Apple and Microsoft corporation a big thanks." These reports may explain why AR motions dominate the movement of DJIA index.

4 ASYMMETRY IN SPEEDS OF PRICE MOTION

Yan et al. argued that a higher degree number k_a indicates a higher possibility that the time series is growing at a superexponential rate at time tick a; hence, the degree number of VG/IVG could be a good indicator for the proximity to the point of a bubble-and-crash regime shift [43]. However, we should point out that such an argument may not be accurate because the high VG degree number can also be a result of a relatively low-speed and smooth growth as long as the time scope is large enough. As the price approaches the critical point in stock markets, the magnitude of fluctuations becomes dramatically large. Therefore, the correlation between degree number k_a

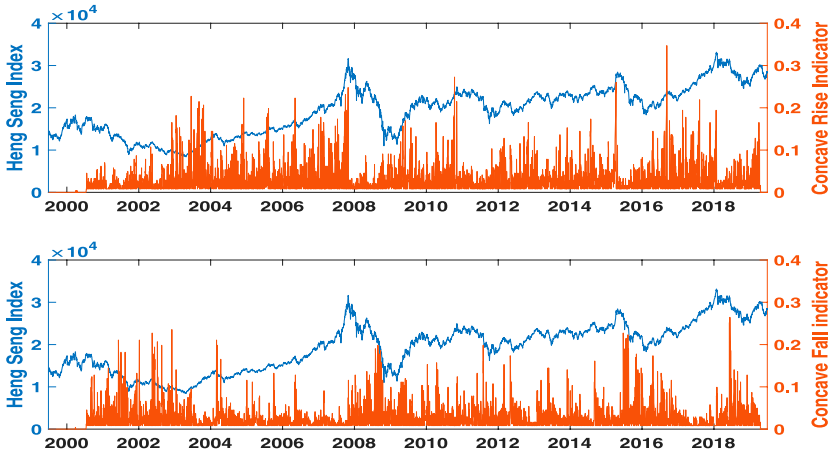


FIGURE 4 | Price motion indicators measure AR and DF in the Hang Seng Index from June 28, 1999, to June 28, 2019. Concave rise and fall indicators are shown in the upper and lower panels, respectively. The Hang Seng Index is plotted using blue lines, while the indicators use red bars. The time scope is set to S = 262, which is equal to the number of trading days per year.

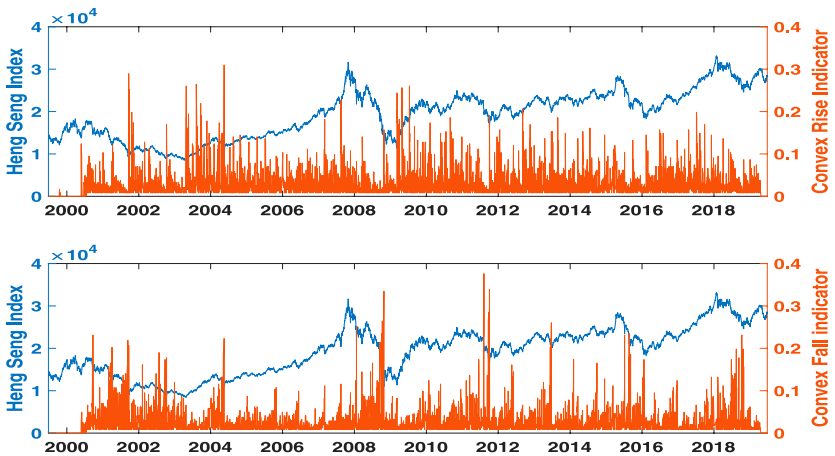


FIGURE 5 | Price motion indicators measure DR and AF in the Hang Seng Index from June 28, 1999, to June 28, 2019. Convex rise and fall indicators are shown in the upper and lower panels, respectively. The Hang Seng Index is plotted using blue lines and indicators are represented using red bars. Here, the time scope is set to S = 262, which is equal to the number of trading days per year.

and price return r_a should be a more appropriate indicator showing the possibility for the stock index to grow/drop at the exponential rate (for AR and DF motions) or at the logarithmic rate (for DR and AF motions). Comparisons of these indicators may reveal the asymmetry in the speeds of price growth and drop.

In order to measure the asymmetry in the speeds of price motion, the expected price return r over a unit time span is calculated conditionally on the node degree k in VG and IVG networks. The conditional expected index return is defined as follows:

$$\langle \boldsymbol{r}|\boldsymbol{k}\rangle = \frac{\sum_a \boldsymbol{r}_a \delta(\boldsymbol{k}_a - \boldsymbol{k})}{\sum_a \delta(\boldsymbol{k}_a - \boldsymbol{k})}, \tag{12}$$

where $r_a = \log P_a - \log P_{a-1}$. P_a is the stock index, k_a is the node degree at time tick a, and $\delta(x)$ is the Kronecker delta function. The conditional expected price rise and fall returns on the node degree in VG and IVG networks are defined as follows in order to illustrate the asymmetry in the speeds of price growth and price drop for the eight aforementioned financial time series:

$$\langle \boldsymbol{r}_{rise}|\boldsymbol{k}\rangle = \frac{\sum_a \boldsymbol{H}(\boldsymbol{r}_a)\boldsymbol{r}_a \delta(\boldsymbol{k}_a - \boldsymbol{k})}{\sum_a \delta(\boldsymbol{k}_a - \boldsymbol{k})}, \tag{13}$$

$$\langle \boldsymbol{r}_{fall}|\boldsymbol{k}\rangle = \frac{\sum_a [\boldsymbol{1} - \boldsymbol{H}(\boldsymbol{r}_a)]\boldsymbol{r}_a \delta(\boldsymbol{k}_a - \boldsymbol{k})}{\sum_a \delta(\boldsymbol{k}_a - \boldsymbol{k})}, \tag{14}$$

where $H(x)$ is a Heaviside step function [12].

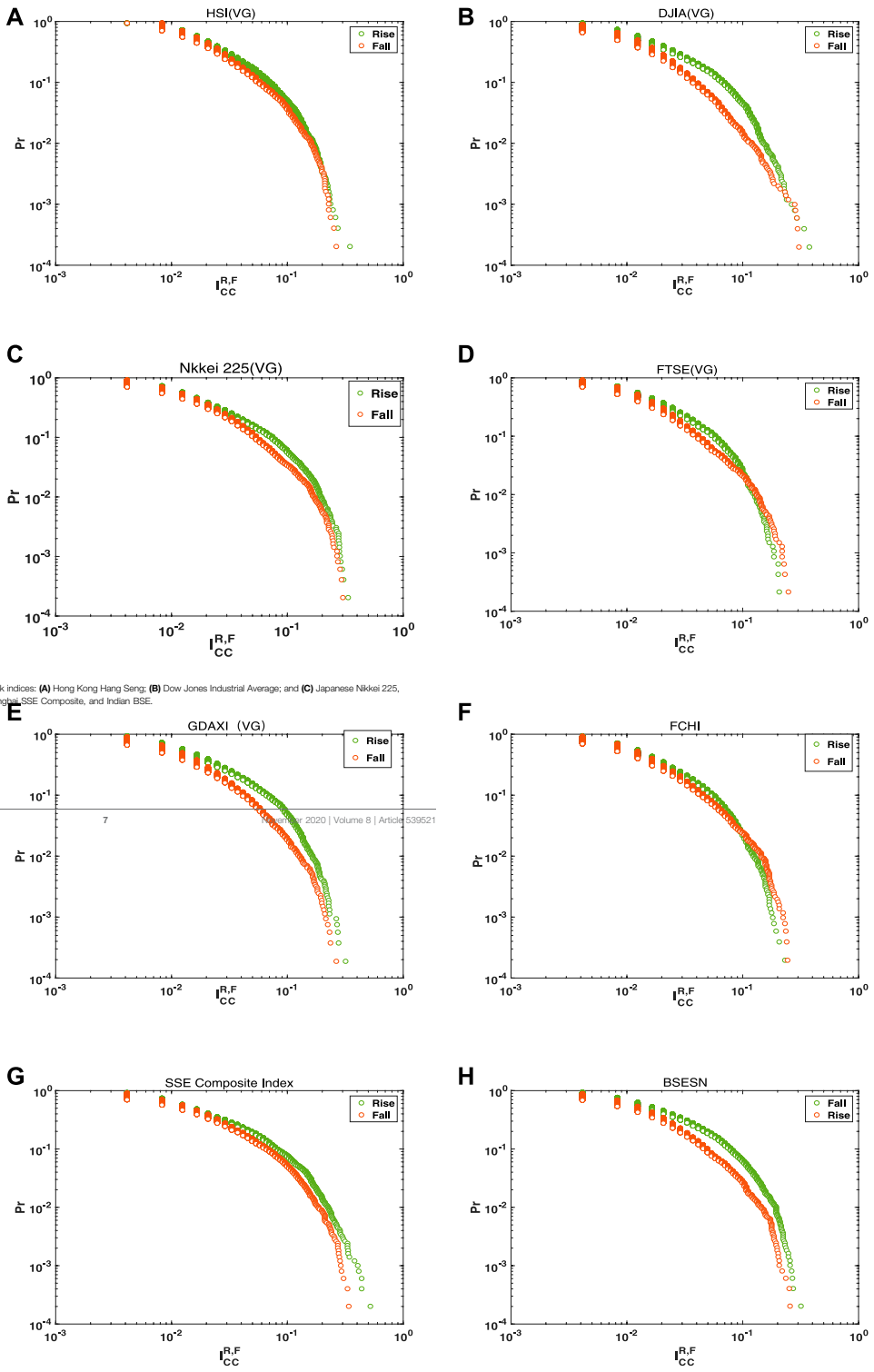


FIGURE 6 | Rise and fall $I_{CC}^{R,F}$ distributions for eight stock indices: **(A)** Hong Kong Hang Seng; **(B)** Dow Jones Industrial Average; and **(C)** Japanese Nikkei 225, London FTSE 100, German DAX, French CAC40, Shanghai SSE Composite, and Indian BSE.

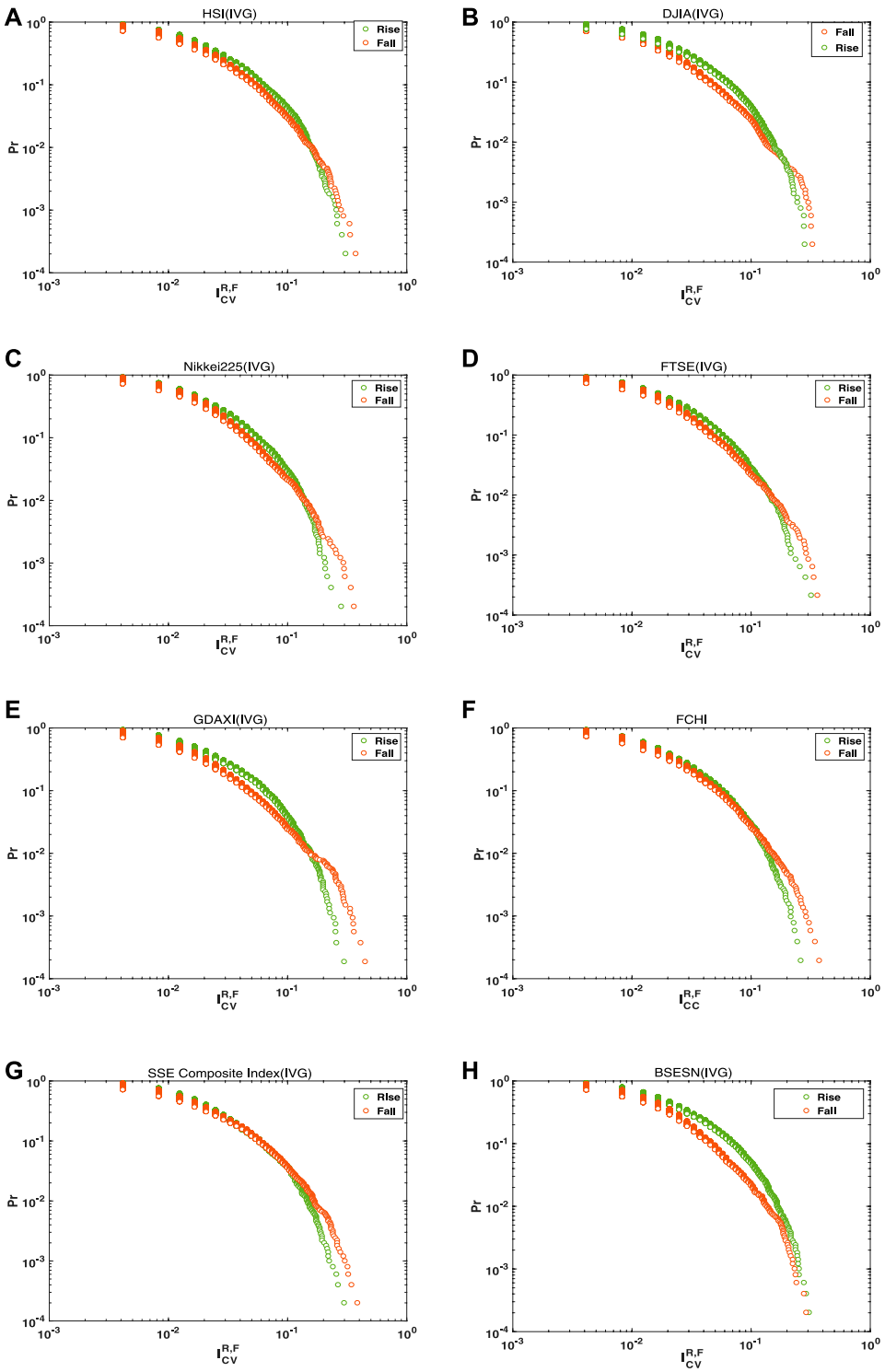


FIGURE 7 | Distributions of rise and fall $I_{CV}^{R,F}$ values for the eight stock indices: **(A)** Hong Kong Hang Seng; **(B)** Dow Jones Industrial Average; and **(C)** Japanese Nikkei 225, London FTSE 100, German DAX, French CAC40, Shanghai SSE Composite, and Indian BSE.

The expected rise and fall returns that are conditioned on the VG node degree for eight stock indices are plotted in **Figure 10**, while those conditioned on the IVG node degree are in **Figure 11**. Expected return data points are fitted linearly and shown as black lines for rise motions and pink lines for fall motions, respectively, in **Figures 10** and **11**. The slopes of the fitting lines show the correlated relationships between expected returns and degree regardless of whether the trajectories are concave or convex. The larger the absolute slope, the more significantly a stock index exhibits a superexponential or a logarithmic motion.

For concave price trajectories in **Figure 10**, stock indices can be classified into four categories. For category I, which includes the DAX and FCHI, the rise in absolute slope is almost the same as the fall in absolute slope, and both the rise and fall slopes are quite small. This implies that AR and DF are not the main forms in which these two stock markets exhibit bubbles and crashes. For category II, which includes the DJIA and FTSE, the absolute rise slope is far smaller than the fall slope. This suggests that it is possible for the price to decrease at a logarithmic fall rate after stock crashes in these two markets. For case III, which includes

TABLE 1 | Rise and fall $\overline{I_{CC}}$ values of eight stock indices.

$\overline{I^X_{CC}}$	HSI	DJIA	Nikkei	FTSE	DAX	FCHI	SSEC	BSESN
Rise	0.0275	0.0276	0.0291	0.0241	0.0251	0.0234	0.0326	0.0316
Fall	0.0243	0.0198	0.0248	0.0208	0.0216	0.0217	0.0279	0.0228

TABLE 2 | Rise and fall $\overline{I_{CV}}$ values of eight stock indices.

$\overline{I^X_{CV}}$	HSI	DJIA	Nikkei	FTSE	DAX	FCHI	SSEC	BSESN
Rise	0.0277	0.0276	0.0252	0.0258	0.0278	0.0252	0.0259	0.0304
Fall	0.0245	0.0225	0.0228	0.0234	0.0244	0.0243	0.0260	0.0228

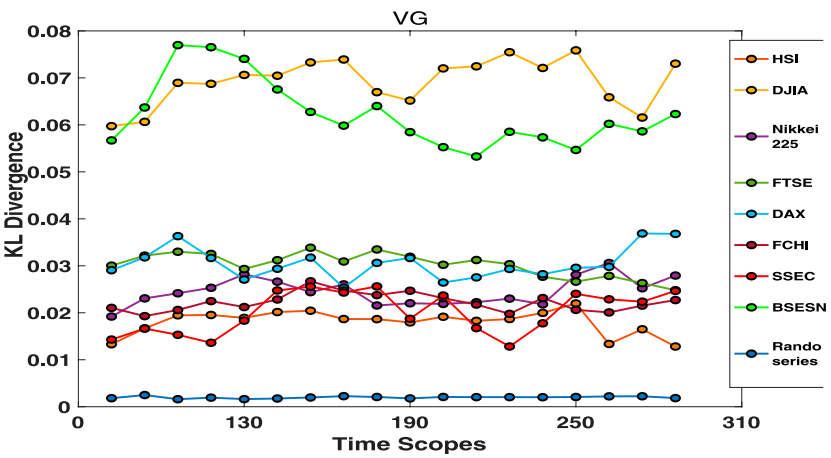


FIGURE 8 | KL divergence between I_{CC} rise and fall distributions as functions of time scopes for eight stock indices and a random series.

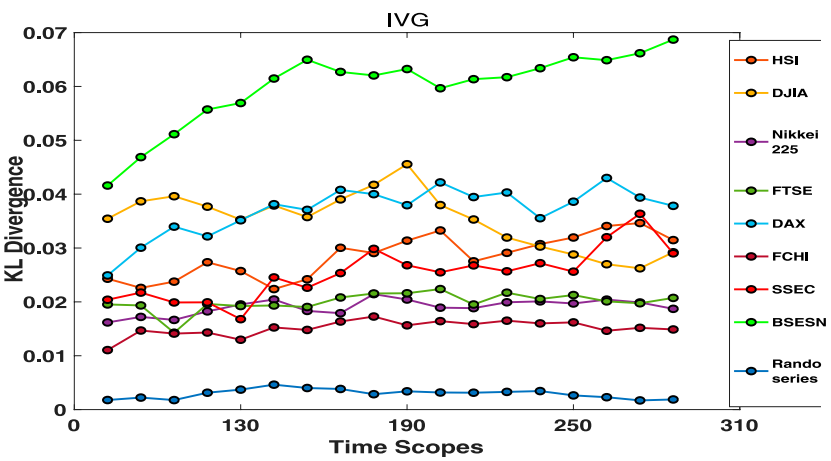


FIGURE 9 | KL divergences between I_{CV} rise and fall distributions as functions of the time scope for eight stock indices and a pure random series.

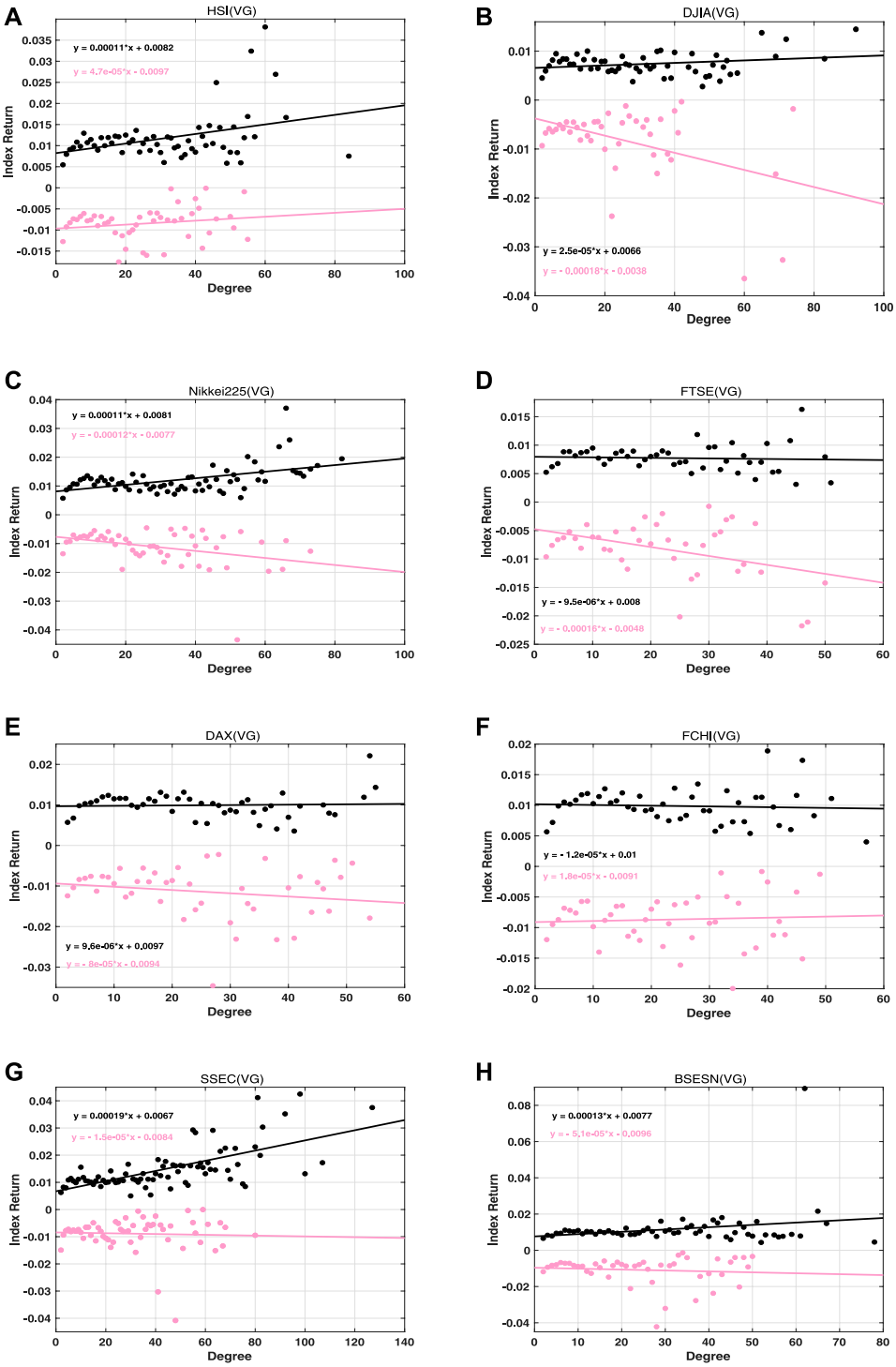


FIGURE 10 | Conditional expected index returns for the degree of rise and fall price motions within VG networks for the eight stock indices. Red indicates a rise and black indicates a fall.

the HSI, SSEC, and BSESN, the absolute rise slope is larger than the fall slope. Obviously, these markets are likely to increase via a superexponential growth rate within bubble regimes. For case IV, which includes the Nikkei 225, the absolute rise slope and absolute fall slope are almost the same. However, their values are bigger than those noted in case I. This means that the Nikkei 225 rise and fall trajectories contain many concave motions in the bubble-and-crash regime. For the convex trajectories in

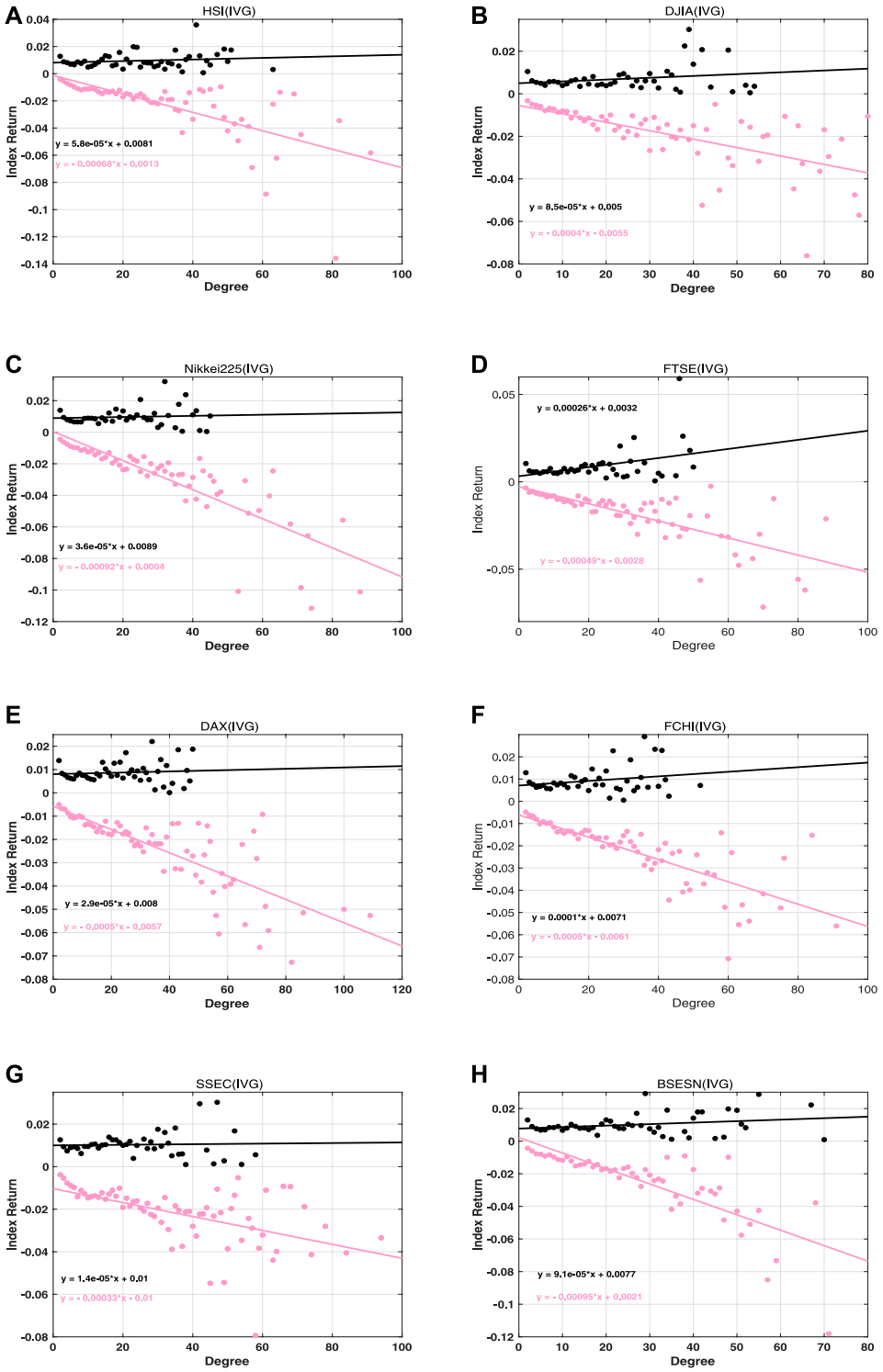


FIGURE 11 | Conditional expected index returns based on the degree of rise and fall price motions in IVG networks for the eight stock indices. Red indicates a rise and black indicates a fall.

Figure 11, the absolute fall slopes exceed the rise slopes for all of the stock indices. This implies that all of the markets crash at logarithmic rates. We also note that the FTSE and FCHI have rather large absolute rise slopes, which means that the price approaches the critical point in the DR way within the bubble regimes.

Overall, the findings in this section agree with the previous studies of gain-loss asymmetry [37–40]. In particular, stock market prices fall faster than they rise in developed countries. The analysis in this study provides a clearer picture regarding the conclusion made in Ref. [40] that the rise speed overtakes the fall speed in developing country stock markets, such as those of India and China. Indeed, the speed of AR price motion is larger than that of DF motion, while the speed of AF price motion exceeds that of DR motion, just as in mature markets.

5 CONCLUSION

In this paper, we developed a new concept of financial time series asymmetry based on the topological distinguishability of price motion trajectories. A new application of VGA and IVGA was developed to capture different types of price motion trajectories. Measures based on VGA and IVGA were employed to analyze asymmetry in price motion trajectories as well as in price motion speeds. To analyze topological asymmetry in price motion trajectories, we compared the distributions of concave and convex indicators for both rise and fall price motions. Deviations in rise and fall indicator distributions among VG and IVG networks showed that AR-DF and DR-AF stock index motions are asymmetric with each other. To investigate the influences of time scopes, the relation between KLD and time scope was also illustrated. Unlike with the random series, the KLD of stock index rise and fall indicator distributions is significant and the dependence of KLD on time scopes is strong. This is especially true for Indian and American stock indices.

Furthermore, we calculated the conditional expected index return on node degree to show asymmetry in price motion speeds. The rise and fall conditional expected index returns on VG or IVG network node degrees were distributed in an asymmetric manner, which indicates that asymmetry is embedded in AR-DF and DR-AF price motion speeds when the stock index approaches a bubble-and-crash regime shift. Our result was in line with gain-loss asymmetry overall. However, it offered details regarding why AF motions in emerging markets (e.g., China and India) contribute to faster rises and slower falls.

As a byproduct of this study, we also get some knowledge of the relationship between the topological symmetry and the time reversibility of a time series. By the definition of time reversibility, we proved, with an artificially combined piece of HSI time series, that the topologically symmetrical time series must be time-reversible. On the other hand, by checking the topological symmetry of a random series, the numerical evidence, which supports the postulation that a time-irreversible series must be topologically asymmetric, has also been found.

Future research will include exploration of topological asymmetry in other empirical data that exhibits chaotic behaviors, such as sunspots, heartbeats, and earthquake waves. The relation between topological symmetry and time reversibility is also to be investigated theoretically. Finally, the most important task is to explore how topological symmetry among financial time series affects the time reversibility. In this sense, we must study the network properties of VG and IVG networks and identify network characteristics right before large-scale price changes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

RL and YC contributed to the conception of the study; RL performed the data analyses and wrote the article; CY helped perform the analysis with constructive discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.539521/full#supplementary-material.

REFERENCES

1. Zhang J, Small M. Complex network from pseudoperiodic time series: topology versus dynamics. *Phys Rev Lett* (2006) 96:238701. doi:10.1103/physrevlett.96.238701.
2. Zhang J, Luo X, Small M. Detecting chaos in pseudoperiodic time series without embedding. *Phys Rev* (2006) 73:016216. doi:10.1103/physreve.73.016216.
3. Zhang J, Sun J, Luo X, Zhang K, Nakamura T, Small M. Characterizing pseudoperiodic time series through the complex network approach. *Phys Nonlinear Phenom* (2008) 237:2856–65. doi:10.1016/j.physd.2008.05.008.
4. Xu X, Zhang J, Small M. Superfamily phenomena and motifs of networks induced from time series. *Proc Natl Acad Sci Unit States Am* (2008) 105: 19601–5. doi:10.1073/pnas.0806082105.
5. Marwan N, Donges JF, Zou Y, Donner RV, Kurths J. Complex network approach for recurrence analysis of time series. *Phys Lett A* (2009) 373: 4246–54. doi:10.1016/j.physleta.2009.09.042.
6. Donner RV, Zou Y, Donges JF, Marwan N, Kurths J. Recurrence networks—a novel paradigm for nonlinear time series analysis. *New J Phys* (2010) 12(3): 033025. doi:10.1088/1367-2630/12/3/033025.
7. Donner RV, Zou Y, Donges JF, Marwan N, Kurths J. Ambiguities in recurrence-based complex network representations of time series. *Phys Rev E* (2010) 81: 015101. doi:10.1103/physreve.81.015101.
8. Donner RV, Small M, Donges JF, Marwan N, Zou Y, Xiang R, et al. Recurrence-based time series analysis by means of complex network methods. *Int J Bifurcation Chaos* (2011) 21:1019–46. doi:10.1142/s0218127411029021.
9. Nicolis G, Cantu AG, Nicolis C. Dynamical aspects of interaction networks. *Int J Bifurcation Chaos* (2005) 15:3467–80. doi:10.1142/s0218127405014167.
10. McCullough M, Small M, Stemler T, Ho-Ching Iu H. Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems. *Chaos* (2015) 25:053101. doi:10.1063/1.4919075.
11. Nuñez AM, Lacasa L, Gomez JP, Luque B. Visibility algorithms: a short review. In: Y Zhang, editor. *New frontiers in graph theory*. Rijeka, Croatia: IntechOpen (2012) p. 119–52.
12. Lacasa L, Luque B, Ballesteros F, Luque J, Nuño JC. From time series to complex networks: the visibility graph. *Proc Natl Acad Sci Unit States Am* (2008) 105:4972–5. doi:10.1073/pnas.0709247105.

13. Lacasa L, Luque B, Luque J, Nuño JC. The visibility graph: a new method for estimating the Hurst exponent of fractional Brownian motion. *Europhys Lett* (2009) 86:30001. doi:10.1209/0295-5075/86/30001.
14. Luque B, Lacasa L, Ballesteros F, Luque J. Horizontal visibility graphs: exact results for random time series. *Phys Rev E* (2009) 80:046103. doi:10.1103/physreve.80.046103.
15. Lacasa L, Toral R. Description of stochastic and chaotic series using visibility graphs. *Phys Rev E* (2010) 82:036120. doi:10.1103/physreve.82.036120.
16. Bezsudnov IV, Snarskii AA. From the time series to the complex networks: the parametric natural visibility graph. *Phys Stat Mech Appl* (2014) 414:53–60. doi:10.1016/j.physa.2014.07.002.
17. Zhou T-T, Jin N-D, Gao Z-K, Luo Y-B. Limited penetrable visibility graph for establishing complex network from time series. *Acta Phys Sin* (2012) 61(3): 030506. doi:10.7498/aps.61.030506.
18. Gao Z-K, Hu L-D, Zhou T-T, Jin N-D. Limited penetrable visibility graph from two-phase flow for investigating flow pattern dynamics. *Acta Phys Sin* (2013) 62(11):110507. doi:10.7498/aps.62.110507.
19. Stephen M, Gu C, Yang H. Visibility graph based time series analysis. *PloS One* (2015) 10:e0143015. doi:10.1371/journal.pone.0143015.
20. Vamvakaris MD, Pantelous AA, Zuev KM. Time series analysis of S&P 500 index: a horizontal visibility graph approach. *Phys Stat Mech Appl* (2018) 497: 41–51. doi:10.1016/j.physa.2018.01.010.
21. Fan X, Li X, Yin J, Tian L, Liang J. Similarity and heterogeneity of price dynamics across China's regional carbon markets: a visibility graph network approach. *Appl Energy* (2019) 235:739–46. doi:10.1016/j.apenergy.2018.11.007.
22. Qian M-C, Jiang Z-Q, Zhou W-X. Universal and nonuniversal allometric scaling behaviors in the visibility graphs of world stock market indices. *J Phys Math Theor* (2010) 43:335002. doi:10.1088/1751-8113/43/33/335002.
23. Wang N, Dong L, Wang Q. Visibility graph analysis on quarterly macroeconomic series of China based on complex network theory. *Phys Stat Mech Appl* (2012) 391:6543–55. doi:10.1016/j.physa.2012.07.054.
24. Long Y. Visibility graph network analysis of gold price time series. *Phys Stat Mech Appl* (2013) 392:3374–84. doi:10.1016/j.physa.2013.03.063.
25. Yao C-zhong, Lin J-N. A visibility graph approach to CNY exchange rate networks and characteristic analysis. *Discrete Dyn Nat Soc* (2017) 2017: 5632374. doi:10.1155/2017/5632374.
26. Zhang R, Ashuri B, Deng Y. A novel method for forecasting time series based on fuzzy logic and visibility graph. *Adv Data Anal Classif* (2017) 11:759–83. doi:10.1007/s11634-017-0300-3.
27. Liu C, Zhou W-X, Yuan W-K. Statistical properties of visibility graph of energy dissipation rates in three-dimensional fully developed turbulence. *Phys Stat Mech Appl* (2010) 389:2675–81. doi:10.1016/j.physa.2010.02.043.
28. Ni X-H, Jiang Z-Q, Zhou W-X. Degree distribution of the visibility graphs mapped from fractional Brownian motions and multifractal random walks. *Phys Lett Sec A* (2008) 373:3822. doi:10.1016/j.physleta.2009.08.041.
29. Liu C, Zhou W-X, Yuan W-K. Statistical properties of visibility graph of energy dissipation rates in three-dimensional fully developed turbulence. *Physica A* (2010) 389:2675–81. doi:10.1016/j.physa.2010.02.043.
30. Siven JV, Todd Lins J. Gain/loss asymmetry in time series of individual stock prices and its relationship to the leverage effect. Preprint repository name [Preprint] (2009) Available from: https://arxiv.org/abs/0911.4679.
31. Tsay RS *Analysis of financial time series*. Hoboken, NJ: John Wiley & Sons (2005) p. 720.
32. Yan J. *Asymmetry, fat-tail, and autoregressive conditional density in financial return data with systems of frequency curves*. Iowa City, Iowa: University of Iowa: Department of Statistics and Actuarial Science (2005) p. 26.
33. Verhoeven P, McAleer M. Fat tails and asymmetry in financial volatility models. *Math Comput Simulat* (2004) 64:351–61. doi:10.1016/s0378-4754(03)00101-0.
34. Jensen MH, Johansen A, Simonsen I. Inverse statistics in economics: the gain–loss asymmetry. *Phys Stat Mech Appl* (2003) 324(1-2):338–43. doi:10.1016/s0378-4371(02)01884-8.
35. Donangelo R, Jensen MH, Simonsen I, Sneppen K. Synchronization model for stock market asymmetry. *J Stat Mech* (2006) 11:L11001. doi:10.1088/1742-5468/2006/11/L11001.
36. Grudziecki M, Gnatowska E, Karpio K, Orłowski A, Załuska-Kotur M. New results on gain-loss asymmetry for stock markets time series. *Acta Phys Pol* (2008) 114:569. doi:10.12693/aphyspola.114.569.
37. Siven J, Lins J, Hansen J. A multiscale view on inverse statistics and gain/loss asymmetry in financial time series. *J Stat Mech Theory Exp* (2008) 2009: P02004. doi:10.1088/1742-5468/2009/02/P02004.
38. Zhou W-X, Yuan W-K. Inverse statistics in stock markets: universality and idiosyncracy. *Phys Stat Mech Appl* (2005) 353:433–44. doi:10.1016/j.physa.2005.02.011.
39. Chen H, Noronha G, Singal V. The price response to S&P 500 index additions and deletions: evidence of asymmetry and a new explanation. *J Finance* (2004) 59:1901–30. doi:10.1111/j.1540-6261.2004.00683.x.
40. Karpio K, Kotur M, Orłowski A. Gain–loss asymmetry for emerging stock markets. *Phys Stat Mech Appl* (2007) 375:599–604. doi:10.1016/j.physa.2006.10.003.
41. Bouchaud J-P, Matacz A, Potters M. Leverage effect in financial markets: the retarded volatility model. *Phys Rev Lett* (2001) 87:228701. doi:10.1103/PhysRevLett.87.228701.
42. Jiang X, Chen T, Zheng B. Time-reversal asymmetry in financial systems. *Phys Stat Mech Appl* (2013) 392:5369–75. doi:10.1016/j.physa.2013.07.006.
43. Yan W, Serooskerken E. Forecasting financial extremes: a network degree measure of super-exponential growth. *PloS One* (2015) 10:e0128908. doi:10.1371/journal.pone.0128908.
44. Lacasa L, Nuñez A, Roldán É, Parrondo JMR, Luque B. Time series irreversibility: a visibility graph approach. *Eur Phys J B* (2012) 85:217. doi:10.1140/epjb/e2012-20809-8.
45. Lacasa L, Ryan F. Time reversibility from visibility graphs of nonstationary processes. *Phys Rev E* (2015) 92:022817. doi:10.1103/PhysRevE.92.022817.
46. Flanagan R, Lacasa L. Irreversibility of financial time series: a graph-theoretical approach. *Phys Lett A* (2016) 380:1689–97. doi:10.1016/j.physleta.2016.03.011.
47. Xiong H, Shang P, Xia J, Wang J. Time irreversibility and intrinsics revealing of series with complex network approach. *Phys Stat Mech Appl* (2018) 499:241–9. doi:10.1016/j.physa.2018.02.041.
48. Wu Z, Shang P, Xiong H. An improvement of the measurement of time series irreversibility with visibility graph approach. *Phys Stat Mech Appl* (2018) 502: 370–8. doi:10.1016/j.physa.2018.02.131.
49. Xie F, Fu Z, Piao L, Mao Z. Time irreversibility of mean temperature anomaly variations over China. *Theor Appl Climatol* (2016) 123:161–70. doi:10.1007/s00704-014-1347-0.
50. Vamvakaris MD, Pantelous AA, Zuev KM. Time series analysis of S&P 500 index: a horizontal visibility graph approach. *Phys Stat Mech Appl* (2018) 497: 41–51. doi:10.1016/j.physa.2018.01.010.
51. Vamvakaris MD, Pantelous AA, Zuev. K. Investors' behavior on S&P 500 index during periods of market crashes: a visibility graph approach. In: F Economou, K Gavriilidis, GN Gregoriou, V Kallinterakis, editors. *Handbook of investors' behavior during financial crises*. San Diego, CA: Academic Press (2017) p. 401–17.
52. Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat* (1951) 22(1):79–86. doi:10.1214/aoms/1177729694.
53. Erven T, Harremoës P. R\'enyi divergence and kullback-leibler divergence (2012) Available from: https://arxiv.org/abs/1206.2459.
54. Klebnikov S. Apple, Microsoft, Amazon, Google and Facebook make up a record chunk of the S&P 500. Here's why that might be dangerous (2020) Available from: https://www.forbes.com/sites/sergeiklebnikov/2020/07/24/apple-microsoft-amazon-google-and-facebook-make-up-a-record-chunk-of-the-sp-500-heres-why-that-might-be-dangerous/#432d745a4f6b
55. Global Market Strategist for US Institutional Equities. How Apple and Microsoft could blow up the stock market (2019) Available from: https://edition.cnn.com/2019/11/01/investing/tech-stocks-faang/index.html.
56. Derrick J. Why the Dow Jones industrial average owes Apple and Microsoft a big thanks (2019) Available from: https://finance.yahoo.com/news/why-dow-jones-industrial-average-183102086.html?soc_src=social-sh&soc_trk=ma.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Liu and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

REVIEW
published: 14 December 2020
doi: 10.3389/fphy.2020.562182

Inequality Measures: The Kolkata Index in Comparison With Other Measures

Suchismita Banerjee[1], *Bikas K. Chakrabarti*[1,2], *Manipushpak Mitra*[1]* *and Suresh Mutuswami*[3]

[1]Economic Research Unit, Indian Statistical Institute, Kolkata, India, [2]Saha Institute of Nuclear Physics, Kolkata, and Economic Research Unit, Indian Statistical Institute, Kolkata, India, [3]School of Business, University of Leicester, Leicester, United Kingdom

We provide a survey of the *Kolkata index* of social inequality, focusing in particular on income inequality. Based on the observation that inequality functions (such as the Lorenz function), giving the measures of income or wealth against that of the population, to be generally nonlinear, we show that the fixed point (like Kolkata index k) of such a nonlinear function (or related, like the complementary Lorenz function) offer better measure of inequality than the average quantities (like Gini index). Indeed the Kolkata index can be viewed as a generalized Hirsch index for a normalized inequality function and gives the fraction k of the total wealth possessed by the rich $1-k$ fraction of the population. We analyze the structures of the inequality indices for both continuous and discrete income distributions. We also compare the Kolkata index to some other measures like the Gini coefficient and the Pietra index. Lastly, we provide some empirical studies which illustrate the differences between the Kolkata index and the Gini coefficient.

Keywords: Lorenz function, complementary Lorenz function, *k*-index and the normalized *k*-index, Gini coefficient, Pietra index

OPEN ACCESS

Edited by:
Damien Challet,
Université Paris-Saclay, France

Reviewed by:
Jordi Duch Gavalda,
University of Rovira i Virgili, Spain
Marcelo Byrro Ribeiro,
Federal University of Rio de Janeiro, Brazil

***Correspondence:**
Manipushpak Mitra
mmitra@isical.ac.in

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *14 May 2020*
Accepted: *27 October 2020*
Published: *14 December 2020*

Citation:
Banerjee S, Chakrabarti BK, Mitra M and Mutuswami S (2020) Inequality Measures: The Kolkata Index in Comparison With Other Measures. Front. Phys. 8:562182. doi: 10.3389/fphy.2020.562182

1. INTRODUCTION

Inequality in a society can broadly be categorized as *inequality of condition* or *inequality of opportunity*. The former refers to disparities in the current status of individuals, whether this be income, wealth or their ownership of different goods and services. The latter refers to disparities in the future potential of individuals. Typically, inequality of opportunity is inferred indirectly through its effects like education level and quality, health status and treatment by the justice system. Though the two types of inequality are interrelated, we are interested in the former type only in this survey. Therefore, in what follows, the term "inequality" will refer exclusively to inequality of condition.

We focus here on one aspect of inequality, viz., the measurement of inequality. Measuring inequality is important for answering a wide range of questions. For instance: is the income distribution more equal than what it was in the past? Are underdeveloped countries characterized by greater inequality than developed countries? Do taxes or other kinds of policy interventions lead to greater equality in the distribution of income or wealth? Since the way inequality is measured also determines how the above questions (among others) are answered, a rigorous discussion of the measurement of inequality is necessary (see, e.g., Refs. 1–5).

A tool that is indispensable in measuring income and wealth inequality is the *Lorenz function* and its graphical representation, the *Lorenz curve* (see Ref. 6). The Lorenz curve plots the percentage of total income earned by various portions of the population when the population is ordered by the size of their incomes. The Lorenz curve is typically depicted as a curve in the unit square with end points at $(0, 0)$ and $(1, 1)$ (see **Figure 1**).[1] The 45° line is the *line of perfect equality* representing a situation where all individuals have the same income.

The Lorenz curve can be used, in a limited way, as a measure of inequality. Since the 45° line is the line of perfect equality, we can say that the "closer" a Lorenz curve is to the 45° line, the more equal is the income distribution. Unfortunately, this does not get us very far because Lorenz curves can intersect and hence, the Lorenz curves cannot be ranked unambiguously using the above criterion (see Ref. 7). We have more to say on this point in **Section 2**.

The existing literature sees two approaches to deal with the problem of intersecting Lorenz curves. The first is to consider ranking criterion that are "weaker" than this dominance criterion meaningful only for non-intersecting Lorenz curves (see Refs. 7–11). The pioneering work in this approach is Ref. 12 which suggested that there is an underlying notion of social welfare associated with any measure of income inequality. It is this concept with which we should be concerned. Furthermore, we should approach the question by considering directly the form of the social welfare function to be employed (see Ref. 13). This is a normative approach and is meaningful when we want to obtain a ranking of income distributions in order to infer something from the social welfare angle like whether "post-tax income is more equally distributed than pre-tax income".

The second approach is to develop summary measures of inequality using the Lorenz functions (see Ref. 7 for details). Here, each Lorenz function is associated with a real number and these numbers are used to compare inequality across different income distributions. This is a descriptive approach where we quantify the difference in inequality between pairs of distributions (see Ref. 13).

An index of income inequality is therefore a scalar measure of interpersonal income differences within a given population. High income inequality means concentration of high incomes in the hands of few and is likely to compress the size of the middle class. A large and rich middle class contributes significantly to the well-being of a society in many ways. In particular, a large and rich middle class contributes in terms of high economic growth, better health status, higher education level, a sizable contribution to the country's tax revenue and a better infrastructure, and more social cohesion resulting from fellow feeling. A society characterized with a small middle class and more persons away from the middle income group may lead to a strained relationship between the subgroups on the two sides of the middle class which can generate unrest (see Ref. 4). Hence, the need for identifying the magnitude of income inequality through different indices is of prime importance.

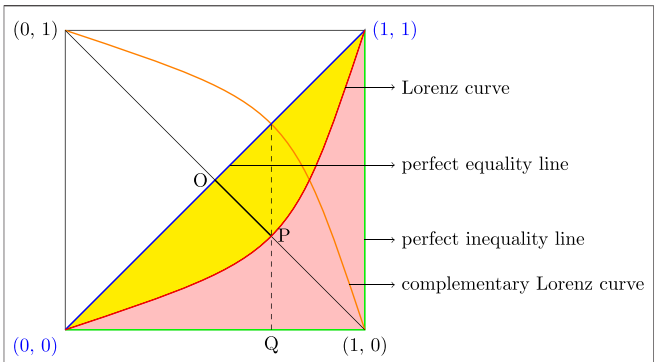


FIGURE 1 | The Lorenz and the complementary Lorenz curves. Q is the k-index of the Lorenz curve. $\overline{OP}$ represents the maximum distance between the perfect equality line and the Lorenz curve.

Except for the unique case of equality, where the Lorenz curve is trivially linear, the Lorenz function is typically nonlinear and it accommodates the essential features of the inequalities involved. However, most of the common inequality indices formulated and used so far studies some of the "average" properties of the Lorenz function. On the other hand, the established observations in statistical physics, for example in developing the Renormalization Group theory of phase transitions (see, e.g., Ref. 14) or the chaos theory (see, e.g., Ref. 15), strongly indicated the richness of the (nontrivial) fixed point structure (and also of the eigen vectors and eigen values for the linearized function near that fixed point) of such non-linear functions to comprehend the physical and mathematical process represented by such nonlinear functions. We noted earlier (see Ref. 1) that, while the Lorenz function has got trivial fixed points, a complementary Lorenz function has a non-trivial point corresponding to an inequality index called the Kolkata index, having several intriguing and useful properties.

Our primary focus in this survey will be on the Kolkata index as a measure of inequality. The *Kolkata index*, first introduced by Ref. 1 and later analyzed in Ref. 2 and in Ref. 3, is that proportion k of the population such that the proportion of income that we can associate with k is $(1 - k)$. Since no single summary statistic can reflect all aspects of inequality exhibited by the Lorenz curve, the importance of using alternative measures of inequality is universally acknowledged (see Ref. 7). We would also discuss two popular indices namely, the *Gini coefficient or index* (see Ref. 16) and the *Pietra index* (see Ref. 17). The Gini index is the ratio of the area between the 45° line and the Lorenz curve to the total area under the 45° line. Equivalently, the Gini index is twice the area between the Lorenz curve and the line of perfect equality. The Pietra index is the maximum value of the gap between the 45° line and the Lorenz curve (also see Ref. 18).

In **Section 2**, we discuss the fundamentals of Lorenz and complementary Lorenz functions, along with some examples extending from continuous to discrete wealth distributions. In

[1]The end points are clear since none of the population possesses none of the income while the entire population possesses all the income.

Section 3, we define the Kolkata index (k-index) and show some example calculation of the k-index for continuous wealth distributions. We also demonstrate an algorithm for calculating the k-index for discrete wealth distribution. We conclude the section by comparing the k-index with various other indices. In **Sections 4 and 5**, we continue this comparison based on rich-poor disparity. In **Section 6**, we measure the k-index from real society data. **Section 7** summarizes and concludes this work.

2. LORENZ FUNCTION AND THE COMPLEMENTARY LORENZ FUNCTION

Let F be the distribution function of a non-negative random variable X which represents the income distribution in a society. The left-inverse of F is defined as $F^{-1}(q) = \inf_x\{x \in X | F(x) \geq q\}$. As long as the mean income $\mu = \int_0^\infty x\mathrm{d}F(x)$ is finite, we obtain an alternative representation of the mean: $\mu = \int_0^1 F^{-1}(q)\mathrm{d}q$. The function associated with the Lorenz curve is the *Lorenz function*, defined as $L_F(p) = (1/\mu)\int_0^p F^{-1}(q)\mathrm{d}q$. The Lorenz function gives the proportion of total income earned by the bottom $100p\%$ of the population for every given $p \in [0,1]$. The advantage of this definition of Lorenz function due to Ref. 19 is that it can be applied to income distributions with both discrete and continuous random variables. The Lorenz function thus defined has the following properties: i) $L_F(p)$ is continuous, non-decreasing and convex in $p \in (0,1)$ and, ii) $L_F(0) = 0$, $L_F(1) = 1$ and $L_F(p) \leq p$ for all $p \in (0,1)$. Moreover, if there exists $p \in (0,1)$ such that $L_F(p) = p$, then for all $p \in [0,1]$, $L_F(p) = p$. If the Lorenz function $L_F(p)$ is differentiable in the open interval $(0,1)$, then the slope of the Lorenz function at any $p \in (0,1)$ is given by $F^{-1}(p)/\mu$. Let M_F be the median as a percentage of the mean. Then M_F is given by the slope of the Lorenz curve at $p = 1/2$, that is, $M_F = F^{-1}(1/2)/\mu$. Since many real life distributions of incomes are skewed to the right, the mean often exceeds the median so that $M_F < 1$. The *complementary Lorenz function* is defined as $\widehat{L}_F(p) = 1 - L_F(p)$. It measures the proportion of the total income earned by the top $100(1-p)\%$ of the population. Therefore,

$$\widehat{L}_F(p) := 1 - L_F(p) = 1 - \frac{\int_0^p F^{-1}(q)\mathrm{d}q}{\mu} = \frac{\int_p^1 F^{-1}(q)\mathrm{d}q}{\mu}. \quad (1)$$

It easily follows that $\widehat{L}_F(0) = 1, \widehat{L}_F(1) = 0$, and $0 \leq \widehat{L}_F(p) \leq 1$ for $p \in (0,1)$. Furthermore, $\widehat{L}_F(p)$ is continuous, non-increasing and concave for $p \in (0,1)$.

Consider any egalitarian income distribution F_e where all agents earn a common positive income so that the associated Lorenz function is $L_{F_e}(p) = p$ for all $p \in (0,1)$. Thus, we have a case of perfect equality where every $p\%$ of the population enjoys $p\%$ of the total income and the Lorenz curve coincides with the diagonal line of perfect equality. In reality, we do not find any society where all individuals have equal income. For all other income distributions the Lorenz curve will lie below the egalitarian line, that is below the Lorenz curve associated with the Lorenz function $L_{F_e}(.)$ for the egalitarian income distribution F_e. Similarly, we also do not find a society where one person has all the income, that is, an income distribution F_I such that $L_{F_I}(p) = 0$ for all $p \in (0,1)$. Specifically, with complete inequality associated with the income distribution F_I, which is characterized by the situation where only one agent has positive income and all other persons have zero income, the Lorenz curve will run through the horizontal axis until we reach the richest person and then it rises perpendicularly (see **Figure 1**). Hence, for any realistic income distribution of a society, Lorenz curve always lie in between the perfect equality line and the perfect inequality line. The Lorenz curve is quite useful because it shows graphically how the actual distribution of incomes differs not only from the perfect equality line associated with the egalitarian income distribution F_e but also from the perfect inequality line associated with the income distribution F_I. The Lorenz curve, complimentary Lorenz curve, perfect equality and perfect inequality lines are shown in **Figure 1** below, where we plot the fraction of population from poorest to richest on the horizontal axis and the fraction of associated income on the vertical axis.

We provide some simple examples of Lorenz functions for which the associated income distribution is a continuous random variable.

- *Uniform distribution:* Consider a society where the income distribution is uniform on some compact interval $[a,b]$ with $0 \leq a < b < \infty$ so that the probability density function is $f_u(x) = 1/(b-a)$ and the distribution function is $F_u(x) = (x-a)/(b-a)$ for every $x \in [a,b]$. Since $\mu_u = (a+b)/2$ and $F_u^{-1}(q) = a + (b-a)q$, we get

$$L_{F_u}(p) = \frac{1}{\mu_u}\int_0^p \{a + (b-a)q\}dq = p\left[1 - \frac{(b-a)}{(a+b)}(1-p)\right],$$

Observe that if $a = 0$, then we have $L_{F_u}(p) = p^2$.

- *Exponential distribution:* Suppose the income distribution is exponential so that the probability density function is given by $f_E(x) = \lambda e^{-\lambda x}$ with $\lambda > 0$ and the distribution function is $F_E(x) = 1 - e^{-\lambda x}$ for any $x \geq 0$. In this case $\mu_E = 1/\lambda$ and $F_E^{-1}(q) = -(1/\lambda)\ln(1-q)$ implying

$$L_{F_E}(p) = \int_0^p -\ln(1-q)\mathrm{d}q = -\int_{t=1-p}^{t=1} \ln(t)\mathrm{d}t$$
$$= p - (1-p)\ln\left(\frac{1}{1-p}\right).$$

- *Pareto distribution:* Consider a society where the income distribution is Pareto so that the density function is $f_{P,\alpha}(x) = \alpha(m)^\alpha/(x)^{\alpha+1}$ and the distribution function is $F_{P,\alpha}(x) = 1 - (m/\mathrm{x})^\alpha$ where $m > 0$ is the minimum income, $\alpha > 1$ and the density and distribution functions are defined for all $x \geq m$.

In this case $\mu_P = \alpha m/(\alpha-1)$ and $F^{-1}_{P,\alpha}(q) = m(1-q)^{-(1/\alpha)}$ implying

$$L_{F_{P,\alpha}}(p) = \frac{(\alpha-1)}{\alpha}\int_0^p (1-q)^{-\frac{1}{\alpha}}dq = \left[t^{\frac{(\alpha-1)}{\alpha}}\right]_{t=1-p}^{t=1} = 1-(1-p)^{1-(1/\alpha).} \quad (2)$$

Hence, if the income distribution is a continuous random variable F, one can calculate the Lorenz function $L_F(p)$ and, using $\widehat{L}_F(p) = 1 - L_F(p)$, we can easily calculate the associated complementary Lorenz function as well.

Example 1. Discrete random variable. To understand the procedure for getting the Lorenz function for income distribution given by discrete random variables, consider an economy with G groups of people where each group $g \in \{1,\ldots,G\}$ has a total of $n_g \geq 1$ people with each person within this group having the same income of x_g and also assume that $0 \leq x_1 < \cdots < x_G$. Define the total population as $N := \sum_{g\in G} n_g$ and the total income of the economy as $M := \sum_{g\in G} n_g x_g$ so that the mean income for this society is $\mu_G = M/N$. This income distribution is a discrete random variable $X = \{x_1,\ldots,x_G\}$ such that the probability mass function is given by $f_G(x_g) = n_g/N$ for all $g \in \{1,\ldots,G\}$ and the distribution function is given by

$$F_G(x) = \begin{cases} 0, & \text{if } x \in [0, x_1), \\ \frac{\{\sum_{t=1}^{g} n_t\}}{N}, & \text{if } x \in \left[x_g, x_{g+1}\right) \text{ for any given } g \in \{1,\ldots,G\}, \\ 1, & \text{if } x \geq x_G, \end{cases}$$

For each $g \in \{1,\ldots,G\}$, define $N(g) := \sum_{t=1}^{g} n_t/N$, $N(0) := 0$, $M(g) := \sum_{t=1}^{g} n_t x_t/M$ and $M(0) := 0$. For any given $g \in \{1,\ldots,G\}$ and any $q_g \in (N(g-1), N(g)]$, one can easily verify that $F_G^{-1}(q_g) = x_g$. Hence, using the Lorenz function formula we have the following: For any given $g \in \{1,\ldots,G\}$ and any $p_g \in (N(g-1), N(g)]$,

$$L_{F_G}(p_g) = M(g-1) + \left(p_g - N(g-1)\right)\left(\frac{Nx_g}{M}\right). \quad (3)$$

The following observations are helpful in this context.

(1) The Lorenz function $L_{F_G}(p)$ is piecewise linear and, for each $g \in \{1,\ldots,G-1\}$, the point $(N(g), L_G(N(g)) = M(g))$ on the coordinate plane of the graph of the Lorenz curve is a kink point.
(2) If $G = 1$ so that $M = Nx_1$, $N(1) = 1$, then from **Eq. 3** we get $L_{F_1}(p) = M(0) + (p - N(0))(Nx_1/Nx_1) = p$ for all $p \in (N(0), N(1)]$, that is, Lorenz curve is associated with the egalitarian distribution and we have $L_{F_1}(p) = L_{F_e}(p) = p$ for all $p \in (0,1)$.

2.1. The Lorenz Function as a Measure of Inequality

The Lorenz curve allows us to rank distributions according to inequality and say that the country with Lorenz curve closer to the perfect equality line has less inequality than the country with Lorenz curve further away. Consider two societies with income distributions given by the distribution functions F_a and F_b. If it so happens that $L_{F_a}(p) \leq L_{F_b}(p)$ for all $p \in [0,1]$, then clearly, the society with income distribution F_a is more unequal compared to the society having the income distribution F_b since for every $p \in (0,1)$ the bottom 100 p% population has a weakly lower percentage share of income under F_a than under F_b. Formally, for any two income distributions F_a and F_b, we say that F_b Lorenz dominates F_a if the Lorenz curve $L_{F_b}(p)$ associated with the income distribution F_b lies nowhere below that of Lorenz curve $L_{F_a}(p)$ associated with the income distribution F_a and at some places (at least) lies above. Thus, we can think of domination relation across pairs of Lorenz curves to infer about inequality and, in particular, in a pairwise Lorenz curve comparison, higher of the Lorenz curves are preferable. However, if the Lorenz curves of the two distributions cross, then such an unambiguous conclusion about inequality ordering cannot be drawn. The next example provides such an instance of intersecting Lorenz curves.

Example 2. Consider a society with four people and consider the following income distribution. Person 1 and Person 2 has an income of 20, Person 3 has an income of 30 and Person 4 has an income of 50. We first try to think of a meaningful representation of such an income distribution. Observe that if we draw a person at random, then with 1/2 probability we will draw a person having an income of 20, with 1/4 probability we will draw a person having an income of 30 and with 1/4 probability we will draw a person having an income of 50. Therefore, we have a probability mass function of a random variable of three possible incomes $X_A = \{20, 30, 50\}$ and the probability mass function is given by $f_A(20) = 1/2$, $f_A(30) = 1/4$ and $f_A(50) = 1/4$. Using **Eq. 3**, the Lorenz function is given by

$$L_{F_A}(p) = \begin{cases} \frac{2p}{3}, & \text{if } p \in \left(0, \frac{1}{2}\right], \\ \frac{6p-1}{6}, & \text{if } p \in \left(\frac{1}{2}, \frac{3}{4}\right], \\ \frac{5p-2}{3}, & \text{if } p \in \left(\frac{3}{4}, 1\right]. \end{cases}$$

Similarly, consider a society with four people and consider the following income distribution. Person 1 and Person 2 has an income of 15, Person 3 has an income of 42 and Person 4 has an income of 48. We have a probability mass function of a random variable $X_B = \{15, 42, 50\}$ and the probability mass function is given by $f_B(15) = 1/2$, $f_B(42) = 1/4$ and $f_B(48) = 1/4$. Again, using **Eq. 3**, the Lorenz function is given by

$$L_{F_B}(p) = \begin{cases} \frac{p}{2}, & \text{if } p \in \left(0, \frac{1}{2}\right], \\ \frac{28p-9}{20}, & \text{if } p \in \left(\frac{1}{2}, \frac{3}{4}\right], \\ \frac{8p-3}{5}, & \text{if } p \in \left(\frac{3}{4}, 1\right]. \end{cases}$$

Now consider the income distribution F_A and compare it with the income distribution F_B. Note that at $p = 1/2$, $L_{F_A}(1/2) =$

$1/3 > L_{F_B}(1/2) = 1/4$ and at $p = 3/4$, $L_{F_A}(3/4) = 7/12 < L_{F_B}(3/4) = 3/5$. Hence, given both $L_{F_A}(p)$ and $L_{F_B}(p)$ are continuous in $p \in [0,1]$, the two Lorenz curves overlap and, in particular, these two Lorenz curve intersects at $p^\star = 17/24$, that is, at $p^\star$ we have $L_{F_A}(p^\star) = L_{F_B}(p^\star)$.

3. INEQUALITY INDICES IN DETAIL

3.1. The Kolkata Index

The k-index for any income distribution F is defined by the solution to the equation $k_F + L_F(k_F) = 1$. It has been proposed as a measure of income inequality (see Refs. 2 and 3, and Ref. 1, for more details). We can rewrite $k_F + L_F(k_F) = 1$ as $\widehat{L}_F(k_F) = k_F$ implying that the k-index is a fixed point of the complementary Lorenz function. Since the complementary Lorenz function maps $[0,1]$ to $[0,1]$ and is continuous, it has a fixed point. Furthermore, since complementary Lorenz function $\widehat{L}_F(p)$ is non-increasing, the fixed point is unique. Since for any F, $p_F^* := L_F^{-1}(1/2) \geq 1/2$ with the equality holding only if we have an egalitarian income distribution, the unique fixed point of $\widehat{L}_F$ lies in the interval $[1/2, p_F^*]$ implying that for any distribution F, $k_F \in [1/2, 1)$. Therefore, k_F lies between 50% population proportion and the population proportion $p_F^* = L_F^{-1}(1/2)$ that we associate with 50% income given the income distribution F. Observe that if $L_F(p) = p$, then $k_F = 1/2$ and for any other income distribution, $1/2 < k_F < 1$. Also note that while the Lorenz curve typically has only two trivial fixed points, that is, $L_F(0) = 0$ and $L_F(1) = 1$, the complementary Lorenz function $\widehat{L}_F(p)$ has a unique non-trivial fixed point k_F.

The Pareto principle is based on Pareto's observation (in the year 1906) that approximately 80% of the land in Italy was owned by 20% of the population. The evidence, though, suggests that the income distribution of many countries fails to satisfy the 80/20 rule (see Ref. 1). The k-index can be thought of as a generalization of the Pareto principle. Note that $L_F(k_F) = 1 - k_F$; hence, the top $100(1 - k_F)\%$ of the population has $100(1 - (1 - k_F)) = 100k_F\%$ of the income. Hence, the "Pareto ratio" for the k-index is $k_F/(1 - k_F)$. Observe, however, that this ratio is obtained endogenously from the income distribution and in general, there is no reason to expect that this ratio will coincide with the Pareto principle. The fact that the k-index generalizes Pareto's 80/20 rule was first pointed out in Ref. 1 and later also in Refs. 20 and 21.

- *Uniform distribution.* If we have the uniform distribution F_u defined on $[a,b]$ where $0 \leq a < b < \infty$. Then

$$k_{F_u} = \frac{-(3a+b) + \sqrt{5a^2 + 6ab + 5b^2}}{2(b-a)},$$
$$\mathcal{K}_{F_u} = \frac{-2(a+b) + \sqrt{5a^2 + 6ab + 5b^2}}{(b-a)}.$$

- *Exponential distribution.* For the exponential distribution F_E, the complementary Lorenz function is given by $\widehat{L}_{F_E}(p) = (1-p)[1 + \ln\{1/(1-p)\}]$. One can show that $k_{F_E} \simeq 0.6822$ and hence $\mathcal{K}_{F_E} \simeq 0.3644$.
- *Pareto distribution.* For the Pareto distribution $F_{P,\alpha}$, the complementary Lorenz function is given $\widehat{L}_{F_{P,\alpha}}(p) = (1-p)^{1-(1/\alpha)}$. The k-index is therefore a solution to (I) $(1 - k_{F_P})^{1-(1/\alpha)} = k_{F_P}$. It is difficult to provide a general solution to (I). However, we an interesting observation in this context.
- If $\widehat{\alpha} = \ln 5/\ln 4 \simeq 1.16$, then $k_{F_{P,\widehat{\alpha}}} = 0.8$ and we get the Pareto principle or the 80/20 rule. Also note that $\mathcal{K}_{F_{P,\widehat{\alpha}}} = 0.6$.

3.1.1. Discrete Random Variable

Consider any discrete random variable with distribution function F_G discussed in Example 1 for which the Lorenz function is given by **Eq. 3**. To obtain the explicit form of the k-index one can first apply a simple algorithm to identify the interval of the form $[N(g-1), N(g))$ defined for $g \in \{1, \ldots, G\}$ in which the k-index can lie.

Algorithm-A:

Step 1: Consider the smallest $g_1 \in \{1, \ldots, G\}$ such that $N(g_1) \geq 1/2$ and consider the sum $N(g_1) + M(g_1)$. If $N(g_1) + M(g_1) \geq 1$, then stop and $k_{F_G} \in (N_{g_1-1}, N(g_1)]$ and, in particular, $k_F = N(g_1)$ if and only if $N(g_1) + M(g_1) = 1$. Instead, if $N(g_1) + M(g_1) < 1$, then go to Step 2 and consider the group $g_1 + 1$ and repeat the process.

⋮

Step t. We have reached Step t means that in Step $(t-1)$ we had $N(g_1 + t - 1) + M(g_1 + t - 1) < 1$. Therefore, consider the sum $N(g_1 + t) + M(g_1 + t)$. If $N(g_1 + t) + M(g_1 + t) \geq 1$, the stop and $k_{F_G} \in [N(g_1 + t - 1), N(g_1 + t))$ and, in particular, $k_F = N(g_1 + t)$ if and only if $N(g_1 + t) + M(g_1 + t) = 1$. If $N(g_1 + t) + M(g_1 + t) < 1$, then go to Step $(t+1)$.

Observe that since $N(G) = M(G) = 1$, if we have $N(G-1) + M(G-1) < 1$ in some step, then, in the next step, this algorithm has to end since $N(G) + M(G) = 2 > 1$.

Suppose for any discrete random variable with distribution function F_G discussed in Example 1, Algorithm-A identifies $g^* \in \{1, \ldots, G\}$ such that $N(g^*) + M(g^*) \geq 1$. If $N(g^*) + M(g^*) = 1$, then $k_{F_G} = N(g^*)$ and if $N(g^*) + M(g^*) > 1$, the k_{F_G} is the solution to the following equation:

$$k_{F_G} + \left\{ M(g^* - 1) + (k_{F_G} - N(g^* - 1)) \left(\frac{N x_{g^*}}{M} \right) \right\} = 1.$$

Thus, to derive the k-index of any discrete random variable with distribution function F_G discussed in Example 1, we first identifying the group $g^* \in \{1, \ldots, G\}$ such that $k_{F_G} \in (N(g^* - 1), N(g^*)]$ (using Algorithm-A) and then, using g^*, we get the following value of k_{F_G}:

$$k_{F_G} = \begin{cases} N(g^*), & \text{if } N(g^*) + M(g^*) = 1, \\ \dfrac{\mu_G + N(g^*) x_{g^*} - M(g^*)}{\mu_G + x_{g^*}}, & \text{if } N(g^*) + M(g^*) > 1. \end{cases}$$

Remark 1. Consider the income distributions F_A and F_B defined in Example 2. Recall that the Lorenz functions and

$L_{F_B}(p)$ are such that $L_{F_A}(p) > L_{F_B}(p)$ for all $p \in (0, 17/24)$ and $L_{F_A}(p) < L_{F_B}(p)$ for all $p \in (17/24, 1)$. However, one can work out that the k-indices for these distributions. Specifically, note that for F_A, $N(1) = 1/2$ and $M(1) = 1/3$ implying that $N(1) + M(1) = 1/6 < 1$ and $N(2) = 3/4$ and $M(1) = 7/12$ implying that $N(2) + M(2) = 4/3 >$. Hence, by Algorithm-A, $k_{F_A} \in (1/3, 3/4)$ and it is a solution to the equation $k_{F_A} + (6k_{F_A} - 1)/6 = 1$ implying that $k_{F_A} = 7/12 \simeq 0.58\dot{3}$ and hence the normalized value is $\mathcal{K}_{F_A} = 1/6 \simeq 0.1\dot{6}$. Similarly, for F_B, $N(1) = 1/2$ and $M(1) = 1/4$ implying that $N(1) + M(1) = 3/4 < 1$ and $N(2) = 3/4$ and $M(1) = 3/4$ implying that $N(2) + M(2) = 3/2 >$. Hence, by Algorithm-A, $k_{F_B} \in (1/2, 3/4)$ and it is a solution to the equation $k_{F_B} + (28k_{F_B} - 9)/20 = 1$ implying that $k_{F_B} = 29/48 \simeq 0.6041\dot{6}$ and hence the normalized value is $\mathcal{K}_{F_B} = 5/24 \simeq 0.208\dot{3}$. Observe that $k_{F_A} < k_{F_B}$ and hence $\mathcal{K}_{F_A} < \mathcal{K}_{F_B}$ implying that according to k-index as a measure of income inequality, the income distribution F_A is less unequal than income distribution F_B.

3.1.2. The Hirsch Index

The physicist Jorge E. Hirsch suggested this index to measure the citation impact of the publications of a research scientist (see Ref. 22). Let $X = (x_1, \ldots, x_m)$ be the set of research papers of a scientist. Let $f : X \to \mathcal{N}$ be the *citation function* of the scientist. The citation function simply gives the number of citations for each publication. Let $X_{()} = (x_{(1)}, \ldots, x_{(m)})$ be a reordering of the elements in the set X such that $f(x_{(1)}) \geq \ldots \geq f(x_{(m)})$. The Hirsch index, or the h-index, is the largest number $H^* \in \{0, 1, \ldots, m\}$ such that $f(x_{(H^*)}) \geq H^*$. Note that if $f(x_{(1)}) = 0$, then $H^* = 0$, and, if $f(x_{(m)}) \geq m$, then $H^* = m$ and for all other cases $H^* \in \{1, \ldots, m-1\}$.

If neither $f(x_{(1)}) = 0$ nor $f(x_{(m)}) \geq m$ holds, then how do we identify the h-index? To see this, suppose that we plot a graph where on the x-axis we plot the ordered set of publications of a research scientist in non-increasing order of citations and on the y-axis we plot the number of citations for each publication. Moreover, if we join the consecutive plotted points like $f(x_{(t)})$ and $f(x_{(t+1)})$ by a straight line for each $t \in \{1, \ldots, m-1\}$, then we get a curve representing a function $\tilde{f} : [1, m] \to [f(x_1), f(x_m)]$, defined on the domain $[1, m]$ with co-domain $[f(x_1), f(x_m)]$, which we call the *generated citation curve*. The generated citation curve is continuous, piecewise linear and has a non-positive slope whenever the slope exists. The generated citation curve resembles a lot like the complementary Lorenz curve that we can associate with any income distribution. Consider the fixed point of the generated citation curve $\tilde{f}$ on the interval $[1, m]$, that is, consider $\tilde{h} \in [1, m]$ such that $\tilde{f}(\tilde{h}) = \tilde{h}$. As long as there is at least one citation and as long as all papers are not cited more than $(m-1)$-times, such a fixed point $\tilde{h}$ exists and is unique with the added property that $\tilde{h} \in [1, m-1]$. Given this fixed point, we can identify the relevant value of the h-index, that is, $H^* \in \{1, \ldots, m\}$ for f by the following procedure: If the fixed point $\tilde{h}$ is an integer, then it is the H^* that we are looking for, that is, $\tilde{h} = H^*$. If, however, $\tilde{h}$ is not an integer, then there exists an integer $\widehat{h}$ such that $\tilde{f}(x_{\widehat{(h)}}) = f(x_{\widehat{(h)}}) > \widehat{h}$ and $\tilde{f}(x_{\widehat{(h+1)}}) = f(x_{\widehat{(h+1)}}) < \widehat{h} + 1$ and then, the relevant value of the h-index is $\widehat{h} = H^*$. Therefore, graphically, the procedure of obtaining the h-index of any research scientist using the generated citation curve is the same as identifying the fixed point of the complementary Lorenz function of any income distribution that yields the k index.

3.2. The Gini Index

The Gini index is the ratio of the area that lies between the line of perfect equality and the Lorenz curve over the total area under the line of perfect equality. If we plot cumulative share of population from lowest income to highest income on the horizontal axis and cumulative share of income on the Vertical axis (as shown in **Figure 1** above), then the Gini index $\mathcal{G}_F(p)$ of any income distribution F is given by $\mathcal{G}_F :=$ areaofAOCPA/areaofAOCBA. If all people have non-negative income (or wealth, as the case may be), the Gini index can theoretically range from 0 (complete equality) to 1 (complete inequality); it is sometimes expressed as a percentage ranging between 0 and 100. In practice, both extreme values are not quite reached. The Gini index is given by the following formula:

$$\mathcal{G}_F = \frac{\int_0^1 (q - L_F(q))\mathrm{d}q}{(1/2)} = 2\int_0^1 (q - L_F(q))\mathrm{d}q = 1 - 2\int_0^1 L_F(q)\mathrm{d}q. \tag{4}$$

It is obvious that if $L_{F_e}(p) = p$ for all $p \in (0, 1)$, then $\mathcal{G}_F = 0$. If the income distribution for a society with n people follows a Power Law distribution, then $L_{F_n}(p) = p^n$. The Gini index is then given by $\mathcal{G}_{F_n} = \{1 - 2/(n+1)\}$. Hence, as $n \to \infty$, we have $\mathcal{G}_{F_\infty} = 1$. Gini index of some standard continuous random variable are provided below.

- *Uniform distribution:* Consider uniform distribution on some compact interval $[a, b]$ with $0 \leq a < b < \infty$. The Gini index is given by

$$\mathcal{G}_{F_u} = 2\int_0^1 \left[q - q\left\{1 - \frac{(b-a)}{(a+b)}(1-q)\right\}\right]\mathrm{d}q = \frac{(b-a)}{3(a+b)} > \mathcal{K}_{F_u}.$$

- *Exponential distribution:* Consider the exponential distribution with distribution function given by $F_E(x) = 1 - e^{-\lambda x}$ for any $x \geq 0$ with $\lambda > 0$. The Gini index is given by

$$\mathcal{G}_{F_E} = 2\int_0^1 [q - L_{F_E}(q)]\mathrm{d}q = 2\int_0^1 (1-q)\ln\left(\frac{1}{1-q}\right)\mathrm{d}q = \frac{1}{2} > \mathcal{K}_{F_E}.$$

- *Pareto distribution:* For Pareto distribution given by the distribution function is $F_{P,\alpha}(x) = 1 - (m/x)^\alpha$ with $m > 0$ as the minimum income and $\alpha > 1$, the Gini index is given by

$$\mathcal{G}_{F_{P,\alpha}} = 2\int_0^1 \left[q - \left\{1 - (1-q)^{1-\frac{1}{\alpha}}\right\}\right]\mathrm{d}q = \frac{1}{2\alpha - 1}.$$

If we plot the Gini index for different values of $\alpha > 1$, then note that as α increases the Gini index decreases, and, as $\alpha \to 1$ we have $\mathcal{G}_{F_{P,\alpha}} \to 1$. Also note that if $\widehat{\alpha} = \ln 5/\ln 4$, then $\mathcal{G}_{F_{P,\widehat{\alpha}}} \simeq 0.7565 > \mathcal{K}_{F_{P,\widehat{\alpha}}} = 0.6$.

3.2.1. Discrete Random Variable

Consider the discrete random variable F_G discussed in Example 1 for which the Lorenz function is given by **Eq. 3**. As show in **Appendix A**, we have the following explicit form of the Gini index.

$$\mathcal{G}_{F_G} = \frac{\sum_{g=1}^{G}\sum_{t=1}^{G} n_t n_g |x_t - x_g|}{2NM}, \tag{5}$$

Note that if $n_g = 1$ for all $g \in \{1, \ldots, G\}$ so that $G = N$ and $M = \sum_{g=1}^{N} x_g$, then from **Eq. 5** it follows that

$$\mathcal{G}_{F_N} = \frac{\sum_{g=1}^{N}\sum_{t=1}^{N} |x_t - x_g|}{2N\sum_{g=1}^{N} x_g}. \tag{6}$$

Remark 2. Consider the income distributions F_A and F_B defined in Example 2. One can work out that the Gini indices are $\mathcal{G}_{F_A} = \mathcal{K}_{F_B} = 5/24 \simeq 0.208\dot{3} > \mathcal{K}_{F_A}$ and $\mathcal{G}_{F_B} = 21/80 = 0.2625 > \mathcal{K}_{F_B}$. Hence, like the normalized k-index, according Gini index the income distribution F_A is less unequal than income distribution F_B.

3.3. The Pietra Index

An interesting index of inequality is the Pietra index (see Pietra [17]) that tries to identify that proportion of total income that needs to be reallocated across the population in order to achieve perfect equality. Given any income distribution F, this proportion is given by the maximum value of $p - L_F(p)$. Therefore, the Pietra index is $\mathcal{P}_F = \max_{p \in [0,1]} (p - L_F(p))$. It is immediate that if $L_F(p) = p$ for all $p \in [0, 1]$, then $\mathcal{K}_F = \mathcal{P}_F = \mathcal{G}_F = 0$. For any other income distribution F, the maximum distance between the perfect equality line and the Lorenz curve is the distance OP in **Figure 1** above. Note that for any random variable X with distribution function F, $p - L_F(p) = p - (\int_0^p F^{-1}(q)\mathrm{d}q)/\mu = \int_0^p \{\mu - F^{-1}(q)\}\mathrm{d}q\}/\mu$. Therefore, maximizing $(p - L_F(p))$ by selecting $p \in [0, 1]$ is equivalent to maximizing the area $\int_0^p \{\mu - F^{-1}(q)\}\mathrm{d}q$ by selecting $p \in [0, 1]$. Since the Lorenz curve plots the percentage of total income earned by various portions of the population when *the population is ordered by the size of their incomes*, it is obvious that $\{\mu - F^{-1}(q)\} > 0$ for all $q \in [0, F(\mu))$, $\{\mu - F^{-1}(q)\} < 0$ for all $q \in (F(\mu), 1]$ and $\{\mu - F^{-1}(q)\} = 0$ at $q = F(\mu)$. Thus, it follows that the maximum value of the integral $\int_0^p \{\mu - F^{-1}(q)\}\mathrm{d}q$ is attained at $p = F(\mu)$. Hence, the Pietra index for any random variable with distribution function F is

$$\mathcal{P}_F = \max_{p \in [0,1]} (p - L_F(p)) = F(\mu) - L_F(F(\mu)). \tag{7}$$

- *Uniform distribution:* For the uniform distribution on some compact interval $[a, b]$ with $0 \le a < b < \infty$, we have $p - L_{F_u}(p) = (b-a)p(1-p)/(a+b)$ for all $p \in [0, p]$. Moreover, $\mu_u = (a+b)/2$ and as a result $F_u(\mu_u) = 1/2$. Hence, the Pietra index is given by

$$\mathcal{P}_{F_u} = \frac{(b-a)}{(a+b)} F_u(\mu_u)(1 - F_u(\mu_u)) = \frac{(b-a)}{4(a+b)},$$

Given $\mathcal{G}_{F_u} = (b-a)/3(a+b)$, we have $\mathcal{P}_{F_u} = (3/4)\mathcal{G}_{F_u} < \mathcal{G}_{F_u}$. Moreover, one can easily check that $\mathcal{P}_{F_u} > \mathcal{K}_{F_u}$.

- *Exponential distribution:* For the exponential distribution $F_E(x) = 1 - e^{-\lambda x}$ defined for any $x \ge 0$ with $\lambda > 0$, we have $p - L_E(p) = (1-p)\ln(1/(1-p))$ for all $p \in [0, 1]$. We also have $\mu_E = 1/\lambda$ and hence $F_E(\mu_E) = 1 - e^{-1}$. The Pietra index is given by

$$\mathcal{P}_{F_E} = (1 - F_E(\mu_E))\ln\left(\frac{1}{1 - F_E(\mu_E)}\right) = \frac{1}{e},$$

Observe that $\mathcal{K}_{F_E} \simeq 0.3644 < \mathcal{P}_{F_E} = 1/e \simeq 0.3679 < \mathcal{G}_{F_E} = 1/2$.

- *Pareto distribution:* For Pareto distribution given by the distribution function is $F_{P,\alpha}(x) = 1 - (m/x)^{\alpha}$ with $m > 0$ as the minimum income and $\alpha > 1$, we have $p - L_P(p) = (1-p)^{1-(1/\alpha)} - (1-p)$ for all $p \in [0, p]$, $\mu_P = \alpha m/(\alpha - 1)$ and $F_{P,\alpha}(\mu_P) = 1 - \{(\alpha-1)/\alpha\}^{\alpha}$. The Pietra index is given by

$$\mathcal{P}_{F_{P,\alpha}} = (1 - F_P(\mu_P))^{1-(1/\alpha)-(1-F_P(\mu_P))} = \frac{(\alpha-1)^{\alpha-1}}{\alpha^{\alpha}},$$

One can verify that $\mathcal{P}_{F_{P,\alpha}} < \mathcal{G}_{F_P} = 1/(2\alpha - 1)$ for all $\alpha > 1$. Also note that if $\widehat{\alpha} = \ln 5/\ln 4$, then $\mathcal{G}_{F_{P,\widehat{\alpha}}} \simeq 0.7565 > \mathcal{P}_{F_{P,\widehat{\alpha}}} \simeq 0.626655 > \mathcal{K}_{F_{P,\widehat{\alpha}}} = 0.6$.

As shown in **Appendix B(i)**, there is an alternative representation of the Pietra index as the ratio of the mean absolute deviation of the income distribution and twice its mean, that is, $\mathcal{P}_F = E(|x - \mu|)/2\mu$.

3.3.1 Discrete Random Variable

Consider the discrete random variable F_G discussed in Example 1 for which the Lorenz function is given by **Eq. 3**. It is shown in Appendix B(ii) that the Pietra index has the following representations:

$$\mathcal{P}_{F_G} = \frac{\sum_{g=1}^{\tilde{g}} n_g(\mu_G - x_g)}{M} = \frac{E(|x - \mu_G|)}{2\mu_G}, \tag{8}$$

where $\tilde{g} \in \{1, \ldots, G-1\}$ is such that $\mu_G \in [x_{\tilde{g}}, x_{\tilde{g}+1})$ implying that $F_G(\mu_G) = N(\tilde{g})$.

Remark 3. Consider the income distributions F_A and F_B defined in Example 2. Observe that for both F_A and F_B the mean is the same and, in particular $\mu_A = \mu_B = 30$. Therefore, $F_A(\mu_A) = 3/4$ and $L_{F_A}(\mu_A) = 7/12$ implying $\mathcal{P}_{F_A} = \mathcal{K}_{F_A} = 1/6 \simeq 0.16\dot{} < \mathcal{G}_{F_A}$, and, we also have $F_B(\mu_B) = 1/2$ and $L_{F_B}(\mu_A) = 1/4$ implying $\mathcal{P}_{F_B} = 1/4 = 0.25 \in (\mathcal{K}_{F_B}, \mathcal{G}_{F_B})$.

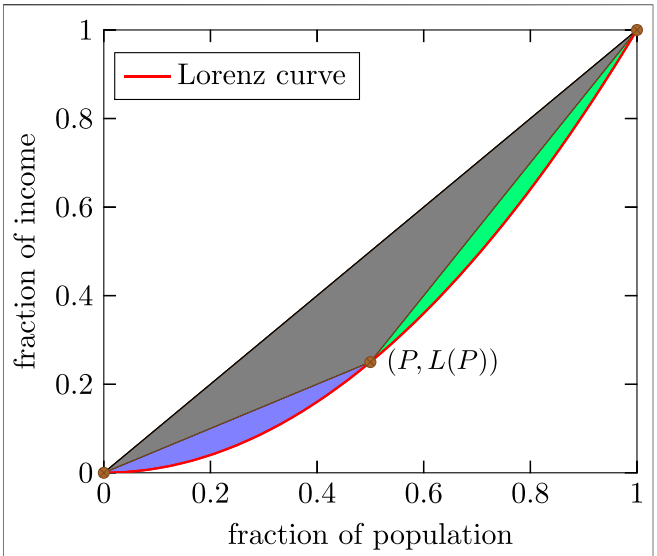


FIGURE 2 | Rich-poor disparity assuming that the poor are p% of the population. The blue-shaded area is the disparity among the poor, the green-shaded area is the disparity among the rich, and the grey-shaded area is the disparity between the rich and the poor.

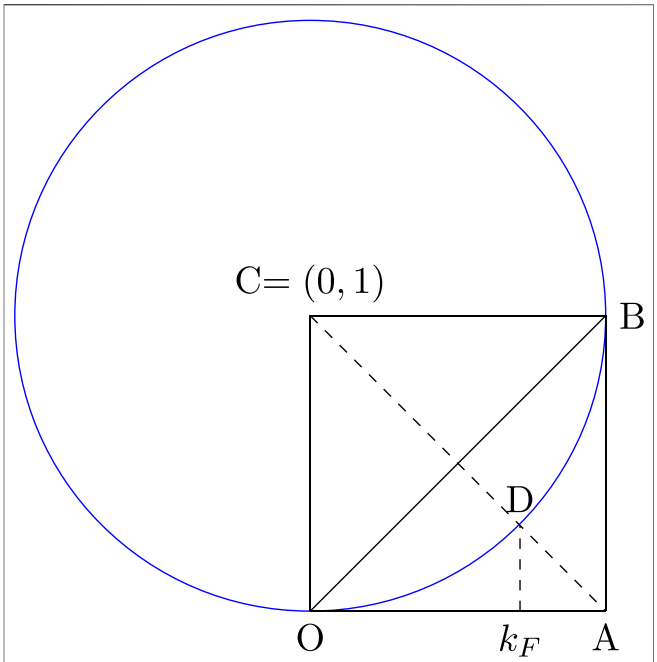


FIGURE 3 | The Lorenz curve as an arc of a unit circle. Here, the normalized k-index and Pietra index are equal but different from the Gini index: $\mathcal{K}_{F_R} = \mathcal{P}_{F_R} = \sqrt{2} - 1 < G = \pi/2 - 1$.

Thus, $\mathcal{P}_{F_A} < \mathcal{P}_{F_B}$ and hence, like the ordering with the k-index as well as the Gini index, according to the Pietra index, the income distribution F_A is less unequal than income distribution.

4. COMPARING THE MEASURES

4.1. Rich-Poor Disparity

The Gini index, as is well-known, measures inequality by the area between the Lorenz curve and the line of perfect equality. For any $p \in [0, 1]$, one can decompose the Gini index into three parts: two representing the *within-group inequality* and one representing the *across-group inequality*. In **Figure 2** below, the unshaded area bounded by the Lorenz curve and the line from $(0, 0)$ to $(p, L_F(p))$ is the within-group inequality of the poor. It represents the extent to which inequality can be reduced by redistributing incomes among the poor. Similarly, the area bounded by the Lorenz curve and the line segment from $(p, L_F(p))$ to $(1, 1)$ represents the within-group inequality of the rich. The shaded area represents the across-group inequality. Easy computation shows that the extent of across-group inequality between the bottom $p \times 100\%$ and top is the (across-group) disparity function $D_F(p) = (1/2)[p - L_F(p)]$. One can ask for what value of p is the across-group inequality maximized? The answer is that this is maximized at the proportion associated with the *Pietra index* given by $\mathcal{P}_F = F(\mu) - L_F(F(\mu))$. Hence, $F(\mu)$ is the proportion where the disparity is maximized. Therefore, the Pietra index is that fraction which splits the society into two groups in a way such that inter-group inequality is maximized.

The discussion to follow shows that interpretation of the k-index is different from that of the Pietra index. Let us divide society into two groups, the "poorest" who constitute a fraction p of the population and the "richest" who constitute a fraction $1 - p$ of the population. Given the Lorenz curve $L_F(p)$, we look at the distance of the "boundary person" from the poorest person on the one hand and the distance of this person from the richest person on the other hand. These distances are given by $\sqrt{p^2 + L_F(p)^2}$ and $\sqrt{(1-p)^2 + (1 - L_F(p))^2}$, respectively. Then, the k-index divides society into two groups in a manner such that the Euclidean distance of the boundary person from the poorest person is equal to the distance from the richest person.

The value of the disparity function at the k-index is $D_F(k_F) = k_F - 1/2$. It measures the gap between the proportion k_F of the poor from the 50 – 50 population split. As long as we do not have a completely egalitarian society, $k_F > 1/2$ and hence it is one way of highlighting the rich-poor disparity with k_F defining the income proportion of the top $(1 - k_F)$ proportion of the rich population. In terms of disparity, the Gini index and Pietra index do not have as nice an interpretation.

4.2. Comparison of Magnitudes

To compare the k-index with other measures of inequality we will use the normalized k-index which is given by $\mathcal{K}_F := 2k_F - 1$. The normalized k-index was first introduced in Ref. 20 and was called the "perpendicular-diameter index" (see Refs. 20, 21, 23). For all income distributions used till the previous section we found that given any F, the value of the normalized k index is no more than the value of the Pietra index and the value of the Pietra index is no more than the value of the Gini index. This is

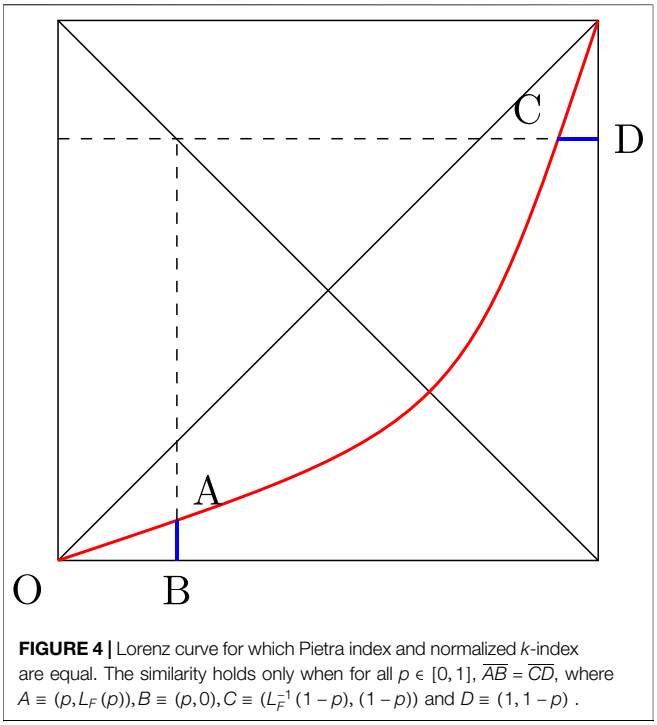


FIGURE 4 | Lorenz curve for which Pietra index and normalized *k*-index are equal. The similarity holds only when for all $p \in [0,1]$, $\overline{AB} = \overline{CD}$, where $A \equiv (p, L_F(p)), B \equiv (p,0), C \equiv (L_F^{-1}(1-p), (1-p))$ and $D \equiv (1, 1-p)$.

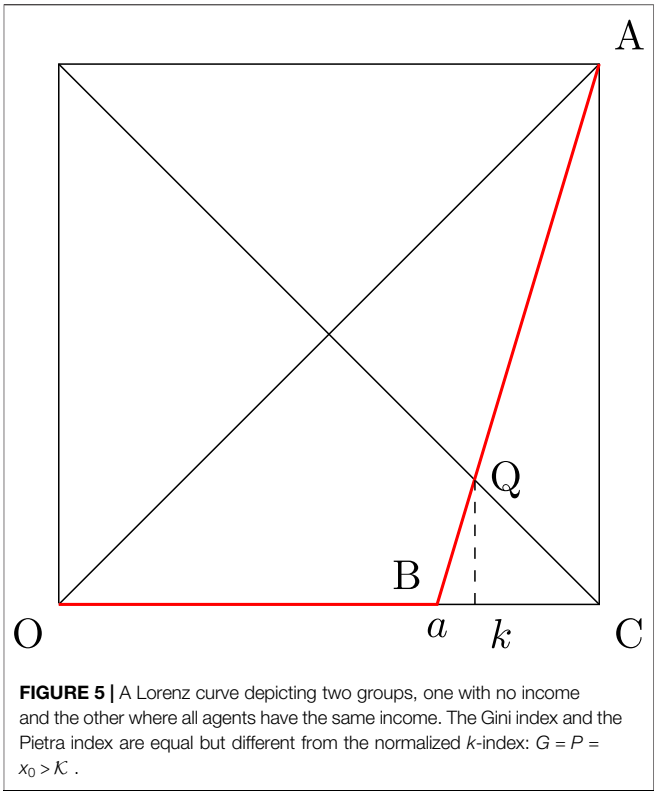


FIGURE 5 | A Lorenz curve depicting two groups, one with no income and the other where all agents have the same income. The Gini index and the Pietra index are equal but different from the normalized *k*-index: $G = P = x_0 > \mathcal{K}$.

not just a coincidence. It was established in Ref. 3 that for any income distribution F, we have $\mathcal{K}_F \leq \mathcal{P}_F \leq \mathcal{G}_F$. It is obvious that since the Pietra index maximizes $p - L_F(p)$, it is obvious that $\mathcal{K}_F = 2k_F - 1 = k_F - L_F(k_F) \leq \mathcal{P}_F$. Moreover, in Ref. 3, it was also established that for any given distribution F and any $p \in [0,1]$, $p - L_F(p) \leq \mathcal{G}_F$ and hence, using this result, it follows that $\max_{p \in [0,1]}\{p - L_F(p)\} \leq \mathcal{G}_F$ and hence we get $\mathcal{P}_F \leq \mathcal{G}_F$.

We first provide an example where the normalized *k*-index coincides with the Pietra index. This example is taken from Ref. 3. Let us consider an arc of a unit circle ODB as a Lorenz curve where OB is one of the diagonal (egalitarian line) of the unit square ABCO (as shown in **Figure 3**) where CD represents the unit radius of the circle, CA is the other diagonal of the unit square ABCO = $\sqrt{2}$. In this case the Lorenz curve is, $L_{F_{kg}}(p) = 1 - \sqrt{1-p^2}$ where F_{kg} is the relevant income distribution. One can verify that $\mathcal{K}_{F_{kg}} = \mathcal{P}_{F_{kg}} = \sqrt{2} - 1 \simeq 0.4142 < \mathcal{G}_{F_{kg}} = (\pi/2) - 1 \simeq 0.571$. Hence, the Gini index is larger than the Pietra index and the normalized *k*-index. Also in this case the maximum distance between perfect equality line and the Lorenz curve is at $k_{F_{kg}} = F(\mu_{kg}) = 1/\sqrt{2}$, hence Pietra index coincides with the normalized *k*-index.

The Lorenz function $L_F(p)$ is *symmetric* if for all $p \in [0,1]$, $L_F(\widehat{L}_F(p)) = 1 - p$ or equivalently $L_F(p) + r_F(p) = 1$, where $r_F(p) = L_F^{-1}(1-p)$. The idea of symmetry is explained in **Figure 4**. One can verify that the Lorenz function $L_{F_{kg}}(p) = 1 - \sqrt{1-p^2}$ is symmetric. It was proved in Banerjee, Chakrabarti, Mitra, and Mutuswami [3] that, in general, if the Lorenz function is symmetric and differentiable, then the proportion $F(\mu)$ associated with the Pietra index coincides with the proportion k_F of the *k*-index. Hence, we also have $\mathcal{K}_F = \mathcal{P}_F$.

The next example is one where the Pietra index coincides with the Gini index. This example is taken from Eliazar and Sokolov [18]. Fix any fraction $x_0 \in (0,1)$ and consider the following Lorenz function:

$$L_{F_{pg}}(p) = \begin{cases} 0 & \text{if } p \in [0, x_0], \\ \dfrac{(p - x_0)}{(1 - x_0)}, & \text{if } p \in (x_0, 1]. \end{cases} \tag{9}$$

Figure 5 depicts this Lorenz function $L_{F_{pg}}(.)$ and in particular the curve OBA represents this Lorenz curve. One can show that $x_0/2 - x_0 = \mathcal{K}_{F_{pg}} < \mathcal{P}_{F_{pg}} = \mathcal{G}_{F_{pg}} = x_0$. Hence, the Gini index coincides with Pietra and the normalized *k*-index has a lower value. Therefore, from this example we can say that *k*-index has different features relative to both the Gini index and the Pietra index.

Finally, when does all the three indices coincide? It was established in Ref. 3 that all three measures will coincide if and only if the Lorenz function has the following form defined for any given $C \in [1/2, 1)$:

$$L_C(p) = \begin{cases} \left(\dfrac{1-C}{C}\right)p & \text{if } p \in [0, C], \\ (1-C) + \dfrac{C}{(1-C)}(p - C) & \text{if } p \in (C, 1]. \end{cases} \tag{10}$$

In **Figure 6**, the straight lines OQ and QB taken together represents the Lorenz curve for $L_C(.)$. One can verify that

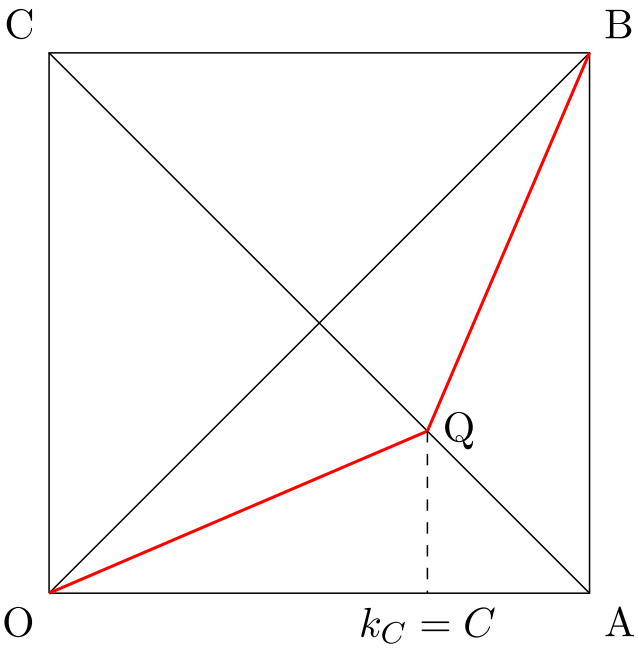


FIGURE 6 | A Lorenz curve depicting two groups with equally distributed incomes but differing average incomes. The Gini, Pietra and normalized k indices are all equal here: $\mathcal{K}_{F_C} = \mathcal{P}_{F_C} = \mathcal{G}_{F_C} = 2k_F - 1$.

$$\mathcal{K}_{F_C} = \mathcal{P}_{F_C} = \mathcal{G}_{F_C} = 2C - 1, \tag{11}$$

Observe that, if $C = 1/2$, then we have $L_{F_{0.5}}(p) = L_{F_e}(p) = p$ for all $p \in (0, 1)$ and in that case the three indices also coincide since $\mathcal{G}_{F_e} = \mathcal{P}_{F_e} = \mathcal{K}_{F_e} = 0$.

It is clear that the Lorenz functions of the form $L_{F_C}(.)$ with $C \in (1/2, 1)$ is valid for any society having two income groups. Therefore, a natural question in this context is the following: What does the coincidence of the three measures mean in terms of discrete random variables? For any discrete random variable F_G such that $G = 2$, we have $N = n_1 + n_2$, $M = n_1x_1 + n_2x_2$ with $x_1 < x_2$ and the associated Lorenz function has the following form:

$$L_{F_2}(p) = \begin{cases} \dfrac{(n_1+n_2)x_1 p}{n_1x_1 + n_2x_2}, & \text{if } p \in \left(0, \dfrac{n_1}{n_1+n_2}\right], \\ \dfrac{n_1x_1}{n_1x_1+n_2x_2} + \left(\dfrac{(n_1+n_2)x_2}{n_1x_1+n_2x_2}\right)\left(p - \dfrac{n_1}{n_1+n_2}\right), & \text{if } p \in \left(\dfrac{n_1}{n_1+n_2}, 1\right). \end{cases} \tag{12}$$

For the coincidence of all the three indices we first require that $C \in (1/2, 0)$ and $C = n_1/(n_1 + n_2)$ implying that $n_1 > n_2$. Moreover, for the coincidence we also require $C = k_{F_2}$, that is, $C + L_{F_2}(C) = 1$ which yields $n_1^2x_1 = n_2^2x_2$. Thus, from the above discussion we have the following result.

- Consider any discrete random variable F_G discussed in Example 1 for which the Lorenz function is given by **Eq. 3**. The normalized k-index coincides with the Gini index and the Pietra index if and only if any one of the following conditions holds:

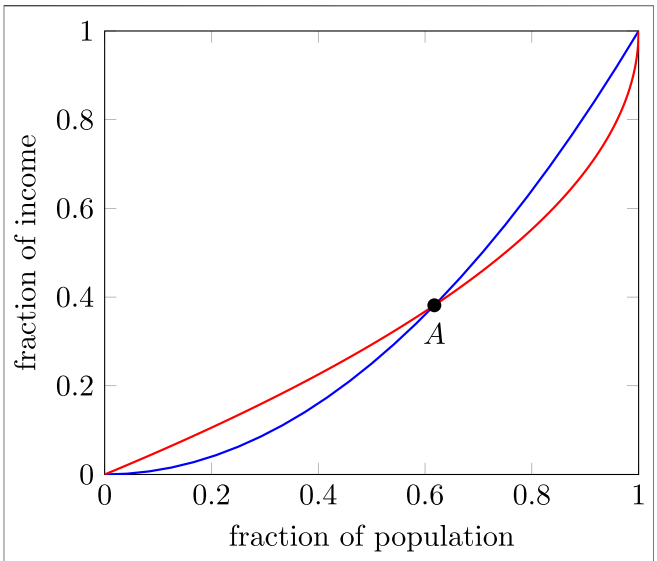


FIGURE 7 | Two Lorenz curves with identical Gini, Pietra and normalized k-indices. The blue curve is $L_{F_{\overline{u}}}(p) = p^2$ and the red curve is $L_{F_{P,2}}(p) = 1 - \sqrt{1-p}$.

(C1) The society has all agents having the same income $x_1 > 0$ so that $L_{F_1}(p) = L_{F_e}(p) = p$ for all $p \in (0, 1)$. For this case we have, $\mathcal{K}_{F_1} = \mathcal{P}_{F_1} = \mathcal{G}_{F_1} = 0$.

(C2) The society has two groups of agents with one group of n_1 agents having an income of x_1 and another group of n_2 agents having an income of x_2 such that $x_1 < x_2$. Moreover, the Lorenz function is $L_{F_2}(p)$ given in **Eq. 12** with the added restrictions that $n_1 > n_2$, $n_1^2x_1 = n_2^2x_2$ and hence $n_1x_1 < n_2x_2$. For this case we have, $\mathcal{K}_{F_2} = \mathcal{P}_{F_2} = \mathcal{G}_{F_2} = 2k_{F_2} - 1 = (n_1 - n_2)/(n_1 + n_2)$.

5. RANKING LORENZ FUNCTIONS

Consider the uniform income distribution $F_{\overline{u}}$ defined on any compact interval $[0, b]$ with $b > 0$. The Lorenz function is given by $L_{F_{\overline{u}}}(p) = p^2$ for all $p \in [0, 1]$ (see **Figure 7**). Here $k_{F_{\overline{u}}}$ is the reciprocal of the Golden ratio, that is, $k_{F_{\overline{u}}} = (\sqrt{5} - 1)/2 = 1/\phi$ where $\phi = (\sqrt{5} + 1)/2 \simeq 0.61803$ is the *Golden ratio*. Moreover, $\mathcal{K}_{F_{\overline{u}}} = \sqrt{5} - 2 \simeq 0.23607$. Similarly, one can derive that the Gini index is $\mathcal{G}_{F_{\overline{u}}} = 1/3$ and the Pietra index is $\mathcal{P}_{F_{\overline{u}}} = 1/4$ with $\mu_{\overline{u}} = 1/2$. Hence, we have $\mathcal{G}_{F_{\overline{u}}} = 1/3 > \mathcal{P}_{F_{\overline{u}}} = 1/4 > \mathcal{K}_{F_{\overline{u}}} = \sqrt{5} - 2$. Similarly, consider the Pareto distribution $F_{P,\alpha}$ with parameter value $\alpha = 2$. The Lorenz function is given by $L_{F_{P,2}}(p) = 1 - \sqrt{1-p}$ so that $\widehat{L}_{F_{P,2}}(p) = \sqrt{1-p}$ and the k-index is again the reciprocal of the Golden ratio, that is, $k_{F_{P,2}} = 1/\phi$ and $\mathcal{K}_{F_{P,2}} = \sqrt{5} - 2$ (see **Figure 7**). Thus, according to the normalized k-index, a society with an income distribution $F_{\overline{u}}$ is equivalent to a society with an income distribution of $F_{P,2}$ in terms of inequality. One can verify that this equivalence between $F_{\overline{u}}$ and $F_{P,2}$ is also preserved under the Gini index

and the Pietra index. Specifically, we have $\mathcal{G}_{F_{P,\alpha}} = \mathcal{G}_{F_{\overline{u}}} = 1/3$ and $\mathcal{P}_{F_{P,2}} = \mathcal{P}_{F_{\overline{u}}} = 1/4$ though $\mu_{P,2} = 3/4 > \mu_{\overline{u}} = 1/2$. Hence, we have

$$\mathcal{G}_{F_{P,\alpha}} = \mathcal{G}_{F_{\overline{u}}} = 1/3 > \mathcal{P}_{F_{P,2}} = \mathcal{P}_{F_{\overline{u}}} = 1/4 > \mathcal{K}_{F_{P,\alpha}} = \mathcal{K}_{F_{\overline{u}}} = \sqrt{5} - 2.$$

Consider the income distributions F_A and F_B defined in Example 2. From Remark 1 it follows that $\mathcal{K}_{F_A} < \mathcal{K}_{F_B}$, from Remark 2 it follows that $\mathcal{G}_{F_A} < \mathcal{G}_{F_B}$ and from Remark 3 it also follows that $\mathcal{P}_{F_A} < \mathcal{P}_{F_B}$. Therefore, all the three measures unambiguously assures that the society with income distribution F_A is less unequal that the society with income distribution F_B.

Given the above examples of this section, one may be tempted to think that ranking Lorenz functions using these three measures always gives the same order, that is, if one measure shows that the income distribution F is equivalent to another income distribution F' in terms of inequality, then the other two measures will also give the same result, and, if one measure shows that the income distribution F is less unequal than the income distribution F', then also the other two measures will establish the same order. However, as argued in Ref. 3, this is not the case. To establish this point [3] provided the following two examples.

In the first example the following Lorenz functions were considered to establish that the normalized k-index yields a different ranking from the Pietra index.

$$L_{F_a}(p) = \begin{cases} \dfrac{3p}{4}, & \text{if } p \in [0, 1/3], \\ \dfrac{9p-1}{8}, & \text{if } p \in (1/3, 1]. \end{cases}$$

$$L_{F_b}(p) = \begin{cases} \dfrac{8p}{9}, & \text{if } p \in [0, 7/8], \\ \dfrac{16p-7}{9}, & \text{if } p \in (7/8, 1]. \end{cases}$$

One can show that $\mathcal{K}_{F_a} = \mathcal{K}_{F_b} = 1/7 < \mathcal{P}_{F_a} = 1/12 < \mathcal{P}_{F_b} = 7/72$, that is, according to the normalized k-index, the society with income distribution F_b is equivalent to the society with income distribution F_b in terms of inequality. However, according to the Pietra index, the society with income distribution F_a is less unequal than the society with income distribution F_b.

In the second example, two Lorenz functions were considered of which the first one is the standard uniform distribution defined on any compact interval of the form $[0, b]$ with $b > 0$, that is, $L_{F_{\overline{u}}}(p) = p^2$ for all $p \in [0, 1]$. The other Lorenz function has the following form:

$$L_{F_S}(p) = \begin{cases} p^2 & \text{if } p \in [0, 3/4], \\ 1 - \left(\dfrac{7(1-p)}{4}\right) & \text{if } p \in (3/4, 1]. \end{cases}$$

$\mathcal{K}_{F_{\overline{u}}} = \mathcal{K}_{F_S} = 2/\phi - 1 < \mathcal{G}_{F_S} = 21/64 < \mathcal{G}_{F_{\overline{u}}} = 1/3$. This example demonstrates an important difference between $\mathcal{K}_F$ and $\mathcal{G}_F$. The Gini index is affected by transfers within a group. In particular, the poor are unaffected but the rich (lying in the interval $[3/4, 1)$) have become more egalitarian while moving from $L_{F_{\overline{u}}}$ to L_{F_S}. The normalized k-index on the other hand is unaffected with such intra-group transfers. Therefore, if we are interested in reducing inequality between groups, then the normalized k-index is a better indicator than the Gini index.

TABLE 1 | The Gini and k-indices for the income distributions of various countries, 1963–1983.

Country	Gini index	k-index
Brazil	0.62	0.73
Denmark	0.36	0.63
India	0.45	0.66
Japan	0.31	0.61
Malaysia	0.50	0.68
New Zealand	0.37	0.63
Panama	0.44	0.66
Sweden	0.38	0.64
Tunisia	0.50	0.69
Uruguay	0.49	0.68
Columbia	0.55	0.70
Finland	0.47	0.67
Indonesia	0.44	0.65
Kenya	0.61	0.73
Netherlands	0.44	0.66
Norway	0.36	0.63
Sri Lanka	0.40	0.65
Tanzania	0.53	0.70
United Kingdom	0.36	0.63
Australia	0.34	0.62
Canada	0.34	0.62
Netherlands	0.31	0.61
Norway	0.31	0.61
Sweden	0.29	0.60
Switzerland	0.38	0.63
Germany	0.31	0.61
United Kingdom	0.34	0.62
United States	0.36	0.63

The maximum error bar in estimated Gini and k *values is* ≃0.01 *[Adapted from Ref. 1].*

6. NUMERICAL OBSERVATIONS

For the purpose of comparison between different inequality indices, we present in **Table 1**, the estimated values of the Gini and k-indices for the income distributions in some countries for the period 1963–1983. **Tables 2** and **3** give the estimated values of these indices along with the Pietra index for citations, for different institutions and universities across the world observed in different years. **Table 4** also shows the comparison between Gini, Pietra and k for inequalities in paper citations for various science journals. All the tables are taken from Ref. 1.

In Ref. 1 it was observed that **Eq. 11** is an approximate result and can differ for large values of G and k. Furthermore, the value of k corresponds to an upper limit beyond which the distribution follows a power law pattern, similar to the celebrated Pareto law

TABLE 2 | The Gini coefficient, Pietra and *k*-indices for citations (up to December 2013) of the papers published from different universities as obtained from ISI web of science.

Inst./Univ	Year	Total papers	Citations	Gini index	Pietra index	k-index
Melbourne	1980	866	16,107	0.67	0.51	0.75
	1990	1,131	30,349	0.68	0.50	0.75
	2000	2,116	57,871	0.65	0.49	0.74
	2010	5,255	63,151	0.68	0.50	0.75
Tokyo	1980	2,871	60,682	0.69	0.52	0.76
	1990	4,196	108,127	0.68	0.51	0.76
	2000	7,955	221,323	0.70	0.53	0.76
	2010	9,154	91,349	0.70	0.52	0.76
Harvard	1980	4,897	225,626	0.73	0.55	0.78
	1990	6,036	387,244	0.73	0.55	0.78
	2000	9,566	571,666	0.71	0.54	0.77
	2010	15,079	263,600	0.69	0.52	0.76
MIT	1980	2,414	101,929	0.76	0.59	0.79
	1990	2,873	156,707	0.73	0.56	0.78
	2000	3,532	206,165	0.74	0.56	0.78
	2010	5,470	109,995	0.69	0.51	0.76
Cambridge	1980	1,678	62,981	0.74	0.56	0.78
	1990	2,616	111,818	0.74	0.56	0.78
	2000	4,899	196,250	0.71	0.54	0.77
	2010	6,443	108,864	0.70	0.52	0.76
Oxford	1980	1,241	39,392	0.70	0.53	0.77
	1990	2,147	83,937	0.73	0.56	0.78
	2000	4,073	191,096	0.72	0.54	0.77
	2010	6,863	114,657	0.71	0.53	0.76

The number of papers and citations give an idea of the data size involved in the analysis [Adapted from Refs. 1 and 2].

TABLE 3 | The Gini, Pietra and *k*-indices for citations (up to December 2013) of the papers published from different Indian universities, as obtained from ISI web of science [Adapted from Ref. 1].

Inst./ Univ	Year	Total papers	Citations	Gini index	Pietra index	k-index
SINP	1980	32	170	0.72	0.49	0.74
	1990	91	666	0.66	0.47	0.73
	2000	148	2,225	0.77	0.57	0.79
	2010	238	1896	0.71	0.52	0.76
IISC	1980	450	4,728	0.73	0.56	0.78
	1990	573	8,410	0.70	0.53	0.76
	2000	874	19,167	0.67	0.50	0.75
	2010	1,624	11,497	0.62	0.45	0.73
TIFR	1980	167	2024	0.70	0.52	0.76
	1990	303	4,961	0.73	0.54	0.77
	2000	439	11,275	0.74	0.55	0.77
	2010	573	9,988	0.78	0.59	0.79
Calcutta	1980	162	749	0.74	0.56	0.78
	1990	217	1,511	0.64	0.48	0.74
	2000	173	2073	0.68	0.50	0.74
	2010	432	2,470	0.61	0.45	0.73
Delhi	1980	426	2,614	0.67	0.49	0.75
	1990	247	2,252	0.68	0.52	0.76
	2000	301	3,791	0.68	0.51	0.76
	2010	914	6,896	0.66	0.49	0.74
Madras	1980	193	1,317	0.69	0.53	0.76
	1990	158	1,044	0.68	0.52	0.76
	2000	188	2,177	0.64	0.47	0.73
	2010	348	2,268	0.78	0.58	0.79

TABLE 4 | The Gini, Pietra and *k*-indices for citations (up to December 2013) of the papers published from different journals, as obtained from ISI web of science [Adapted from Ref. 1].

Journals	Year	Total papers	Citations	Gini index	Pietra index	k-index
Nature	1980	2,904	178,927	0.80	0.63	0.81
	1990	3,676	307,545	0.86	0.72	0.85
	2000	3,021	393,521	0.81	0.65	0.82
	2010	2,577	100,808	0.79	0.63	0.81
Science	1980	1,722	111,737	0.77	0.60	0.80
	1990	2,449	228,121	0.84	0.70	0.84
	2000	2,590	301,093	0.81	0.66	0.82
	2010	2,439	85,879	0.76	0.60	0.79
PNAS(USA)	1980	-	-	-	-	—
	1990	2,133	282,930	0.54	0.39	0.70
	2000	2,698	315,684	0.49	0.35	0.68
	2010	4,218	116,037	0.46	0.33	0.66
Cell	1980	394	72,676	0.54	0.39	0.70
	1990	516	169,868	0.50	0.36	0.68
	2000	351	110,602	0.56	0.41	0.70
	2010	573	32,485	0.68	0.51	0.75
PRL	1980	1,196	87,773	0.66	0.48	0.74
	1990	1904	156,722	0.63	0.47	0.74
	2000	3,124	225,591	0.59	0.43	0.72
	2010	3,350	73,917	0.51	0.37	0.68
PRA	1980	639	24,802	0.61	0.45	0.73
	1990	1922	54,511	0.61	0.45	0.72
	2000	1,410	38,948	0.60	0.44	0.72
	2010	2,934	26,314	0.53	0.38	0.69
PRB	1980	1,413	62,741	0.65	0.49	0.74
	1990	3,488	153,521	0.65	0.48	0.74
	2000	4,814	155,172	0.59	0.44	0.72
	2010	6,207	70,612	0.53	0.38	0.69
PRC	1980	630	19,373	0.66	0.49	0.75
	1990	728	15,312	0.63	0.46	0.73
	2000	856	19,143	0.57	0.42	0.71
	2010	1,061	11,764	0.56	0.40	0.70
PRD	1980	800	36,263	0.76	0.59	0.80
	1990	1,049	33,257	0.68	0.52	0.76
	2000	2061	66,408	0.61	0.45	0.73
	2010	3,012	40,167	0.54	0.39	0.69
PRE	1980	—	—	—	—	—
	1990	—	—	—	—	—
	2000	2,078	51,860	0.58	0.42	0.71
	2010	2,381	16,605	0.50	0.36	0.68

[24]. For the inequality in citation data, if n is the fraction of papers and w is the cumulative fraction of citations, then for $n \geq k$, $1 - w \sim (1 - n)^{\alpha}$ with $\alpha = 0.50 \pm 0.10$ which implies $n = 1 - c*(1 - w)^{\nu}$ for $\nu = 2.0 \pm 0.5$ and c is a proportionality constant. This is illustrated in **Figures 8** and **9**.

7. SUMMARY AND DISCUSSION

For the nonlinear Lorenz function ($L_F(p)$), the traditional measures like Gini index measures some "average property", while the Kolkata index (k) identifies the non-trivial fixed point of the complementary Lorenz function ($\widehat{L}_F(p) = 1 - L(p)$; note that $L_F(p)$ has trivial fixed points at $p = 0$ and 1, while $\widehat{L}_F(p)$ has a nontrivial fixed point at $p = k$). This k-index

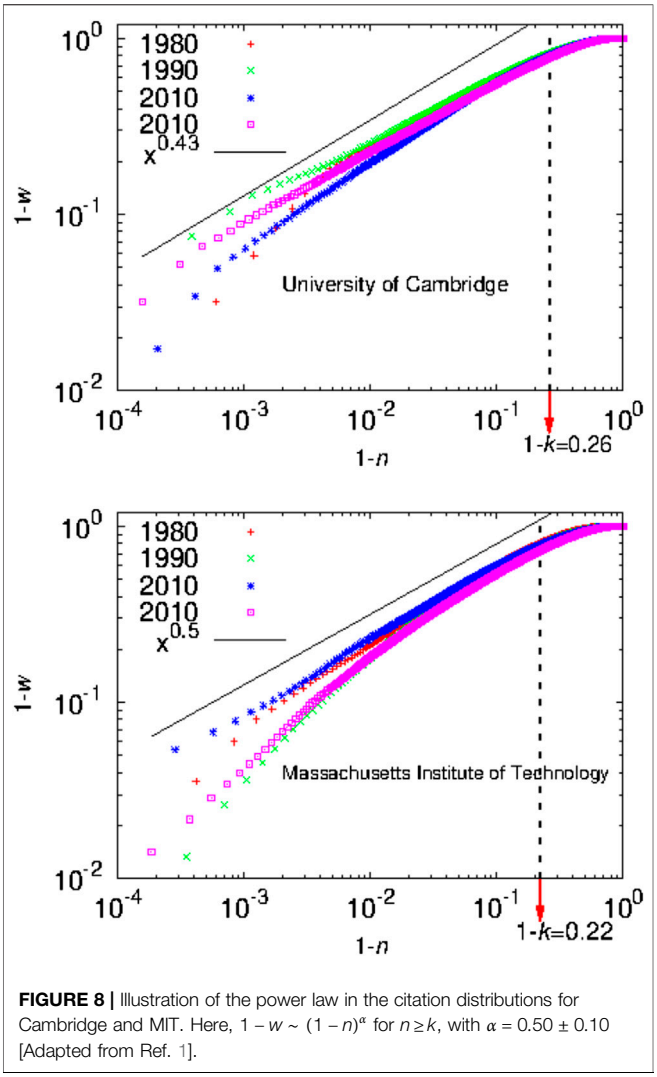


FIGURE 8 | Illustration of the power law in the citation distributions for Cambridge and MIT. Here, $1 - w \sim (1 - n)^{\alpha}$ for $n \geq k$, with $\alpha = 0.50 \pm 0.10$ [Adapted from Ref. 1].

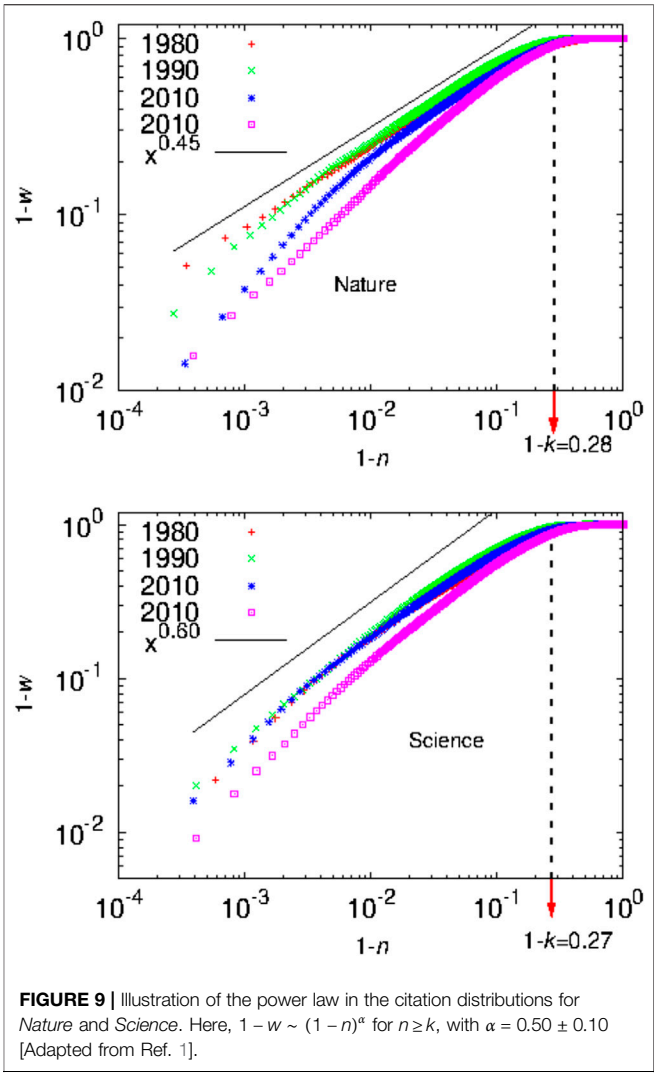


FIGURE 9 | Illustration of the power law in the citation distributions for *Nature* and *Science*. Here, $1 - w \sim (1 - n)^{\alpha}$ for $n \geq k$, with $\alpha = 0.50 \pm 0.10$ [Adapted from Ref. 1].

apart from capturing the essential character of the nonlinear Lorenz function (as inspired by the major developments of renormalization group theory in statistical physics [14] or in identifying the universal characters corresponding to the onset of chaos in nonlinear systems [15]), also gives us a very tangible one, giving that $(1 - k)$ fraction of the population possess k fraction of the total wealth in the society. In Ref. 25 the k-index is used to define a generalized Gini index. In a recent study, the k-index has been used to quantify the inequality for spreading of the Covid-19 infection in urban neighbourhoods and slums in a society (see Ref. 26).

After a general introduction in **Section 1**, we discuss in **Section 2**, some structural features of the Lorenz function and introduce the Complementary Lorenz function, which has a nontrivial fixed point (namely the Kolkata index) as mentioned above. In **Sections 3 and 4**, we try to demonstrate the uniqueness of the k-index, compared to Gini and Pietra indices in ranking the rich-poor disparity, assuming some typical income distributions. we have argued (in **Section 3**) that the procedure of obtaining the h-index of any research scientist using the generated citation curve is the same as identifying the fixed point of the complementary Lorenz function of any income distribution that yields the k index. While comparing the normalized k-index with the Pietra index and with the Gini index, one can show that for any given distribution the normalized k-index is no more than the Pietra index and the Pietra index is no more than the Gini index. We have also argued (in **Section 4.2**) that for any given distribution the normalized k-index, the Pietra index and the Gini index coincide only if either the society is such that all agents have equal income or there are only two income groups in a society with some added restrictions (see condition C2 in this subsection). We have also argued (in **Section 5**) that if we are interested in reducing inequality between the rich and poor groups of the society, then the normalized k-index is a better indicator than the Gini index. In **Section 6**, we can see that while the Gini index value typically ranges from 0.30 to 0.62, the Kolkata index value ranges from 0.60 to 0.73 at any particular time or year for income or wealth data across the countries of the world. It may be mentioned here that income inequality data are not easily available from reliable sources. On

the other hand, the (paper) citations may be considered as a measure of the wealth created by the respective University or Institution and the resulting inequality data are abundantly available in accurate digital formats (say from the ISI Web of Science). We estimated the Gini, Pietra, and Kolkata index values for the citations earned by the yearly publications of various academic institutions from such data sources. We find that while Gini and Pietra index values range from 0.65 to 0.75 and 0.50 to 0.60, respectively, the Kolkata index remains around 0.75 ± 0.05 value for Institutions or Universities across the world. As mentioned already, k-index is the social equivalent to the h-index for an individual researcher or academician. Also we find that the value for k-index gives an estimate of the crossover point beyond which the growth of income (or citations) with the fraction of population (or publications) enters a power law (Pareto) region (see **Figures 8** and **9**).

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Ghosh A, Chattopadhyay N, Chakrabarti BK. Inequality in societies, academic institutions and science journals: Gini and k-indices. *Phys A* (2014) 410:30–4. doi:10.1016/j.physa.2014.05.026
2. Chatterjee A, Ghosh A, Chakrabarti BK. Socio-economic inequality: relationship between Gini and Kolkata indices. *Phys A* (2017) 466:583–95. doi:10.1016/j.physa.2016.09.027
3. Banerjee S, Chakrabarti BK, Mitra M, Mutuswami S. On k-index as a measure of income inequality. *Phys A* (2019) 545:123178. doi:10.1016/j.physa.2019.123178
4. Chakravarty SR. *Inequality, polarization and conflict*. New York, NY: Springer (2015).
5. Jenkins SP, Van Kerm P. The measurement of economic inequality. In: B Nolan, W Salverda, TM Smeeding, editors *The oxford handbook of economic inequality*. Oxford, UK: Oxford University Press (2011).
6. Lorenz M. Methods of measuring the concentration of wealth. *Publ Am Stat Ass* (1905). 9:209–19. doi:10.2307/2276207
7. Aaberge R. Characterizations of Lorenz curves and income distributions. *Soc Choice Welfare* (2000) 17:639–53. doi:10.1007/s003550000046
8. Dardanoni V, Lambert PJ. Welfare rankings of income distributions: a role for variance and some insights for tax reforms. *Soc Choice Welfare* (1988). 5:1–17. doi:10.1007/bf00435494
9. Lambert PJ. *The distribution and redistribution of income: a mathematical analysis*. Manchester, UK: Manchester University Press (1993).
10. Shorrocks AF, Foster JE. Transfer sensitive inequality measures. *Rev Econ Stud* (1987). 54(3):485–97. doi:10.2307/2297571
11. Zoli C. Intersecting generalized Lorenz curves and the Gini index. *Soc Choice Welfare* (1999) 16:183–96. doi:10.1007/s003550050139
12. Dalton H. The measurement of the inequality of incomes. *Econ J* (1920) 30: 348–461. doi:10.2307/2223525
13. Atkinson AB. On the measurement of inequality. *J Econ Theor* (1970). 2: 244–63. doi:10.1016/0022-0531(70)90039-6
14. Fisher ME. Renormalization group theory: its basis and formulation in statistical physics. *Rev Mod Phys* (1998) 70:653–81. doi:10.1103/revmodphys.70.653
15. Feigenbaum MJ. Universal behavior in nonlinear systems. *Phys Nonlinear Phenom* (1983) 7:16–39. doi:10.1016/0167-2789(83)90112-4
16. Gini CW. *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche*. Bologna, Italy: C. Cuppini (1912).
17. Pietra G. Delle relazioni tra gli indici di variabilita. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti* (1915) 74:775–804. Note I, II.
18. Eliazar II, Sokolov IM. Measuring statistical heterogeneity: the Pietra index. *Phys A* (2010) 389:117–25. doi:10.1016/j.physa.2009.08.006
19. Gastwirth JP. A general definition of the Lorenz curve. *Econometrica* (1971) 39(6):1037–9. doi:10.2307/1909675
20. Eliazar II. The sociogeometry of inequality: Part I. *Phys A* (2015) 426:93–115. doi:10.1016/j.physa.2015.01.016
21. Eliazar II. Harnessing inequality. *Phys Rep* (2016) 649:1–29. doi:10.1016/j.physrep.2016.07.005
22. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* (2005) 102:16569–72. doi:10.1073/pnas.0507655102
23. Eliazar II. The sociogeometry of inequality: Part II. *Phys A* (2015). 426:116–37. doi:10.1016/j.physa.2014.12.021
24. Pareto V. *Cours d'économie politique*. Vol. 1. Paris, France: Librairie Droz (1964).
25. Subramanian S. Further tricks with the Lorenz curve. *Econ Bull* (2019) 39(3): 1677–86.
26. Sahasranaman A, Jensen HJ. *Spread of Covid-19 in urban neighbourhoods and slums of the developing world*. arxiv [physics.soc-ph], 2010.06958 (2020).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Banerjee, Chakrabarti, Mitra and Mutuswami. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

8. APPENDICES

8.1. Appendix A

We formally show that for the discrete random variable F_G with the Lorenz function is given by **Eq. 3**, the Gini index has the following explicit form:

$$\mathcal{G}_{F_G} = \frac{\sum_{g=1}^{G}\sum_{t=1}^{G} n_t n_g |x_t - x_g|}{2NM}.$$

Observe first that

$$\begin{aligned}
&\int_0^1 L_{F_G}(q)\mathrm{d}q = \sum_{g=1}^{G}\left\{\int_{N(g-1)}^{N(g)} L_{F_G}(p_k)dp_k\right\}\\
&= \sum_{g=1}^{G}\left\{\int_{N(g-1)}^{N(g)} \left\{M(g-1)+\left(p_g - N(g-1)\right)\left(\frac{Nx_g}{M}\right)\right\}dp_g\right\} \qquad \text{(A1)}\\
&= -\frac{\sum_{g=1}^{G}\sum_{t=1}^{g-1} n_g n_t\left(x_g - x_t\right)}{NM} + \frac{\sum_{g=1}^{G}\left(2\sum_{t=1}^{g-1} n_t + n_g\right)n_t x_g}{2NM}.
\end{aligned}$$

Thus, using $2\sum_{g=1}^{G}(\sum_{t=1}^{g-1} n_t - \sum_{t=g+1}^{G} n_t)n_g x_g = \sum_{g=1}^{G}\sum_{t=1}^{G} n_g n_t |x_g - x_t|$ and using **Eq. A1** we get

$$\begin{aligned}
\mathcal{G}_{F_G} &= 1 - 2\int_0^1 L_{F_G}(q)\mathrm{d}q\\
&= 1 - \frac{\sum_{g=1}^{G}\left(2\sum_{t=1}^{g-1} n_t + n_g\right)n_g x_g}{NM} + \frac{2\left\{\sum_{g=1}^{G}\sum_{t=1}^{g-1} n_g n_t\left(x_g - x_t\right)\right\}}{NM}\\
&= \frac{\sum_{g=1}^{G}\left(\sum_{t=g+1}^{G} n_t - \sum_{t=1}^{g-1} n_t\right)n_g x_g}{NM} + \frac{\sum_{g=1}^{G}\sum_{t=1}^{G} n_g n_t |x_g - x_t|}{NM} \qquad \text{(A2)}\\
&= -\frac{\sum_{g=1}^{G}\sum_{g=1}^{G} n_g n_t |x_g - x_t|}{2NM} + \frac{\sum_{g=1}^{G}\sum_{t=1}^{G} n_g n_t |x_g - x_t|}{NM}\\
&= \frac{\sum_{g=1}^{G}\sum_{t=1}^{G} n_g n_t |x_g - x_t|}{2NM}.
\end{aligned}$$

Hence, from the last inequality in **Eq. A2** the result follows.

8.2. Appendix B

8.2.1. Appendix B (i)

The following derivation shows why $\mathcal{P}_F = E(|x-\mu|)/2\mu$ this is true.

$$\begin{aligned}
\mathcal{P}_F &= F(\mu) - L_F(\mu)\\
&= F(\mu) - \frac{\int_0^{F(\mu)} F^{-1}(q)\mathrm{d}q}{\mu}\\
&= \frac{\int_0^{F(\mu)}\{\mu - F^{-1}(q)\}\mathrm{d}q}{\mu} = \frac{2\int_0^{F(\mu)}\{\mu - F^{-1}(q)\}\mathrm{d}q}{2\mu} \qquad \text{(B1)}\\
&= \frac{\int_0^{F(\mu)}\{\mu - F^{-1}(q)\}\mathrm{d}q + \int_{F(\mu)}^{1}\{F^{-1}(q) - \mu\}\mathrm{d}q}{2\mu}\\
&= \frac{\int_0^1 |F^{-1}(q) - \mu|\mathrm{d}q}{2\mu}\\
&= \frac{E(|x-\mu|)}{2\mu}.
\end{aligned}$$

8.2.2. Appendix B (ii)

We formally show that for the discrete random variable F_G with the Lorenz function is given by **Eq. 3**, the Pietra index has the following explicit form:

$$\mathcal{P}_{F_G} = \frac{\sum_{g=1}^{\tilde{g}} n_g\left(\mu_G - x_g\right)}{M} = \frac{E(|x-\mu_G|)}{2\mu_G},$$

where $\tilde{g} \in \{1,\ldots,G-1\}$ is such that $\mu_G \in [x_{\tilde{g}}, x_{\tilde{g}+1})$ implying that $F_G(\mu_G) = N(\tilde{g})$.

For the first equality, observe that there exists $\tilde{g} \in \{1,\ldots,G-1\}$ such that $\mu_G \in [x_{\tilde{g}}, x_{\tilde{g}+1})$ implying that $F_G(\mu_G) = N(\tilde{g})$. Thus, using $\sum_{g=1}^{G} n_g(x_g - \mu_G) = 0$ and using $F_G(\mu_G) - N(\tilde{g}) = 0$ we get

$$\begin{aligned}
\mathcal{P}_{F_G} &= F_G(\mu_G) - L_{F_G}(\mu_G)\\
&= F_G(\mu_G) - M(\tilde{g}-1) - \{(F_G(\mu_G) - N(\tilde{g}-1))\}\left(\frac{Nx_{\tilde{g}}}{M}\right)\\
&= F_G(\mu_G)\left(\frac{M - Nx_{\tilde{g}}}{M}\right) - \left\{M(\tilde{g}-1) - N(\tilde{g}-1)\left(\frac{Nx_{\tilde{g}}}{M}\right)\right\}\\
&= \frac{F_G(\mu_G)\left(\sum_{g=1}^{G} n_g\left(x_g - x_{\tilde{g}}\right)\right)}{M} + \frac{\sum_{g=1}^{\tilde{g}} n_g\left(x_{\tilde{g}} - x_g\right)}{M} \qquad \text{(B2)}\\
&= \frac{\sum_{g=1}^{G} n_g\left(x_g - \mu_G\right)}{M} + \frac{\sum_{g=1}^{\tilde{g}} n_g\left(\mu_G - x_g\right)}{M} + \frac{(F_G(\mu_G) - N(\tilde{g}))N\left(\mu_G - x_{\tilde{g}}\right)}{M}\\
&= \frac{\sum_{g=1}^{\tilde{g}} n_g\left(\mu_G - x_g\right)}{M}.
\end{aligned}$$

Given **Eq. B2** it follows that the Pietra index of the distribution F_G with $\mu_G \in [x_{\tilde{g}}, x_{\tilde{g}+1})$ is

$$\mathcal{P}_{F_G} = \frac{\sum_{g=1}^{\tilde{g}} n_g\left(\mu_G - x_g\right)}{M}. \qquad \text{(B3)}$$

Given **Eq. B3**, we can also derive second equality by using $\mu_G \in [x_{\tilde{g}}, x_{\tilde{g}+1})$ and by using $\sum_{g=1}^{\tilde{g}} n_g(\mu_G - x_g) = \sum_{g=\tilde{g}+1}^{n} n_g(x_g - \mu_G)$. Specifically,

$$\begin{aligned}
\mathcal{P}_{F_G} &= \frac{\sum_{g=1}^{\tilde{g}} n_g\left(\mu_G - x_g\right)}{M}\\
&= \frac{\sum_{g=1}^{\tilde{g}} n_g\left(\mu_G - x_g\right) + \sum_{g=\tilde{g}+1}^{G} n_g\left(x_g - \mu_G\right)}{2M} \qquad \text{(B4)}\\
&= \frac{\sum_{g=1}^{G}\left(n_g/N\right)|x_g - \mu_G|}{2\left(M/N\right)}\\
&= \frac{E(|x-\mu_G|)}{2\mu_G}.
\end{aligned}$$

frontiers in Physics

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fphy.2020.581994

Lending Diversification and Interconnectedness of the Syndicated Loan Market

Gabjin Oh[1] *and A-young Park*[1,2]*

[1]Division of Business Administration, Chosun University, Gwangju, South Korea, [2]Department of Anesthesiology, University of Michigan, Ann Arbor, MI, United States

We investigate the effects of syndicated loan network centrality on bank performance. Syndicated loan network centrality measures the similarity and influence of the other banks within a given banks network. The network centrality constructed by syndicated loans can allow banks to gather and transfer valuable information and can thus facilitate profit-making acquisition in loan investment decisions. We use a planar maximally filtered graph to construct an interbank network using syndicated loan portfolios at the industry level. We show that the syndicated loan portfolios of high-centrality banks exhibit a higher level of portfolio diversification than those of low-centrality banks. We also document that our composite centrality measure of the bank network showed statistical significance in terms of bank performance even after controlling for the financial variables of market size, loan allocation, total asset, and loan diversification. Our findings suggest that the performance of a bank in a syndicated loan hierarchy is related to its position in this hierarchy.

Keywords: performance, connectedness, diversificiation, planar maximally filtered graph network, bank network, syndicated loan market

OPEN ACCESS

Edited by:
Haroldo V. Ribeiro,
State University of Maringá, Brazil

Reviewed by:
Satyam Mukherjee,
Indian Institute of Management Udaipur, India
Higor Sigaki,
State University of Maringá, Brazil

***Correspondence:**
A-young Park
ayppark@gmail.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *10 July 2020*
Accepted: *29 October 2020*
Published: *12 January 2021*

Citation:
Oh G and Park A (2021) Lending Diversification and Interconnectedness of the Syndicated Loan Market. Front. Phys. 8:581994. doi: 10.3389/fphy.2020.581994

1. INTRODUCTION

The connectivity between banks demonstrates the ways in which the contagious nature of high levels of risk among financial institutions can cause financial crizes and affect future economic conditions [1–4]. The network structure of the interbank market created by the syndicated loan market suggests that connections between banks should be an important channel of contagion among financial institutions [5–7]. Information contagions between banks represent a significant channel that might explain how information travels through financial systems. Recently, the application of complex networks to solve this challenging problem has become increasingly widespread in diverse areas [8–10].

In this paper, we study interbank networks in the form of common exposures among financial institutions to analyze bank performance based on banks' exposure to large syndicated loans. Syndicated loans represent one of the crucial sources of external financing for many firms and provide an ideal experimental setting for studying the interconnectedness of banks. In this study, the network between banks is constructed from data sets that contain information regarding both the borrowers and lenders of syndicated loans. The common exposures of banks are able to measure bank's investment strategies in this market in terms of loan portfolio diversification.

Prior research provides evidence that interconnectedness has a considerable impact on the economy from the perspective of risk exposure. Interconnection between companies or industries amplifies and propagates shock within an economy [11]. Negative shock and financial distress contribute to asset fire sales [12]. Consistent with these concepts, credit concentration tends to lead to

a cascade effect of shock in an economy [13]. [5] defined market connectedness using banks' loan specializations in a syndicated loan market that reflected systemic risk. Furthermore, prior studies that have examined the role of diversification have focused on performance. For example, banks with a greater number of geographically-concentrated mortgage loans performed better than others with fewer of these loans [14]. In terms of mergers and acquisitions, diversification is correlated with fluctuations in external market friction [15].

Based on the social exchange theory as proposed by [16]; we present different perspectives to understand the banking industry in the United States; these perspectives recognize the complex and rich social relationships that define interbank network. When the economy is growing, banks actually benefit from promoting the sharing of information with network members for business expansion; as a result of this sharing, they are able increase their profits. Nonetheless, during periods of economic contraction, banks cannot force network members to restructure because they may be subject to strict constraints due to their obligations. Banks are expected to expend effort monitoring and screening their borrowers to mitigate risk exposure. Additionally, bank performance is negatively affected within a contracting economy.

To assess the level of connectedness between the banks of syndicated loan portfolios, we establish a measure of interconnectedness that utilizes the similarity between bank's syndicated loan portfolios at the industry level as proposed by [5]. An advantage provided by the use of loan portfolios is the ability to investigate the response of banking systems via direct connections. To extract meaningful information from all-to-all connected networks, we employ the planar maximally filtered graph (PMFG) [17]. We utilize centrality measures to drive an important component that may affect whether a bank's centrality in the interbank network created in the financial sector is related to its performance. In this paper, the centrality is measured by the principal component analysis (PCA) method based on four common measures of centrality in the context of networks: degree, eigenvector, closeness, and betweenness.

To date, only the lending relationship between banks and firms has been studied through analyzing the characteristics of individual banks or firms using corporate loan data. The aim of this paper is to study an interbank network, namely, the syndicated loan market. We investigate the evolution of several types of syndicated loans over time using a Dealscan database, with a special emphasis on the amount of syndicated loans that have been extended. More interestingly, the syndicated loan data used in this study allows us to investigate the effect of the centrality of interbank networks on bank performance.

We show that banks with a higher level of network centrality are more likely to pursue diversification and that this diversification is more likely to increase during market instability. To extend our examination of the relationship between interbank networks and bank performance, we move beyond bank-to-firm lending by studying interbank networks in the context of the syndicated loan market. We further find that banks with a high level of centrality have higher returns than do banks with a low level of centrality. Since a bank's centrality within the network plays an important role in its loan portfolio strategy, it also plays a significant role for lending market participants. We also found that in the core group, there was a negative correlation between diversification and centrality; however, a positive relation was observed in the peripheral group.

The paper is organized as follows. **Section 2** explains the methodology that we employed. **Section 3** presents a description of the database used, and **Section 4** contains an empirical analysis. **Section 5** concludes this paper.

2. METHODOLOGY

In this section, we explain the network construction and regression variables. For each month, we define an interconnectedness based on the similarities between syndicated loan portfolios. The results are not qualitatively sensitive to bank performance measures, e.g., we obtain essentially the same results even if we use different financial variables to measure bank performance.

2.1. Network Construction

In this subsection, we explain the way in which we estimate the distance between two banks based on their loan portfolios. We then describe the way in which we construct an interbank network. To map our interbank network, we obtain information on the relationships between banks and firms between 1990 and 2017 from the DealScan database.

First, we investigate bank syndicated loans in the United.States. lending industry classified using two-digit SIC industry codes. This measure was developed by [5] and uses the Euclidean distance between two banks. For each month, we calculate the distance between bank i and bank k by quantifying the similarity of these two banks in a n-dimensional space as follows.

$$Distance_{i,k,t} = \frac{1}{\sqrt{2}} \times \sqrt{\Sigma_{j=1}^{n}\left(w_{i,j,t} - w_{k,j,t}\right)^2} \tag{1}$$

where $w_{i,j,t} = \frac{L_j}{\sum_{j=1}^{n} L_j}$, with syndicated loan of bank i invested in industry j, L_j, within the 12 months prior to month t. The distance is normalized between 0 and 1; 0 refers to perfectly matched portfolios and one refers to portfolios that do not overlap at all. We then construct a filtered network that connects all the banks so that a planar maximally filtered graph (PMFG) can be used [17]. The most common method of forming a stock network is based on the correlation of stock returns using threshold [18, 19]. This method has a problem in which correlation coefficient only assumes a linear relationship and lead to neglect some information. In addition, the minimum spanning tree (MST), a tree formed by a subset of edges of a given undirected graph, is also a common method in complex network analysis [20]. However, this method reflects hierarchical clustering with information loss to generate a efficient network. To address these issues, we use PMFG measure to construct a network based on the syndicated loans.

2.2. Main Dependent and Independent Variables

We investigate how network structure affects bank performance using the banks in the United.States. between January 1, 1990 and December 31, 2017. We use the Return on asset (ROA) variable as the dependent variable to measure bank's performance and employ several financial variables, such as the bank size, an amount of syndicated loan, etc. as control variables to examine network effect on bank's performance.

2.2.1. Diversification

In information theory, following [21]; the entropy of a discrete random variable X is denoted as

$$H(X) = \frac{-p(x_i)\Sigma_i^n log\left(p(x_i)\right)}{-p^m(x_i)\Sigma_i^n log\left(p^m(x_i)\right)} \quad (2)$$

where $p(x_i)$ is the probability distribution of outcome X and $p^m(x_i)$ is defined by $1/n$. x_i is the proportion of the total loan amount of industry i held by a bank and n is the number of industries invested by the bank. It is well known that entropy is viewed as a measure of the uncertainty of a random variable. Entropy have manifested useful across a wide range of fields, so it is remarkable they have begun to make noticeable effect into economics and finance. It has also been a popular diversity index in previous literature. In this paper, we use the concept of diversification that corresponds to the above measure within the range of zero to one. When H is zero, the bank has concentration of loan portfolio. Otherwise, when H is one, the bank has perfect diversification of loan portfolio.

2.2.2. Network Centrality

The effect of bank network centrality on bank performance is due to the importance of bank-firm lending structure in the context of information asymmetry. A bank's network created by bank-to-firm loan information should affect the profit of lending banks. Generally, centrality refers to a bank's location in a network compared to that of others. The four indices of centrality are frequently discussed in the social network literature [22]. These four indices are degree centrality, eigenvector centrality, closeness centrality, and betweenness centrality. These indices represent different dimensions of connectedness that affect information sharing via a network. Degree centrality is the sum of the first-degree connections of an entity in a network. The raw score is divided by the total number of nodes in the network minus 1, because the size of the interbank network changes each month [23]. Eigenvector centrality measures an individual bank's ability to obtain or influence information within the network. This measure increases as connections with other highly connected neighbors are added. The raw score is divided by the total number of nodes in the network minus one because the size of the interbank network changes each month. Closeness centrality is the inverse of the mean of the shortest path length between an individual bank and all the other reachable banks in the network. The raw score is multiplied by the total number of nodes in the network minus one because the size of the interbank network changes each month. Betweenness centrality describes the extent to which an individual bank is connected to the other banks in the network. When the shortest path of all bank pairs passes through a bank, the betweenness centrality of that bank is high; this is the reason why it is important to control the flow of the entire network. The raw score is divided by the total number of the connected nodes because the size of the interbank network changes each month.

To generate our composite centrality index (CCI) in **Table 1**, we standardize the centrality indices to a mean of 0 and a standard deviation of 1. Consistent with [24–26]; we use the factor score to aggregate CCI using the first principal component for each bank with four centrality indices in the PMFG network.

TABLE 1 | The effect of diversification and network centrality on bank performance.

Variables	Coefficient	t-value	Coefficient	t-value
Intercept	-2.1×10^{10}	(–0.4797)	-1.3×10^{10}	(–0.2783)
CCI			0.0084***	(10.5008)
CCI×Dummy			–0.0170***	(–6.6791)
DIV	0.0344***	(5.0023)	0.0216***	(3.0677)
Market size	0.0820***	(4.3706)	0.0850***	(4.5429)
Market share	–0.0140***	(–5.8422)	–0.0197***	(–8.0319)
Bank size	–0.0623***	(–29.2561)	–0.0626***	(–29.3747)
Observations	33,289		33,289	
Year FEs	Yes		Yes	
Adj. R^2	0.2224		0.2255	

*This table reports the regressions of diversification and centrality on ROA. ROA is defined as the net income divided by total assets. DEGREE is the degree centrality. Diversification (DIV) is measured by the Shannon entropy of bank portfolio calculated as the amount of loans extended to ten industries by each bank. Composite centrality index (CCI) is calculated by using principal component analysis of four centrality measures pertaining to the PMFG network, namely, degree centrality (DEGREE), eigenvector centrality (EIGEN), betweenness centrality (BTWN), and closeness centrality (CLOSE). Market size is defined as the log of the sum of all outstanding loans. Market share is defined as the log of the amount of loans extended by each bank. Bank size is defined by log of total assets of each bank. Year fixed effects are included to account for time characteristics. The t-statistic is reported in brackets. The symbols *, **, and *** indicate statistical significance at 10%, 5%, and, 1%, respectively.*

2.2.3. Bank Performance Measure

Return on assets (ROA) is an indicator of how well a company generates profit from its total assets. We calculated ROA by dividing firms' profit or loss before taxes by their total assets in month t and converted this figure to a percentage. The previous studies related to the current research area show that ROA is the best measure of performance when comparing similar companies with the same industry.

3. DATA DESCRIPTION

To test the hypotheses outlined in **Section 1**, we construct a sample of syndicated loans matched according to firm and bank characteristics. Below, we describe the sample construction and summarize the sample characteristics.

3.1. Data Source

We build our datasets from a comprehensive sample of syndicated loans and the associated lender and borrower

TABLE 2 | Pearson correlation of regression variables.

	ROA	Market size	Market share	Bank size	DIV	CCI	DEGREE	EIGEN	BTWN	CLOSE
ROA	1.00	−0.05	−0.12	−0.27	−0.04	0.02	-0.09	−0.09	0.02	−0.13
Market size		1.00	0.37	0.57	0.15	0.00	−0.07	−0.06	−0.02	0.03
Market share			1.00	0.38	0.08	0.57	0.45	0.57	0.36	0.63
Bank size				1.00	0.16	0.03	0.08	0.11	0.02	0.18
DIV					1.00	0.62	0.44	0.62	0.36	0.69
CCI						1.00	0.84	0.88	0.88	0.76
DEGREE							1.00	0.82	0.78	0.67
EIGEN								1.00	0.70	0.86
BTWN									1.0	0.56
CLOSE										1.0

*This table presents correlation coefficient of two variables. The value in **Table 2** has statistical significance ($p < 0.01$). ROA is defined as the net income divided by total assets. Market size is defined as the natural logarithm of the sum of all outstanding loans. Market share is defined as the natural logarithm of the amount of loans extended by each bank. Bank size is defined by natural logarithm of total assets of each bank. Diversification (DIV) is measured by the Shannon entropy of bank portfolio calculated as the amount of loans extended to ten industries by each bank. Composite centrality index (CCI) is calculated by using principal component analysis of four centrality measures pertaining to the PMFG network, namely, degree centrality (DEGREE), eigenvector centrality (EIGEN), betweenness centrality (BTWN), and closeness centrality (CLOSE).*

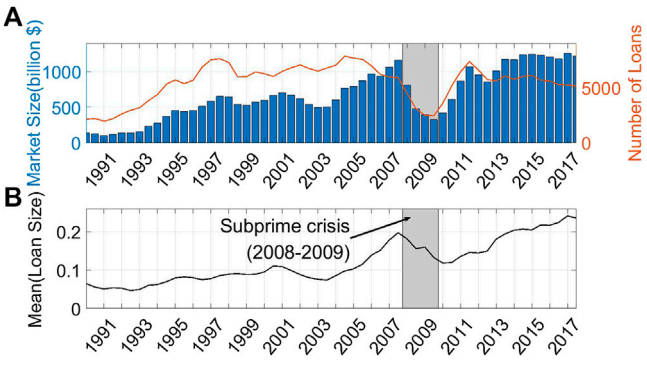


FIGURE 1 | This figure is related to the syndicated loan market in the United States from 1990 to 2017. **(A)** describes market size and the number of loans extended by lead arrangers to borrowers every quarter. Market size is defined as the sum of the loan amounts extended by each bank. The number of loans is defined as the total number of loans extended during each quarter. **(B)** represents the average loan size, which is the market size divided by the number of loans during each quarter. Gray shadows represent recessions as measured as the subprime morgage crisis periods during 2008-2009.

information by merging data derived from Standard & Poor's Compustat and from Thomson Reuters' LPC Dealscan from 1990 to 2017. The Compustat database is free of survival bias, as it contains the monthly historical accounting data of borrowing companies, and data regarding syndicated loans are included in the Dealscan database. Our starting points are the DealScan-Compustat Link [27] and the Lender link [28].[a]

Syndicated loans play a crucial role in the American corporate loan market. These loans are typically offered by a group of lenders. The lenders in a syndicate are large banks that fall into two categories of lenders: lead arrangers and participants. In this study, following the work of [5]; we classify lenders as lender-to-lead arrangers and participants. We designate a lender as a lead arranger if lead arranger credit of it is yes or lender role of it is administrative agent, agent, arranger, book runner, coordinating arranger, lead bank, lead manager, mandated arranger, or mandated lead arranger. We designate a lender as a participant if it is not the lead arranger. We refer to lead arrangers as banks from now on, but we do not refer to participants in this way. Following the literature, we exclude loans made to financial companies (i.e., SIC codes between 6,000 and 6,999) as well as classified companies belonging to the Fama-French 12th industrial classification (i.e., others).[b]

The use of syndicated loan data allows us to explore the activities of the financial intermediaries in the loan market. Our loan data, with 52,685 facilities and 35,632 packages, comprises a complex structure. After excluding banks with negative total assets, the study sample is composed of banks listed in the United States during the period 1990–2017.

3.2. Sample Characteristics

Table 2 summarizes the composition of the sample in terms of diversification, centrality indices, and the control variables described in **Section 2.2**. The correlation coefficients of the variables are reported at the lead-arranger level. Our sample is consisted of 33,386 matched lead arranger-month sets drawn from U.S. institutions heavily invested in the U.S. syndicated loan market. Diversification (DIV) is highly correlated with the composite centrality index (CCI) (0.62) in **Table 2** and **Figure 1**. In terms of multicollinearity, we control the effect of dummy variables related to 2008–2009 financial crisis in the centrality variables.

4. EMPIRICAL RESULTS

In this section, we first empirically explore the degree distribution of the PMFG network in the U.S. syndicated loan market. We then examine the ways in which network topology and investment characteristics impact bank performance. We investigate the effect of

[a]The lenders in our sample have at least $10 billion in outstanding loans or at least 50 outstanding loans, following [28].

[b]We downloaded the 12 classification data at Fama-French website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html).

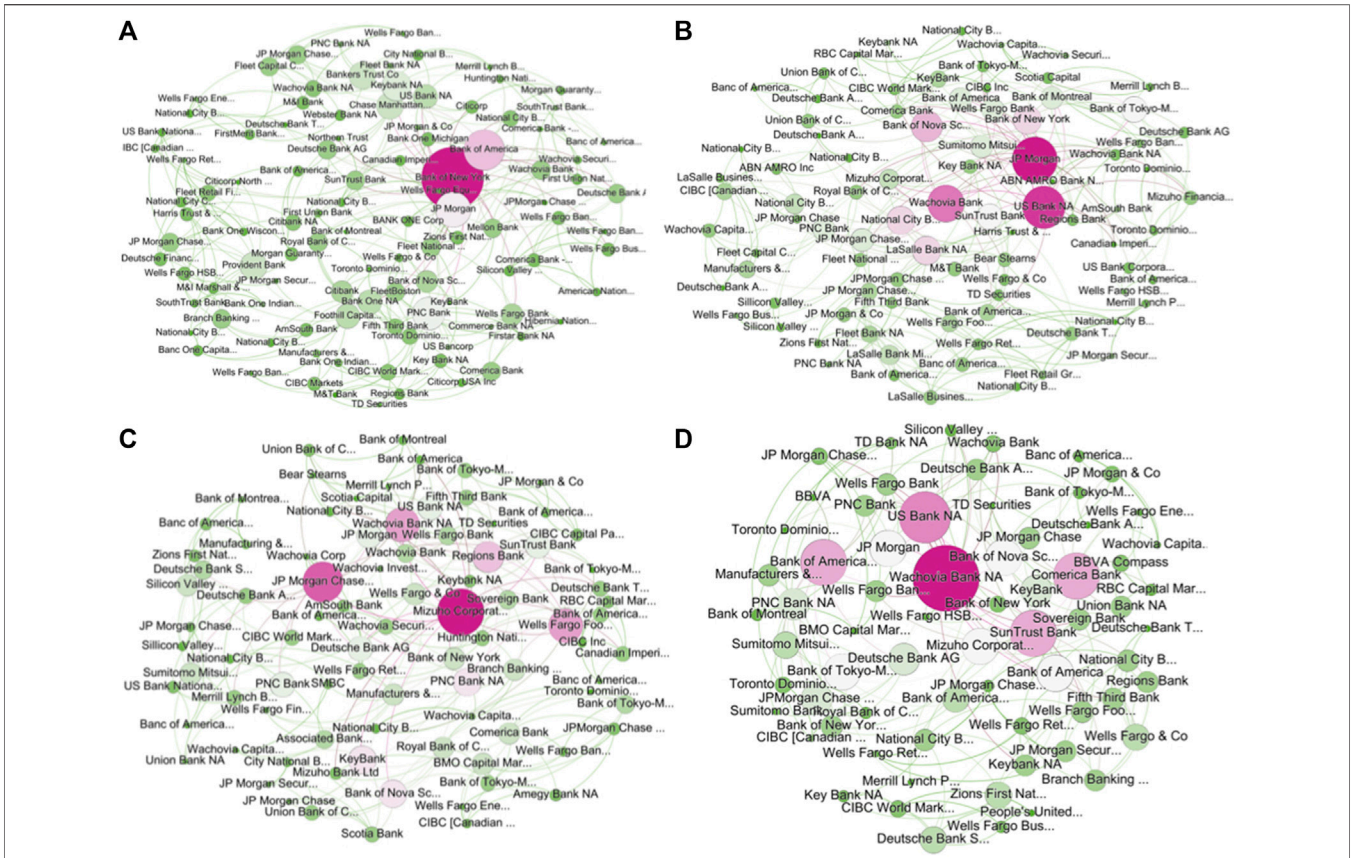


FIGURE 2 | PMFG network **(A)** 2002 **(B)** 2006 **(C)** 2008 **(D)** 2010. The nodes represent each bank, and the node size is determined by the corresponding bank's degree centrality. A node with a higher degree centrality is colored pink and one with a lower degree centrality is colored light green.

bank network centrality on bank performance because of the importance of the bank-firm lending structure in terms of information asymmetry. The structure of an interbank network should affect bank performance. Interbank networks, which are created by the degree of information asymmetry during the bank-firm lending process, should affect the performance of lending banks. A bank with a higher level of information asymmetry might mimic the loan portfolio structure of a bank with a lower level of information asymmetry to reduce this asymmetry and generate profits. The systemic risk research has identified network connectivity and centrality as channels that transmit contagions related to negative events [1, 2, 5, 29]. This implies that a highly interconnected structure can increase systemic risk. Ultimately, increased connectivity and rapid propagation in bank-to-bank networks can allow high-centrality banks to address market instability. In summary, we expect that well-connected banks should experience lower levels of information asymmetry than do poorly connected banks and that they should also experience higher levels of market performance.

4.1. The Analysis of Interbank Network

Since the amount of syndicated loans is related to exposure to assets, a decline in asset prices should affect the stability of the banking system. We analyze syndicated loans issued during each quarter from 1990 to 2017. A visual inspection of the amount of syndicated loans over time suggests that this figure reflects the state of the financial market. **Figure 1A** shows the amount of syndicated loans as a measure of overall banking loans and the number of syndicated loans. We measure the total amount of syndicated loans in each quarter. First, we find that both the overall amount and the number of syndicated loans follow a similar pattern. The total amount of syndicated loans started to increase in 2003 and continued to rise until Q4 of 2007, finally decreasing in 2009. After the subprime crisis, these loans rapidly increased until 2012. Second, the mean amount of syndicated loans is calculated as follows: Mean (Loan) = Market size/number of loans. **Figure 1A** shows a pattern similar to that of the results in **Figure 1A**.

The main goal of this paper is to conduct more rigorous tests on the relationship between the interconnectivity of banks and bank performance. To test the validity of our hypothesis, we construct an interbank network using the PMFG method developed by [17] based on loan portfolio data in **Figure 2**. In January 2002 (2006), this interbank network for the normal market status consisted of 513 (428) connections and 105 (88) nodes. The interbank network during and after the financial market crisis consisted of 423 (328) connections and 87 (68) nodes in January 2008 (2010). If the loan portfolio of each bank tended to have a distinct and unique investment strategy, then the

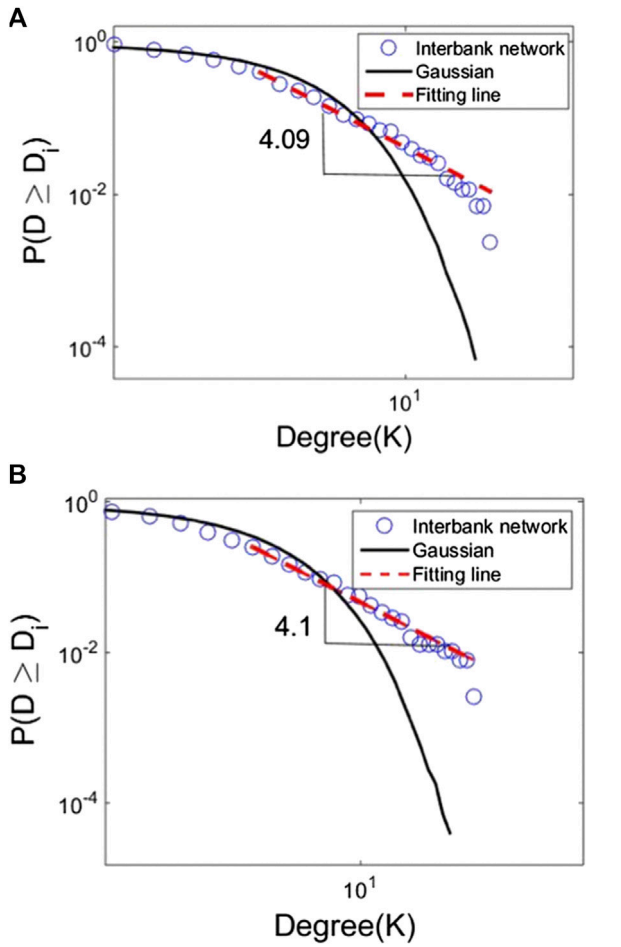


FIGURE 3 | The CDF for the degree of the interbank network is plotted with a double logarithmic scale. The cumulative distribution function for the degree of network during four years **(A)** from 2006 to 2009 and **(B)** from 2010 to 2013, the Gaussian distribution, and the fitted line are denoted using dotted blue lines, a black line, and dashed red lines, respectively.

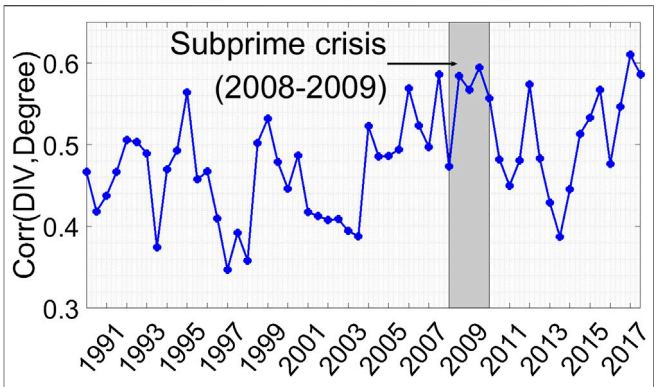


FIGURE 4 | This figure shows the correlation between diversification (DIV) and the degree of the PMFG network during the sample period of six months. Gray shadows represent recessions as measured as the subprime mortgage crisis periods during 2008-2009.

interbank network would be disconnected, and each bank would correspond to a random network. We construct interbank networks for normal and abnormal periods based on the banks' loan portfolio structures to test whether the characteristics of the network are related to the market status. The obtained interbank network, shown in **Figure 2 A–D**, displays the banks with higher connections between banks, regardless of market status, suggesting that the syndicated loan portfolios of banks are shared with other banks.

The degree (k) distribution of the interbank network indicates that most of the banks are linked to a few other banks, whereas a few banks with a large amount of capital are connected to many individual banks. As shown in **Figure 3**, the degree distribution in 2006 (2010) follows the power-law distribution with an exponent of 4.09 (4.1). Consistent with [30, 31]; **Table 3** compiles the results of the likelihood ratio test and includes judgments supported by statistical methods for the power-law hypothesis for each distribution over four years. We find that the degree distributions follow a power-law when comparing to exponential, stretched exponential, power law with cutoff, and log normal distributions. The power-law exponents of degree distributions of PMFG network are in the range 3.49 and 4.43. As a result, we think that there are the influential banks with a lot of connections in the interbank network.

The diversification of loan portfolios has important implications of the role that banks' investment strategies play in the syndicated loan market. Is this loan portfolio strategy, i.e., the diversification of syndicated loans at the industry level, related to the interbank network? We estimate the correlation between the diversification of portfolios and network structure to test whether the investment strategy of a bank is related to the other banks in the network. **Figure 4** shows the correlation between diversification and the degree of network centrality for each year. Overall, there is a positive correlation between diversification and degree of centrality, regardless of the subperiod observed. In particular, the correlation value starts to increase in 2002 and continues to rise until 2007 before the subprime crisis; after this, it decreases in 2011, suggesting that the correlation between the loan portfolio strategies of banks and the centrality of the network connectivity among banks should be understood as indicators of the financial crisis.

To observe the relationship between the degree of network centrality and portfolio strategies, we divided the whole sample into three groups according to centrality: G (high), G (middle), and G (low). **Figure 5** displays the distribution function of these three groups using box plots and calculates the similarity of each distribution function using the Kolmogorov-Smirnov test (K-S test) [32]. The results are reported in **Table 3**. In addition, we calculate the average diversification of the three groups over time. **Figure 6** shows the time evolution of the average diversification of these three groups defined according to their degrees of network centrality from January 1990 to December 2017. The diversification of the three groups is calculated based on the loan portfolios using the entropy method. The red circles, blue diamonds, and black triangles indicate the high-, middle-,

TABLE 3 | Comparisons of the fitted power-law behavior to alternatives.

			Exponential		Stretched exp		Power law + cutoff		Log-normal	
Year	**Power law *p***	**Est. *α***	**LR**	***p***	**LR**	***p***	**LR**	***p***	**LR**	***p***
1990	0.43	4.23	2.32	0.02	1.03	0.30	4.44	0.00	−0.14	0.89
1994	0.34	4.43	2.74	0.01	1.22	0.22	5.85	0.00	−0.47	0.64
1998	0.10	3.76	4.29	0.00	0.87	0.06	8.60	0.00	−0.02	0.99
2002	0.02	3.57	2.41	0.02	1.74	0.08	5.77	0.00	0.29	0.77
2006	0.41	4.10	3.27	0.00	2.13	0.03	5.94	0.00	0.75	0.45
2010	0.30	4.11	1.86	0.06	1.36	0.17	4.85	0.00	−0.31	0.76
2014	0.22	3.49	1.50	0.13	1.17	0.24	4.49	0.00	0.05	0.96

We checked the power law test proposed by [30, 31] for the degree distribution of PMFG networks during four years. Statistically significant values are given in bold. Estimated α is the power-law exponent of the degree distribution.

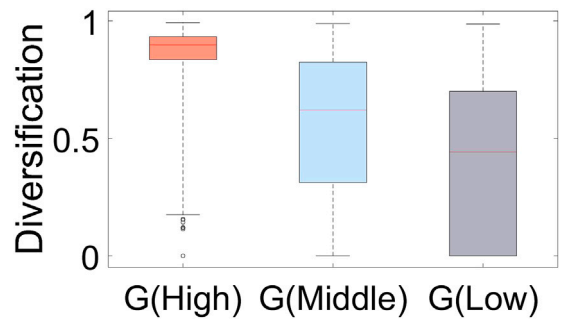


FIGURE 5 | We divide banks into three groups: high, middle, and low-centrality. The banks corresponding to the highest (lowest) 10% in terms of degree centrality are designated as the core (peripheral) of banks in this paper. The core banks have higher levels of diversification than middle and low-centrality groups.

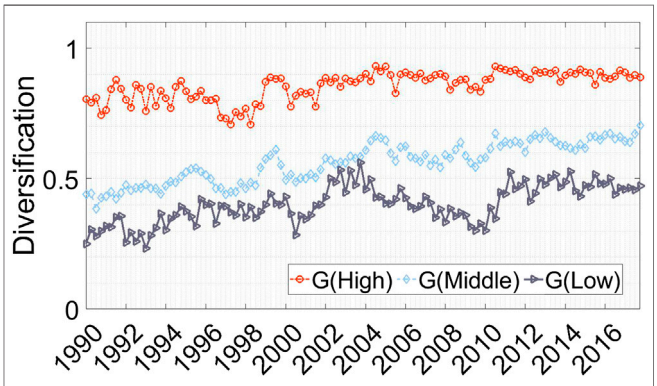


FIGURE 6 | Time series of diversification of three groups according to their degree centrality. This figure shows the time series of the monthly diversification of syndicated loan portfolios from January 1990 to December 2017. The diversification of the three groups is computed by using the entropy method based on their loan portfolios. We divided sample into three groups. The red circles, blue diamonds, and black triangles indicate the high, middle, and low-centrality groups, respectively.

and low-centrality groups, respectively. As shown in **Figure 6**, we find that since 2004, the diversification levels of low-centrality groups have moved more volatile than high-centrality groups.

4.2. The Effect of Centrality and Diversification on Bank Performance

To the extent that interbank networks in the United States have heterogeneous characteristics, we suggest that the strategic behaviors of banks and the central characteristics of banks have impacts on performance. We focus on two properties of banks: structural properties and strategic properties. We use the four measures of centrality as structural properties in the PMFG network. The relationships between lenders and borrowers are likely to mitigate the problem of information asymmetry because lending banks collect a considerable amount of information about the corporate management of their borrowers and have stable and long-term relationships with the managers of these organizations [33]. Sometimes, banks place their directors on borrower's boards of directors to improve the quantity and quality of information regarding operations that they receive [25]. We found that capitalized banks tend to centralize their networks. Therefore, we assume that banks with a high level of centrality in their networks have the unique abilities of quickly obtaining resources through the members of their network and of reducing the level of information asymmetry between lenders and borrowers.

Based on our assumption, centralized banks would feel more secure when expanding their business. In this context, we would expect to see that these banks hold portfolios that are more diverse. Diversification in the syndicated loan market creates the potential advantage of reducing credit risk exposure [5]. Banks become more resilient to common shocks such as exposure to risk when holding diversified portfolios. We estimate the following regression with pooled data:

$$ROA_{i,t} = \alpha + \beta_1 DIV_{i,t} + \beta_2 Centrality_{i,t} + \beta_3 Centrality_{i,t} \times Dummy + \beta_4 Marketsize_t + \beta_5 Marketshare_{i,t} + \beta_6 Banksize_{i,t} + e_{i,t}, \quad (3)$$

where the dependent variable $ROA_{i,t}$ is a financial indicator of profitability during month t $DIV_{i,t}$ measures the diversification of bank i based on its syndicated loan portfolio during the twelve months prior to month t and dummy as an indicator

TABLE 4 | Dimensions of connectedness and likelihood of performance.

Variable	(1)		(2)		(3)		(4)	
Intercept	-7.2×10^{10}	(−0.1475)	-1.85×10^{11}	(−0.3832)	-1.1×10^{11}	(−0.2178)	-6.7×10^{10}	(−0.1377)
DEGREE	0.3970***	(9.3934)						
DEGREE × Dummy	−1.1628***	(−9.1937)						
EIGEN			22.2780***	(10.6610)				
EIGEN × Dummy			−50.1001***	(−10.2535)				
CLOSE					0.3078***	(9.8138)		
CLOSE × Dummy					−0.8945***	(−10.9380)		
BTWN							0.0853***	(8.7560)
BTWN × Dummy							−0.1878***	(−5.1339)
DIV	0.0339***	(4.9161)	0.0196***	(2.7594)	0.0094	(1.2388)	0.0338***	(4.9143)
Market size	0.0920***	(4.8797)	0.0979***	(5.1879)	0.0927***	(4.9039)	0.085***	(4.5681)
Market share	−0.0193***	(−7.8407)	−0.0193***	(7.9123)	−0.0167***	(−6.9458)	−0.0180***	(−7.3890)
Bank size	−0.0622***	(−29.1980)	−0.0628***	(−29.5153)	−0.0623***	(−29.3096)	−0.0623***	(−29.2473)
Observations	33,289		33,289		33,289		33,289	
Year FEs	Yes		Yes		Yes		Yes	
Adj.R^2	0.2256		0.2263		0.2244		0.2266	

This table reports the regressions of four dimension of connectedness and diversification on ROA: degree centrality (DEGREE), eigenvector centrality (EIGEN), closeness centrality (CLOSE), and betweenness centrality (BTWN). ROA is defined as the net income divided by total assets. Consistent with ***Section 2.2****, the centrality indices of the banks are measured for each month. Diversification (DIV) is measured by the Shannon entropy of bank portfolio calculated as the amount of the loans extended to ten industries by each bank. Market size is defined as the log of the sum of all outstanding loans. Market share is defined as the log of the amount of loans extended by each bank. Bank size is defined by log of total assets of each bank. Year fixed effects are included to account for time characteristics. The t-statistic is reported in brackets. The symbols *, **, and *** indicate statistical significance at 10%, 5%, and, 1%, respectively.*

variable as follow: Dummy is one if the observation is from financial crisis period, otherwise 0. As a proxy for structural importance in the PMFG network, $centrality_{i,t}$ is replaced by four representative types of centrality: degree centrality, eigenvector centrality, closeness centrality, and betweennes centrality.

By including the variables market size, market share, and bank size in this regression, we control for the systematic and idiosyncratic effects that we cannot directly observe. Market share is measured by the natural logarithm of the amount of outstanding loans held by each bank [34]. use that as a proxy for a lead arranger's reputation in terms of market participants' perceptions of its screening and monitoring of borrowers. We control for market share to identify the effects of banks' reputations. Market size is calculated as the natural logarithm of the sum of the loan amounts of newly originated syndicated loans in billions of United States. dollars. Controlling high performance of bank with higher asset, bank size is estimated by the natural logarithm of total assets of each bank. In all regressions, we include market size and year fixed effects to remove the time characteristics.

We report the results related to diversification and four centrality measures of the interbank networks. In all models, the regression coefficients of the measures of diversification are statistically highly significant, and they indicate a positive relationship $(0.3970, p < 0.01; 22.2780, p < 0.01; 0.3078, p < 0.01; 0.0853, p < 0.01)$ in **Table 4**. These findings are in line with the results of the descriptive studies by [35]; which report that product-diversified firms have high levels of performance and innovation. There are simply too many results and perspectives about the agency theory of diversification to include them in this paper. Our results support the existing evidence regarding diversification and profitability in terms of lead arrangers' loan portfolios. Each type of centrality represents a different aspect of a bank's structural position in the network. These findings allow us to determine whether each type of centrality is able to represent a factor of composite centrality index (CCI) in **Table 1**. Overall, our results suggest that higher levels of the individual dimensions of centrality based on loan portfolio similarities are related to increases in the profitability of banks.

Next, we estabilish dummy variable with centrality indices to exclude the financial crisis effect in 2008–2009. They are statistically significant with negative coefficients of DEGREE × Dummy, EIGEN × Dummy, CLOSE × Dummy, BTWN × Dummy $(-1.1628, p < 0.01; -50.1001, p < 0.01; -0.8945, p < 0.01; -0.1878, p < 0.01)$ in **Table 4**. As shown in columns 1–4 of **Table 4**, although the dummy variable has a negative sign, the main effect for the dimension of centrality and diversification is positive and significant. It means that the impact of network centrality on performance is negative during 2008–2009 financial crisis and positive during the normal period. We then show the results of the regression using our composite centrality index (CCI) through principal component

TABLE 5 | The relation of the diversification of the subsets of banks to degree centrality.

	High	Middle	Low
High	1	0.7906*** ($7.82E^{-51}$)	0.5488*** ($7.82E^{-51}$)
Middle		1	0.7072*** ($7.70E^{-22}$)
Low			1

The table represents the Pearson correlation among the three groups of banks. We construct two groups from the sample bank. One is the core as designated by High with the highest 10% degree centrality and another is the peripheral as designated by Low with the lowest 10% degree centrality. The other group of banks is the Middle in table. A two-sample Kolmogorov-Smirnov test asymptotic significance value (2-tailed) is shown in the bracket. (P < 0.01) rejects the null hypothesis of the other population distributions.

TABLE 6 | The effect of diversification on the bank performance of core and peripheral banks.

Variables	Core		Peripheral	
Intercept	-1.1×10^{12}	(−0.5745)	-2.1×10^{12}	(01175)
DIV	−0.0635*	(−1.7079)	00,651***	(52336)
Market size	00735	(13409)	01548***	(40456)
Market share	00372**	(37565)	−0.0380***	(−8.1899)
Bank size	−0.1200***	(−15.6491)	−0.0650***	(−14.8352)
Observations	4,241		7,330	
Year FEs	Yes		Yes	
Adj. R^2	02315		02427	

We investigate the effect of diversification on ROA for (1) the core of banks and (2) peripheral of banks. ROA is defined as net income divided by total assets. Diversification (DIV) is measured by the Shannon entropy of bank portfolio calculated as the amount of loans extended to ten industries by each bank. The control variable is consistent with ***Eq. 3****. Year fixed effects are included to account for time characteristics. The t-statistic is reported in brackets. The symbols *, **, and *** indicate statistical significance at 10%, 5%, and, 1%, respectively.*

analysis, including degree centrality, eigenvector centrality, closeness centrality, and betweenness centrality based on the results shown in **Table 4**. The results of the regression including CCI are reported in **Table 1** using equation model 3. Consistent with the preceding regressions, we use the dummy variable with CCI to remove the recession trends. We find a negative and significant coefficient for the CCI × Dummy $(-0.0170, p < 0.01)$, whereas the coefficients of CCI and DIV are positive and significant $(0.0084, p < 0.01; 0.0216; p < 0.01)$, consistent with the results in **Table 1**. Together, these results suggest that overall centrality consistently moderates the increase in a bank's profitability when it holds a diversified portfolio.

4.3. The Effect of Diversification on Bank Performance According to the Level of Centrality

In this section, we examine the different ways in which the structural importance of the PMFG network affects bank's strategic actions. We also consider the way in which the relationship between strategic actions and relative profitability identified in the full sample may vary based on banks' degree of centrality. Several papers have highlighted the likelihood that board interlocking between banks has more power and information in the market when they reduce financial risk [22, 34, 36, 37]. Because the importance of each bank in the network is not homogeneous, we group the banks by their degrees centrality into groups consisting of core banks and of peripheral banks. We designated the upper (lower) 10% of banks in terms of degree centrality as high (low) groups to define the cores and peripheral in the PMFG network. **Table 5** represents the Pearson correlation of diversification between each subset of banks. The high- and middle-centrality groups have positive correlations (0.7906), and the low-centrality groups also have positive correlations with the other groups $(0.5488, 0.7072)$. Additionally, we investigate a two-sample Kolmogorov-Smirnov test to assess the distribution of the two samples in brackets. This test implies a heterogeneous distribution of diversification among the three groups of banks. As a result, we conclude that the three groups classified by degree centrality could have investment strategies with differing characteristics. Our interpretation is consistent with the results in **Figure 5** and **Figure 6**. Specifically, we run the following regression on two sets of banks; core and peripheral.

$$ROA_{i,t} = \alpha + \beta_1 DIV_{i,t} + \beta_2 Marketsize_t + \beta_3 Marketshare_{i,t} + \beta_4 Banksize_{i,t} + e_{i,t}, \quad (4)$$

Table 6 shows the results of the linear regressions regarding bank diversification using the same explanatory variables we used for the subset of banks. These results indicate that core banks could obtain better private information than peripheral banks. This result is consistent with the study of [14]; who insist that concentrated lenders had higher profits than diversified lenders during the financial crisis. Additionally [38], find that the diversification of bank assets is not guaranteed to produce superior return performances or greater safety for banks. These findings are different from the comprehensive perspectives of the market power view and the resource view in terms of profit maximization. Note, however, that these studies do not control for network centrality. Consistent with the systemic risk literature [5], we consider core banks to have high levels of risk exposure, and concentrated lenders have high levels of performance during our sample periods $(-0.0635, p < 0.1)$. As shown in column 2 of **Table 6**, the group composed of peripheral banks has a statistically significant positive effect on performance $(0.0651, p < 0.01)$. This means that the subsets of banks in the interbank network reflect the different risk cultures among banks.

5. CONCLUSION

Banks that are centrally located in a syndicated loan network have access to better information and more influence in the syndicated loan market. Adding to the previous studies on the role of network centrality among banks, we employ a network centrality measure to test the connection between bank performance and network structure. In terms of the diversification of loan portfolios, we show that banks with higher levels of network centrality are more likely to pursue diversification, and that this diversification is more likely to

increase during periods of market instability. The evidence shows that sample banks' lending strategies exhibited a significant relationship with these banks' degrees of network centrality, regardless of the market status. We further find that banks with a high level of centrality have higher returns than banks with a low level of centrality. We then test whether the diversification of the syndicated loan portfolios of individual banks is related to the performance of these banks according to their centrality position in the interbank network. Since a bank's centrality in the network plays an important role in its loan portfolio strategy, this centrality also plays a significant role for lending market participants. We found that in the core group, diversification showed a negative correlation with centrality; however, a positive relation was observed in the peripheral group.

We contribute to the literature on the bank-firm lending process in the field of finance by introducing the interbank network based on the syndicated loan market. Our findings extend the existing literature on the lending mechanisms between banks and firms and show that banks' centrality within the interbank network influences their portfolios in the syndicated loan market. Future studies can help to shed light on bank performance and lending mechanisms.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work has been supported by NRF (National Research Foundation of Korea) Grant funded by the Korean Government (NRF-2017R1A2B4009349).

REFERENCES

1. Battiston S, Puliga M, Kaushik R, Tasca P, Caldarelli G. Debtrank: too central to fail? financial networks, the fed and systemic risk. *Sci Rep* (2012) 2:541. doi:10.1038/srep00541
2. Demirer M, Diebold FX, Liu L, Yilmaz K. Estimating global bank network connectedness. *J Appl Econom* (2018) 33:1–15. doi:10.1002/jae.2585
3. Corsi F, Lillo F, Pirino D, Trapin L. Measuring the propagation of financial distress with granger-causality tail risk networks. *J Financ Stabil* (2018) 38: 18–36. doi:10.1016/j.jfs.2018.06.003
4. Acharya VV, Pedersen LH, Philippon T, Richardson M. Measuring systemic risk. *Rev Financ Stud* (2017) 30:2–47. doi:10.1093/rfs/hhw088
5. Cai J, Eidam F, Saunders A, Steffen S. Syndication, interconnectedness, and systemic risk. *J Financ Stabil* (2018) 34:105–20. doi:10.1016/j.jfs.2017.12.005
6. Ivashina V, Scharfstein D. Bank lending during the financial crisis of 2008. *J Financ Econ* (2010) 97:319–38. doi:10.2139/ssrn.1297337
7. Fahlenbrach R, Prilmeier R, Stulz RM. This time is the same: using bank performance in 1998 to explain bank performance during the recent financial crisis. *J Finance* (2012) 67:2139–85. doi:10.3386/w17038
8. Rajpal H, Rosas F, Jensen HJ. Tangled worldview model of opinion dynamics. *Frontiers in Physics* (2019) 7:163. doi:10.3389/fphy.2019.00163
9. Tomassini M, Antonioni A. Public goods games on coevolving social network models. *FrP* (2020) 8:58. doi:10.1063/1.4991679
10. Wen S, Tan Y, Li M, Deng Y, Huang C. Analysis of global remittance based on complex networks. *Front Phys* (2020) 8:85. doi:10.3389/fphy.2020.00085
11. Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A. The network origins of aggregate fluctuations. *Econometrica* (2012) 80:1977–2016. doi:10.3982/ECTA9623
12. Shleifer A, Vishny R. Fire sales in finance and macroeconomics. *J Econ Perspect* (2011) 25:29–48. doi:10.1257/jep.25.1.29
13. Cont R, Moussa A, Santos EB. (2010). Network structure and systemic risk in banking systems. Available at SSRN: https://ssrn.com/abstract=1733528 (Accessed December 1 2010).
14. Loutskina E, Strahan PE. Informed and uninformed investment in housing: the downside of diversification. *Rev Financ Stud* (2011) 24:1447–80. doi:10.1093/rfs/hhq142
15. Matvos G, Seru A, Silva RC. Financial market frictions and diversification. *J Financ Econ* (2018) 127:21–50. doi:10.1016/j.jfineco.2017.09.006
16. Blau P. *Exchange and power in social life (Routledge)*. London and New York: Taylor & Francis (2017). p. 372.
17. Tumminello M, Aste T, Di Matteo T, Mantegna RN. A tool for filtering information in complex systems. *Proc Natl Acad Sci USA* (2005) 102:10421–6. doi:10.1073/pnas.0500298102
18. Onnela J-P, Kaski K, Kertész J. Clustering and information in correlation based financial networks. *Eur Phys J B* (2004) 38:353–62. doi:10.1140/epjb/e2004-00128-7
19. Chi KT, Liu J, Lau FC. A network perspective of the stock market. *J Empir Finance* (2010) 17:659–67. doi:10.1016/j.jempfin.2010.04.008
20. Onnela JP, Chakraborti A, Kaski K, Kertész J, Kanto A. Dynamics of market correlations: taxonomy and portfolio analysis. *Phys Rev E–Stat Nonlinear Soft Matter Phys* (2003) 68, 056110. doi:10.1103/PhysRevE.68.056110
21. Shannon CE. A mathematical theory of communication. *Bell Sys Tech J* (1948) 27:379–423. doi:10.1002/j.1538-7305.1948.tb01338.x
22. Newman ME. The structure and function of complex networks. *SIAM Review* (2003) 45:167–256. doi:10.1137/S003614450342480
23. Wasserman S, Faust K. *Social network analysis: methods and applications*, Vol. 8. Cambridge, United Kingdom: Cambridge University Press (1994) p. 825.
24. Omer TC, Shelley MK, Tice FM. Do well-connected directors affect firm value?. *J Appl Finance* (2014) 24:17–32. doi:10.2139/ssrn.2167354
25. Omer TC, Shelley MK, Tice FM. Do director networks matter for financial reporting quality? Evidence from audit committee connectedness and restatements. *Manag Sci* (2019) 66:3361–88. doi:10.1287/mnsc.2019.3331
26. Larcker DF, So EC, Wang CC. Boardroom centrality and firm performance. *J Account Econ* (2013) 55:225–50. doi:10.1016/j.jbankfin.2018.11.002
27. Chava S, Roberts MR. How does financing impact investment? the role of debt covenants. *J Finance* (2008) 63:2085–121. doi:10.1111/j.1540-6261.2008.01391.x
28. Schwert M. Bank capital and lending relationships. *J Finance* (2018) 73: 787–830. doi:10.1111/jofi.12604
29. Elliott M, Golub B, Jackson MO. Financial networks and contagion. *Am Econ Rev* (2014) 104:3115–53. 10.1257/aer.104.10.3115
30. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. *SIAM Review* (2009) 51:661–703. doi:10.1137/070710111
31. Virkar Y, Clauset A. Power-law distributions in binned empirical data. *Ann Appl Stat* (2014) 89–119. doi:10.1214/13-AOAS710
32. Chakravarti I, Laha R, Roy J. Kolmogorov-smirnov (ks) test. *Handbook Methods Appli Statis* (1967) 1:392–4.
33. Sufi A. Information asymmetry and financing arrangements: evidence from syndicated loans. *J Finance* (2007) 62:629–68. doi:10.1111/j.1540-6261.2007.01219.x
34. Gopalan R, Nanda V, Yerramilli V. Does poor performance damage the reputation of financial intermediaries? Evidence from the loan syndication market. *J Finance* (2011) 66:2083–120. 10.2307/41305185

35. Hitt MA, Hoskisson RE, Kim H. International diversification: effects on innovation and firm performance in product-diversified firms. *Acad Manag J* (1997) 40:767–98. doi:10.5465/256948
36. Mariolis P. Interlocking directorates and control of corporations: the theory of bank control. *Soc Sci Q* (1975) 425–39.
37. Berger AN, Kick T, Schaeck K. Executive board composition and bank risk taking. *J Corp Finance* (2014) 28:48–65. doi:10.1016/j.jcorpfin.2013.11.006
38. Acharya VV, Hasan I, Saunders A. Should banks be diversified? evidence from individual bank loan portfolios. *J Bus* (2006) 79:1355–412. doi:10.2139/ssrn.293295

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Oh and Park. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 18 January 2021
doi: 10.3389/fphy.2020.586561

Analysis of the Global Banking Network by Random Matrix Theory

Ali Namaki[1,2], Reza Raei[1,2], Jamshid Ardalankia[3,4], Leila Hedayatifar[4,5], Ali Hosseiny[6], Emmanuel Haven[7,8] and G. Reza Jafari[6]**

[1]Department of Finance, University of Tehran, Tehran, Iran, [2]Iran Finance Association, Tehran, Iran, [3]Department of Financial Management, Shahid Beheshti University, Tehran, Iran, [4]Center for Complex Networks and Social Datascience, Department of Physics, Shahid Beheshti University, Tehran, Iran, [5]New England Complex Systems Institute, Cambridge, MA, United States, [6]Department of Physics, Shahid Beheshti University, Tehran, Iran, [7]Faculty of Business Administration, Memorial University, St. John's, NL, Canada, [8]IQSCS, Leicester, United Kingdom

Since the financial crisis of 2008, the network analysis of financial systems has attracted a lot of attention. In this paper, we analyze the global banking network via the method of Random Matrix Theory. By applying that method on a cross border lending network, it is shown that while the connectivity between different parts of the network has risen and the profile of transactions has diversified, the role of hubs remains important in the weighted perspective. The largest eigenvalue of the transaction matrix as the leading mode of the system shows sharp growth since 2002. As well, it is observed that its growth has diminished since 2008. This indicates that the crisis of 2008 has left a long-lasting footprint on the financial system. Analyzing the mean value of the participation ratio reveals the fact that the role of countries in forming small modes, has increased since 2002. In our final analysis, we provide snapshots of the hubs in the network over time. We observe that the share of countries in total transactions is not equal to their share in shaping the eigenvector of the largest eigenvalue. In 2018 for example, while the United Kingdom leads the share of transactions, it is the United States that has the largest value in the leading eigenvector. The proposed technique in the paper can be useful for analyzing different types of interaction networks between countries.

Keywords: global banking network, complex systems, random matrix theory, financial contagion, collective behavior

OPEN ACCESS

Edited by:
Anirban Chakraborti, Jawaharlal Nehru University, India

Reviewed by:
Sunil Kumar, University of Delhi, India
Suchetana Sadhukhan, National Autonomous University of Mexico, Mexico

****Correspondence:***
Ali Namaki
alinamaki@ut.ac.ir
G. Reza Jafari
g_jafari@sbu.ac.ir

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *23 July 2020*
Accepted: *26 November 2020*
Published: *18 January 2021*

Citation:
Namaki A, Raei R, Ardalankia J, Hedayatifar L, Hosseiny A, Haven E and Jafari GR (2021) Analysis of the Global Banking Network by Random Matrix Theory. Front. Phys. 8:586561. doi: 10.3389/fphy.2020.586561

1 INTRODUCTION

Since the recent global financial crisis, cross-border lending and financial contagions have gained in importance. The propagated effects [1, 2] of financial crises on political and economic systems [3, 4] are not to be underestimated. Those developments have prompted a lot of research on the systemic dependence of the international banking sector [5–13].

The field of complexity can aid in understanding better such systemic dependence [5, 14–22]. Complex networks are useful instruments for describing a large number of financial systems [23–31].

Most of the networks have different topological properties such as small-world and scale-free characteristics [24–26, 32–39].

The purpose of complexity science in finance focusses on the analysis of the structure and the dynamics of entangled systems.

Many scholars have applied complexity techniques for the analysis of financial contagion [6, 9, 10, 40–42].

Their findings suggested that the connectivity of financial institutions is the source of potential contagions.

For example, Glasserman and Young [40] reviewed the extensive literature on the network's structures and their interactions with other key variables such as leverage, size, and short-term funding. They emphasized that the network connections expand the firms' risk exposures, and through different routes, the shocks can be proliferated via contagion.

Random Matrix Theory is one of the useful methods for analyzing the behavior of complex systems [16, 43–54].

This theory was developed to describe the energy levels of quantum systems [55, 56].

It is the universality regime of the eigenvalue statistics which provides for the success factor of Random Matrix Theory [57–59]. Based on previous studies, it is shown that when the size of the matrix is very large, the eigenvalue distribution tends toward a specific distribution [59].

Random Matrix Theory has been applied to analyze the behavior of coupling matrices [16]. This technique divides the contents of the coupling matrix into noise and information parts. The noise part of the coupling matrix conforms to the Random Matrix Theory findings, and the information part deviates from them. This concept stems from the idea of solving the problem of non-stationary cross-correlation and measurement noise which result from market conditions and the finite length of time series [57, 59].

A system which can be analyzed by the complexity approach is the global banking network [60].

Minoiu and Reyes [60] have analyzed the global banking network from 1978 to 2009. They have applied network metrics such as centrality, connectivity, and clustering for analyzing financial interconnectedness. They have shown that during and after systemic banking crises (and sovereign debt crises), the connectivity drops. Also, it was shown that the 2008–2009 financial crisis provided for an unusually large perturbation to the global banking network. For more research on this, please see [61–69].

In this paper, by applying Random Matrix Theory on bilateral locational statistics data provided by the *Bank for International Settlements (BIS)* [70] from 1978 until 2019, we aim to analyze the global banking network. This data includes all 'core' countries (the qualifier 'core' is used by many researchers such as [60], for countries which regularly report their financial data to BIS).

Our paper is organized as follows. In **Section 2** we present our methods and, in **Section 3** we apply Random Matrix Theory on the global banking network and present our findings. Then, in **Section 4** we conclude.

2 METHODS

Random Matrix Theory has been presented by some scholars in nuclear physics such as Mehta [55, 56], for analyzing the energy levels of complex quantum systems. Subsequently, the method has helped to address specific issues in other fields, such as finance [45, 57–59, 71, 72].

From random matrix theory, we know that the eigenvalues–in the real matrix–which deviate from the range of the eigenvalues–in the random matrix–possess relatively more complete information from the system [51, 58, 59]. It can be shown that the majority of the eigenvalues of coupling matrices, agree with the random matrix predictions, but the largest eigenvalue has deviations from those estimations [50, 57, 58, 73]. In essence, this eigenvalue develops an energy gap that separates it from the other eigenvalues [45]. The largest eigenvalue is related to a strongly delocalized eigenvector that represents the collective evolution of the system. This is called market mode. From this perspective, the largest eigenvalue's magnitude reflects the coupling strength of the system [45].

In Random Matrix Theory, there is a parameter named *Inverse Participation Ratio* IPR [74]. Its inverse provides a measure for the number of components which significantly participate in each eigenvector. This notion shows the effect of components of each eigenvector and specifically indicates how the largest eigenvalues deviate from the bulk region which is densely occupied by eigenvalues of the random matrix. Based on previous papers [45, 75], IPR can be applied as an indicator for measuring the collective behavior of the networks. The formula of this concept is as follows:

$$IPR(k) = \sum_{l=1}^{n} \left(u_l^k\right)^4; \tag{1}$$

where $l = 1, \ldots, n$ and u_l^k is the l^{th} element of the k^{th} eigenvector (lk). To further clarify the concept, one may consider examples below:

i. In case all elements of a certain eigenvector are equal to $1/\sqrt{N}$, IPR will be equal to $1/N$. This implies that whole elements are significantly influential on the systems' behavior.
ii. On the other hand, if just a single element is equal to one and the others are equal to 0, IPR would be equal to 1. This implies that only this component is effective in the corresponding eigenvector. Hence, one can perceive that *IPR* clarifies the number of influential elements in a certain eigenvector.

3 ANALYSIS OF GLOBAL BANKING NETWORK BY RANDOM MATRIX THEORY

The banking industry is one of the most important sectors in finance. Given this importance, it is not surprising that a significant aspect of financial contagion shows that the banking network is the conduit, through which the emergence and transmission of crises occurs. In this paper, we create a weighted and directed financial transaction network corresponding to each quarter from 1978 until 2018. Each link corresponds to a loan given by a certain country to another one. Previous studies have shed light on a country's dependency

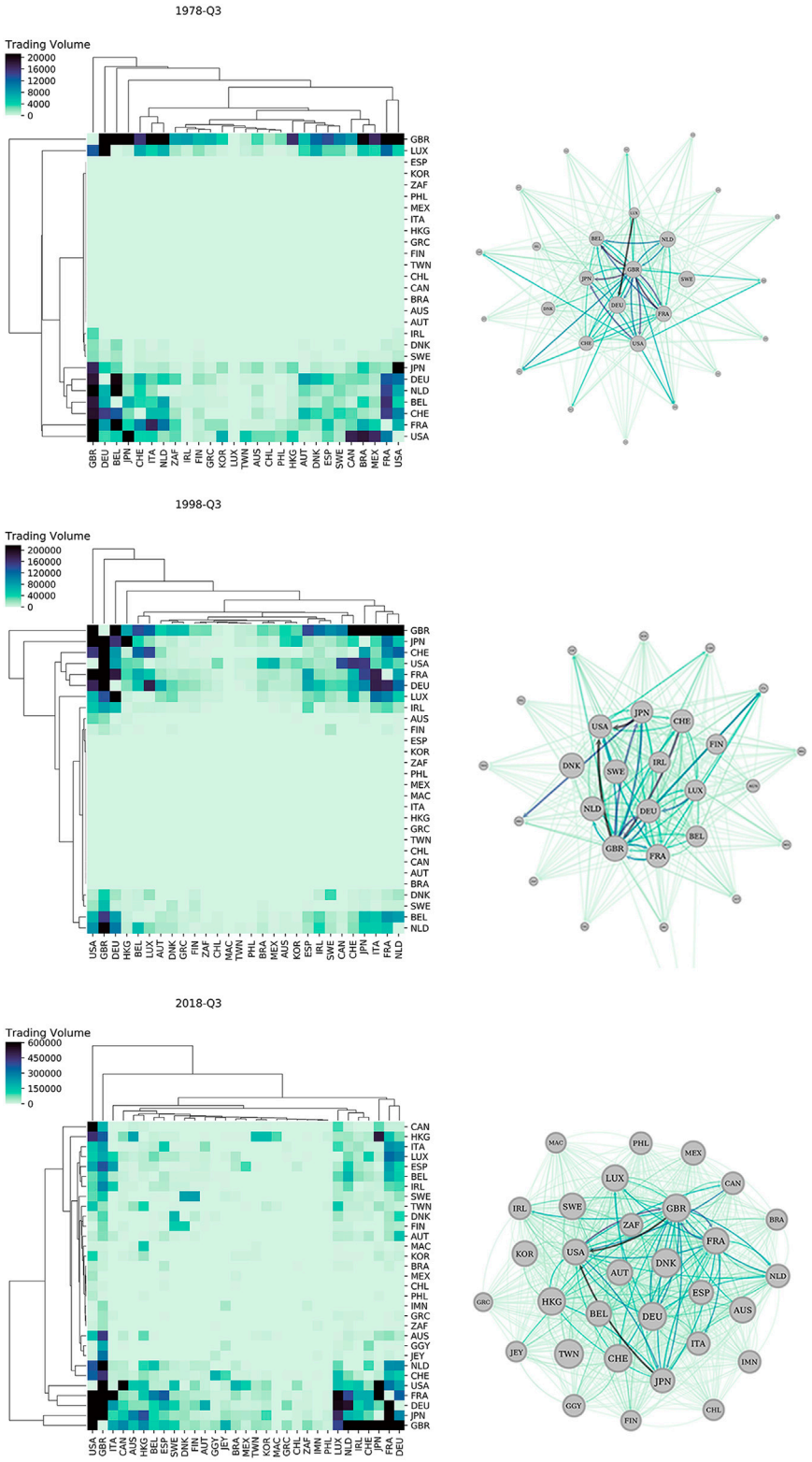


FIGURE 1 | The evolution of the global banking network is demonstrated for three snapshots of 1978-Q3, 1998-Q3, and 2018-Q3. Left) shows the dendrogram structure of communities for trading weighted matrices. Right) the network topology is graphed. The size of each node represents the degree.

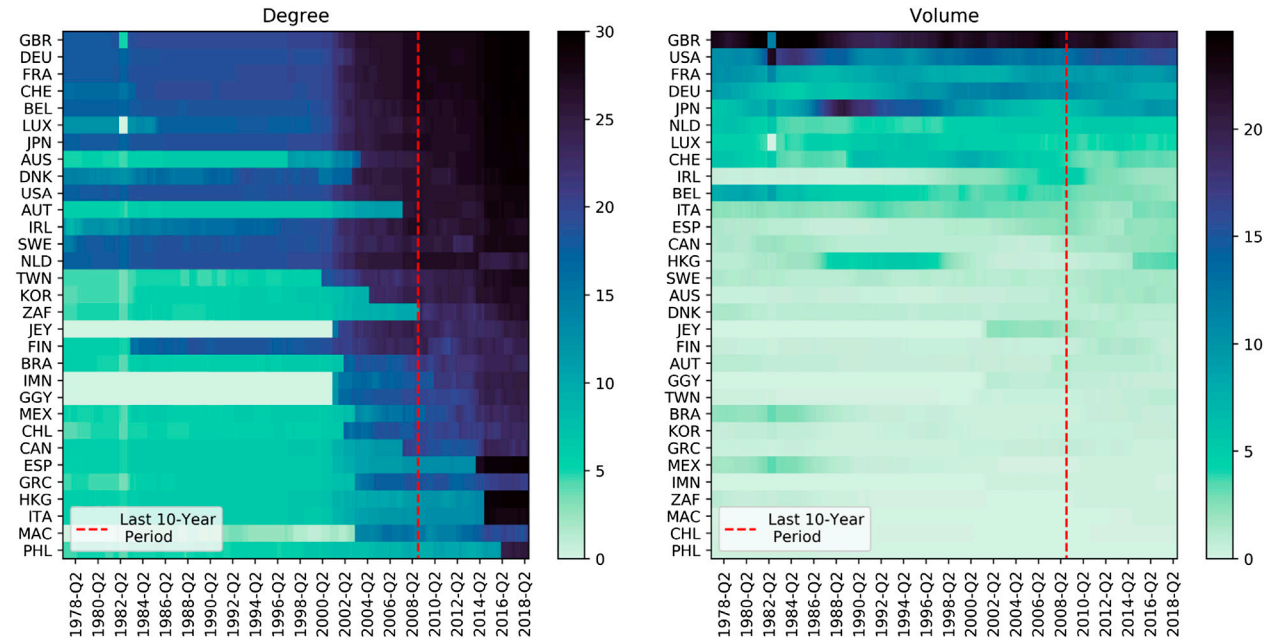


FIGURE 2 | Temporal evolution of: **(Left)** the degree, and **(Right)** the volume of each country over time. The sorting order of countries is based on the average of the last 10-year period.

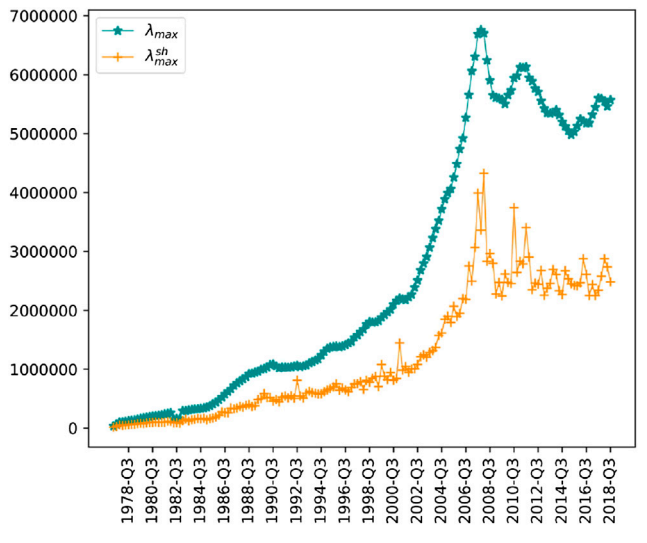


FIGURE 3 | The evolution of the largest eigenvalue, λ_{max}, of the global banking network and its shuffled, λ^{sh}_{max}, are depicted.

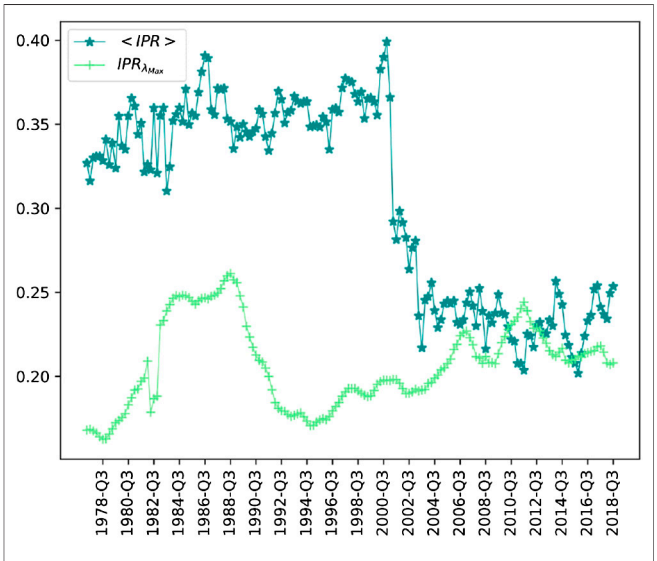


FIGURE 4 | The average of the Inverse participation ratio of all eigenvectors $\langle IPR \rangle$ and $IPR_{\lambda_{max}}$ has been depicted. The decline of the $\langle IPR \rangle$ implies that the contribution of countries in all modes has been increased which can be a consequence of the growth of the connectivity in the banking network.

network and they showed an increase, over time, of the dependency structure of the network [7, 60].

In **Figure 1**, the evolution of the global banking network in three snapshots (1978-Q3, 1998-Q3, and 2018-Q3) has been depicted. The left panel in **Figure 1** shows the dendrogram structure of communities for trading weighted matrices. Furthermore, the right panel shows the evolution, over time, of the network topology and the size of nodes stands for the degree. As depicted, not only the size of the network has grown but also transactions have become more diversified. It is obvious that, over time, the degree of all countries has grown and has become more homogeneous. If the size of degrees is considered, a few countries can be distinguished as hubs, and this will be

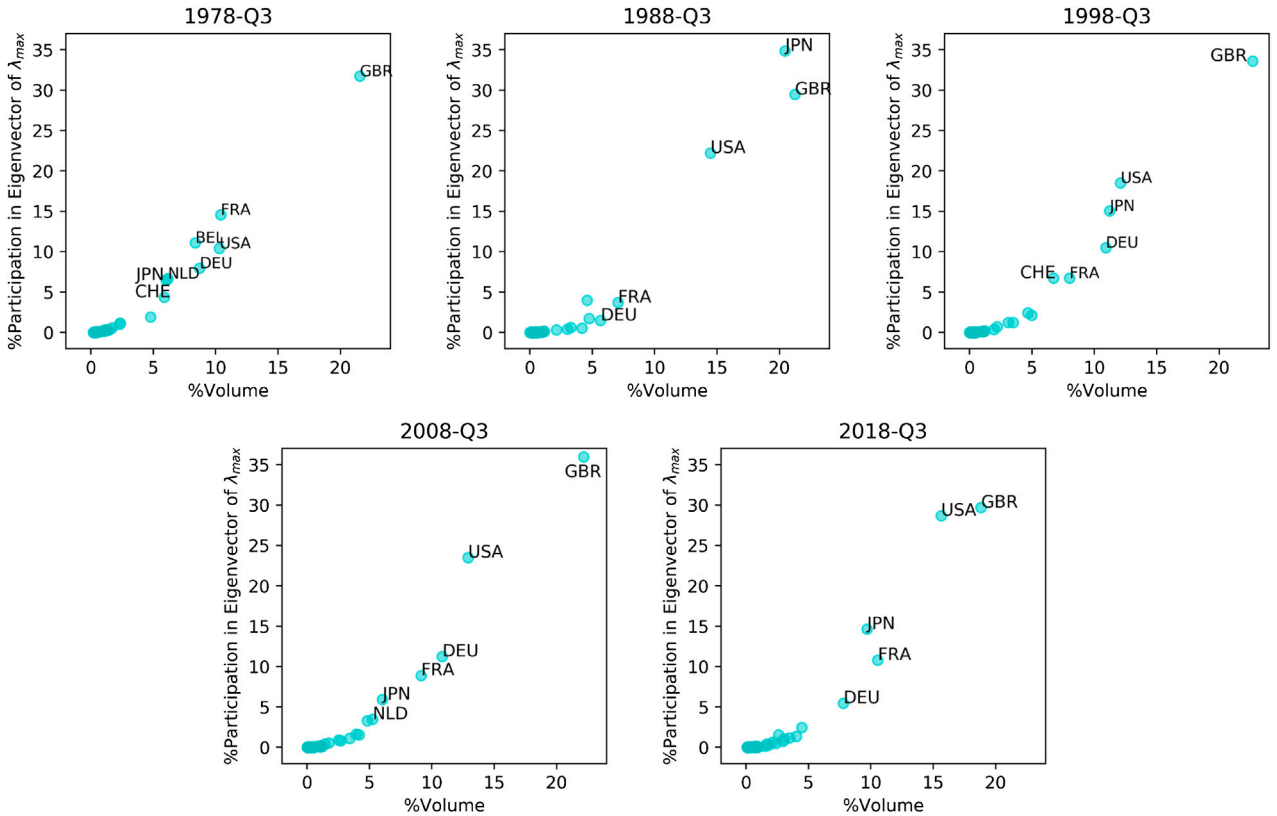


FIGURE 5 | The percentage of the participation of each country in the eigenvector of λ_{max} vs. the percentage of the transaction of each country from the total transaction ($\%volume_i/\sum_i^N Volume_i$). (Note: Countries possessing more than 5% of total volume are annotated by name.)

discussed later in the paper. For the continuous monitoring of networks during the period of study, **Figure 2** is plotted. The results show the same outcomes as **Figure 1**. The left panel of **Figure 2** represents the evolution of the degree for each country over time. As can be seen at the beginning of the period, only a small portion of countries has a high degree. But over time, both the degree across more countries and the average degree rise. It means that the sparseness has declined and connectivity has risen. The right panel of **Figure 2** shows this fact, i.e., that only a small number of countries are in charge of a big portion of transactions.

To move further into our analysis, we now apply random matrix techniques. The global banking network possesses an adjacency matrix. In random matrix theory, we have learned that the largest eigenvalue is important and addresses the global trend of a system [45, 57, 58, 76]. In **Figure 3** we have depicted the evolution of the largest eigenvalue (λ_{max}) overtime. As can be seen, λ_{max} has grown significantly before the financial crisis of 2008. To figure out whether the growth of the eigenvalue is a mere consequence of either the growth of the transaction or the change of the structure of the network, we have compared the growth of the λ_{max} of the original matrix with λ_{max} of the shuffled network. If the growth of the largest eigenvalue is a consequence of the growth of transactions, then we expect that its value will be close to its counterpart in the shuffled network. In the shuffling technique, we rewire the network. We do so as follows. Pairs of links are chosen randomly and their values are exchanged. Over the course of such a process, the information concerning the structure of the network is lost. All remains are the size of the network and the profile of transactions.

The difference between λ_{max} of the network itself and its shuffled counterpart, implies the existence of information content which is embedded in the largest eigenvalue of the banking interaction matrix. This will be discussed further below. The fast growth of the λ_{max} of the shuffled network from 2002 to 2004, is the consequence of the fast pace of transaction volumes.

The fact that the largest eigenvalues of both the banking network and the shuffled network, have lost their growth trends after 2008, means that the financial crisis has left a long-lasting footprint on the network. Since the obtained eigenvalue does not describe all the details and properties of the collective behavior, one should investigate other quantities in the network.

As already discussed in the method section, one should keep in mind that the IPR possesses the ability of information extraction from the collective the behavior of systems. **Figure 4** represents the evolution of the $\langle IPR \rangle$ and $IPR_{\lambda_{max}}$. In this context, by

focusing on the mean inverse participation ratio of all eigenvectors, $\langle IPR \rangle$, and also, the inverse participation ratio of the largest eigenvalue corresponding to the largest eigenvector, we investigate banking behaviors of countries and their influences on the network structure and the market trend.

In a network of size N, IPR could have a value within $[1/N - 1]$. Values close to the lower end will imply that almost all nodes play a role in the leading mode. Values close to one, indicate that a few nodes play an important role in shaping the eigenvector. As can be seen, for the largest eigenvalue, IPR has kept a value much higher than its possible minimum, i.e., $1/N$. This means that a few countries lead the network. Disparities have been even stronger in small modes in the early years of the studied period. However, from 2002 to 2004, following the fast growth of global transactions, the average of IPR of all modes, has tended to the IPR of the largest mode. This means that the participation of countries in shaping the small modes has grown.

The sustainability of the relatively high rate of IPR in the largest eigenvalue leads us to investigate the share of countries in shaping its eigenvector. We expect the countries which have a higher share of transactions, to play a more important role in shaping the trend of the system embedded in this eigenvector. **Figure 5** visualizes the contribution of countries in the structure of the leading mode vs. their contribution to trading volume in five snapshots since 1978. A couple of interesting results can be inferred from the figure.

In all snapshots, the share of hubs in the leading mode has been higher than their share in transactions. For example, in 1978 while the share of the United Kingdom in total transactions has been 21.5 percent, its share in the leading eigenvector has been 31.73 percent. Hence, this means that the role of the United Kingdom in shaping the leading eigenvalue has been larger than its share in total transactions. The same scenario works for other hubs such as France and the United States.

The interesting observation of 1988 is that, while the United Kingdom holds the lead in the share of transactions, Japan has the largest component of the leading eigenvector. On the eve of the economic downturn, Japan has not repeated its leading role in any other snapshots.

Within the last 2 decades, the United States has become closer to the United Kingdom in shaping the eigenvector of the largest eigenvalue. However, for both countries, their share in the largest eigenvalue is bigger than their share in the total transactions. Such an effect could be a matter of the country's role in the structure of the leading mode in the network.

4 CONCLUSION

In this paper, by applying Random Matrix Theory, the global banking network is analyzed. For this purpose, we consider the matrix of the interaction of the banking sectors of BIS countries. We first focus on the largest eigenvalue which defines the leading mode in a system. We observe that the largest eigenvalue grows over time. By making a comparison with the largest eigenvalue of the network itself and the shuffled network, we conclude that the growth of the largest eigenvalue originates from two sources. The first source is the growth of the transaction volume and the other source is the network structure. We observe that the growth of the largest eigenvalue has vanished after 2008.

By focusing on the temporal behavior of the IPR of the largest eigenvalue, we observe that it has kept a sustainable value far above its possible minimum. This emphasizes the role of a few countries as hubs in the system. In comparison, the mean value of the IPR of all eigenvectors has declined sharply after 2002. This leads us to conclude, that the contribution of countries to shape small modes and possibly local structures, has grown. This phenomenon has occurred in tandem with the fast growth of transactions from 2002 to 2004. In comparing the share of countries in total transactions with their share in the leading mode, we observe that usually the share of the leading countries in shaping the market mode, is larger than their share in total transactions.

In this work, we analyzed the network of the international banking system. Our work sheds light on some features of this network. We suggest future research where financial networks are studied along with other variables such as commercial interactions in a multi-layer scheme.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.bis.org/.

AUTHOR CONTRIBUTIONS

All authors contribute to all sections. All authors have read and agreed to the published version of the manuscript. GJ and RR supervise the project; AN administrates the project.

REFERENCES

1. Iori G, Mantegna RN, Marotta L, Miccichè S, Porter J, Tumminello M. Networked relationships in the e-MID interbank market: a trading model with memory. *J Econ Dynam Contr* (2015) 50:98–116. doi:10.1016/j.jedc.2014.08.016
2. Haldane AG, May RM. Systemic risk in banking ecosystems. *Nature* (2011) 469(7330):351–5. doi:10.1038/nature09659
3. Reinhart C, Rogoff K. *This time is different: eight centuries of financial folly*. 1st ed. Princeton, NJ: Princeton University Press (2009).
4. Contreras MGA, Fagiolo G. Propagation of economic shocks in input-output networks: a cross-country analysis. *Phys Rev* (2014) 90(6). doi:10.1103/physreve.90.062812
5. Rama C, Moussa A, Bastos e Santos E. Network structure and systemic risk in banking systems. Available at: https://ssrn.com/abstract=1733528 (2010).
6. Etesami J, Habibnia A, Kiyavash N. Econometric modeling of systemic risk: going beyond pairwise comparison and allowing for nonlinearity (2017). The London School of Economics and Political Science. Technical Paper 70769.
7. Battiston S, Gatti DD, Gallegati M, Greenwald B, Stiglitz JE. Liaisons dangereuses: increasing connectivity, risk sharing, and systemic risk. *J Econ Dynam Contr* (2012) 36(8):1121–41. doi:10.1016/j.jedc.2012.04.001

8. Betz F, Hautsch N, Peltonen TA, Schienle M. Systemic risk spillovers in the european banking and sovereign network. *J Financ Stabil* (2016) 25:206–24. doi:10.1016/j.jfs.2015.10.006
9. D'Errico M, Battiston S, Peltonen T, Scheicher M. How does risk flow in the credit default swap market? Available at: https://ssrn.com/abstract=3723379 (2018).
10. Battiston S, Caldarelli G. Systemic risk in financial networks. *J Finan Manag Markets Inst* (2013) 2013:129–54. doi:10.12831/75568
11. Maluck J, Glanemann N, Donner RV. Bilateral trade agreements and the interconnectedness of global trade. *Front Phys* (2018) 6:134. doi:10.3389/fphy.2018.00134
12. Almog A, Bird R, Garlaschelli D. Enhanced gravity model of trade: reconciling macroeconomic and network models. *Front Phys* (2019) 7:55. doi:10.3389/fphy.2019.00055
13. Atyabi F, Buchel O, Hedayatifar L. Driver countries in global banking network. *Entropy* (2020) 22(8):810. doi:10.3390/e22080810
14. Arthur WB. Complexity and the economy. *Science* (1999) 284(5411):107–9. doi:10.1126/science.284.5411.107
15. Turalska M, West BJ. Fractional dynamics of individuals in complex networks. *Front Phys* (2018) 6:110. doi:10.3389/fphy.2018.00110
16. Jafari G, Shirazi AH, Namaki A, Raei R. Coupled time series analysis: methods and applications. *Comput Sci Eng* (2011) 13(6):84–9. doi:10.1109/mcse.2011.102
17. Ardalankia J, Askari J, Sheykhali S, Haven E, Jafari GR. Mapping coupled time-series onto complex network (2020). Available at: https://arxiv.org/abs/2004.13536.
18. Martnez-Jaramillo S, Prez OP, Avila FE, Dey FLG. Systemic risk, financial contagion and financial fragility. *J Econ Dynam Contr* (2010) 34(11):2358–74. doi:10.1016/j.jedc.2010.06.004
19. Lee D, Park H. Measuring global financial linkages: a network entropy approach. *Sustainability* (2019) 11(17):4691. doi:10.3390/su11174691
20. Berardi S, Tedeschi G. From banks strategies to financial (in)stability. *Int Rev Econ Finance* (2017) 47:255–72. doi:10.1016/j.iref.2016.11.001
21. Raei R, Namaki A, Vahabi H. Analysis of collective behavior of Iran banking sector by random matrix theory. *Iranian J Finan* (2019) 3(4):60–75. doi:10.22034/ijf.2019.111729
22. Namaki A, Raei R, Asadi N, Hajihasani A. Analysis of Iran banking sector by multi-layer approach. *Iranian Journal of Finance* (2019) 3(1):73–89. doi:10.22034/ijf.2019.101367
23. Albert R, Barabási A-L. Statistical mechanics of complex networks. *Rev Mod Phys* (2002) 74(1):47–97. doi:10.1103/revmodphys.74.47
24. Newman MEJ. The structure and function of complex networks. *SIAM Rev* (2003) 45(2):167–256. doi:10.1137/s003614450342480
25. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature* (1998) 393(6684):440–2. doi:10.1038/30918
26. Barabási A-L, Albert R. Emergence of scaling in random networks. *Science* (1999) 286(5439):509–12. doi:10.1126/science.286.5439.509
27. Hosseiny A, Bahrami M, Palestrini A, Gallegati M. Metastable features of economic networks and responses to exogenous shocks. *PloS One* (2016) 11(10):e0160363. doi:10.1371/journal.pone.0160363
28. Hosseiny A. A geometrical imaging of the real gap between economies of China and the United States. *Phys Stat Mech Appl* (2017) 479:151–61. doi:10.1016/j.physa.2017.02.079
29. Faizliev A, Balash V, Petrov V, Grigoriev A, Melnichuk D, Sidorov S. Stability analysis of company co-mention network and market graph over time using graph similarity measures. *J Open Innov* (2019) 5(3):55. doi:10.3390/joitmc5030055
30. Balash V, Chekmareva A, Faizliev A, Grigoriev A, Sidorov S. Analysis of financial network topological dynamics of the Russian stock market from 2012 to 2019. *J Phys Conf* (2020) 1564:012030. doi:10.1088/1742-6596/1564/1/012030
31. Afsharizand B, Chaghoei PH, Kordbacheh AA, Trufanov A, Jafari G. Market of stocks during crisis looks like a flock of birds. *Entropy* (2020) 22(9):1038. doi:10.3390/e22091038
32. Newman M, Watts D. Renormalization group analysis of the small-world network model. *Phys Lett* (1999) 263(4-6):341–6. doi:10.1016/s0375-9601(99)00757-4
33. Barrat A, Weigt M. On the properties of small-world network models. *Eur Phys J B* (2000) 13(3):547–60. doi:10.1007/s100510050067
34. Ravasz E, Barabási A-L. Hierarchical organization in complex networks. *Phys Rev* (2003) 67(2). doi:10.1103/physreve.67.026112
35. Li X, Chen G. A local-world evolving network model. *Phys Stat Mech Appl* (2003) 328(1-2):274–86. doi:10.1016/s0378-4371(03)00604-6
36. Huang W-Q, Zhuang X-T, Yao S. A network analysis of the Chinese stock market. *Phys Stat Mech Appl* (2009) 388(14):2956–64. doi:10.1016/j.physa.2009.03.028
37. Lux T, Marchesi M. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature* (1999) 397(6719):498–500. doi:10.1038/17290
38. Tang Y, Xiong JJ, Luo Y, Zhang Y-C. How do the global stock markets influence one another? evidence from finance big data and granger causality directed network. *Int J Electron Commer* (2019) 23(1):85–109. doi:10.1080/10864415.2018.1512283
39. Bahrami M, Chinichian N, Hosseiny A, Jafari G, Ausloos M. Optimization of the post-crisis recovery plans in scale-free networks. *Phys Stat Mech Appl* (2020) 540:123203. doi:10.1016/j.physa.2019.123203
40. Glasserman P, Young HP. Contagion in financial networks. *J Econ Lit* (2016) 54(3):779–831. doi:10.1257/jel.20151228
41. Bardoscia M, Battiston S, Caccioli F, Caldarelli G. Pathways towards instability in financial networks. *Nat Commun* (2017) 8(1):14416. doi:10.1038/ncomms14416
42. Lenzu S, Tedeschi G. Systemic risk on different interbank network topologies. *Phys Stat Mech Appl* (2012) 391(18):4331–41. doi:10.1016/j.physa.2012.03.035
43. Potters M, Bouchaud J-P. Financial applications of random matrix theory: a short review (2009). Available at: https://arxiv.org/abs/0910.1205.
44. Jiang XF, Chen TT, Zheng B. Structure of local interactions in complex financial dynamics. *Sci Rep* (2014) 4(1):5321. doi:10.1038/srep05321
45. Namaki A, Shirazi A, Raei R, Jafari G. Network analysis of a financial market based on genuine correlation and threshold method. *Phys Stat Mech Appl* (2011) 390(21-22):3835–41. doi:10.1016/j.physa.2011.06.033
46. MacMahon M, Garlaschelli D. Community detection for correlation matrices. *Phys Rev X* (2015) 5(2):021006. doi:10.1103/physrevx.5.021006
47. Sandoval L, Franca IDP. Correlation of financial markets in times of crisis. *Phys Stat Mech Appl* (2012) 391(1-2):187–208. doi:10.1016/j.physa.2011.07.023
48. Shirazi AH, Jafari GR, Davoudi J, Peinke J, Tabar MRR, Sahimi M. Mapping stochastic processes onto complex networks. *J Stat Mech Theor Exp* (2009) 2009(07):P07046. doi:10.1088/1742-5468/2009/07/p07046
49. Jurczyk J, Rehberg T, Eckrot A, Morgenstern I. Measuring critical transitions in financial markets. *Sci Rep* (2017) 7(1):11564. doi:10.1038/s41598-017-11854-1
50. Kwapień J, Drożdż S. Physical approach to complex systems. *Phys Rep* (2012) 515(3-4):115–226. doi:10.1016/j.physrep.2012.01.007
51. Namaki A, Jafari G, Raei R. Comparing the structure of an emerging market with a mature one under global perturbation. *Phys Stat Mech Appl* (2011) 390(17):3020–5. doi:10.1016/j.physa.2011.04.004
52. Nie C-X, Song F-T. Analyzing the stock market based on the structure of kNN network. *Chaos Solitons Fractals* (2018) 113:148–59. doi:10.1016/j.chaos.2018.05.018
53. Jalan S, Solymosi N, Vattay G, Li B. Random matrix analysis of localization properties of gene coexpression network. *Phys Rev* (2010) 81(4):046118. doi:10.1103/physreve.81.046118
54. Jagannath A, Trogdon T. Random matrices and the New York city subway system. *Phys Rev* (2017) 96(3):030303. doi:10.1103/physreve.96.030101
55. Mehta ML. *Random matrices*. Amsterdam, Netherlands: Elsevier (1991).
56. Mehta ML. Preface to the third edition. In: *Random matrices*. Amsterdam, Netherlands: Elsevier (2004). p. xiii–xiv.
57. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, Stanley HE. Random matrix approach to cross correlations in financial data. *Phys Rev* (2002) 65(6). doi:10.1103/physreve.65.066126
58. Laloux L, Cizeau P, Potters M, Bouchaud J-P. Random matrix theory and financial correlations. *Int J Theor Appl Finance* (2000) 03(03):391–7. doi:10.1142/s0219024900000255
59. Namaki A, Raei R, Jafari GR. Comparing Tehran stock exchange as an emerging market with a mature market by random matrix approach. *Int J Mod Phys C* (2011) 22(04):371–83. doi:10.1142/s0129183111016300
60. Reyes JA, Minoiu C. A network analysis of global banking:1978-2009. *IMF Working Papers* (2011) 11(74):1. doi:10.5089/9781455227051.001

61. Park C-Y, Shin K. Contagion through national and regional exposures to foreign banks during the global financial crisis. *J Financ Stabil* (2020) 46: 100721. doi:10.1016/j.jfs.2019.100721
62. Gong X-L, Liu X-H, Xiong X, Zhang W. Financial systemic risk measurement based on causal network connectedness analysis. *Int Rev Econ Finance* (2019) 64:290–307. doi:10.1016/j.iref.2019.07.004
63. Yamamoto S. Banking network multiplier effects on cross-border bank inflows. *Int Rev Econ Finance* (2020) 70:493–507. doi:10.1016/j.iref.2020.06.003
64. Choi S, Furceri D. Uncertainty and cross-border banking flows. *J Int Money Finance* (2019) 93:260–74. doi:10.1016/j.jimonfin.2019.01.012
65. Minoiu C, Reyes JA. A network analysis of global banking: 1978–2010. *J Financ Stabil* (2013) 9(2):168–84. doi:10.1016/j.jfs.2013.03.001
66. Minoiu C, Kang C, Subrahmanian V, Berea A. Does financial connectedness predict crises? *Quant Finance* (2014) 15(4):607–24. doi:10.1080/14697688.2014.968358
67. Hale G, Kapan T, Minoiu C. Shock transmission through cross-border bank lending: credit and real effects. *Rev Financ Stud* (2019) 33(10):4839–82. doi:10.1093/rfs/hhz147
68. Hale G, Kapan T, Minoiu C. *Crisis transmission through the global banking network*. San Francisco, CA: Federal Reserve Bank of San Francisco (2016).
69. Cerutti E, Hale G, Minoiu C. Financial crises and the composition of cross-border lending. *J Int Money Finance* (2015) 52:60–81. doi:10.1016/j.jimonfin.2014.11.013
70. BIS. Bank for international settlements (bis) (1930). Available at: https://stats.bis.org/.
71. Livan G, Novaes M, Vivo P. *Introduction to random matrices*. Cham, Switzerland: Springer International Publishing (2018).
72. Pan RK, Sinha S. Collective behavior of stock price movements in an emerging market. *Phys Rev* (2007) 76(4):046116. doi:10.1103/physreve.76.046116
73. Wang G-J, Xie C, Chen S, Yang J-J, Yang M-Y. Random matrix theory analysis of cross-correlations in the US stock market: evidence from pearson's correlation coefficient and detrended cross-correlation coefficient. *Phys Stat Mech Appl* (2013) 392(17):3715–30. doi:10.1016/j.physa.2013.04.027
74. Lim G, Kim S, Kim J, Kim P, Kang Y, Park S, et al. Structure of a financial cross-correlation matrix under attack. *Phys Stat Mech Appl* (2009) 388(18):3851–8. doi:10.1016/j.physa.2009.05.018
75. Saeedian M, Jamali T, Kamali M, Bayani H, Yasseri T, Jafari G. Emergence of world-stock-market network. *Phys Stat Mech Appl* (2019) 526:120792. doi:10.1016/j.physa.2019.04.028
76. Utsugi A, Ino K, Oshikawa M. Random matrix theory analysis of cross correlations in financial markets. *Phys Rev* (2004) 70(2): 026110. doi:10.1103/physreve.70.026110

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Namaki, Raei, Ardalankia, Hedayatifar, Hosseiny, Haven and Jafari. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 19 January 2021
doi: 10.3389/fphy.2020.574770

Community Detection of Dynamic Complex Networks in Stock Markets Using Hybrid Methods (RMT-CN-LPAm+ and RMT-BDM-SA)

Acep Purqon and Jamaludin*

Physics of Earth and Complex Systems, Institute of Technology Bandung, Bandung, Indonesia

A stock market represents a large number of interacting elements, leading to complex hidden interactions. It is very challenging to find a useful method to detect the detailed dynamical complex networks involved in the interactions. For this reason, we propose two hybrid methods called RMT-CN-LPAm+ and RMT-BDM-SA (RMT, random matrix theory; CN, complex network; LPAm+, advanced label propagation algorithm; BDM, block diagonal matrix; SA, simulated annealing). In this study, we investigated group mapping in the S&P 500 stock market using these two hybrid methods. Our results showed the good performance of the proposed methods, with both the methods demonstrating their own benefits and strong points. For example, RMT-CN-LPAm+ successfully identified six groups comprising 485 involved nodes and 17 isolated nodes, with a maximum modularity of 0.62 (identified more groups and displayed more maximum modularity). Meanwhile, RMT-BDM-SA provided useful detailed information through the decomposition of matrix *C* into *Cm* (market-wide), *Cg* (group), and *Cr* (noise). Both hybrid methods successfully performed very detailed community detection of dynamic complex networks in the stock market.

Keywords: random matrix theory, complex networks, advanced label propagation algorithm, block diagonal matrix, simulated annealing, hybrid methods

OPEN ACCESS

Edited by:
Siew Ann Cheong,
Nanyang Technological University, Singapore

Reviewed by:
Chengyi Xia,
Tianjin University of Technology, China
Jie Cao,
Nanjing University of Finance and Economics, China

****Correspondence:***
Acep Purqon
acep@fi.itb.ac.id

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *21 June 2020*
Accepted: *28 September 2020*
Published: *19 January 2021*

Citation:
Purqon A and Jamaludin (2021) Community Detection of Dynamic Complex Networks in Stock Markets Using Hybrid Methods (RMT-CN-LPAm+ and RMT-BDM-SA). Front. Phys. 8:574770. doi: 10.3389/fphy.2020.574770

1. INTRODUCTION

Physics is the study of the structure and dynamics of various systems that exist in nature. In its current form, the scope of the subject encompasses not only physical systems, but all complex systems. A complex system is one that comprises parts or agents interacting with each other to produce a new macroscopic collective behavior without a central control [1]. Such systems are easily observed in econophysics and social physics (sociophysics).

An example of a complex system in the field of econophysics is the financial market, especially the stock market. It has numerous investors and companies interacting with each other, exchanging assets in their possession to determine the best price for each of them. In general, there are several scientific reasons for physicists to be interested in learning the dynamics that underlie the stock market system [1].

In physical systems, the basis for each agents interactions with another in the system is known; for example, the electrostatic system, where the interaction between charges is based on Coulomb forces. However, in the stock market, the mechanisms underlying the interactions between each agent are not yet clearly known [2]. A starting point for the study of stock markets can be the analysis of the correlation between stocks. A review of the relations between agents in the system is the easiest way to determine the linearity of such relationships in the system without the need to know their underlying cause. From this efficient market hypothesis, it follows that all agents in the stock market get information simultaneously, and every time the information enters the stock market, the stocks respond with changes in the price of the shares so that the share prices reflect the current market conditions. Therefore, the correlation between the stocks can be seen from fluctuations in the share prices.

Recently, a few excellent studies have been published on community detection based on local information and dynamic expansion [3]; the application of random matrix theories, and graphs or networks [4, 5]. Because each method has its strong points and weaknesses, we propose to combine the strong points and reduce the limitations or the weak points and use a combination of these methods for finding the correlations between agents in the stock market system. For example, group mapping involves two different approaches: advanced label propagation algorithm (LPAm+) and simulated annealing (SA). Both methods have limitations, as shown in a few studies. Our objective was to combine LPAm+ with a complex network (CN) and SA with a block diagonal matrix (BDM) for improved effectiveness. LPAm+ with CN determines the group of each node based on the most frequent label of their neighbor. Meanwhile, SA with BDM provides the dual benefit of constructing a block diagonal matrix and finding a global minimum, showing an annealing concept similar to that seen while constructing a crystal. However, the correlations still contain noise and need to be preprocessed using an efficient method. One of the eligible candidates to clean the stock data and remove noise is the random matrix theory (RMT).

As stock market conditions change all the time, the correlations among shares also change. Therefore, the correlations contained in the stock market do not fully describe the relationship between actual stocks. This implies the possibility of noise in the correlation between stocks [6]. In this study, we investigated a method for separating noise from data that contain real information using RMT. The main concept behind RMT is a comparison of the distribution of eigenvalues and eigenvectors in the correlation matrix data owned by a random correlation matrix. Any part of the data that does not display the characteristics of a random correlation matrix is the part that actually contains the real information (non-noise) of the stock market system; vice versa, if any part of the data displays characteristics similar to a random correlation matrix, it is noise.

An analysis of the eigenvalues and eigenvectors of the stock matrix correlation structure has shown that a few of the largest eigenvector components are localized; for example, components with the greatest contribution to each eigenvector are found in the same sector [2]. However, these results are not sufficiently significant to be adopted as a method for analyzing groups in the stock market because each eigenvector is not independent of each other (a few sectors overlap in one eigenvector). Moreover, during the analysis of eigenvector components, only a few vector components were observed to have the greatest contribution [7]. Therefore, in this study, another approach was used to analyze the stock market groups and a few candidates were found. We used a CN as the first approach and a BDM as the second.

In the CN approach, each share in the stock market is seen as a node, and the correlation between the shares is analogous to the connecting side between the nodes. To form a stock market network, the LPAm+ method is used, which determines the group (community) label of a node based on the majority of its neighbor labels; nodes with the same label are considered to be in the same group or community. Conversely, in the BDM approach, the stock correlation matrix is converted to a BDM, where each block represents a group in the stock market; the method is chosen to create the BDM as an SA algorithm, which mimics the annealing process in crystal formation. The data used in the study were the daily closing price of the shares listed on the S&P 500 from January 1, 2007 to October 28, 2016.

Our purpose was to: 1) generate a correlation filtering data filtering program using the RMT method; 2) develop a program for mapping the groups in the stock market using the LPAm+ and SA algorithms, and 3) compare the results of the mapped groups in the stock market by employing the CN approach using LPAm+ (namely RMT-CN-LPAm+) and the BDM approach using the SA algorithm (namely RMT-BDM-SA).

2. METHODS

2.1. Random Matrix Theory

The application of RMT assumes a matrix whose elements are random or not bound to one another. A random matrix has zero average value and one variance [8]. RMT was first introduced by Wigner to explain energy-level statistics in complex quantum systems. Wigner created a matrix model with random elements to explain the Hamiltonian mechanics of a heavy nucleus that fit the experimental results [9]. In complex quantum systems, RMT predictions can explain all the possibilities that can occur in system [10]; in subsequent developments, it was concluded that parts incompatible with RMT predictions can provide clues about the interactions that underlie the system [8].

In the late 1990s, Laloux et al. and Pelrou et al. applied RMT to correlation data based on changes in stock prices on the American stock markets [6]. Subsequently, several physicists tried to apply RMT to different stock markets. The results showed similarity to the extent that RMT could identify the noise part contained in the correlation data between stocks, and proved that most of the stock data followed the random correlation matrix pattern [2, 7, 10–12]. RMT can distinguish between noise and the real information part by comparing the data held in a random correlation matrix. When a part of the data does not follow the properties or characteristics of a random

correlation matrix, it is ensured that the particular part contains real information from the stock market system; and if a part of data has the same characteristics as a random correlation matrix, then that part is noise.

There are several characteristics of a random correlation matrix used in RMT. For example, if **A** is a random matrix with dimensions $N \times L i^{th}\ i^{th}$, the average value of the element is zero and the variance is one. Then, the random correlation matrix **R** is calculated using the following equation:

$$R = \frac{1}{L} AA^T \tag{1}$$

When the values $N \to \infty, L \to \infty$ are chosen and qualify $Q = \frac{L}{N} > 1$, the distribution of eigenvalues from the R matrix follows the Marchenko– Pastur distribution:

$$P_{rm}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \tag{2}$$

Here, λ_+ and λ_- are the minimum and maximum values of the eigenvalues of matrix **R**. Then, the distribution of the matrix eigenvector component $Ru_l^k; l = 1, \ldots, N$ follows the Gaussian distribution given in the following equation:

$$\rho_{rm}(\lambda) = \frac{1}{\sqrt{2\pi}} exp\left(\frac{-u^2}{2}\right) \tag{3}$$

2.2. Complex Network

A network can be defined as a set of objects called vertices (nodes or vertices); the relationships between vertices are called lines or sides (edges or links) [13]. Suppose a network $G = (N, L)$ consists of two sets of N and L, where N is a set of network nodes $N \equiv n_1, n_2, \ldots, n_N$ and L are the network side sets $L \equiv l_1, l_2, \ldots, l_K$, which are non-sequential pairs of N elements. A network can be represented as a matrix, usually called an adjacency matrix A. An adjacency matrix informs if there are sides (connected or unconnected) between each two nodes in a network.

The degree of a network is defined as the number of sides passing through a node. The degree of node i can be calculated using the following equation:

$$k_i = \sum_{j=1}^{N} A_{ij} \tag{4}$$

Then, the total degree of a network can be calculated as follows:

$$K = \sum_{j=1}^{N} k_i \tag{5}$$

The shortest path that connects two vertices is commonly called the geodesic path. Take for example, a matrix D whose elements are geodesic distances between vertices i and j or d_{ij}. From the shortest distance parameter, we obtain another parameter, that is the diameter of the network, which is defined as the maximum value of the matrix D. Then, other network characteristics geodesic distance is the average between vertices obtained from the following equation:

$$L = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{ij} \tag{6}$$

The *node betweenness* parameter measures the effect of a node in a network by counting the several geodesic paths through that node. Mathematically, it is expressed as

$$b_{ij} = \sum_{i,j \in N, i \neq j} \frac{n_{jk}(i)}{n_{jk}} \tag{7}$$

Here, n_{jk} is the number of geodesic paths connecting vertices j and k; $n_{jk}(i)$ is the number of geodesic paths through node i.

The *cluster coefficient* parameter measures the tendency of n from node i to become a group or cluster in a network. The *cluster coefficient* of node i is calculated by the ratio between the number of sides (e_i) in the subgraph G_i to the maximum number of sides that might form on G_i as follows:

$$c_i = \frac{2e_i}{k_i(k_i - 1)} = \frac{\sum_{j,k} a_{ij} a_{jk} a_{ki}}{k_i(k_i - 1)} \tag{8}$$

Then, the average cluster coefficient of each node, also called the network *cluster coefficient*, is calculated as follows:

$$C = \langle c \rangle = \frac{\sum_{i \in N} c_i}{N} \tag{9}$$

The quality of the grouping of communities in a network can be measured from the relationship between the intra-community and inter-community nodes. When the relationships between the intra-community nodes are dense and those between the inter-community nodes are rare, then the grouping of networks, as well as the parameters that measure the relationships, are considered good. This is called *modularity*; a term first introduced by Newman [14]. The extent of modularity in a network can be calculated using the following equation:

$$Q = \frac{1}{\sum_{ij} A_{ij}} \sum_{ij} \left[A_i j - \frac{k_i k_j}{\sum_{ij} A_{ij}} \right] \delta(c_i, c_j) \tag{10}$$

Here, k_i and k_j are the degrees of nodes i and j, c_i is the community label of node i, and c_j is the community label of node j. Practically, a modularity value above 0.3 is considered a good grouping.

Based on their degree distribution, networks can be classified into two most common types: exponential and scale-free. In exponential networks, the degree distribution follows the Poisson distribution, which means that most of the nodes in the network have the same degree (they are homogeneous). In scale-free networks, the distribution of degrees in heterogeneous networks follows the power-law distribution, that is, most vertices have a small degree; a few or a small proportion of them have a large degree. Examples of exponential and scale-free networks can be seen in Ref. [15] and their distribution in Ref. [16].

2.3. Block Diagonal Matrix

Noh proposed a diagonal block matrix model and demonstrated that for stocks that belong to one group, the diagonals of the formed correlation matrix have a value of one and the remaining

entries have the value zero $C_{ij}^{g} = \delta_{\alpha_i \alpha_j}$, where α_i denotes the group where shares i are Ref. [17].

Here, 1_i is a matrix with $N_i \times N_i$ dimensions (N_i is the number of shares incorporated in group i). More than one diagonal matrix array can be formed in a single correlation matrix. Therefore, an optimal arrangement of stocks is needed to produce a good BDM.

Kim and Jeong proposed an optimization of the BDM by analyzing the correlation between stocks as a force that binds to particles (in this case, stocks) [7]. Because of the binding force between the shares, there is total energy in the system. The equation that calculates the system energy is given in **Eq. 11**, and the most stable BDM form is obtained when the energy in the system is the minimum. An example of the BDM calculated by Kim and Jeong using the New York Stock Exchange (NYSE) stock data for the 1993–2003 period can be found in Ref. [7].

$$E_{total} = \sum_{i<j} C_{ij}^{g} \left| l_i - l_j \right| \otimes \left(C_{ij}^{g} - c_c \right) \tag{11}$$

Here, C^g is the correlation matrix filtered by RMT, l_i is the location where stock i is in stock order, and c_c is a *cutoff* that eliminates the remaining *noise* in C^g (usually $c_c = 0.1$) [7].

2.4. Matrix Decomposition

To separate *noise* from the information in the correlation data through several stages, namely, during the distribution comparison between the correlation matrix **C** and the random correlation matrix **R** to calculate the correlation of each share, the return for each stock is calculated as $i = 1, \ldots, N$ within a certain period (t).

$$G_i(t) = lnS_i(t + \Delta t) - S_i(t) \tag{12}$$

Here, $S_i(t)$ and $G_i(t)$ are the price and stock return i at time Δt, respectively. Because each stock has a different volatility value, a normalized return is defined by

$$g_i(t) = \frac{G_i(t) - \langle G_i(t) \rangle}{\sigma_i}, \tag{13}$$

where $\sigma_i = \sqrt{\langle G_i^2 \rangle - \langle G_i \rangle^2}$ is the standard deviation of G_i. Then, the correlation matrix **C** is calculated by

$$C_{ij} \equiv \langle g_i(t) g_j(t) \rangle \tag{14}$$

In matrix representations, it is expressed by

$$C = \frac{1}{L} G G^T, \tag{15}$$

where G is a matrix $N \times L$, with the element $\{g_i(m\Delta t); i = 1, \ldots, N; m = 0, \ldots, L-1\}$ and $\mathbf{G}^T$ are the transpose matrices of **G**.

To compare the eigenvalue distribution of the correlation matrix **C** and the random correlation matrix **R**, the eigenvector interpretation of the correlation matrix **C** that is outside the predicted RMT tests the stability of each eigenvector of the correlation matrix C. First, we divide the stock price data (matrix **S**) into two parts (the first half S_1 and the other half S_2; each of them is calculated using the correlation matrices C_1 and C_2. Then, the overlap matrix is calculated as follows:

$$C = C_1 C_2^T \tag{16}$$

A matrix can be decomposed into a linear combination of from a collection of matrices. To find a noise-free correlation matrix, the decomposition is expressed by

$$C = \sum_{\alpha=1}^{N} \lambda_\alpha |\alpha\rangle\langle\alpha|, \tag{17}$$

where N is the number of shares and λ is the eigenvalue of the **C** matrix sorted.

2.5. Percolation Theory

In the CN approach, the **C** correlation matrix (which is *noise free*) can be treated as an adjacency matrix [_Jeong_and_Kim_2005] demonstrated that to find a clear definition for each group (community) in the network, a weighted network needs to be chosen for the group analysis in the stock market [7]. Because the value of the elements in the C_g matrix is not binary (1 or 0), the percolation theory is used to set it as the adjacency matrix. The basic idea of percolation theory is to use a *threshold value* to determine whether two nodes are connected in the network. If the correlation coefficient is greater than the boundary value, the adjacency coefficient is 1, and if it is below the boundary value, then the coefficient is 0.

2.6. Advanced Label Propagation Algorithm

LPAm+ is a method for developing the label propagation algorithm (LPA) method. The main idea of the LPA method is to determine the community label of a node based on the majority of labels from its neighbors; the nodes that have the same label are grouped into one community (group) [18]. At the beginning of the algorithm, different (unique) labels are given for each node; then, during propagation, a node changes its label to follow the majority of its neighboring labels, and in case of a tie (there is more than one label with the same number), the label is determined randomly. The iteration stops if there is no longer a label propagation process in the network. In mathematical form, the label update process can be written according to the following equation:

$$C_x' = argmax_c \left(\sum A_{ix} \delta(c_i, c) \right) \tag{18}$$

Because there is a random aspect to the labeling during series conditions as described previously, this LPA method does not produce a unique solution for each run. As a result, more than one community structure can exist even if they originate from the same initial conditions. Therefore, the LPA method is generally performed several times and a community structure that has the greatest modularity value is taken. The main advantage of the LPA method is its very high speed compared with other methods [18].

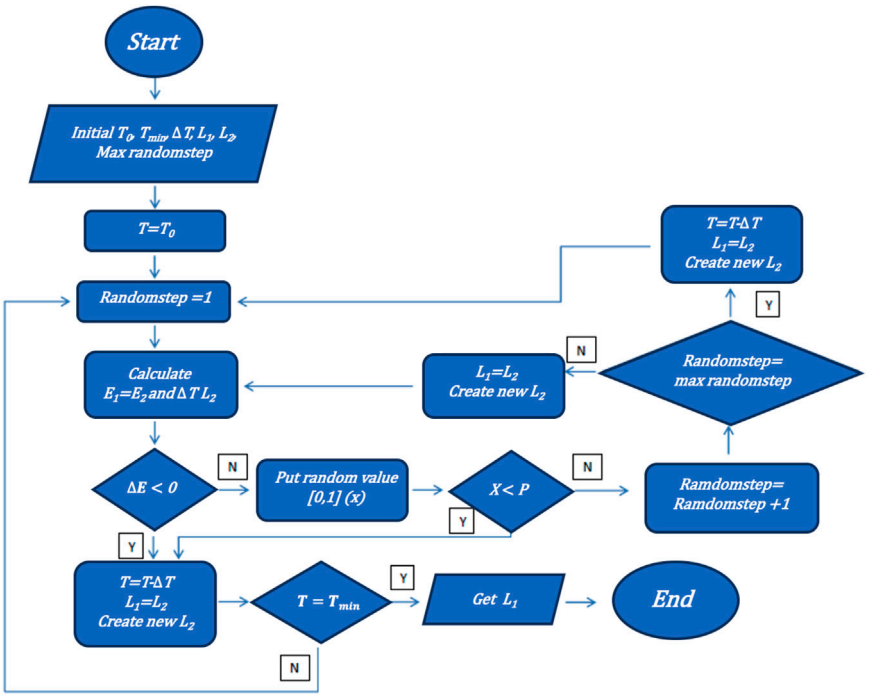


FIGURE 1 | Flowchart of the simulated annealing (SA) algorithm to form a block diagonal matrix (BDM). The most stable BDM is obtained when the energy in the system is of the minimum value. To find the most stable BDM and avoid the local minimum, we propose to combine it with simulated annealing, bringing the concept of annealing to the formation of crystals in optimization problems.

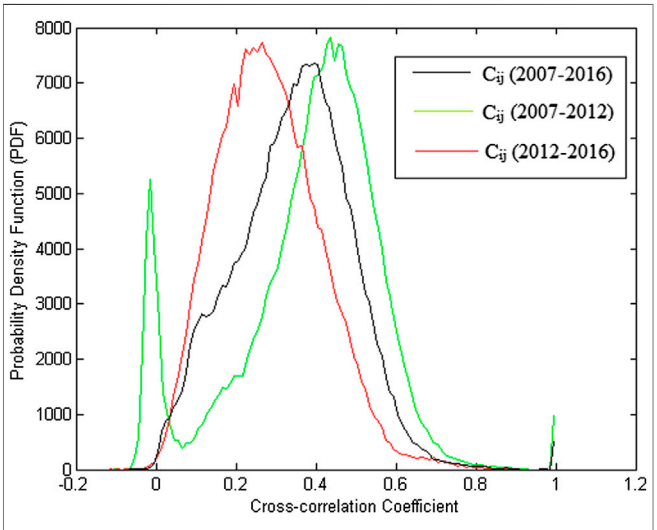


FIGURE 2 | Distribution of correlation coefficients of matrix **C** at three different times in the S&P 500 daily stock price data for the period of January 1, 2007 until October 28, 2016. The black color is for the total period 2012–2016 and it is decomposed into the green one for the period 2007–2012 and the red one for the period 2012–2016. It shows how the black one contributes to the different distribution of the green and red ones.

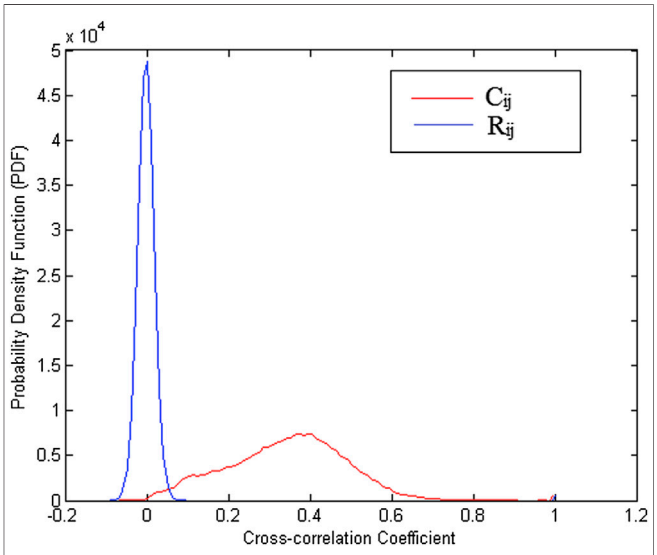


FIGURE 3 | The Probability Density Function (PDF) of the correlation coefficient **C** matrix and the **R** matrix. The red color shows the distribution of the **C** matrix, and the blue one indicates the distribution of the **R** matrix on the S&P 500 daily stock price data for the period of January 1, 2007 until October 28, 2016. It clearly shows the different distribution groups separately for the correlation coefficient **C** matrix and **R**.

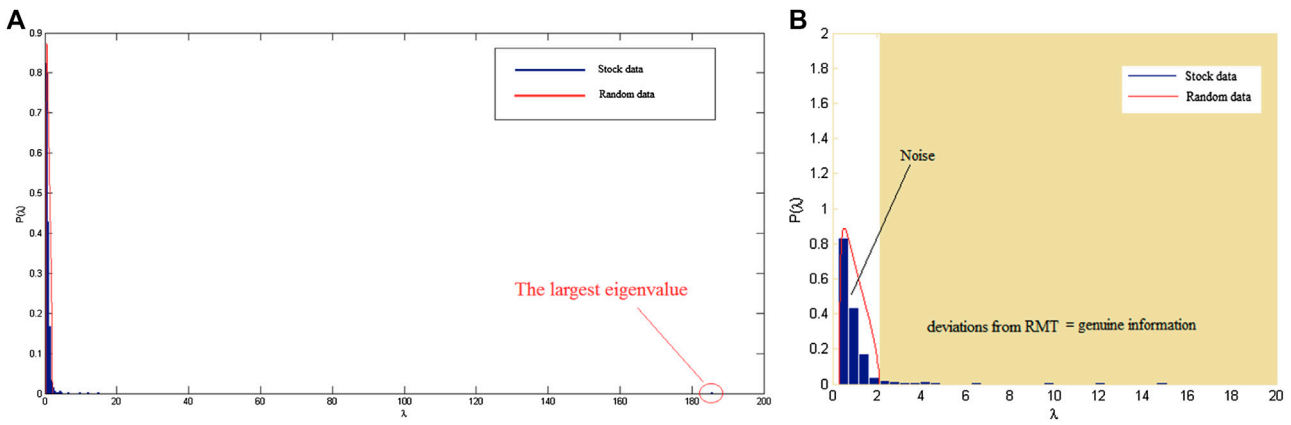


FIGURE 4 | The distribution of eigenvalues from the **C** matrix and the **R** matrix. The blue color represents the distribution of the eigenvalues of the matrix **C**, and the red one shows the distribution of the eigenvalues of the matrix **R** on the S&P 500 daily stock price data for the period of January 1, 2007 until October 28, 2016. **(A)** eigenvalue in the scale for λ of 1–200 and **(B)** The eigenvalue in the zoomed scale for λ of 1-20.

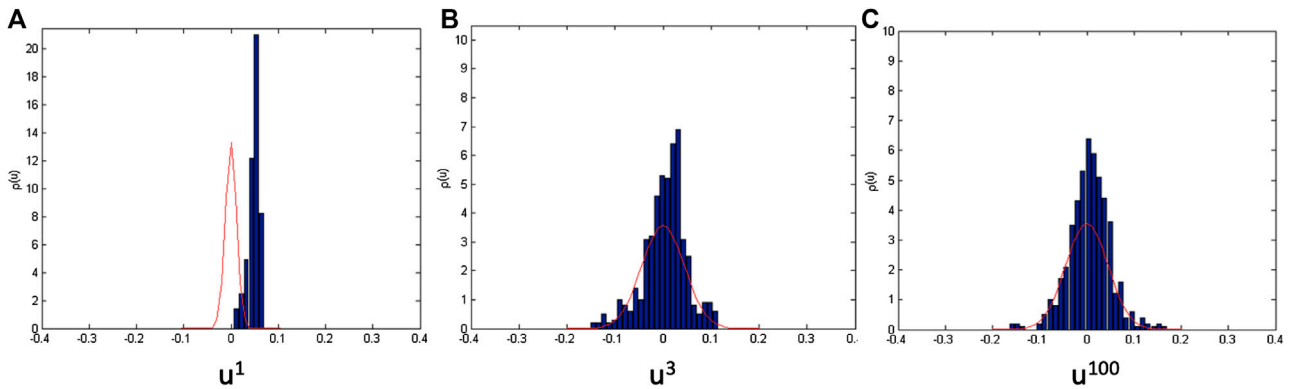


FIGURE 5 | Comparison among the distribution of eigenvector components. For example, **(A)** is the largest u^1, **(B)** is the third largest u^3, and **(C)** is the hundredth largest u^{100}. The blue color represents the distribution of the **C** matrix, and the red one represents the distribution of the **R** matrix in the S&P 500 daily share price data for the period of January 1, 2007 until October 28, 2016.

Barber and Clark modified the LPA method by increasing the monotonous value of the rising modularity in each iteration [19]. The modularity equation can be rearranged as follows:

$$Q = \frac{1}{\sum_{ij} A_{ij}} \left(\sum_{i \neq j \neq x} B_{ij} \delta\left(c_i, c_j\right) - B_{xx} \right) - \frac{2}{\sum_{ij} A_{ij}} \left(\sum_i B_{ix} \delta\left(c_i, c_x\right) \right). \tag{19}$$

$$B_{ij} = A_{ij} - \frac{k_i k_j}{\sum_{ij} A_{ij}} \tag{20}$$

The aforementioned equation denotes the separation of the terms containing the label of node x from the previous modularity equation. To maximize the modularity value, the writer must maximize the 2^{nd} term of **Eq. 19**. Therefore, the equation for label updates becomes

$$C_x' = argmax_c \left(\sum B_{ix} \delta\left(c_i, c\right) \right) \tag{21}$$

However, the LPAm method still has a shortcoming of possibly getting trapped in the local maximum in the modularity space; thus, Liu and Murata modified the LPAm method by applying agglomeration techniques to combine each of the two groups (communities) and avoid any changes in values. The modularity then chooses which results in the largest change in modularity value. The combination of these methods is called LPAm+ [20].

Regardless of the first local maximum value, the LPAm steps are repeated to reach the next local maximum value. The aforementioned two methods (LPAm and agglomeration) are repeated until there are no more modularity changes.

2.7. Simulated Annealing Algorithm

To find the stock arrangement that provides the most system energy, the SA algorithm is used in Monte Carlo simulations to

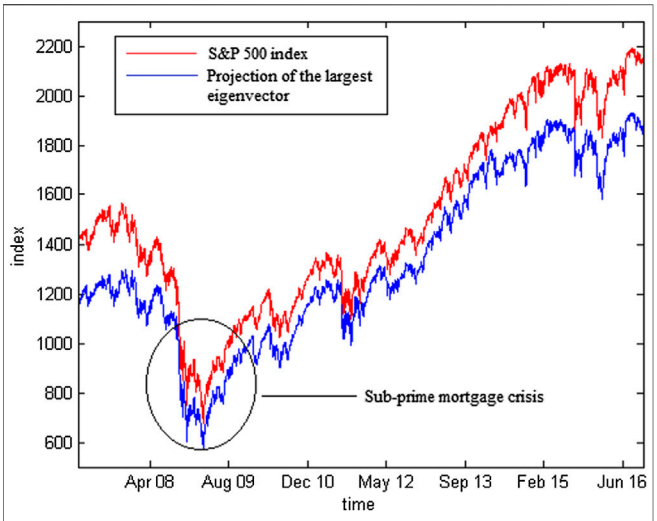


FIGURE 6 | To validate how good the Eigen vectors, we can perform a comparison of projections of the largest eigenvector component (blue) from matrix **C** with the S&P 500 index (red) on the S&P 500 daily share price data for the period of January 1, 2007 until October 28, 2016. This indicates that the method mainly follows the patterns successfully.

avoid brute force. The SA algorithm was first introduced by Metropolis. Furthermore, SA was first applied to the optimization issue by Kirkpatrick et al. to avoid local drinking conditions [21]. This algorithm is analogous to the annealing (cooling) process that is applied while producing glassy materials (comprising crystalline grains).

The annealing process can be defined as a regular or constant temperature drop on a previously heated solid object until it reaches the ground state or freezing point. The temperature is reduced continuously and carefully so that a thermal balance is attained at each level. If the temperature is not reduced stepwise, the solid object acquires structural defects due to the formation of only optimal local structures. This type of process that produces only an optimal local structure is called rapid quenching. The search for a solution with SA is similar to the hill-climbing concept where the solution tends to change continuously until the final temperature is reached.

In the SA algorithm, we introduce the concept of annealing to the formation of crystals in optimization problems. The objective function, that is to search for the minimum value in the optimization problem, is compared with the energy of the material in the case of the annealing process. Then, a control parameter, which is the temperature, is used for each iteration.

The SA algorithm uses the concept of a neighborhood search or local search in each iteration to find conditions that provide the lowest objective function. For each iteration, if the surrounding conditions (in the case of a BDM, the composition of shares) provides an objective function value smaller than the original objective function, then the initial condition is updated (the condition of the neighbor is set as the new initial condition). However, when the condition of the neighbor outputs a value greater than the original objective function, the result can still be accepted (the initial condition is enhanced by the condition of the neighbor) with certain conditions of probability.

Classical particle probability is used in this case, which follows Maxwell-Boltzmann statistics $\left(P = e^{\frac{-\Delta E}{T}}\right)$, where ΔE is an objective function and T is the temperature control parameter. The iteration is performed until the objective function no longer changes or has reached its ground state [22]. Here, L_1 is the first arrangement guess , L_2 is the second arrangement guess, T is temperature, E_1 and E_2 are the system energies for L_1 and L_2, $\Delta E = E_2 - E_1$, respectively, and $\left(P = e^{\frac{-\Delta E}{T}}\right)$ are classical particle probabilities that follow Maxwell–Boltzmann statistics. In this study, we performed the calculations using the flowchart shown in **Figure 1**.

3. RESULTS AND DISCUSSION

3.1. Random Matrix Theory

3.2.1. Distribution of Correlation Matrix C and Random Correlation Matrix R

As mentioned in the previous section, the correlation value between shares has no fixed over time, and a plot was drawn for three different conditions of the stock correlation data. The first is the correlation matrix extracted from the 2007 to 2016 data (black line), the second is the correlation matrix for the data from 2007 to 2012 (green line), and the third is or the data from 2012 to 2016 (red line) using **Eq. 12** through **Eq. 15**. The results are shown in **Figure 2**. According to the figure, there is an increase in the correlation coefficient between the stocks. The average correlation coefficient between the shares in the data for the periods 2007–2012 and 2012– 2016 is 0.3832 and 0.2642, respectively. Then, for the whole period (2007–2016), the average correlation coefficient is 0.3451.

Next, the distribution of the **C** correlational matrix was compared with that of the **R** random correlation matrix. The results in **Figure 3** show that the distribution of the **R** matrix follows a Gaussian trend, whereas the **C** matrix has a positive leaning distribution, indicating that the relationship between the stocks on the S&P 500 dominant correlates with each other compared to those who have anti-correlation relationships.

3.2.2. Eigenvalue Distribution of Correlation Matrix C and Random Correlation Matrix R

Figure 4 shows the eigenvalue distributions of the **C** matrix and the random correlation **R** matrix following **Eq. 1**. It can be seen that most (97%) eigenvalues of the **C** matrix are in the vulnerable boundary of the random **R** matrix, which indicates that most of the stock data are noise. Only 3% eigenvalues of the **C** matrix are outside the boundary of the random matrix **R**, and represent the real information from the stock market. The largest eigenvalue produced is 185.38, which is more than 90 times the upper limit of the eigenvalue matrix **R** ($\lambda_+ = 2.075$).

3.2.3. Distribution of Eigenvector Components in Correlation Matrix C and Random Correlation Matrix R

Apart from looking at the distribution of eigenvalues from the **C** correlation matrix, we also tested for the presence of noise in the data by looking at the distribution of the eigenvector components. **Figure 5** shows the different distributions of eigenvector

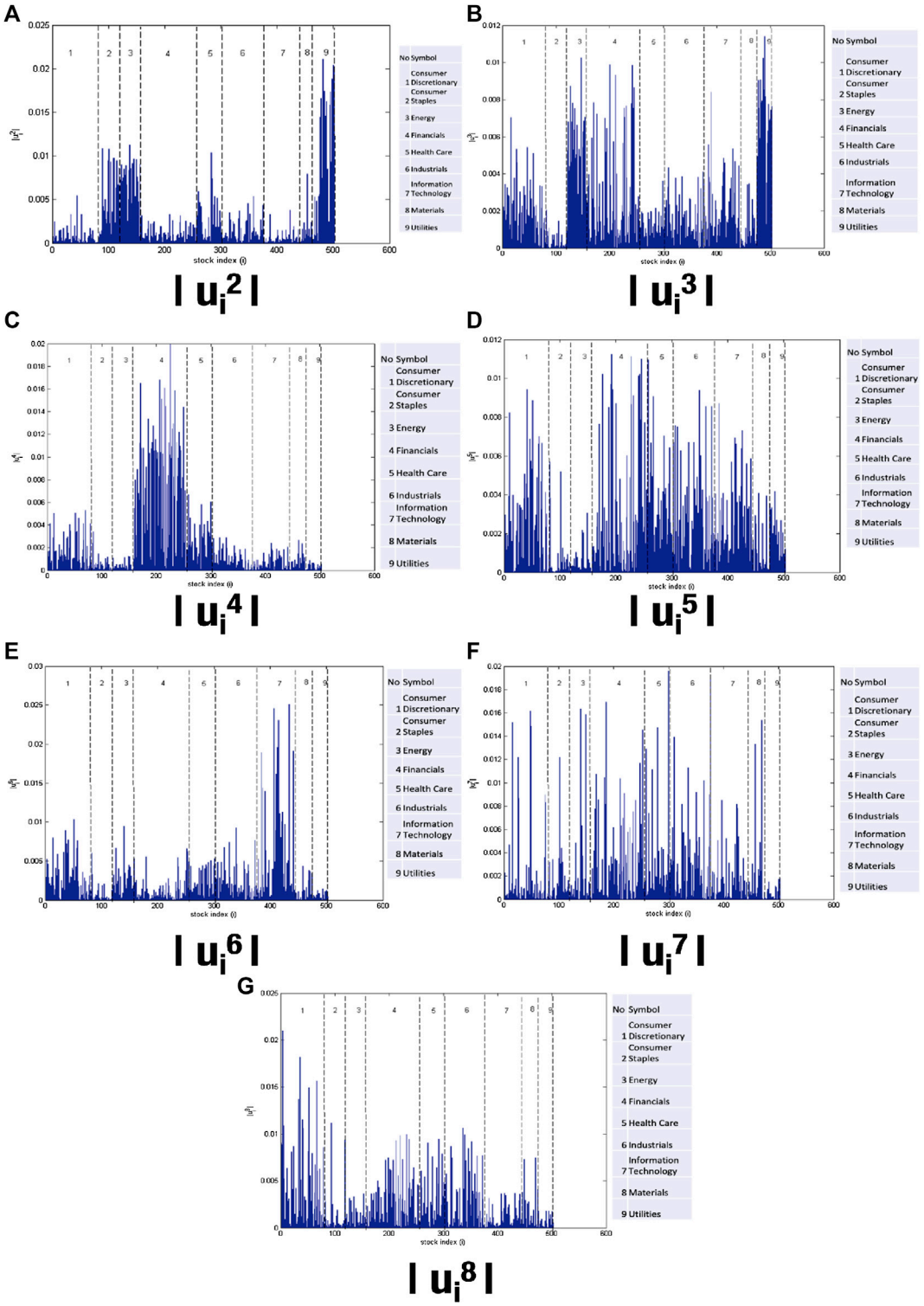


FIGURE 7 | There are many information for each Eigen vectors. Some sectors overlap in several Eigen vectors. This indicates that we require a method to reveal community detection. The vector component of u^2–u^8 from the matrix of **C**, the arrangement of the stock based on the sectors, 1: consumer discretionary; 2: consumer staples; 3: energy; 4: financials; 5: health care; 6: industrial; 7: information technology; 8: materials; dan 9:utilities; from the daily stocks of S&P 500 for the period of January 1, 2007 until October 28, 2016.

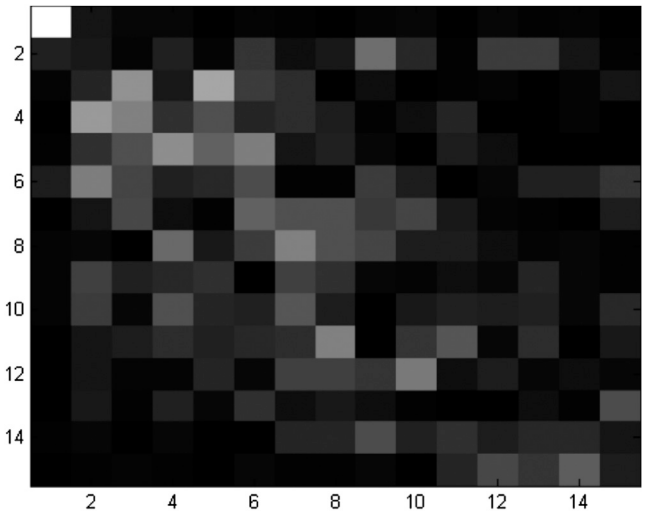


FIGURE 8 | Stability of the eigenvectors from the largest eigenvectors (u^1) to (u^{15}) of **C** in the S&P 500 daily stock price data for the period of January 1, 2007 until October 28, 2016. White blocks indicate stable areas, and black blocks indicate unstable regions.

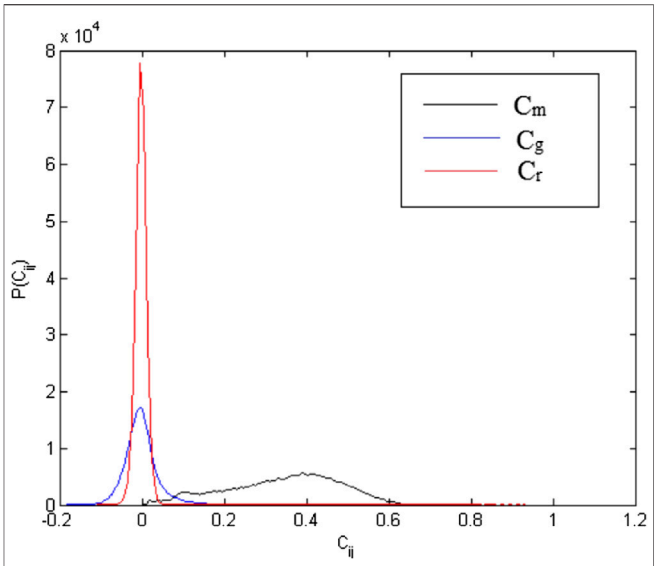


FIGURE 9 | Matrix decomposition shows successfully decompose the group into three different parts. Distribution of **Cm** (black), **Cg** (blue), and **Cr** (red) on the S&P 500 daily share price data for the period of January 1, 2007 until October 28, 2016. This indicates that the method successfully shows the decomposition and contribution from each part.

components outside and within the boundary of a random matrix (using RMT). The eigenvector component within the boundary of the random matrix in **Figure 5C** follows the Gaussian distribution as given in **Eq. 3**. This shows that this part is noise, whereas the distribution of the eigenvector component outside the boundary of the random matrix in **Figures 5A,B** is heavy or leaning toward one side.

3.2.4. Interpretation of the Largest Eigenvalue and Eigenvector (u^1)

After successfully distinguishing between noise and data containing real information, we then identified each C eigenvalue that was outside the boundary of the random matrix. The uniqueness of the largest eigenvalue can be observed easily when compared with other eigenvalues because of its greater value than others, as seen in **Figures 4**, **5A**, which shows that all the components of the eigenvector are positive. This demonstrates that the largest eigenvalue has a very significant influence on the dynamics of the stock market, commonly referred to as the market-wide effect [2].

To test the assumption that the largest eigenvalue has a market-wide effect, a comparison between the projections of the eigenvector components was calculated using **Eq. 22** with an S&P 500 index value. **Figure 6** shows that the projections of the largest eigenvector components and S&P 500 have the same movement patterns. These results reinforce that the largest eigenvalue is a representation of the movement of the stock market itself. The equation is as follows:

$$S^1(t) = \sum_{ix}^{N} u_j^1 S_j(t) \tag{22}$$

3.2.5. Interpretation of Eigenvalues and Other Eigenvectors That Are Still Outside Random Matrix Theory Predictions

After successfully identifying the largest eigenvalue, we also performed identification on other eigenvalues that are still outside the boundary of the random matrix. However, before doing that, the largest eigenvalue must be removed first owing to its market-wide effects. As the results in the previous section have shown, the largest eigenvalue is a representation of the market movement itself and has a very significant effect on the components of other eigenvectors and constrains the other eigenvectors [2]. To eliminate the market-wide effects, an ordinary least square is expressed as follows:

$$G_t(t) = \alpha_i + \beta_i G^{(1)}(t) + \epsilon_i(t), \tag{23}$$

where $G^{(1)}(t)$ is similar to **Eq. 22**, which is $G^{(1)} \equiv \sum_{j=1}^{N} u_j^1 G^{(1)}(t)$, α_i dan β_i are constants. Then, the correlation matrix **C** is re-created using $\epsilon_i(t)$; then each eigenvector component can be seen.

The greater the value of an eigenvector component in its eigenvector, the greater is its contribution to the eigenvector. **Figure 7** shows the values of each component of the eigenvectors u^2 through u^8. It can be seen that a few eigenvectors are localized to the largest components. For example, in eigenvectors u^2 and u^3, the largest components are dominated by utilities in u^4 and u^5 by the financial sector, and in u^6 by the information technology sector. However, there is no dominant sector in u^7. In the last eigenvector u^8, the largest component is dominated by the consumer discretionary sector.

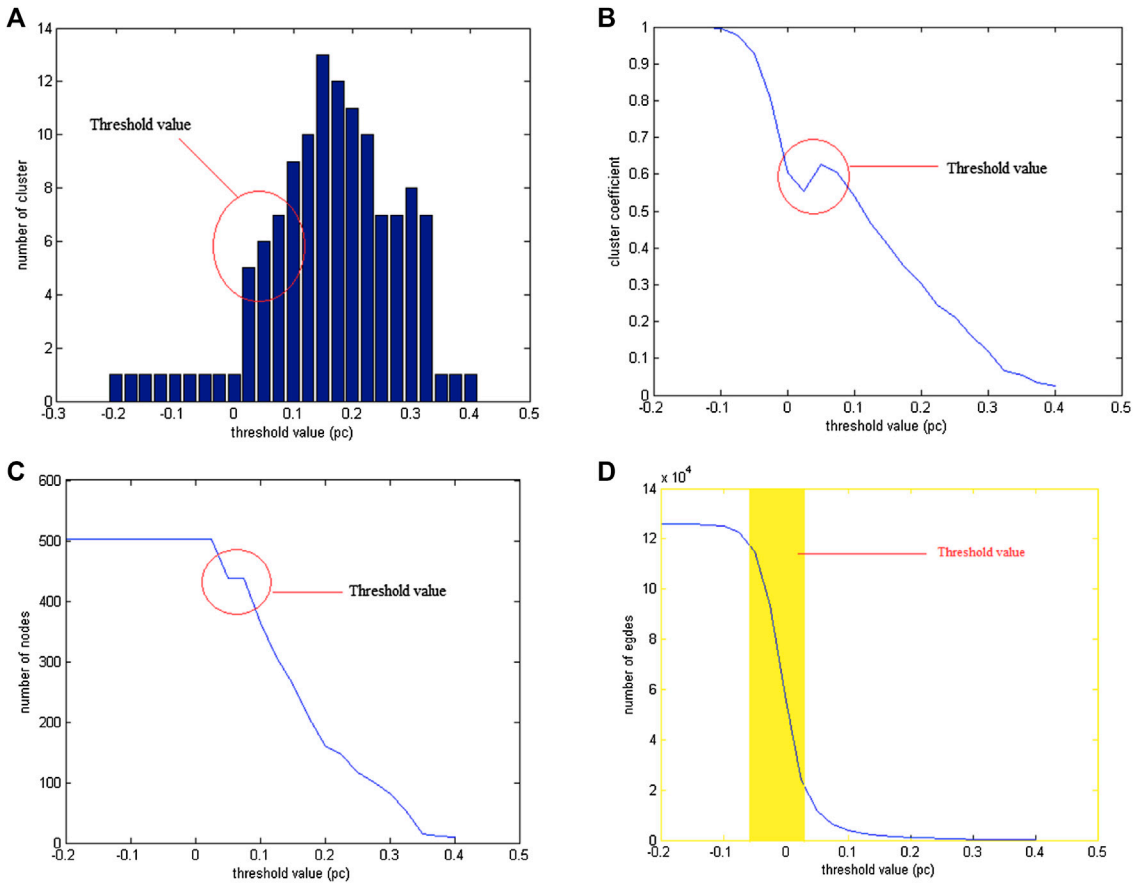


FIGURE 10 | The value **(A)** number of clusters formed, **(B)** number of vertices, **(C)** cluster coefficients and **(D)** number of sides when boundary values vary from –0.2 to 0.4. The method successfully shows the threshold value. This value is important for analyzing the group effectively.

After an analysis of the eigenvector components, it can be concluded that the groups identified are not yet comprehensive owing to the absence of sectors such as energy, materials, industrials, or consumer staples. This is because during the analysis of the eigenvectors, only the components with the greatest value are noted. Therefore, we cannot use only this method for group identification in the stock market. For the next analysis, we used a CN approach and BDM to identify groups in the stock market.

3.2.6. Eigenvector Stability of Correlation Matrix C

The results of mapping the stability of the eigenvectors of the **C** matrix can be seen in **Figure 8**. The results not only show that only the largest eigenvectors are stable over time but also reinforce previous results that eigenvector analysis cannot be used to determine groups in the stock market because only stable eigenvectors can be interpreted [23].

3.2. Matrix Decomposition

From the results of the RMT method, it is evident that the stock market data contains not only noise but market-wide effects also; therefore, before analyzing the stock correlation data with CNs and the BDM, it must be cleaned from noise and market-wide effects. Matrix decomposition is used for cleaning, where matrix **C** is decomposed into three parts, namely market-wide **(Cm)**, group **(Cg)**, and noise **(Cr)**, using **Eq. 24**. To be used as an adjacency **A** matrix in CNs and BDM analysis, only **(Cg)** is used. The equation is as follows:

$$N_g = C_m + C_g + C_r = \lambda_1 |1\rangle\langle 1| + \sum_{n=2}^{N_g} \lambda_n |n\rangle \langle n| + \sum_{n=N_g+1}^{N} \lambda_n |n\rangle\langle n| \tag{24}$$

Here, N_g is the sequence of the last eigenvalue, which is still beyond RMT prediction ($\lambda_{N_g} > \lambda_+$). Our results show the distribution of **Cm** (black), **Cg** (blue), and **Cr** (red) on the S&P 500 daily share price data for the period between January 1, 2007 and October 28, 2016, as shown in **Figure 9**.

3.3. Determining the Threshold Value

In a non-weighted network, the determination of the boundary value is very important because each different boundary value forms a different group structure. If the selected boundary value is too small, all the nodes are connected, which means there is only one large group, and if the chosen boundary value is too large,

FIGURE 11 | The method successfully performs community detection. The S&P 500 stock market network using Gephi software, there are six main clusters, cluster 1 (black) is dominated by the financial sector, cluster 2 (pink) is dominated by the information technology sector, cluster 3 (orange) is dominated by the energy sector, industrial and materials, cluster 4 (green) is dominated by the consumer staples, utilities, and industry sectors; cluster 5 (blue) is dominated by the consumer discretionary sector, and finally cluster 6 (purple) is dominated by the health care sector.

only a small number of nodes are still connected in the network; most of them are isolated.

Therefore, to determine the boundary value in this study, four parameters were considered: the number of groups formed, number of vertices, number of sides, and average cluster coefficient [5, 11]. **Figure 10** shows that the optimal boundary value is 0.05.

3.4. Group Mapping Results With Advanced Label Propagation Algorithm

The LPAm+ program on the S&P 500 network successfully identified the groups on the network. As many as six groups were observed, with the number of involved nodes reaching up to 485 out of a total 502 nodes (17 nodes were isolated from the network) and the maximum modularity value obtained was 0.6164.

The results obtained with LPAm+ show that the shares that belong to the same group are dominated by certain sectors (according to the results of the eigenvector analysis with RMT). Group mapping with LPAm+ can be visualized using Gephi software. The results obtained using the Gephi software are slightly different than those after using the LPAm+ method. In Gephi, there are six large clusters (groups) of the stock market network with a total of 483 (96%) nodes out of the total, 5579 sides, and a modular value of 0.619 (there is a difference of 0.003 with the LPAm+ results). **Figure 11** shows the results of the animation using Gephi.

3.5. Group Mapping Results With Simulated Annealing Algorithm

The results obtained using the SA algorithm and the **Cg** correlation matrix data show accordance with the concept of the BDM, namely the condition of the stock arrangement that

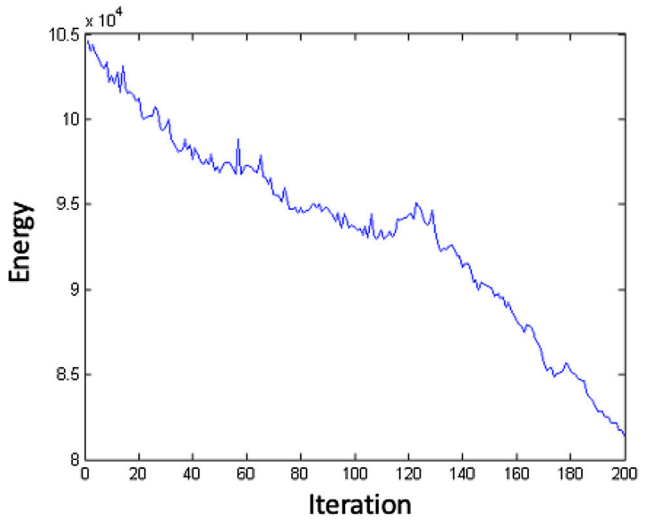


FIGURE 12 | How energy decrease by iteration in Simulated annealing method results, for example with the initial guess set l_1 is random with T_0 = 10, ΔT = 0.01 and the maximum random step is 20.

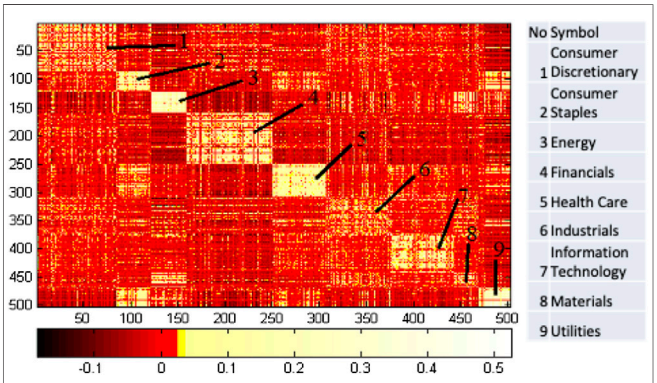


FIGURE 13 | The community detection results from the **Cg** correlation matrix mapping using the stock structure generated by the Simulated Annealing (SA) algorithm with parameters T_0 = 0 and ΔT = 0.01 and the maximum randomstep is 10. Group 1 is the consumer discretionary sector; group 2, namely the consumer staples sector; group 3, the energy sector; group 4, the financial sector; group 5, namely the healthcare sector; group 6, the industrial sector; group 7, namely the information technology sector; group 8, the materials sector; and finally group 9, the utilities sector.

provides the minimum system energy from group blocks in the correlation matrix.

In the SA algorithm, if the initial arrangement of the selected l_1 stock is random, the results obtained are far different from the ideal conditions (never reaching minimum energy when using the initial order that was sorted). For example, in **Figure 12**, a sorted initial arrangement provides a minimum energy of 6.5×10^4, whereas for a randomly selected initial arrangement guess, the minimum energy achieved at the minimum temperature is equal to 8.1×10^4.

Figure 13 shows the result obtained when the initial stock layout guess l_1 follows the order given in the eigenvector analysis, the initial temperature T_0 is 10, ΔT is 0.01, and the maximum

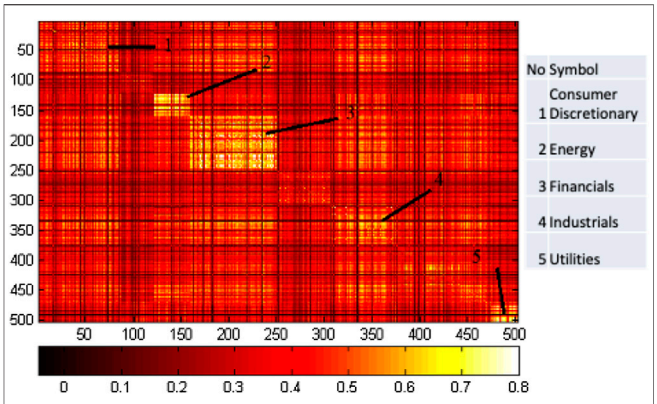


FIGURE 14 | Community detection results from the **Cg** + **Cm** correlation matrix mapping with the same stock structure as in **Figure 13**. The sectors are group 1 consumer, 2 energy, 3 financials, 4 industrials, and 5 utilities.

random step used is 10. However, if **Cg** is replaced with **Cg** + **Cm**, the detected group blocks are less than those using **Cg** only, although the parameters used are the same as in **Figure 14**. These results show that it is very important to perform filtering for **Cm** in addition to **Cr** in group analysis [7].

4. CONCLUDING REMARKS

We investigated complex networks in the S&P 500 stock market using two approaches, namely, a CN approach using an LPAm+ algorithm and a BDM approach using an SA algorithm. Before applying the two approaches, the data of the **C** stock correlation matrix were filtered using the RMT. RMT succeeded in separating the noise from non-noise data and showed that most of the data contained in the correlation matrix **C** were *noise*; an analysis of the distribution of eigenvector components in the RMT indicated that stock movements were driven by groups where each group was dominated by a particular sector. We called this analysis as simply RMT-CN-LPAm+ and RMT-BDM-SA.

In the first approach, the noise-free correlation matrix and *market-wide* **(Cg)** effects were analyzed using the CN approach with a threshold value of 0.05 and an LPAm+ network structure comprising six main groups with 485 out of a total 502 nodes involved (17 nodes were isolated from the network) and an obtained modularity value of 0.62. Then, in the second approach, which is a BDM with the same data, namely **Cg** using a *simulated annealing* algorithm, the stock structure provided a minimum energy system, and from this arrangement, nine groups of shares were produced. The decomposition of matrix **C** into **Cm** (market-wide), **Cg** (group), and **Cr** (noise) was also accomplished. The combination provides useful information to identify group classifications.

The difference between RMT-CN-LPAm+ and RMT-BDM-SA results is that in RMT-CN-LPAm+, a group contains not only the shares of the same sector but also of other minority sectors, whereas in RMT-BDM-SA, a group contains shares of the same sector. The second difference is that in MT-CN-LPAm+, a few shares still remain that have not joined any group, whereas in RMT-BDM-SA, not all shares have a group. In general, both

hybrid methods successfully show good performance to reveal detailed community detections.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

The authors would like to thank Hawoong Jeong, Kim Soo Yong (KAIST), and many colleagues with valuable discussions.

REFERENCES

1. Mantegna RN, Stanley HE. *Introduction to econophysics: correlations and complexity in finance.* Cambridge, United Kingdom, Cambridge University Press (1999).
2. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, Stanley HE. Random matrix approach to cross correlations in financial data. *Phys Rev* (2002) 65:066126. doi:10.1103/physreve.65.066126
3. Luo Y, Wang L, Sun S, Xia C. Community detection based on local information and dynamic expansion. *IEEE Access* (2018) 7:142773–86.
4. Wang Z, Xia C, Chen Z, Chen G. Epidemic propagation with positive and negative preventive information in multiplex networks. *IEEE Trans Cyber* (2020) [Epub ahead of print]. doi:10.1109/itoec49072.2020.9141736
5. Zhang Z, Xia C, Chen S, Yang T, Chen Z. Reachability analysis of networked finite state machine with communication losses: a switched perspective. *IEEE J Select Areas Commun* (2020) 38:845–53. doi:10.1109/jsac.2020.2980920
6. Plerou V, Gopikrishnan P, Rosenow B, Nunes Amaral LA, Stanley HE. Universal and nonuniversal properties of cross correlations in financial time series. *Phys Rev Lett* (1999) 83:1471. doi:10.1103/physrevlett.83.1471
7. Kim D-H, Jeong H. Systematic analysis of group identification in stock markets. *Phys Rev* (2005) 72:046133. doi:10.1103/physreve.72.046133
8. Mehta M. *Random matrices.* New York, NY: Academic (1991).
9. Wigner E. *On a class of analytic functions from the quantum theory of collisions annals of mathematics.* Berlin, Heidelberg: Springer (1951).
10. Namaki A, Shirazi AH, Raei R, Jafari GR. Network analysis of a financial market based on genuine correlation and threshold method. *Phys Stat Mech Appl* (2011) 390:3835–41. doi:10.1016/j.physa.2011.06.033
11. Sharifi S, Crane M, Shamaie A, Ruskin H. Random matrix theory for portfolio optimization: a stability approach. *Phys Stat Mech Appl* (2004) 335:629–43. doi:10.1016/j.physa.2003.12.016
12. Utsugi A, Ino K, Oshikawa M. Random matrix theory analysis of cross correlations in financial markets. *Phys Rev* (2004) 70:026110. doi:10.1103/physreve.70.026110
13. Newman MEJ. The structure and function of complex networks. *SIAM Rev* (2003) 45:167–256. doi:10.1137/s003614450342480
14. Newman ME, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E* (2004) 69:026113. doi:10.1103/physreve.69.026113
15. Jeong H. Complex scale-free networks. *Phys Stat Mech Appl* (2003) 321: 226–37. doi:10.1016/s0378-4371(02)01774-0
16. Wang XF, Chen G. Complex networks: small-world, scale-free and beyond. *IEEE Circuits Syst Magazine* (2003) 3:6–20
17. Noh J. D. (2000). Model for correlations in stock markets. *Phys Rev Rev E* 61: 5981. doi:10.1103/physreve.61.5981
18. Raghavan UN, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks. *Phys Rev E* (2007) 76:036106. doi:10.1103/physreve.76.036106
19. Barber MJ, Clark JW. Detecting network communities by propagating labels under constraints. *Phys Rev* (2009) 80:026129. doi:10.1103/physreve.80.026129
20. Liu X, Murata T. Advanced modularity-specialized label propagation algorithm for detecting communities in networks. *Phys Stat Mech Appl* (2010) 389:1493–1500. doi:10.1016/j.physa.2009.12.019
21. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science* (1983) 220:671–80. doi:10.1126/science.220.4598.671
22. Dréo J, Pétrowski A, Siarry P, Taillard E. *Metaheuristics for hard optimization: methods and case studies.* Berlin, Germany: Springer Science & Business Media (2006).
23. Jolliffe I. Principal component analysis *Encyclopedia Stat Behav Sci* (2002) 30(3):487, 2002. [Dataset].

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Purqon and Jamaludin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 12 February 2021
doi: 10.3389/fphy.2020.624373

Network-centric Indicators for Fragility in Global Financial Indices

Areejit Samal[1]*, *Sunil Kumar*[2], *Yasharth Yadav*[3] and *Anirban Chakraborti*[4,5,6]*

[1]The Institute of Mathematical Sciences (IMSc), Homi Bhabha National Institute (HBNI), Chennai, India, [2]Department of Physics, Ramjas College, University of Delhi, New Delhi, India, [3]Indian Institute of Science Education and Research (IISER), Pune, India, [4]School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India, [5]Centre for Complexity Economics, Applied Spirituality and Public Policy, Jindal School of Government and Public Policy, O. P. Jindal Global University, Sonipat, India, [6]Centro Internacional de Ciencias, Cuernavaca, México

Over the last 2 decades, financial systems have been studied and analyzed from the perspective of complex networks, where the nodes and edges in the network represent the various financial components and the strengths of correlations between them. Here, we adopt a similar network-based approach to analyze the daily closing prices of 69 global financial market indices across 65 countries over a period of 2000–2014. We study the correlations among the indices by constructing threshold networks superimposed over minimum spanning trees at different time frames. We investigate the effect of critical events in financial markets (crashes and bubbles) on the interactions among the indices by performing both static and dynamic analyses of the correlations. We compare and contrast the structures of these networks during periods of crashes and bubbles, with respect to the normal periods in the market. In addition, we study the temporal evolution of traditional market indicators, various global network measures, and the recently developed edge-based curvature measures. We show that network-centric measures can be extremely useful in monitoring the fragility in the global financial market indices.

Keywords: econophysics, correlation, networks, minimum spanning tree, market index

OPEN ACCESS

Edited by:
Mahdi Jalili,
RMIT University, Australia

Reviewed by:
Gholamreza Jafari,
Shahid Beheshti University, Iran
Chengyi Xia,
Tianjin University of Technology, China

***Correspondence:**
Areejit Samal
asamal@imsc.res.in
Anirban Chakraborti
anirban@jnu.ac.in

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *31 October 2020*
Accepted: *21 December 2020*
Published: *12 February 2021*

Citation:
Samal A, Kumar S, Yadav Y and Chakraborti A (2021) Network-centric Indicators for Fragility in Global Financial Indices. Front. Phys. 8:624373. doi: 10.3389/fphy.2020.624373

1 INTRODUCTION

It is possible to describe a financial market using the framework of complex networks such that the nodes in a network represent the financial components and an edge between any two components indicates an interaction between them. A correlation matrix constructed using the cross-correlations of fluctuations in prices can be utilized to identify such interactions. However, a network resulting from the correlation matrix contains densely connected structures. A growing amount of research is focused on methods devised to extract relevant correlations from the correlation matrix and study the topological, hierarchical, and clustering properties of the resulting networks. Mantegna et al. [1, 2] introduced the minimum spanning tree (MST) to extract networks from the correlation matrices computed from the asset returns. Dynamic asset trees, introduced by Onnela et al. [3, 4], were analyzed to monitor the evolution of financial stock markets using the hierarchical clustering properties of such trees. Boginski et al. [5] constructed threshold networks by extracting the edges with correlation values exceeding a chosen threshold and analyzed degree distribution, cliques, and independent sets on the threshold network. Tumminello et al. [6] introduced planar maximally filtered graph (PMFG) as a tool to extract important edges from the correlation matrix, which contains more information than the MST, while also preserving the hierarchical structure induced by

MST. Triangular loops and four-element cliques in PMFG could provide considerable insights into the structure of financial markets.

Network-based analysis has been widely used to study not only particular stock market structures but also the complex networks of correlations among different financial market indices across the globe. For example, MST has been used on stock markets to detect underlying hierarchical organization [7–9]. Bonanno et al. [10] studied the correlations of 51 global financial indices and showed that the corresponding MST was clustered according to the geographical locations of the indices. In addition, the changes in the topological structure of MST could help understand the evolution of financial systems [11–13]. MST and threshold networks have been used to analyze the indices during the global financial crisis of 2008 [14–16]. It has also been shown that geography is one of the major factors which govern the hierarchy of the global market [17, 18]. Also, Eryğit and Eryğit [19] had investigated the temporal evolution of clustering networks (MST and PMFG) of 143 financial indices corresponding to 59 countries across the world from the period 1995–2008 and once again found that the clustering in the networks of financial indices was according to their geographical locations. From the time-dependent network and centrality measures, they showed that the integration of the global financial indices has increased with time. Further, Chen et al. [20] analyzed dynamics of threshold networks of regional and global financial markets from the period 2012–2018, proposed a model for the measurement of systemic risk based on network topology, and then concluded that network-based methods provide a more accurate measurement of systemic risk compared to the traditional absorption technique. Silva et al. [21] studied the average criticality of countries during different periods in the crisis and found that the United States is the most critical country, followed by European countries, Oceanian and Asian countries, and finally Latin American countries and Canada. They also found a decrease in the network fragility after the global financial crisis. It has been also shown that financial crises can be captured using networks of volatility spillovers [22, 23]. Wang et al. [24] constructed and analyzed the dynamical structure of MSTs and hierarchical trees computed from the Pearson correlations as well as partial correlations, among 57 global financial markets from the period 2005–2014, and concluded that MST based on partial correlations provided more information when compared to MST based on the Pearson correlations. The market indices from different stock markets across the globe comprise assets that are very different; apart from stocks of the big multinational companies that are traded across markets, the stock markets would have little in common and hence would be expected to behave independently, in contrary to the previously reported empirical observations.

In this brief research report, we study the evolution of correlation structures among 69 global financial indices through the years 2000–2014. To ensure that we consider only the most relevant correlations, we construct the network by creating an MST (which connects all the nodes) and then add extra edges from the correlation matrix exceeding a certain threshold, which gives modular structures. Our findings corroborate the earlier results of geographical clustering [17, 25]. We then study the changes occurring in the market by analyzing the fluctuations in various global network measures and the recently developed edge-based geometric measures. Since there are complex interactions that occur among groups of three or more nodes, which cannot be described simply by pairwise interactions, the higher-order architecture of complex financial systems captured by the geometrical measures can help us in the betterment of systemic risk estimation and give us an indication of the global market efficiency. To the best of our knowledge, the present work is the first investigation of discrete Ricci curvatures in networks of global market indices. Thus, we find that this approach along with all these network measures can be used to monitor the fragility of the global financial network and as indicators of crashes and bubbles occurring in the markets. This could in turn relate the health of the financial markets to the development or downturn of the global economy, as well as gauging the impact of certain market crises on the multilevel financial-economic phenomena.

2 METHODS

2.1 Data Description

This study is based on a dataset collected from Bloomberg which comprises the daily closing prices of 69 global financial market indices from 65 countries, and this information was compiled for a period of $T = 3{,}513$ days over 14 years from January 11, 2000, to June 24, 2014. Note that the working days for different markets are not the same due to differences in holidays across countries. To overcome any inconsistencies due to this difference in working days, we filtered the data by removing days on which $> 30\%$ of the markets were not operative. Conversely, if $< 30\%$ of the markets were not operative on a day, we used the closing price of such markets on the previous day to complete the dataset. **Supplementary Table S1** lists the 69 global market indices considered here, along with their countries and geographical regions.

2.2 Cross-correlation Matrix and Market Indicators

Given the daily closing price $g_j(t)$ for market index j on day t, wherein $j = 1, 2, \ldots, N$ with $N = 69$ indices, we construct a time series of logarithmic returns as $r_j(t) = \ln g_j(t) - \ln g_j(t-1)$. Then, we construct the equal time Pearson cross-correlation matrix as

$$C_{ij}^{\tau}(t) = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sigma_i \sigma_j}, \tag{1}$$

where the mean and standard deviation are computed over a period of $\tau = 80$ days with end date as t. We also construct the ultrametric distance matrix with elements $D_{ij}^{\tau}(t) = \sqrt{2[1 - C_{ij}^{\tau}(t)]}$ that take values between 0 and 2. To study the temporal dynamics of the global market indices, we computed the correlation matrices for overlapping windows of $\tau = 80$ days with a

rolling shift of $\Delta\tau = 20$ days. Thence, we obtained 172 correlation frames between January 11, 2000, to June 24, 2014.

We have computed three market indicators from these correlation matrices. Firstly, the mean correlation gives the average of the correlations in the matrix $\mathbf{C}^{\tau}(t)$. Secondly, we have computed the eigenentropy [26] which involves the calculation of the Shannon entropy using the eigenvector centralities of the correlation matrix $\mathbf{C}^{\tau}(t)$ of market indices. Both mean correlation and eigenentropy have been shown to detect critical events in financial markets [26–28]. Thirdly, we have computed the risk corresponding to the Markowitz portfolio of the market indices, which is a proxy for the fragility or systemic risk of the global financial network [29]. A detailed description of the Markowitz portfolio optimization is given in the **Supplementary Material**.

2.3 Threshold Network Construction and Characteristics

The distance matrix for the time frame ending on t can be viewed as a complete, undirected, and weighted graph $\mathbf{D}^{\tau}(t)$, where the element $D_{ij}^{\tau}(t)$ is the weight of the edge between market indices i and j. To extract the important edges from $\mathbf{D}^{\tau}(t)$, we first construct its minimum spanning tree (MST) $\mathbf{M}^{\tau}(t)$ using Prim's algorithm [30]. As MST is an oversimplified network without cycles, it may lose crucial information on clusters or cliques. To overcome this, we add edges with correlation $C_{ij}^{\tau} \geq 0.65$ in $\mathbf{D}^{\tau}(t)$ to $\mathbf{M}^{\tau}(t)$ and obtain the threshold graph $\mathbf{S}^{\tau}(t)$. Thereafter, we study the temporal evolution of different network measures in $\mathbf{S}^{\tau}(t)$.

Firstly, we have computed standard global network measures such as the number of edges, edge density, average degree, average weighted degree [31], average shortest path length, diameter, average clustering coefficient [32], modularity [33, 34], communication efficiency [35], global reaching centrality (GRC) [36], network entropy [37], global assortativity [38, 39], and clique number. Note that the chosen set of global network measures studied here are by no means exhaustive and also depend very much on the specific questions of interest; see, for example, Wang et al. [40] for several gravitational centrality measures. Secondly, we have also computed four edge-centric curvature measures, namely, Ollivier-Ricci (OR) curvature [29, 41, 42], Forman-Ricci (FR) curvature [42–45], Menger-Ricci (MR) curvature [46, 47], and Haantjes-Ricci (HR) curvature [46, 47]. A detailed description of these network measures along with the appropriate natural weight, strength, or distance to use in each case is included in the **Supplementary Material**.

2.4 Multidimensional Scaling Map

The multidimensional scaling (MDS) technique tries to embed N objects in high-dimensional space into a low-dimensional space (typically, 2 or 3 dimensions), while preserving the relative distance between pairs of objects [48]. Here, we construct the (average) correlation matrix $\mathbf{C}^{T}$ between the 69 market indices for the complete period of $T = 3{,}513$ days between January 11, 2000, and June 24, 2014, using **Eq. 1**. Then, we compute the distance matrix $\mathbf{D}^{T}$ from $\mathbf{C}^{T}$ for the complete period. Thereafter, we use MDS to map the 69 market indices into a 2-dimensional space such that the distances between pairs of indices in $\mathbf{D}^{T}$ are preserved. To create the MDS plot, we used the in-built function cmdscale.m in MATLAB. Moreover, we also construct the MST $\mathbf{M}^{T}$ starting from the distance matrix $\mathbf{D}^{T}$ and then the threshold network $\mathbf{S}^{T}$ for the complete period from 2000 to 2014 by adding edges with $C_{ij}^{T} \geq 0.65$ to $\mathbf{M}^{T}$.

3 RESULTS AND DISCUSSION

The primary goal of this investigation is to evaluate different network measures for their potential to serve as indicators of fragility or systemic risk and monitor the health of the global financial system. For this purpose, we compiled a dataset of the daily closing prices of 69 global financial market indices from 65 different countries for a 14-year period from 2000 to 2014 (**Section 2**). Thereafter, we use the time series of the logarithmic returns of the daily closing prices for 69 global market indices to compute the Pearson cross-correlation matrices $\mathbf{C}^{\tau}(t)$ with a window size of $\tau = 80$ days with an overlapping shift of $\Delta\tau = 20$ days, and ending on trading days t (**Section 2**). Subsequently, we employ a minimum spanning tree (MST) based approach to construct 172 threshold networks $\mathbf{S}^{\tau}(t)$ corresponding to the cross-correlation matrices $\mathbf{C}^{\tau}(t)$ spanning the 14-year period (**Section 2**). Here, we study the temporal evolution of the structure of these correlation-based threshold networks $\mathbf{S}^{\tau}(t)$ of global market indices using several network measures and moreover contrast the evolution of network properties with generic market indicators such as mean correlation and minimum risk obtained using the Markowitz framework.

We reiterate that the threshold networks $\mathbf{S}^{\tau}(t)$ are constructed by computing the MST of the cross-correlation matrices $\mathbf{C}^{\tau}(t)$ followed by the addition of edges with correlation $C_{ij}^{\tau} \geq 0.65$ (**Section 2**). Intuitively, this network construction procedure ensures that each threshold network is a connected graph and captures the most relevant edges (correlations) between market indices. Since the obtained results may depend on the choice of the threshold (0.65) used for network construction, we present the temporal evolution of properties in networks constructed using 0.65 as threshold in the main text and in networks constructed using 0.75 or 0.85 as threshold in **Supplementary Material**. In the sequel, we will show that the qualitative nature of the obtained results is not very sensitive to the choice of 0.65, 0.75, or 0.85 as thresholds to construct the networks of global market indices.

In **Figures 1**, **2** and **Supplementary Figure S1**, we show the temporal evolution of generic indicators and network measures in the threshold networks of global market indices over the 14-year period (2000–2014). Moreover, the four shaded regions in **Figure 1** highlight four periods of the financial crisis, namely, US housing bubble, Lehman brothers crash, Dow Jones flash crash, and August 2011 stock markets fall. From **Figure 1**, it is seen that the mean correlation between market indices increases during periods of the financial crisis. Also, the eigenentropy

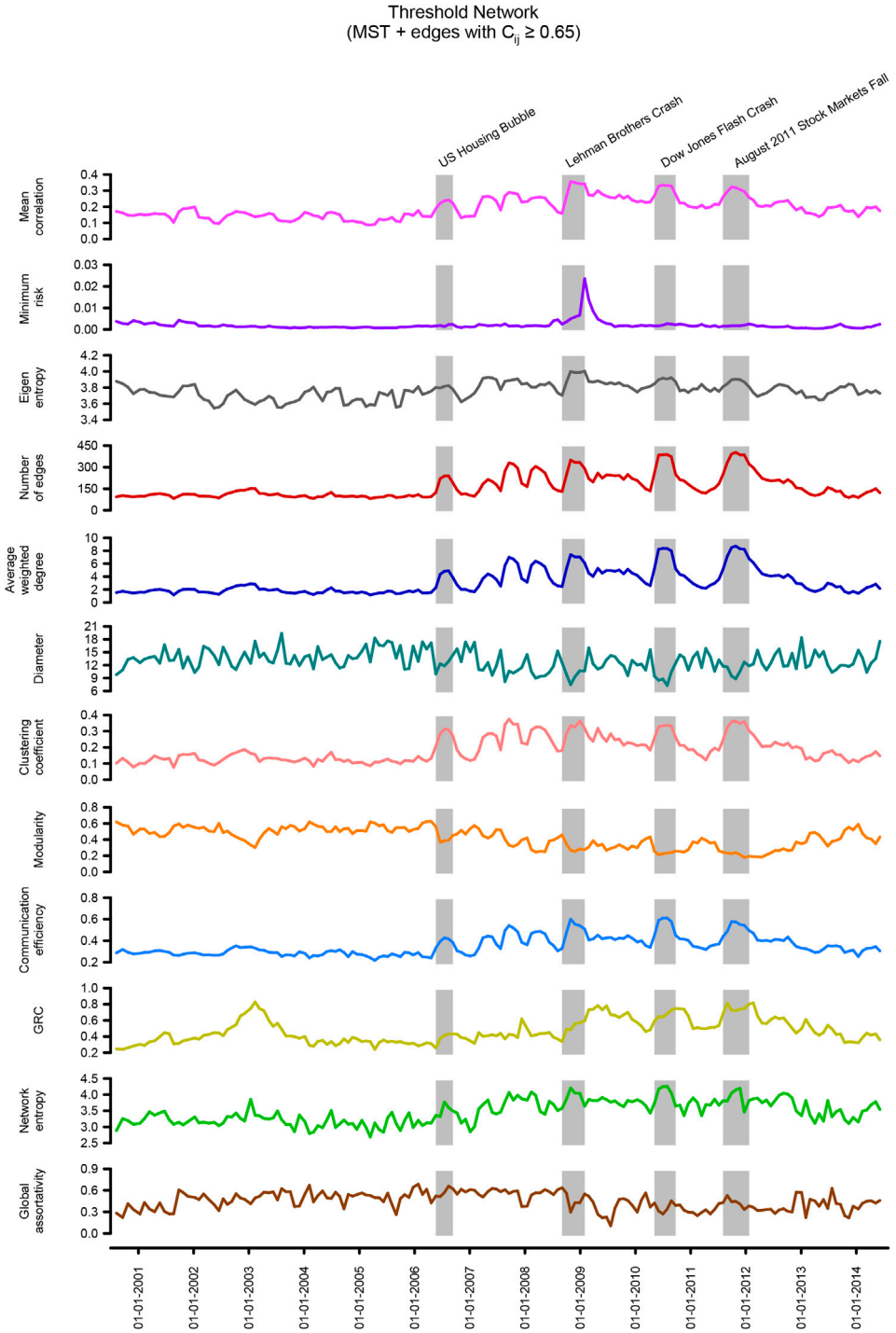


FIGURE 1 | Evolution of generic indicators and network characteristics for the global market indices networks $\mathbf{S}^{\tau}(t)$, constructed from the correlation matrices $\mathbf{C}^{\tau}(t)$ of window size $\tau = 80$ days and an overlapping shift of $\Delta\tau = 20$ days over a period of 14 years (2000–2014). The threshold networks $\mathbf{S}^{\tau}(t)$ were constructed by adding edges with correlation $C_{ij}^{\tau}(t) \geq 0.65$ to the minimum spanning trees (MST). From top to bottom, we compare the plots of mean correlation among market indices, minimum risk corresponding to the Markowitz portfolio optimization, eigenentropy, number of edges, average weighted degree, diameter, clustering coefficient, modularity, communication efficiency, global reaching centrality (GRC), network entropy, and global assortativity. The four shaded regions correspond to the epochs around the four important market events, namely, US housing bubble, Lehman brothers crash, Dow Jones flash crash, and August 2011 stock markets fall.

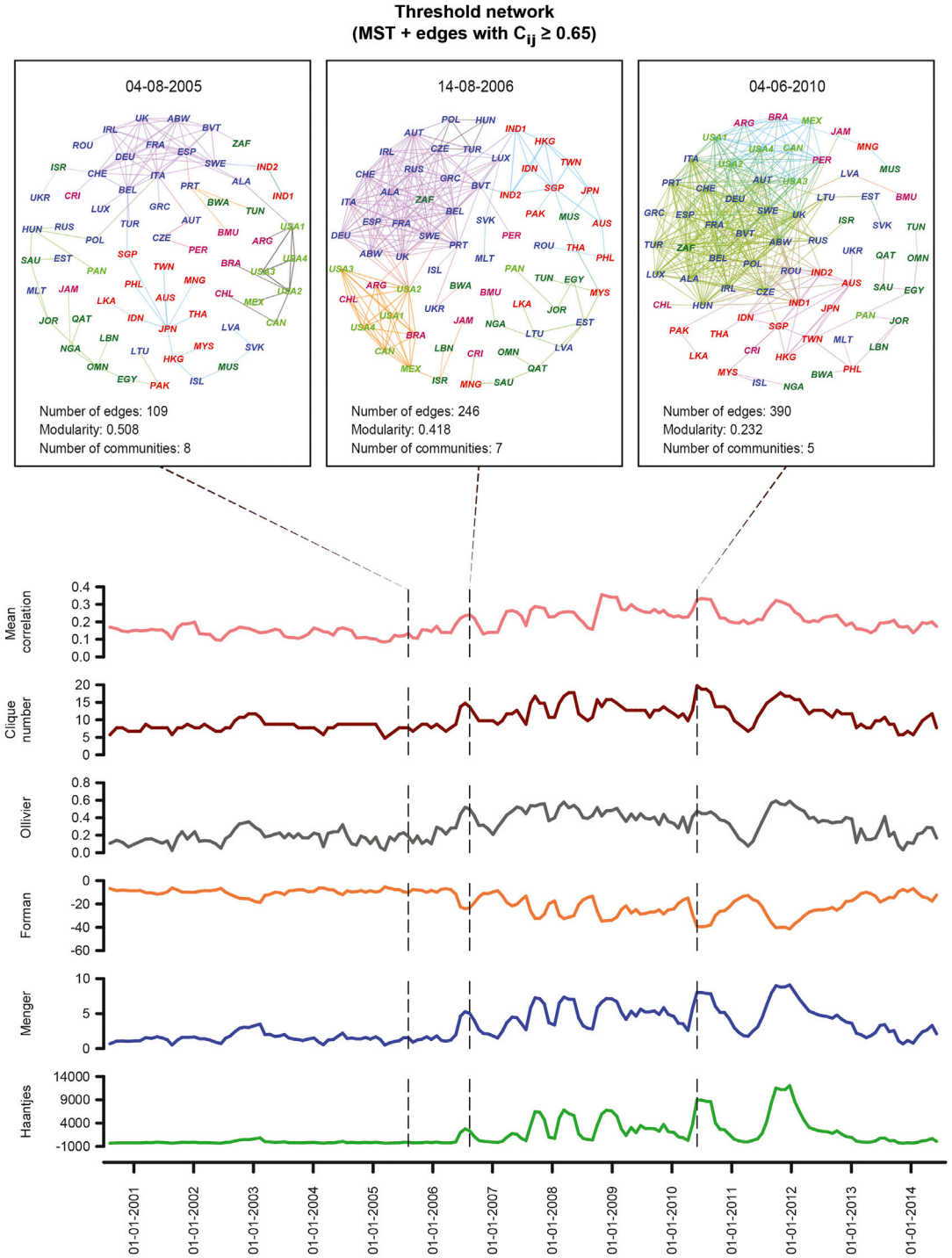


FIGURE 2 | Evolution of network characteristics and visualization of the threshold networks $\mathbf{S}^{\tau}(t)$ of market indices with a window size of $\tau = 80$ days and an overlapping shift of $\Delta\tau = 20$ days, constructed by adding edges with correlation $C_{ij}^{\tau}(t) \geq 0.65$ to the MST. (Lower panel) Comparison of the plots of mean correlation among market indices, clique number, average of Ollivier-Ricci (OR), Forman-Ricci (FR), Menger-Ricci (MR), and Haantjes-Ricci (HR) curvature of edges in threshold networks over the 14-year period. (Upper panel) Visualization of the threshold networks at three distinct epochs of $\tau = 80$ days ending on trading days t equal to August 4, 2005 (normal), August 14, 2006 (US housing bubble), and June 4, 2010 (Dow Jones flash crash). Threshold networks show a higher number of edges and a lower number of communities during the crisis. Correspondingly, there is an increase in mean correlation, clique number, average OR, MR, and HR curvature, and a decrease in average FR curvature of threshold networks during the financial crisis. Node colors and labels are based on geographical region and country, respectively, of the indices and edge colors are based on the communities determined by the Louvain method. The four United States market indices, NASDAQ, NYSE, RUSSELL1000, and SPX, are labeled as USA1, USA2, USA3, and USA4, respectively, while the two Indian indices, namely, NIFTY and SENSEX30, are labeled as IND1 and IND2, respectively.

which is directly computed from the correlation matrix $\mathbf{C}^{\tau}(t)$ increases during the crisis. Earlier works have shown that mean correlation and eigenentropy are indicators of instabilities in the stock market network [26, 28], and we show here that these measures can also serve as indicators of crisis in the network of global financial indices. In **Figure 1**, we also show the temporal evolution of the minimum risk corresponding to the portfolio comprising the market indices using the Markowitz framework. Moving on to widely used network properties, it is seen that the number of edges, edge density, average degree, average weighted degree, clustering coefficient, communication efficiency, and network entropy increase while diameter, average shortest path length, and modularity decrease during periods of the financial crisis (**Figure 1**; **Supplementary Figure S1**). In **Figure 1**, we also show the evolution of two other network measures, global reaching centrality (GRC) and global assortativity. In **Figure 2**, we also visualize the threshold network at three distinct time windows of $\tau = 80$ days ending on trading days t corresponding to August 4, 2005 (normal period), August 14, 2006 (US housing bubble crisis), and June 4, 2010 (Dow Jones flash crash), where the node colors are based on geographical regions of the market indices and edge colors are based on modules determined by Louvain method [34] for community detection. The identified communities in the three networks corresponding to the normal period, the US housing bubble, and the Dow Jones flash crash typically reflect the geographical proximity of financial market indices. For example, the indices of the United States, Canada, Mexico, Argentina, Brazil, and Chile form a single community in the threshold network for the normal period (**Figure 2**). It is evident that the number of edges in threshold networks corresponding to the US housing bubble (246 edges) or Dow Jones flash crash (390 edges) is much higher in comparison to that for the normal period (109 edges). In contrast, the modularity of threshold networks corresponding to the crisis periods, US housing bubble (0.418), or Dow Jones flash crash (0.232) is lower in comparison to that for normal period (0.508). In **Figure 2**, it is clearly seen that the clique number or size of the largest clique in threshold networks increases during financial crisis, and this is also evident from the network visualizations for normal period, US housing bubble, and Dow Jones flash crash. Note that bubbles are not easy to detect. In fact, our proposition is that holistic approaches with network measures, both node- and edge-based measures, including geometric curvatures, may help us to better detect and distinguish the bubbles from market crashes, as also pointed out in recent contributions [26, 49]. In sum, we find that during a normal period the network of global market indices is less connected, very modular, and heterogeneous, whereas, during a fragile period, the network is highly connected, less modular, and more homogeneous.

In addition to the node-centric global network measures described in the preceding paragraph, we have also studied edge-centric network measures, specifically, four discrete Ricci curvatures [Olivier-Ricci (OR), Forman-Ricci (FR), Menger-Ricci (MR), and Haantjes-Ricci (HR)] in threshold networks of global market indices. From **Figure 2**, it is seen that the average OR, MR, or HR curvature of edges increases during crisis periods in comparison to normal periods. In contrast, the average FR curvature of edges decreases during crisis periods in comparison to the normal periods. Notably, Sandhu et al. [29] have shown that OR curvature can serve as an indicator of fragility in stock market networks. However, to our knowledge, the present work is the first investigation of discrete Ricci curvatures in networks of global market indices. Note that different discretizations of Ricci curvature do not capture the entire features of the classical definition for continuous spaces, and thus, the four discrete Ricci curvatures studied here can capture different aspects of analyzed networks [42]. Overall, our results suggest that discrete Ricci curvatures can serve as indicators of fragility and monitor the health of the global financial system.

In **Figure 3**, we show the correlation between generic market indicators and different characteristics of the threshold networks $\mathbf{S}^{\tau}(t)$ of global market indices computed across the 14-year period from 2000 to 2014. From this figure, it is seen that eigenentropy and several network measures have a very high (absolute) Pearson correlation (≈ 0.9) with generic indicator, mean correlation of market indices. Such network measures include the number of edges, average weighted degree (strength), clustering coefficient, communication efficiency, clique number, FR curvature, and MR curvature. In contrast to mean correlation of market indices, there is moderate to no correlation between minimum risk corresponding to the portfolio comprising the market indices and eigenentropy or network measures (**Figure 3**). In sum, these results indicate that network measures including edge-centric FR curvature can be used to forecast crisis and monitor the health of the global financial system. To the best of our knowledge, our work is the largest survey of network measures to identify potential network-centric indicators of fragility in global financial market indices.

We must mention that though in the preceding paragraphs we have described only the results obtained from networks constructed using a threshold of 0.65, we have shown in **Supplementary Figures S2–S9** that the qualitative conclusions remain unchanged even when networks with a threshold of 0.75 and 0.85 are considered. In other words, our results are robust to the choice of the threshold used to construct the networks of global market indices.

In previous works, the econophysics community has employed either minimum spanning tree (MST) [7, 9–13, 15, 19] or planar maximally filtered graph (PMFG) [12, 19] or threshold networks [11, 14, 20] to study the correlation structure between global financial market indices. As far as we know, this work is the first to use threshold networks of MST plus edges with a correlation higher than a specified threshold, to study the temporal evolution of relationships between global financial market indices. In contrast, such threshold networks based on MST have been used earlier to study the structure of stock market networks [29, 49]. While MST has a tree structure without loops or cycles, PMFG or threshold network permits loops or cycles. In **Supplementary Text** and **Supplementary Figures S10–S13**, we also display the temporal evolution and correlation between generic market indicators and network measures in PMFG of global market indices constructed from

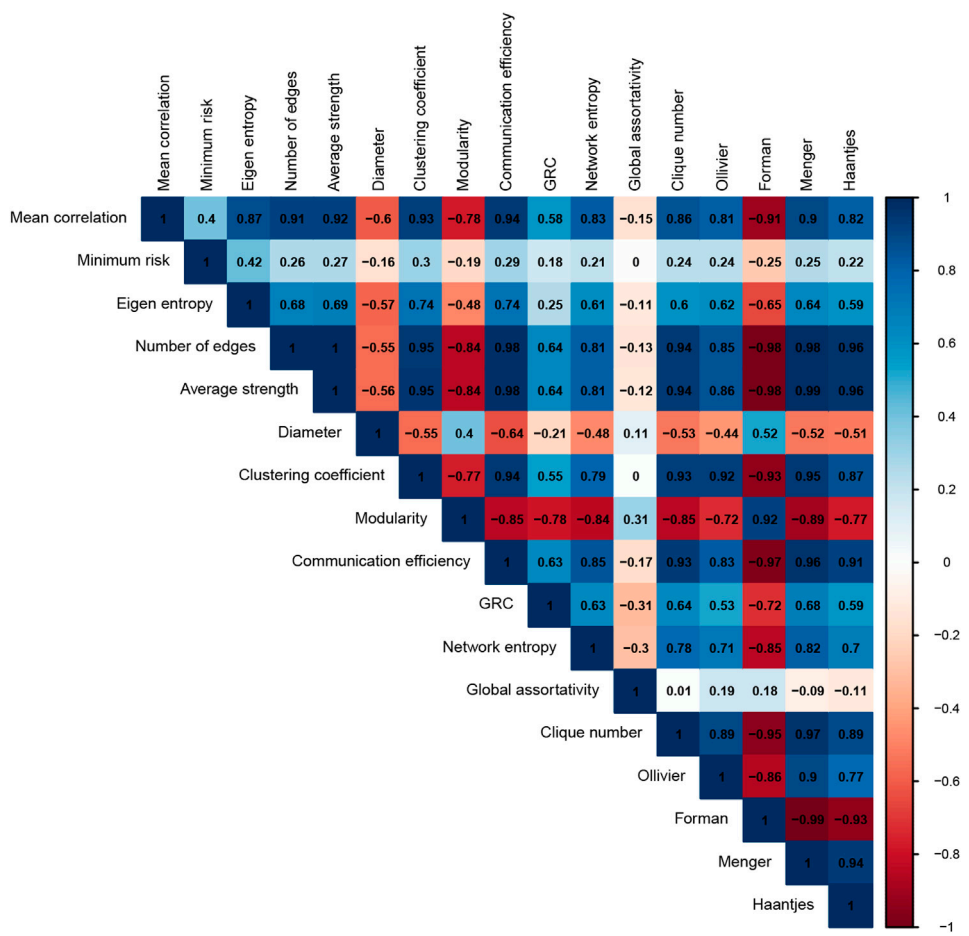


FIGURE 3 | Correlations between generic indicators and network characteristics of the global market indices networks $\mathbf{S}^{\tau}(t)$, constructed from the correlation matrices $\mathbf{C}^{\tau}(t)$ of window size $\tau = 80$ days and an overlapping shift of $\Delta\tau = 20$ days over a period of 14 years (2000–2014). The threshold networks $\mathbf{S}^{\tau}(t)$ were constructed by adding edges with correlation $C_{ij}^{\tau}(t) \geq 0.65$ to the minimum spanning tree (MST).

cross-correlation matrices $\mathbf{C}^{\tau}(t)$. While the construction of PMFG unlike threshold networks is independent of any specific choice of the threshold, the number of edges (thus, edge density and average degree) is fixed in case of PMFG (**Supplementary Figures S10, S11**). Due to this reason, we find that most of the network measures studied here are not correlated with the generic market indicator, mean correlation of market indices, in the PMFG case (**Supplementary Figure S13**). Still, we find that average weighted degree (strength), clustering coefficient, and communication efficiency have a very high correlation with the mean correlation of market indices in PMFG-based networks (**Supplementary Figure S13**). Based on these results, the threshold network construction based on MST plus edges with high correlation seems to be a better framework to monitor the state of the global financial system.

Finally, we have also studied the average correlation structure between global market indices over the 14-year period by computing the correlation matrix $\mathbf{C}^{T}$ between the 69 market indices by taking window size as the complete period of T days between 2000 and 2014 (**Section 2**). Subsequently, we have constructed a threshold network $\mathbf{S}^{T}$ corresponding to $\mathbf{C}^{T}$ by combining MST plus edges with a correlation above the chosen threshold of 0.65 (**Section 2**). In **Figure 4A**, we visualize this overall threshold network $\mathbf{S}^{T}$ of market indices for the complete 14-year period of T days. In this figure, the node colors are based on geographical regions of the market indices and edge colors are based on communities obtained from the Louvain method. In **Figure 4B**, we have separated the communities in this overall threshold network $\mathbf{S}^{T}$ of market indices by removing the intermodule edges in the visualization. From **Figures 4A,B**, it is clear that the market indices form communities in this overall threshold network based on their geographical proximity. Moreover, we have also employed multidimensional scaling (MDS) technique to map

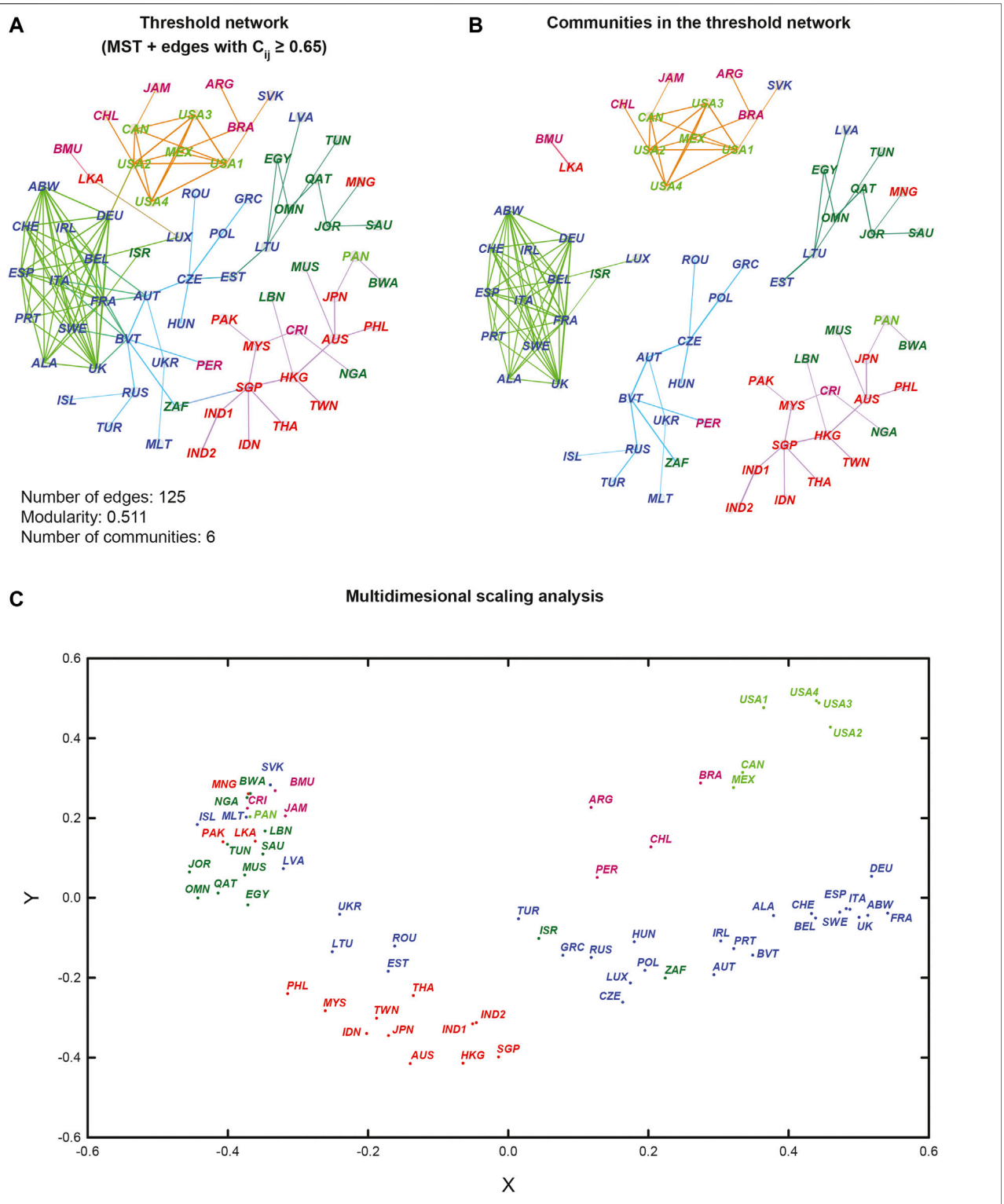


FIGURE 4 | The average correlation structure between 69 market indices over the 14-year period is visualized based on the correlation matrix $\mathbf{C}^T$ for the complete period of T = 3513 days between 2000 to 2014. **(a)** Visualization of the overall threshold network $\mathbf{S}^T$ corresponding to $\mathbf{C}^T$ obtained by combining MST plus edges with correlation ≥0.65. The node colours are based on geographical regions of the market indices and edge colours are based on communities obtained from Louvain method. **(b)** Visualization of the communities in the overall threshold network $\mathbf{S}^T$ after removing the inter-module edges. It is evident that the market indices form communities in this network based on their geographical proximity. **(c)** Multidimensional scaling (MDS) map in 2-dimensions of the 69 market indices. In this figure, the indices are labelled in different colours based on their geographical region and country, respectively. The four USA market indices, NASDAQ, NYSE, RUSSELL1000 and SPX, are labelled as USA1, USA2, USA3 and USA4, respectively, while the two Indian indices, NIFTY and SENSEX30, are labelled as IND1 and IND2, respectively.

the 69 market indices into a 2-dimensional space such that the distances between pairs of indices are preserved (**Figure 4C**; **Section 2**). It can be seen that the MDS map is able to partition the 69 market indices into groups based on their geographical proximity, and further, the structure in the MDS map has a close resemblance to the community structure of the overall threshold network (**Figure 4**). For example, the grouping of indices from the United States, Canada, Mexico, Argentina, Brazil, and Chile can be seen in both the threshold network and MDS map (**Figure 4**). Interestingly, when we plotted in **Supplementary Figure S14** the evolution of the eigenvector centralities of the nodes (market indices), as well as their OR and FR curvature, we found that there exist certain periods of time when some of the countries in close geographical proximity display high (absolute) values and others display low values, indicative of the changes in the complex interactions and community structures.

4 SUMMARY AND CONCLUDING REMARKS

In summary, we have investigated the daily closing prices of 69 global financial indices over a 14-year period using various techniques of cross-correlations based network analysis. We have been able to continuously monitor the complex interactions among the global market indices by using a variety of network-centric measures, including recently developed edge-centric discrete Ricci curvatures. In the present study of the global market indices, the novelty lies in the following: i) Construction of the threshold network $\mathbf{S}^{\tau}(t)$, as superposition of the MST of the cross-correlation matrix and the network of edges with correlations $C_{ij}^{\tau} \geq 0.65$, which ensures that each threshold network is a connected graph and captures the most relevant edges (correlations) between market indices. In **Supplementary Material**, we have also reported the results for networks constructed using MST and two other threshold values, i.e., $C_{ij}^{\tau} \geq 0.75$ and $C_{ij}^{\tau} \geq 0.85$. Besides, we have also reported results for networks constructed using PMFG method. ii) The usage of discrete Ricci curvatures in networks of global market indices, which capture the higher-order architecture of the complex financial system. To the best of our knowledge, this is the first study employing edge-based discrete Ricci curvatures to networks of global financial indices. Our recent work underscores the utility of edge-based curvature measures in the analysis of networks of stocks [49] or global financial indices. In future, curvature measures may also find application in other financial networks including Banking networks [50]. iii) The largest yet by no means exhaustive survey of network measures to identify potential network-centric indicators of fragility and systemic risk in the system of global financial market indices.

The global financial system has become increasingly complex and interdependent and thus prone to sudden unpredictable changes like market crises. Our results, compared to the traditional market indicators, do provide a deeper understanding of the system of global financial markets. Specially, we find that the four discrete Ricci curvatures can be effectively used as indicators of fragility in global financial markets. We reiterate that the methods used in this work can detect instabilities in the market and can be used as early warning signals so that policies can be made in order to prevent the occurrence of such events in the future.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are subject to the following licenses/restrictions. The codes used to construct the networks from correlation matrices and compute the different network measures are publicly available via the GitHub repository: https://github.com/asamallab/FinNetIndicators. The relevant correlation matrices constructed from proprietary data from Bloomberg are available upon reasonable requests. Requests to access these datasets should be directed to anirban@jnu.ac.in.

AUTHOR CONTRIBUTIONS

AS, SK, and AC conceived the project. AS, SK, YY, and AC performed the computations. SK compiled the dataset. YY and AS prepared the figures and tables. AS and AC analyzed the results. AS, SK, YY, and AC wrote the manuscript. All authors have read and approved the manuscript.

FUNDING

We have received no specific research funding for this work. AS acknowledges financial support from Max Planck Society Germany through the award of a Max Planck Partner Group in Mathematical Biology and a Ramanujan fellowship (SB/S2/RJN006/2014) from the Science and Engineering Research Board (SERB), India.

ACKNOWLEDGMENTS

AC acknowledges support from the project UNAM-DGAPA-PAPIIT AG 100819 and CONACyT Project FRONTERAS 201.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.624373/full#supplementary-material.

REFERENCES

1. Mantegna RN. Information and hierarchical structure in financial markets. *Comput Phys Commun* (1999) 121–122:153–6. doi:10.1016/S0010-4655(99)00302-1
2. Mantegna RN. Hierarchical structure in financial markets. *Eur Phys J* (1999) 11:193–7. doi:10.1007/s100510050929
3. Onnela JP, Chakraborti A, Kaski K, Kertsz J, Kanto A. Asset trees and asset graphs in financial markets. *Phys Scr* (2003) T106:48. doi:10.1238/physica.topical.106a00048
4. Onnela JP, Kaski K, Kertész J. Clustering and information in correlation based financial networks. *Eur Phys J* (2004) 38:353–62. doi:10.1140/epjb/e2004-00128-7
5. Boginski V, Butenko S, Pardalos PM. Statistical analysis of financial networks. *Comput Stat Data Anal* (2005) 48:431–43. doi:10.1016/j.csda.2004.02.004
6. Tumminello M, Aste T, Di Matteo T, Mantegna RN. A tool for filtering information in complex systems. *Proc Natl Acad Sci USA* (2005) 102:10421–6. doi:10.1073/pnas.0500298102
7. Bonanno G, Caldarelli G, Lillo F, Mantegna RN. Topology of correlation-based minimal spanning trees in real and model markets. *Phys Rev E* (2003) 68:046130. doi:10.1103/PhysRevE.68.046130
8. Onnela JP, Chakraborti A, Kaski K, Kertész J, Kanto A. Dynamics of market correlations: taxonomy and portfolio analysis. *Phys Rev E* (2003) 68:056110. doi:10.1103/PhysRevE.68.056110
9. Bonanno G, Caldarelli G, Lillo F, Micciché S, Vandewalle N, Mantegna RN. Networks of equities in financial markets. *Eur Phys J* (2004) 38:363–71. doi:10.1140/epjb/e2004-00129-6
10. Bonanno G, Vandewalle N, Mantegna RN. Taxonomy of stock market indices. *Phys Rev E* (2000) 62:R7615–8. doi:10.1103/PhysRevE.62.R7615
11. Nobi A, Lee S, Kim DH, Lee JW. Correlation and network topologies in global and local stock indices. *Phys Lett* (2014) 378:2482–9. doi:10.1016/j.physleta.2014.07.009
12. Wang J, He J. Correlation and interdependence structure in stock market: based on information theory and complex networks. 2017 17th international conference on control, automation and systems (ICCAS); 18–21 Oct 2017; Jeju, South Korea. IEEE (2017). 498–504. doi:10.23919/ICCAS.2017.8204288
13. Coelho R, Gilmore CG, Lucey B, Richmond P, Hutzler S. The evolution of interdependence in world equity markets—evidence from minimum spanning trees. *Phys A* (2007) 376:455–66. doi:10.1016/j.physa.2006.10.045
14. Kumar S, Deo N. Correlation and network analysis of global financial indices. *Phys Rev E* (2012) 86:026101. doi:10.1103/PhysRevE.86.026101
15. Junior L, Mullokandov A, Kenett D. Dependency relations among international stock market indices. *J Risk Financ Manag* (2015) 8:227–65. doi:10.3390/jrfm8020227
16. Lee JW, Nobi A. State and network structures of stock markets around the global financial crisis. *Comput Econ* (2018) 51:195–210. doi:10.1007/s10614-017-9672-x
17. León C, Kim G, Martínez C, Lee D. Equity markets' clustering and the global financial crisis. *Quant Finance* (2017) 17:1–18. doi:10.1080/14697688.2017.1357970
18. Saeedian M, Jamali T, Kamali MZ, Bayani H, Yasseri T, Jafari GR. Emergence of world-stock-market network. *Phys A* (2019) 526:120792. doi:10.1016/j.physa.2019.04.028
19. Eryiğit M, Eryiğit R. Network structure of cross-correlations among the world market indices. *Phys A* (2009) 388:3551–62. doi:10.1016/j.physa.2009.04.028
20. Chen L, Han Q, Qiao Z, Stanley HE. Correlation analysis and systemic risk measurement of regional, financial and global stock indices. *Phys A* (2020) 542:122653. doi:10.1016/j.physa.2019.122653
21. Silva TC, de Souza SRS, Tabak BM. Structure and dynamics of the global financial network. *Chaos Solitons Fractals* (2016) 88:218–34. doi:10.1016/j.chaos.2016.01.023
22. Baumohl E, Evžen Kocenda E, Lyocsa S, Vyrost T. Networks of volatility spillovers among stock markets. *Phys A* (2018) 490:1555–74. doi:10.1016/j.physa.2017.08.123
23. Mensi W, Boubaker FZ, Al-Yahyaee KH, Kang SH. Dynamic volatility spillovers and connectedness between global, regional, and gipsi stock markets. *Finance Res Lett* (2018) 25:230–8. doi:10.1016/j.frl.2017.10.032
24. Wang GJ, Xie C, Stanley HE. Correlation structure and evolution of world stock markets: evidence from pearson and partial correlation-based networks. *Comput Econ* (2018) 51:607–35. doi:10.1007/s10614-016-9627-7
25. Sharma K, Chakrabarti AS, Chakraborti A. Multi-layered network structure: relationship between financial and macroeconomic dynamics. In: F Abergel, BK Chakrabarti, A Chakraborti, N Deo, K Sharma, editors *New perspectives and challenges in econophysics and sociophysics* Gewerbestr, Switzerland: Springer (2019). p 117–31. doi:10.1007/978-3-030-11364-3_9
26. Chakraborti A, Unni HN, Sharma K, Pharasi HK. Phase separation and scaling in correlation structures of financial markets. *J Phys Complexity* (2020) 2:015002 doi:10.1088/2632-072X/abbed1
27. Chakraborti A, Sharma K, Pharasi HK, Shuvo Bakar K, Das S, Seligman TH. Emerging spectra characterization of catastrophic instabilities in complex systems. *New J Phys* (2020) 22:063043. doi:10.1088/1367-2630/ab90d4
28. Kukreti V, Pharasi HK, Gupta P, Kumar S. A perspective on correlation-based financial networks and entropy measures. *Front Phys* (2020) 8:323. doi:10.3389/fphy.2020.00323
29. Sandhu RS, Georgiou TT, Tannenbaum AR. Ricci curvature: an economic indicator for market fragility and systemic risk. *Sci Adv* (2016) 2:e1501495. doi:10.1126/sciadv.1501495
30. Prim RC. Shortest connection networks and some generalizations. *Bell Syst Tech J* (1957) 36:1389–401 doi:10.1002/j.1538-7305.1957.tb01515.x
31. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. *Proc Natl Acad Sci USA* (2004) 101:3747–52. doi:10.1073/pnas.0400087101
32. Onnela JP, Saramäki J, Kertész J, Kaski K. Intensity and coherence of motifs in weighted complex networks. *Phys Rev E* (2005) 71:065103. doi:10.1103/PhysRevE.71.065103
33. Girvan M, Newman ME. Community structure in social and biological networks. *Proc Natl Acad Sci USA* (2002) 99:7821–6. doi:10.1073/pnas.122653799
34. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theor Exp* (2008) 2008:P10008. doi:10.1088/1742-5468/2008/10/P10008
35. Latora V, Marchiori M. Efficient behavior of small-world networks. *Phys Rev Lett* (2001) 87:198701. doi:10.1103/PhysRevLett.87.198701
36. Mones E, Vicsek L, Vicsek T. Hierarchy measure for complex networks. *PLos One* (2012) 7:e33799 doi:10.1371/journal.pone.0033799
37. Solé RV, Valverde S. Information theory of complex networks: on evolution and architectural constraints. In: E Ben-Naim, H Frauenfelder, Z Toroczkai, editors *Complex networks* Berlin, Germany: Springer (2004). p 189–207. doi:10.1007/978-3-540-44485-5_9
38. Newman MEJ. Mixing patterns in networks. *Phys Rev E* (2003) 67:026126. doi:10.1103/PhysRevE.67.026126
39. Leung CC, Chau H. Weighted assortative and disassortative networks model. *Phys Stat Mech Appl* (2007) 378:591–602. doi:10.1016/j.physa.2006.12.022
40. Wang J, Li C, Xia C. Improved centrality indicators to characterize the nodal spreading capability in complex networks. *Appl Math Comput* (2018) 334:388–400. doi:10.1016/j.amc.2018.04.028
41. Ollivier Y. Ricci curvature of metric spaces. *Compt Rendus Math* (2007) 345:643–6. doi:10.1016/j.crma.2007.10.041
42. Samal A, Sreejith RP, Gu J, Liu S, Saucan E, Jost J. Comparative analysis of two discretizations of Ricci curvature for complex networks. *Sci Rep* (2018) 8:8650. doi:10.1038/s41598-018-27001-3
43. Forman R. Bochner's method for cell complexes and combinatorial Ricci curvature. *Discrete Comput Geom* (2003) 29:323–74. doi:10.1007/s00454-002-0743-x
44. Sreejith RP, Mohanraj K, Jost J, Saucan E, Samal A. Forman curvature for complex networks. *J Stat Mech Theor Exp* (2016) 2016:P063206. doi:10.1088/1742-5468/2016/06/063206
45. Saucan E, Sreejith RP, Vivek-Ananth RP, Jost J, Samal A. Discrete Ricci curvatures for directed networks. *Chaos Solitons Fractals* (2019) 118:347–60. doi:10.1016/j.chaos.2018.11.031
46. Saucan E, Samal A, Jost J. A simple differential geometry for networks and its generalizations. In: H Cherifi, S Gaito, JF Mendes, E Moro, LM Rocha, editors *International conference on complex networks and their applications* Gewerbestr, Switzerland: Springer (2019b) p 943–54. doi:10.1007/978-3-030-36687-2_78

47. Saucan E, Samal A, Jost J. A simple differential geometry for complex networks. *Network Sci* (2020) 1–28. doi:10.1017/nws.2020.42
48. Borg I, Groenen PJF. *Modern multidimensional scaling: theory and applications* New York, NY: Springer Science & Business Media (2005).
49. Samal A, Pharasi HK, Ramaia SJ, Kannan H, Saucan E, Jost J, et al. Network geometry and market instability. *R Soc Open Sci* (2021) 8:201734. doi:10.1098/rsos.201734
50. Namaki A, Ardalankia J, Raei R, Hedayatifar L, Hosseiny A, Haven E, et al. Analysis of the global banking network by random matrix theory. *arXiv: 2009.12335* (2020).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Samal, Kumar, Yadav and Chakraborti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 04 March 2021
doi: 10.3389/fphy.2021.572216

Using Topological Data Analysis (TDA) and Persistent Homology to Analyze the Stock Markets in Singapore and Taiwan

Peter Tsung-Wen Yen[1] *and Siew Ann Cheong*[2,3]*

[1]Energy Research Institute @ NTU (ERI@N), Singapore, Singapore, [2]Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore, [3]Complexity Institute, Nanyang Technological University, Singapore, Singapore

In recent years, persistent homology (PH) and topological data analysis (TDA) have gained increasing attention in the fields of shape recognition, image analysis, data analysis, machine learning, computer vision, computational biology, brain functional networks, financial networks, haze detection, etc. In this article, we will focus on stock markets and demonstrate how TDA can be useful in this regard. We first explain signatures that can be detected using TDA, for three toy models of topological changes. We then showed how to go beyond network concepts like nodes (0-simplex) and links (1-simplex), and the standard minimal spanning tree or planar maximally filtered graph picture of the cross correlations in stock markets, to work with faces (2-simplex) or any k-dim simplex in TDA. By scanning through a full range of correlation thresholds in a procedure called filtration, we were able to examine robust topological features (i.e. less susceptible to random noise) in higher dimensions. To demonstrate the advantages of TDA, we collected time-series data from the Straits Times Index and Taiwan Capitalization Weighted Stock Index (TAIEX), and then computed barcodes, persistence diagrams, persistent entropy, the bottleneck distance, Betti numbers, and Euler characteristic. We found that during the periods of market crashes, the homology groups become less persistent as we vary the characteristic correlation. For both markets, we found consistent signatures associated with market crashes in the Betti numbers, Euler characteristics, and persistent entropy, in agreement with our theoretical expectations.

Keywords: topological data analysis, econophysics, applied topology, financial market, STI, TAIEX

OPEN ACCESS

Edited by:
Sitabhra Sinha,
Institute of Mathematical Sciences, Chennai, India

Reviewed by:
Indrava Roy,
Institute of Mathematical Sciences, India
Zbigniew R. Struzik,
The University of Tokyo, Japan

***Correspondence:**
Siew Ann Cheong
cheongsa@ntu.edu.sg

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *13 June 2020*
Accepted: *11 January 2021*
Published: *04 March 2021*

Citation:
Yen PT-W and Cheong SA (2021) Using Topological Data Analysis (TDA) and Persistent Homology to Analyze the Stock Markets in Singapore and Taiwan. Front. Phys. 9:572216. doi: 10.3389/fphy.2021.572216

INTRODUCTION

The earliest success of econophysics is the application of random matrix theory (RMT, which is a theory combining nuclear physics and statistical mechanics) to the stock market [1–4]. In RMT, one treats noise as a kind of symmetry, and thus information represents some form of symmetry breaking. RMT thus allows physicists to discriminate between noise and signal. The next significant milestone in econophysics is the realization that stock returns follow heavy-tailed Levy distributions [5] instead of a normal distribution. Also, their dynamical properties can be described in terms of fractals (in terms for example, of the Hurst exponent) and multifractals [6, 7] instead of the random walk proposed by Bachelier to model price movements in the stock market. Physicists also love to

strip problems down to their simplest essence, using information filtering approaches such as the minimal spanning tree (MST) [8], planar maximally filtered graph (PMFG) [9], triangular maximally filtered graph (TMFG) [10], etc. These represent some of the methodological contributions by econophysicists.

In the PMFG method, important correlations are projected onto a sphere, which has genus $g = 0$. This is a good starting point for understanding the correlated price movements between different stocks. However, it is possible that the pattern of dynamic correlations may be explainable more naturally in terms of some nontrivial geometrical structure with $g > 0$. Therefore, the determination of the optimum genus represents a gap in our understanding of correlations in the stock market. This best genus can change with time for the same window length. It can change with window length over the same period, and it can also depend on which market we are looking at. There is also a second gap in our understanding of these dynamic correlations, and that is the problem of overlapping communities. Information filtering methods like the MST and PMFG are not clustering algorithms, but there are clustering algorithms based off them. There are also standard clustering algorithms like k-means and hierarchical clustering that can be used to study the correlation structure in a market. However, all clustering algorithms assume that a stock can be a member-only of one cluster. Ultimately, classifying stocks into clusters help us better imagine the geometry of the correlations, but we do not claim that clusters are independent of each other. We know that within clusters, the interactions are stronger, and between clusters, the interactions are weaker. Recently, researchers started to realize that in many cases, nodes can belong to more than one cluster, giving rise to the problem of overlapping clusters. Currently, the identification of the correct overlapping structures without sacrificing accuracy and speed remains a daunting challenge. These hinder a deeper understanding of co-authorship networks, protein-protein yeast networks, and word association networks. Topological data analysis (TDA) is a method that will kill both birds with one stone. It is ideally suited to 1) identify geometrical structures that are like clusters, and 2) elucidate the weak connections between them.

So how do TDA concepts like *simplicial complexes* and *persistent homology* help in filling these gaps? First, once the size of the sliding window is decided, TDA can be quite robust in deciding which topological space or genus to use for projecting the correlation matrices. Indeed, TDA lets the data speak for itself on choosing the optimal topological space and the value of the genus. Second, by appealing to persistency, we do not presuppose which correlation threshold value to use. Instead, we scan through a full range of correlation threshold values, to determine which range the topological structure is most persistent. Third, TDA homologies are very robust to random noises, and as a result, we can avoid technical nuisances such as 'accumulation of noises' or 'overfitting the data' when clustering data in higher dimensions. Lastly, persistent homology can be presented in the form of persistence barcodes, persistence diagrams, persistence landscapes.

In this paper, our research problem is to use TDA to understand topological changes accompanying crashes in the Singapore and Taiwan stock markets in terms of simplicial complexes, persistent homology, and other metrics. Our hypothesis is that in different market states, different topological features emerge, and TDA can be effective in elucidating these changes. This paper is organized as follows: In **Data** Section, we briefly introduce how to collect data on the daily returns of the Straits Times Index (STI), the Taiwan Capitalization Weighted Stock Index (TAIEX), and how to preprocess them. In **Topology, and Persistent Homology** Section we introduce the mathematical background of simplicial complexes, persistent homology, and filtration. In **TDA Toolkits** Section, we introduce TDA toolkits like barcodes, persistent diagrams, *Betti numbers*, and *Euler characteristics*. In **TDA of Toy Models and Hypothesis on Real Markets** Section, we introduce toy models of TDA and the hypothesis on real markets. In **Results and Discussion** Section, we show our numerical results computed by TDA and discuss how they confirm our hypothesis. Finally, in **Conclusion** Section, we give concluding remarks and perspectives.

MATERIALS AND METHODS

Data

Data Collection

First, we show how to collect price data from stocks in the Singapore Exchange (SGX) [Taiwan Stock Exchange (TWSE)] using Python pandas, and its function web. DataReader, and use the Yahoo Finance API option. Second, to use this option, we need to prepare all the tickers symbols in SGX (TWSE). The procedure is as follows: 1) we go to the 'My Screeners' tab in https://sg.finance.yahoo.com/, and choose 'Singapore' in the 'Saved Screeners/Region' tab, before choosing 'Find Stocks' to see a list of ticker symbols. For SGX, there are 672 ticker symbols; 2) copy and save all of them into a file, and 3) using this file of ticker symbols and pandas. web.DataReader's Yahoo API to fetch historical data between January 1, 2017 and April 30, 2019 from the Yahoo Finance database and save as a CSV file. The Python code to do so is shown in Code 1, and this code can be modified to the TWSE (January 1, 2017 to March 31, 2020) or other markets.

Data Cleaning and Preprocessing

After we collected the raw data, the data needed to be cleaned. First, some ticker symbols are duplicated, so we keep only one copy. Second, for some ticker symbols, the Yahoo Finance API gave an error and caused the program to halt, so we needed to identify these and removed them from the ticker symbol list. Finally, some of the data may include 'NaN's and we needed to replace them with '0's. However, if the time series contains more than 50% '0's, we also remove this ticker symbol from the list. After cleaning, we ended up with the times series data for 560 distinct stocks.

Before we computed the cross correlations between stocks from the time-series data to obtain the correlation matrices, three procedures are necessary. First, we standardized the daily prices, which is $\delta x_i = \frac{x_i - \overline{x_i}}{\sqrt{\sum (x_i - \overline{x_i})^2/(t-1)}}$, where x_i is the raw stock price for

```
def get_data(dataname, start_year,start_date, end_date):
    companies = pd.read_csv("csv/tickers_"+dataname+".csv")
    companies["Ticker"] = companies["Tickers"]
    companies = companies.set_index("Tickers")
    start_date, end_date = datetime.datetime(start_year, 1, 1),
    datetime.datetime(2019,6,30)

    #get Adj Close prices and save
    data = web.DataReader(companies["Ticker"], 'yahoo', start_date,
    end_date)["Adj Close"];
    data.to_csv('SGX_raw.csv')
```

CODE 1. | A Python code that implements the data collection procedure.

the ith stock, $\overline{x_l}$ is the average stock price for the ith stock, $t = 120$ is the number of trading days over a six-month period. Second, we smoothed the time series by averaging over a sliding time window of 15 days (for a detailed explanation on why we choose a 15-days window, please see **Supplementary Figure S1**). Lastly, we converted the daily stock prices to their derivatives, i.e. $\Delta x_i(t) = \delta x_i(t) - \delta x_i(t-1)$. In **Figure 1A** we show the stock price derivatives within a 6-month period after pre-processing, and in **Figure 1B**, we show the correlation matrix generated from the derivative data. We converted the correlation matrix to a distance matrix using the formula $d_{ij} = \sqrt{2(1-\rho_{ij})}$. Finally, we generated distance matrices for successive 6-month periods that are one month apart, to use as input data for subsequent TDA calculations. Other data formats acceptable for TDA include point clouds, networks, or digital images. To be more clear, the procedures are shown in a flowchart (**Figure 2**).

Topology, and Persistent Homology

TDA is a mathematical apparatus developed by Herbert Edelsbrunner, Afra Zomorodian, Gunnar Carlsson, and his graduate student Gurjeet Singh [11–13]; it was popularized by Carlsson's paper [14] that later turned TDA into a hot field in applied mathematics, and also found many applications in data analytics. The foundations of TDA had been laid years before by others in the fields of topology [15–19], group theory [20, 21], linear algebra [22, 23], and graph theory [24–26].

To explain the concept of persistent homology, imagine we have collected a bunch of data points that we refer to as a data cloud. Next, imagine that there is a control parameter called the proximity parameter ϵ, which defines the radius of an imaginary ball centered at each of these data points. When we gradually increase ϵ, the balls will grow outwards and eventually touch other balls. The overlapping of these balls form a unique topological characteristic that is unique to this dataset, and hence we can use this unique topological characteristic to differentiate nuances in the topologies of different point clouds. This *filtration* process can be demonstrated and visualized in **Figure 3**.

Through this encoding process, we can convert a point cloud that is made from brain functional signals, or a correlation matrix from financial time series data, to filtration diagrams. From these filtration diagrams, we can calculate barcodes, persistence diagrams, and other TDA metrics for further applications.

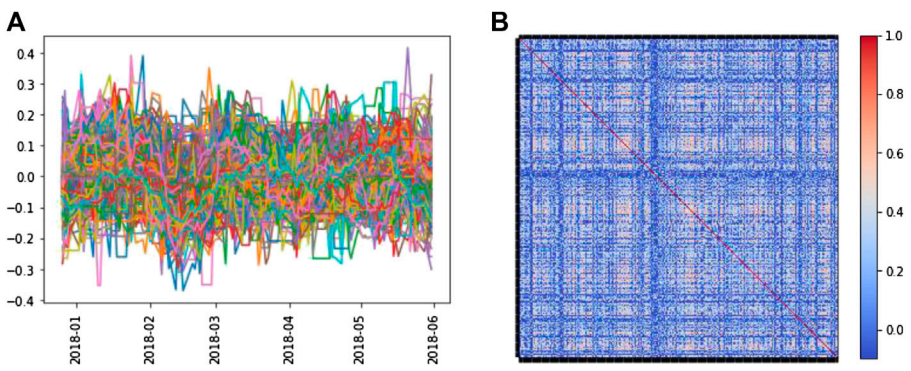


FIGURE 1 | (A) The derivative data of a 6-month period collected from STI. **(B)** The cross correlation matrix is generated from the derivative data in **(A)**. For the time series and correlation matrices in this work, we will not show error bars to not distract the readers from the overall features.

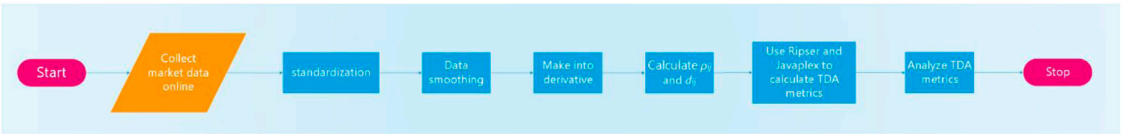


FIGURE 2 | Flowchart of the procedure implemented. There are two parts; the first part involves data collection and pre-processing. The second part regards TDA-related computations.

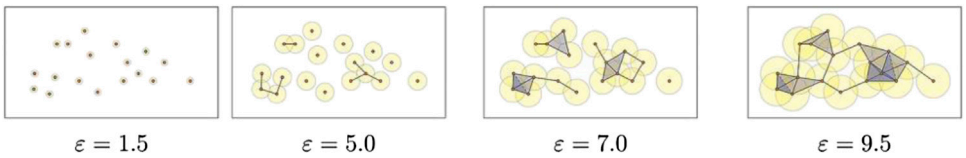


FIGURE 3 | A schematic diagram showing a data cloud, and how the filtration process results in outcomes of various overlapping of balls from different proximity parameters ϵ (shown in the upper column). The bottom column is barcodes scanning through a full-range of proximity parameter ϵ values. β_0 and β_1 denote the 0-dim and 1-dim Betti numbers, which can be deduced from the subfigures to be roughly $18 \rightarrow 11 \rightarrow 4 \rightarrow 1$, and $0 \rightarrow 0 \rightarrow 1 \rightarrow 2$, respectively.

Simplicial Complexes

A simplicial complex is an abstract collection of entities, which consists of nodes $(i, j, k, \ldots)$, or sets of nodes $(\{i\}, \{i, j\}, \{i, j, k\}, \ldots)$. These collections of nodes or sets can then be used to construct links, surfaces, and higher-dimensional objects. For example, we can decompose an arbitrary simplicial complex into its 0-simplexes (nodes), 1-simplexes (links), 2-simplexes (faces), 3-simplexes (tetrahedrons) components. In other words, simplexes are generalizations of a triangle in arbitrary dimensions, and a simplicial complex is an outcome of performing triangulation in arbitrary dimensions of the raw data. The simplicial complex is a unique signature that characterizes the topological structure of the data. Some of the common simplicial complexes include Vietoris-Rips (VR) complexes [27], Čech complexes [16, 18], Delaunay complexes [28], Alpha complexes [29], witness complexes [30], as well as others. In this work, we used the VR complex for our TDA calculations. VR is appealing because it approximates the more exact Čech complexes but is more efficient to calculate [31].

Suppose we collected two sets of time-series data of the same duration from a stock market. After we encode them into simplicial complexes, these may be different in terms of their local and global topologies. We then can use TDA metrics such as Betti numbers, Euler characteristics, barcodes, persistence diagrams, persistence landscapes, and Wasserstein distance as topological descriptors to quantify these differences. In the following subsections, we introduce some of these terminologies, their respective definitions, and elaborate on them.

Filtration

Here let us formalize the definition of the filtration procedure, which is commonly done to obtain barcodes. By changing the proximity parameter ϵ, we control the size of the balls and thus their overlaps. At a specific ϵ value, some balls overlap while others do not, and therefore we have a collection of 0-simplexes (isolated nodes), 1-simplexes (pairs of linked nodes), 2-simplexes (triangles), 3-simplexes (tetrahedrons), and so on.

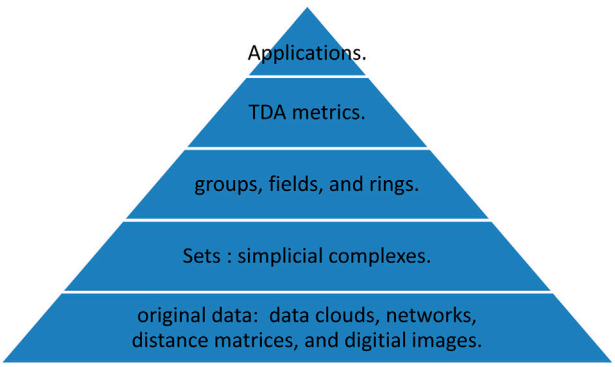


FIGURE 4 | A pyramid illustrating sequential procedures of how we make use of the original data and convert them into different forms of sets, i.e. simplicial complexes, and to groups, fields, and rings. At the topmost stage, we can use them for various applications, such as ML, and statistical learning, etc.

Such a collection is called a *sub-complex*. If we increase ϵ further, the sub-complex (and its topological features) may or may not change. This procedure resembles what we see in physics: by changing the external fields, e.g. temperature, or magnetic fields, the system changes from one symmetry group to another. We call this *symmetry breaking*. A filtration is conceptually similar to varying the external fields, and observe how they result in different symmetry groups. The difference is that in performing filtrations, we look at how the topological features evolve. Mathematically, a filtration can be described as a sequence

$$\varnothing = \Sigma^0 \subseteq \Sigma^1 \subseteq \Sigma^2 \cdots \subseteq \Sigma^m = \Sigma,$$

where $\Sigma^m = \Sigma$ is the simplicial complex, $\varnothing = \Sigma^0$ is the empty set, and $\Sigma^k \subseteq \Sigma^{k+1}$ indicates that the kth sub-complex is included in the $(k+1)$th sub-complex. In performing the filtration, we witness

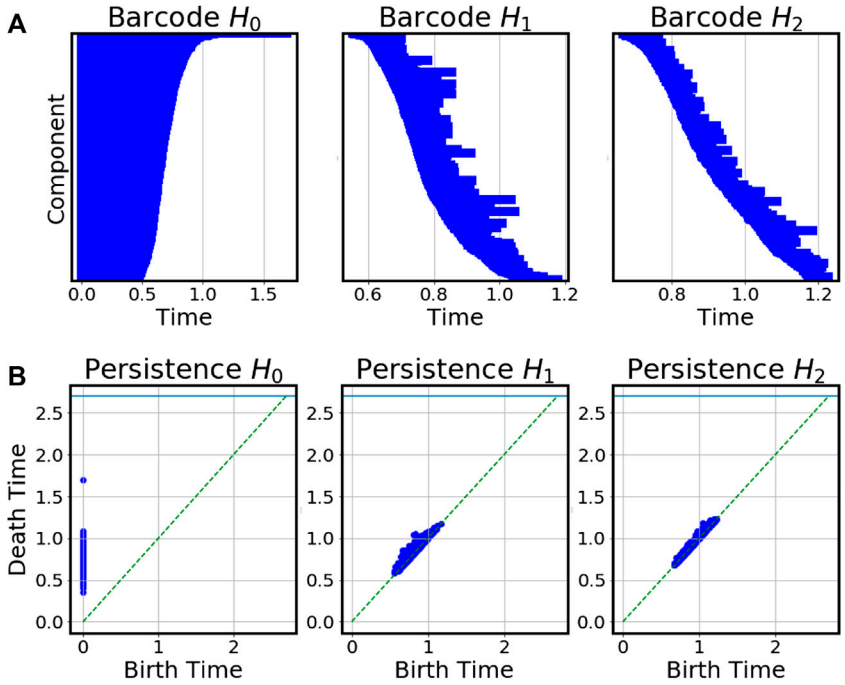


FIGURE 5 | (A) Barcodes in 0, 1, and 2 dimensions. Each bar represent a generator of the homology group, i.e. $H_n^{\{p,q\}}(\Sigma)$, where {p, q} marks a lifetime, the rank of $H_n^{\{p,q\}}(\Sigma)$ equals Betti number of homology groups in the *n*th dimension, the length of {p, q} signifies the persistence of the *n*th Betti number. **(B)** The barcodes can be converted into persistence diagrams, where one of the bars in **(A)** is equivalent to one point in the persistence diagram. The lifetime of each bar in **(A)** can be transformed into a perpendicular distance concerning the diagonal lines in **(B)**. If A point that is farther away from the diagonal line implies a more persistent topological feature, whereas a point that is closer to the diagonal line represents a less persistent feature. We found that the barcodes are rather robust and do not show fluctuations once the dataset is fixed, thus we do not include error bars in all barcodes appearing in this paper.

at that at $\epsilon = \epsilon_k$, there is a topological transition from Σ^k to Σ^{k+1}. By tracking all these ϵ_{kS}, we know how the simplicial complex's topology changes. We can then characterize these topological changes in terms of Betti numbers, Euler characteristics, barcodes, persistence diagrams, and persistence landscapes.

TDA Toolkits

Homology Group

In a filtration process, one can imagine that for smaller ϵ, the data points will have lesser overlaps; while we increase ϵ further, balls start to grow in size and eventually touch other balls, resulting in more overlaps; this continues until ϵ become so large that all the balls overlap with each other, leaving no space for holes to persist. Thus, for an intermediate ϵ, we expect to see balls making some overlaps but not too much, and the extents of these overlaps constitute different topological characteristics in terms of '*n*-dimensional holes'. Homology is a mathematical theory for studying these *n*-dimensional holes that exist in simplicial complexes by identifying which entities constitute these *n*-holes, and how many there are.

As mentioned before, SCs are obtained from performing a triangulation in arbitrary dimensions of the input data, or a way to represent the data in terms of 'sets'. But looking at sets is sometimes hard to develop an overall, comprehensive picture of the data, and also less convenient for executing mathematical operations on them. For this reason, mathematicians convert SCs, and other topological sets into groups, rings, or fields, so that in these constructs, they not only can discern between different sets, but also can impose structures like associative binary operations, the identity element, and the inverse element on them. See **Figure 4** for the procedures of encoding the raw data into sets, TDA metrics, allowing for further applications.

Unlike manifolds, which are continuous sets of points, SCs comprise discrete points instead. Although both can describe topological features in the data space, there is one advantage in using SCs, and that is one uses a triangulated (coarse-grained) surface instead of a continuous one. Practically, we are interested only in the *n*-holes and their numbers, and we only have limited data. In this sense, SCs and their homology are adequate to fulfill these goals.

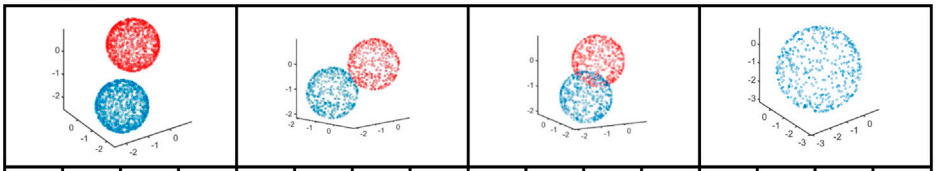
Betti Numbers and Euler Characteristics

An Euler characteristic is used to classify different polyhedrons, it reads:

$$\chi = V - E + F = 2(1 - g),$$

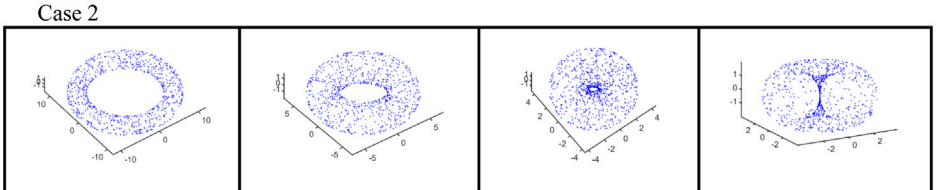
where χ is the Euler characteristic of a polyhedron, V, E, and F are the numbers of nodes, links, and surfaces, respectively. In this formula, g is the genus of the polyhedron. χ can also be calculated as a sum of Betti numbers,

A Case 1

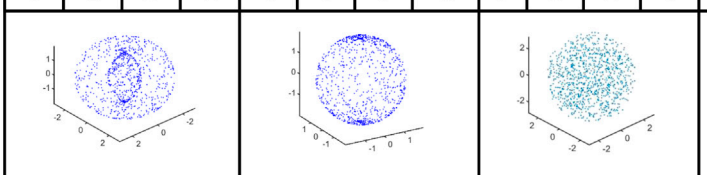


χ	β_0	β_1	β_2	χ	β_0	β_1	β_2	χ	β_0	β_1	β_2	χ	β_0	β_1	β_2
4	2	0	2	3	1	0	2	2	1	0	1	2	1	0	1

B Case 2

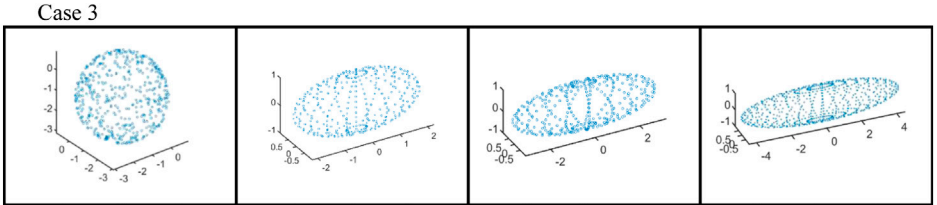


χ	β_0	β_1	β_2	χ	β_0	β_1	β_2	χ	β_0	β_1	β_2	χ	β_0	β_1	β_2
0	1	2	1	0	1	2	1	0	1	2	1	1	1	1	1



χ	β_0	β_1	β_2	χ	β_0	β_1	β_2	χ	β_0	β_1	β_2
2	1	1	2	2	1	0	1	1	1	0	0

C Case 3



χ	β_0	β_1	β_2	χ	β_0	β_1	β_2	χ	β_0	β_1	β_2	χ	β_0	β_1	β_2
2	1	0	1	2	1	0	1	2	1	0	1	2	1	0	1

FIGURE 6 | The toy models illustrating three different sequences of topological changes. **(A)** We start with two spherical shells of radius one and move them closer until their surfaces touch, overlap, and finally merged into a larger single spherical shell. **(B)** We generate a sequence of surfaces of revolution that starts with a torus, then one with a smaller hole, then a horn torus, a spindle torus, and finally a spherical shell. We also show a solid sphere. **(C)** We start with a spherical shell and then deform it into ellipsoids, whose semi-axis we increase from $5 \rightarrow 10 \rightarrow 20$.

$$\chi = \sum_{\{n \geq 0\}} (-1)^n \beta_n(\Sigma),$$

where the nth dimensional Betti number β_n is the dimension of the nth homology group $H_n(\Sigma)$ of the SC Σ. These are important metrics that characterize the topology of the data.

Barcodes and Persistence Diagrams

Barcodes help us visualize the n-dimensional homology group $H_n(\Sigma)$ in terms of its generators. We understand that each bar represents a generator of the persistent homology group $H_n^{\{p,q\}}(\Sigma)$. This representation tell us that the number of bars that are born at or before the pth filtration stage that are still alive at the qth filtration stage is precisely the rank of $H_n^{\{p,q\}}(\Sigma)$, which includes the essential classes that do not die with filtration [32].

The rank of the homology group in nth dimensions equals the nth Betti numbers, which we use to calculate the Euler characteristics χ. For more persistent bars, their topological features are more important, whereas the topological features of those that are less persistent can be treated as noises. Here, we convert the barcodes into persistence diagrams in **Figure 5**. Persistence diagrams carry similar topological information as barcodes. It is more useful in constructing statistical topological models that can be used to design weighted kernels.

Computational Methods

TDA Toolkits

Numerically, we used two softwares to perform the TDA calculations. The first software is called Ripser [33], which is included inside the Python package TDA [34]. Another is a Java program called Javaplex [35], which we used to calculate Betti numbers and Euler characteristics. Javaplex supports parallel computation in MPI and OpenMP, shortening the computing time for calculating persistent homology in higher dimensions. We installed Javaplex on Nanyang Technological University High-Performance Computing Centre's NYA2, equipped with Sandy-Bridge Processors-cores (Intel(R) Xeon(R) CPU E5-2680 @ 2.70 GHz), and 64 GB RAM per Node. NYA2 runs Red Hat Enterprise Linux Server release 6.3 (Santiago) and manages job queues using the Load Sharing Facility (LSF).

For some of the time windows, the Javaplex calculations failed with the error message "OutOfMemoryError: Java heap space" and "OutOfMemoryError: GC overhead limit exceeded". Here we offer two solutions. The first is to utilize more than 100 GB of memory on client computers, which can be switched on by adding a line "#BSUB -q MEM128G-S" in the LSF script. This option allows the submitted jobs to access up to 128 GB of memory. If the first solution fails, a second solution is to reduce the upper limit of the filtration value, say a value near 1.0. These two options can in general solve the problem of the memory shortage issue. On average, a job submitted to NYA2 accessing 16 CPU cores requires 1–3 days to finish. For each of the calculations, we saved barcodes figures, n-dimensional Betti numbers, Euler characteristics in separate folders for further analysis.

TDA OF TOY MODELS AND HYPOTHESIS ON REAL MARKETS

Before we analyze the SGX and TWSE data, and discuss their results, as a proof-of-concept we first digress to demonstrate the main idea behind our work by applying TDA to three toy models with definite topological changes. In these three cases, we randomly sampled data points on the surfaces or in the volumes and then saved these data points in separate files. Then, we use the Javaplex software to read in the files and calculate the persistent homology and respective Betti numbers up to dim 2. Finally, we use $\chi = \sum_{\{n \geq 0\}} (-1)^n \beta_n(\Sigma)$ to calculate the Euler characteristic. These results are shown in **Figure 6**.

In the first case (**Figure 6A**), we started with two spherical shells of radius one that do not overlap. We then moved the two shells closer until their surfaces touch, before we moved them even closer that they overlap. For this sequence of configurations, we saved the data points and thereafter invoked the Javaplex software. In the third, overlapping configuration, we manually deleted those data points that lie inside the spherical shells. Finally, we compared this sequence of configurations against a larger spherical shell. We found that χ went from $4 \rightarrow 3 \rightarrow 2 \rightarrow 2$, which was consistent with the analytical results. The sequences of Betti numbers provided even more information. As we went through the sequence of configurations, β_0 changed from $2 \rightarrow 1 \rightarrow 1 \rightarrow 1$, which agrees with what we expected, since β_0 tells us how many connected components there are in the configuration. We also found $\beta_1 = 0$ throughout the sequence, since it is the number of irreducible closed loops, and in all configurations, we can always shrink a closed loop to a point. Finally, we found β_2 changing from $2 \rightarrow 2 \rightarrow 1 \rightarrow 1$, since it is the number of voids enclosed within the different surfaces, so this becomes 1 after the two spherical shells overlap.

For the second case (**Figure 6B**), we went through a sequence of surfaces of revolution of two circles at increasing closer distances. When the two generating circles were far apart, we obtained a torus with a big hole, and when the two generating circles were closer but still non-overlapping, we obtained a torus with a small hole. When the two generating circles touched each other, we ended up with a *horn torus*, which is a critical surface with no holes but is pinched at a point. When the two generating circles overlapped each other, we obtained a *spindle torus*, which has an inner as well as an outer surface. Finally, when the two generating circles overlapped completely, we obtained a spherical surface. This last configuration is then compared against a solid sphere. For this sequence, we found χ going from $0 \rightarrow 0 \rightarrow -4 \rightarrow -1 \rightarrow 2 \rightarrow 1$, which is the result of an interesting interplay between the Betti numbers. Going through the sequence, we found $\beta_0 = 1$ throughout, because there is only one connected object. In contrast, β_1 went from $2 \rightarrow 2 \rightarrow 5 \rightarrow 3 \rightarrow 0 \rightarrow 0$ and $\beta_2 = 1$ for all configurations, except for the spindle torus ($\beta_2 = 2$), and the solid sphere ($\beta_2 = 0$). Since β_2 is the number of voids enclosed, we understand why $\beta_2 = 1$ for the spherical shell configurations, and why $\beta_2 = 2$ for the spindle torus. In this sequence, the most interesting change occurred in β_1.

For the final case (**Figure 6C**), we started with a spherical shell and increased its eccentricity to get longer ellipsoids, with semi-axis a going from $5 \rightarrow 10 \rightarrow 20$. For all these different ellipsoids, we found that $\chi = 2$, confirming the fact that deformation alone cannot change the topology or the Euler characteristic. We also found $\beta_0 = 1$, $\beta_1 = 0$, and $\beta_2 = 1$ for all these surfaces, as expected.

For all cases, we also computed the corresponding barcodes and persistent diagrams for better insights into how they evolve with topological deformations. These are shown in **Supplementary Figure S2**.

Generally speaking, the cross correlations in a stock market will be in the form of a high-dimension topological space, with more complicated features than those shown above. Nevertheless, we believe the insights derived from the toy models can help us grasp the topological changes that occur during a stock market crash (shown schematically in **Figure 7**). Just before the market crash (**Figure 7A**), we show the cross correlations of the stock market as a single giant cluster with four holes, which tells us that $\beta_0 = 1$ and $\beta_2 = 4$, while β_1 will depend on the detail shape of the topological surface. This strongly interconnected situation is typically generated by a bubble in the market and can be viewed as the starting point of a market crash [36, 37]. When the market crash starts (**Figure 7B**), parts of the surface will break (red circles in **Figure 7B**) but overall the giant cluster remains. The breaking of these two handles results in β_2 going from 4 to 2, while β_0 remains 1. For every handle broken, β_1

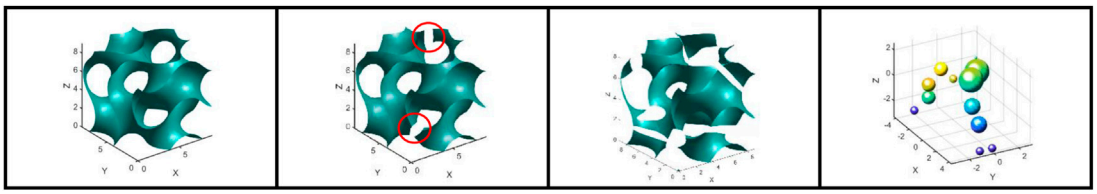


FIGURE 7 | A schematic diagram illustrating different states in the stock market across a market crash. **(A)** All stock components are interconnected and form a single giant cluster with holes. **(B)** As the market starts to crash, some of the connections are broken, but the single giant cluster remains as in case **(A)**. **(C)** As the market crash progresses, the giant cluster remains, but part of it has fragmented into four smaller clusters. **(D)** At the end of the market crash, the stocks are now organized into many small and disjoint clusters.

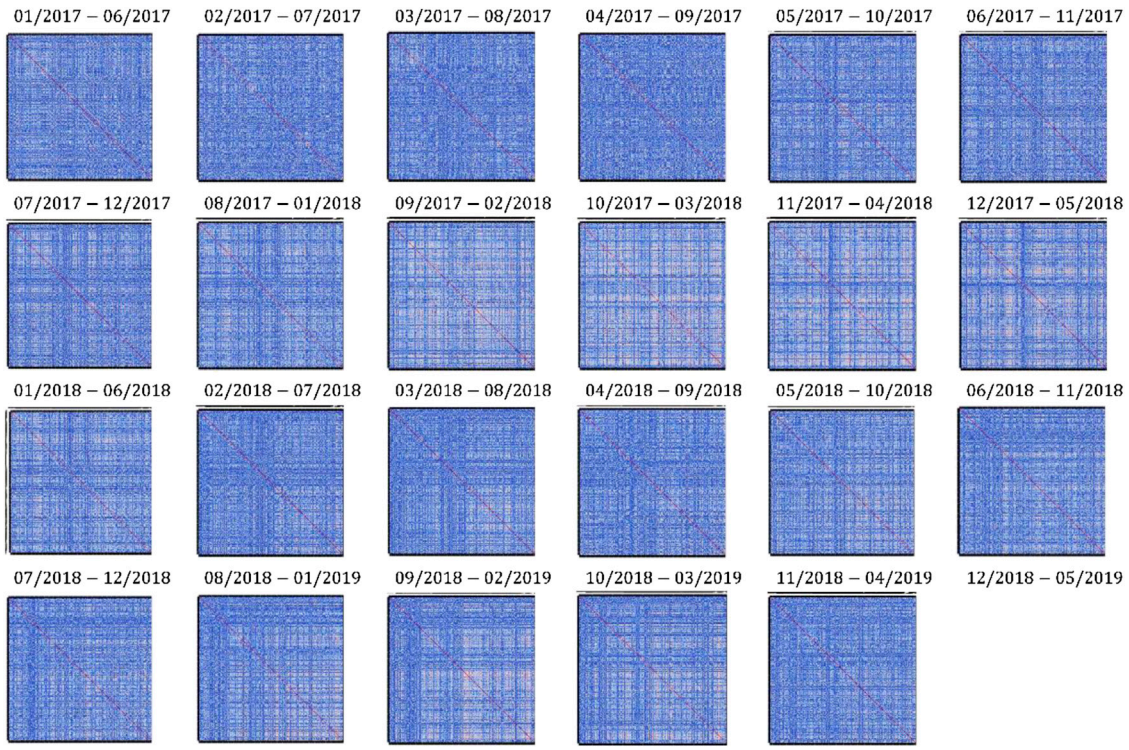


FIGURE 8 | The correlation matrices of STI from Jan 2017 to Apr 2019.

also decreases by 2. As the market crash progresses, the giant cluster starts to crumble, giving rise to additional small clusters like the ones shown in **Figure 7C**. When the number of connected components increases, β_0 goes from 1 to 5, and β_2 decreases further to 1, because there is only one hole remaining. The small clusters do not contribute to β_2 if they are homomorphic to spheres. Finally, at the end of the market crash, many small clusters are produced by the dissociation of the giant cluster, so β_0 increases dramatically, but β_1 and β_2 become small. Such a cluster fusion-fission scenario has been proposed previously [38, 39], but we suspect TDA will provide additional information regarding subtle topological changes that these models cannot provide.

Armed with these insights, we proceed next to the research question, that is to use TDA to examine the topological changes associated with market crashes in the SGX and TWSE, to see how well our hypothesis holds out.

RESULTS AND DISCUSSION

In this work, we examined two stock markets, i.e. the Singapore Stock Market (STI), and the Taiwan Stock Exchange (TAIEX). Both markets consist of roughly 600 stock components, and the economic scales of Taiwan and Singapore are comparable. The time durations that we collect data are from Jan 2017 to Apr 2019 for STI, and from Jan 2017 to March 2020 for TAIEX. For TAIEX, there is a small market crash from Sep 2018 to Jan 2019, and a major crash in Mar

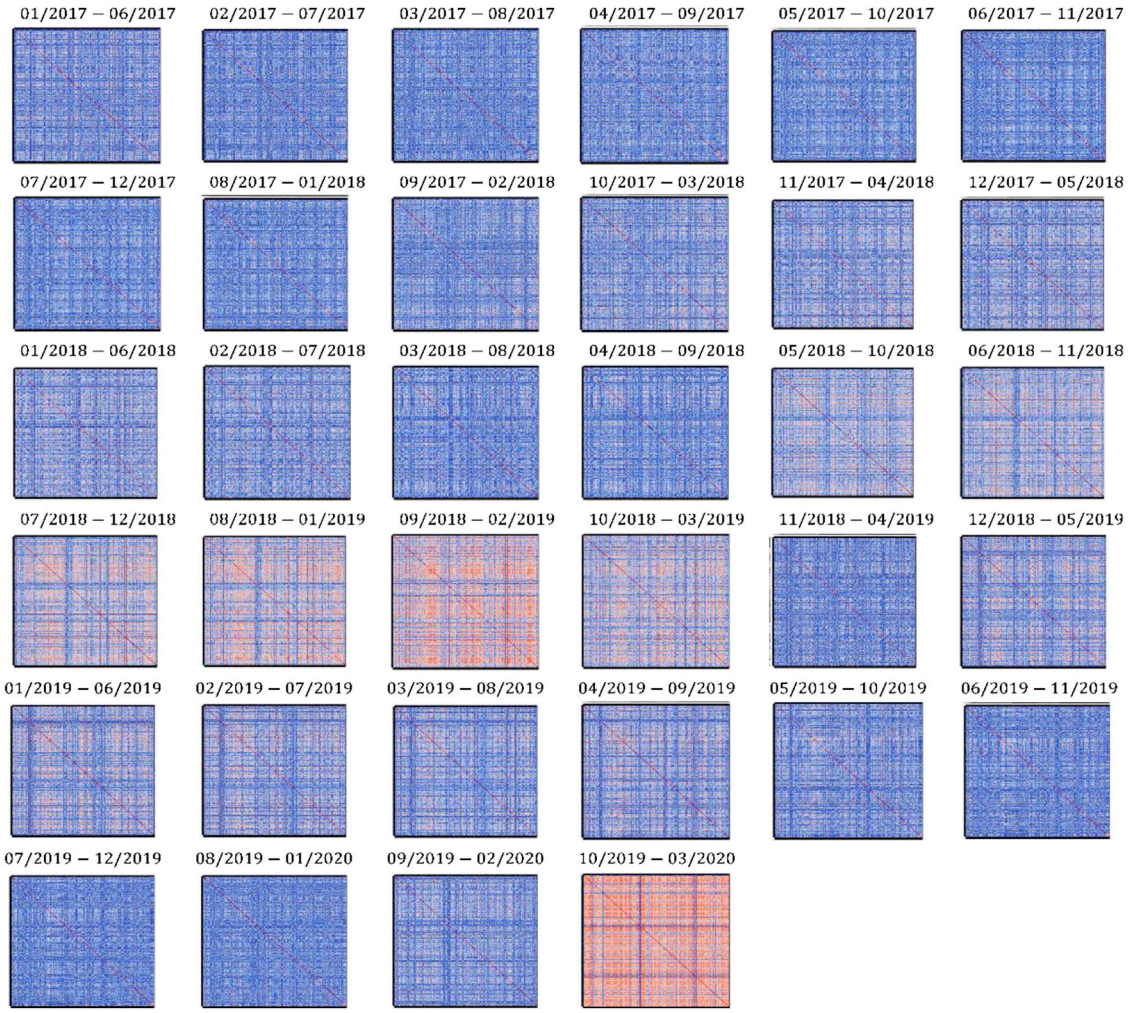


FIGURE 9 | The correlation matrices of TAIEX from Jan 2018 to Mar 2020.

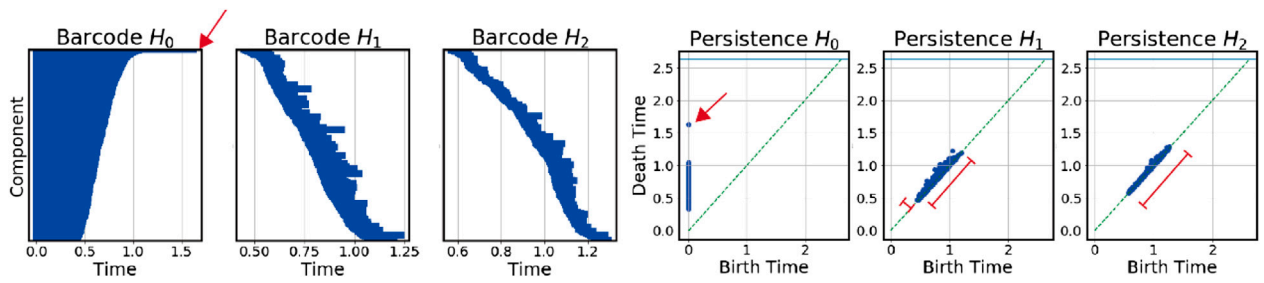


FIGURE 10 | The barcodes, and corresponding persistence diagrams for data collected from Apr 01, 2019 to Sep 30, 2019 in TAIEX.

2020 that is caused by the COVID-19 pandemic, whereas no market crash was found for the STI.

Correlation Matrices

We visualize the complex dynamics in the two stock markets by computing correlation matrices over six-month periods that are one month apart. We used a heat map color scheme, where the highest correlation value of 1 is red, and the most negative correlation of −0.1 is blue. These are shown in **Figures 8** and **9** for STI, and TAIEX.

In **Figure 8** for STI, it is clear that the average correlation is low over most periods. The exceptions are the periods (Sep 2017, Feb 2018), (Oct 2017, Mar 2018), (Nov 2017, Apr 2018), and (Dec

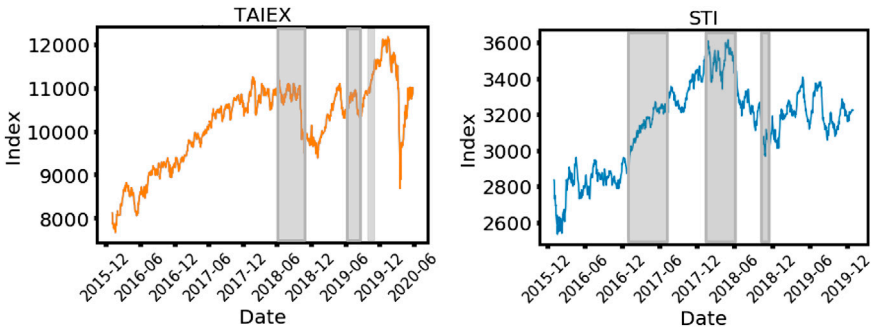


FIGURE 11 | The **(A)** TAIEX and **(B)** STI index for the past five years, which include the period we collected our data. For TAIEX, the period is from Jan 01, 2017 to Mar 31, 2020. For STI, the period is from Jan 01, 2017 to Apr 30, 2019. In this figure, the gray bands are periods seen from **Figure 14** where the Euler characteristic is positive.

2017, May 2018). In **Figure 9** for TAIEX, however, we observe more drastic changes. The correlation matrix first becomes reddish for the (May 2018, Oct 2018) period, and remains reddish until the (Oct 2018, Mar 2019). It then became reddish again in the (Oct 2019, Mar 2020) period because of the COVID-19 pandemic. Particularly, the few correlation matrices preceding the COVID-19 crash were blue, making the reddening very sudden.

In the literature, spectral reddening can be used as early warning signals to inform critical transitions [40–43]. Before market crashes, the co-movement among stocks becomes stronger, variations become increasingly concentrated at low wavenumbers, and result in a reddish color in the spectral density. Although inspecting different properties, both show early warning signals by turning into red colors when approaching these critical transition points.

Barcodes and Persistence Diagrams

The barcodes and their corresponding persistence diagrams for TWSE data between Apr 2019 and Sep 2019 are shown in **Figure 10**. In **Figure 10**, the left three figures are the barcodes in 0-dim, 1-dim, and 2-dim, respectively, whereas the right three figures are the corresponding persistence diagrams. In the right three figures, the x-axis refers to the time of birth, while the y-axis refers to the time of death for each homology group, represented by a dot on the figures. We also use red arrows to indicate which bar in the left figures corresponds to which dot in the right ones. When the period is varied, the shape of these figures also changes, revealing the dynamics of the topological structures.

For the persistence diagrams, the dots in the 0-dim figure only move vertically in time, whereas for those in 1-dim, and 2-dim, the data points can cluster together forming a small bump, flatten out along the diagonal line, or translate toward or away from the origin along the diagonal line. During market crashes, intriguing dynamical properties can be seen in these figures. To make a clearer comparison, we show in **Figure 11** the aggregated STI and TAIEX 5-years historical data and discuss the features seen in **Supplementary Appendix Figure A1** and **Supplementary Appendix Figure A2**, where we show all the barcodes and persistence diagrams for the data collected from SGX and TWSE.

In **Figure 11A**, we find a local market minimum from Sep 2018 to Jan 2019, spanning roughly five months following a small crash in Sep 2018. From the barcodes and persistence diagrams in **Supplementary Appendix Figure A1**, we discover an interesting feature related to this small crash. When we compare the 1-dim and 2-dim persistence diagrams for the (Mar 2018, Aug 2018) period (not including the crash) against those of the (Apr 2018, Sep 2018) (including the crash) in **Supplementary Appendix Figure A1**, the data points flatten out along the diagonal line, suggesting that in these two dimensions, the persistence of the homology groups weakens. However, the 0-dim result shows no signs of change when we compare these two subfigures. This episode of a persistence-weakening in 1-dim and 2-dim continued until the (Oct 2018, Mar 2019) period in **Supplementary Appendix Figure A1** when the flattening-out phenomenon disappears. Looking at the barcodes in the same period, we witnessed that the 1-dim and 2-dim bars, which are generally wider before the (Mar 2018, Aug 2018) period, becoming visibly shorter in the period (Apr 2018, Sep 2018) to (Oct 2018, Mar 2019). To aid visualization, we used red-shaded windows in **Supplementary Appendix Figure A1** to identify those barcodes manifesting persistence weakening. In the **Supplementary Figure S4**, we also show schematically how bars in the barcodes become dots in the persistent diagram during a normal market phase and a market crash phase.

We observed an even stronger persistence weakening for the (Oct 2019, Mar 2020) period than for the small crash. Going back to the barcodes, we found the widths of the bars becoming smaller as the distribution of data points flatten in the persistence diagram. We also found a large gap of $1.2 < \epsilon < 1.4$ between the death of one bar, and the birth of the next bar in the 2-dim barcode. To unravel how this persistence-weakening phenomenon occurs, we reduced the time windows' sizes to 2, 3, 4, and 5 months, and show the results in **Figure 12**.

In **Figure 12**, we witness some interesting features. First, in the 0-dim persistence diagram, the dots seem to be lower compared to those periods without market crashes. This corresponds to a shorter life expectancy for the homology groups, which can also be observed in the barcodes. In the 1-dim and 2-dim barcodes, we

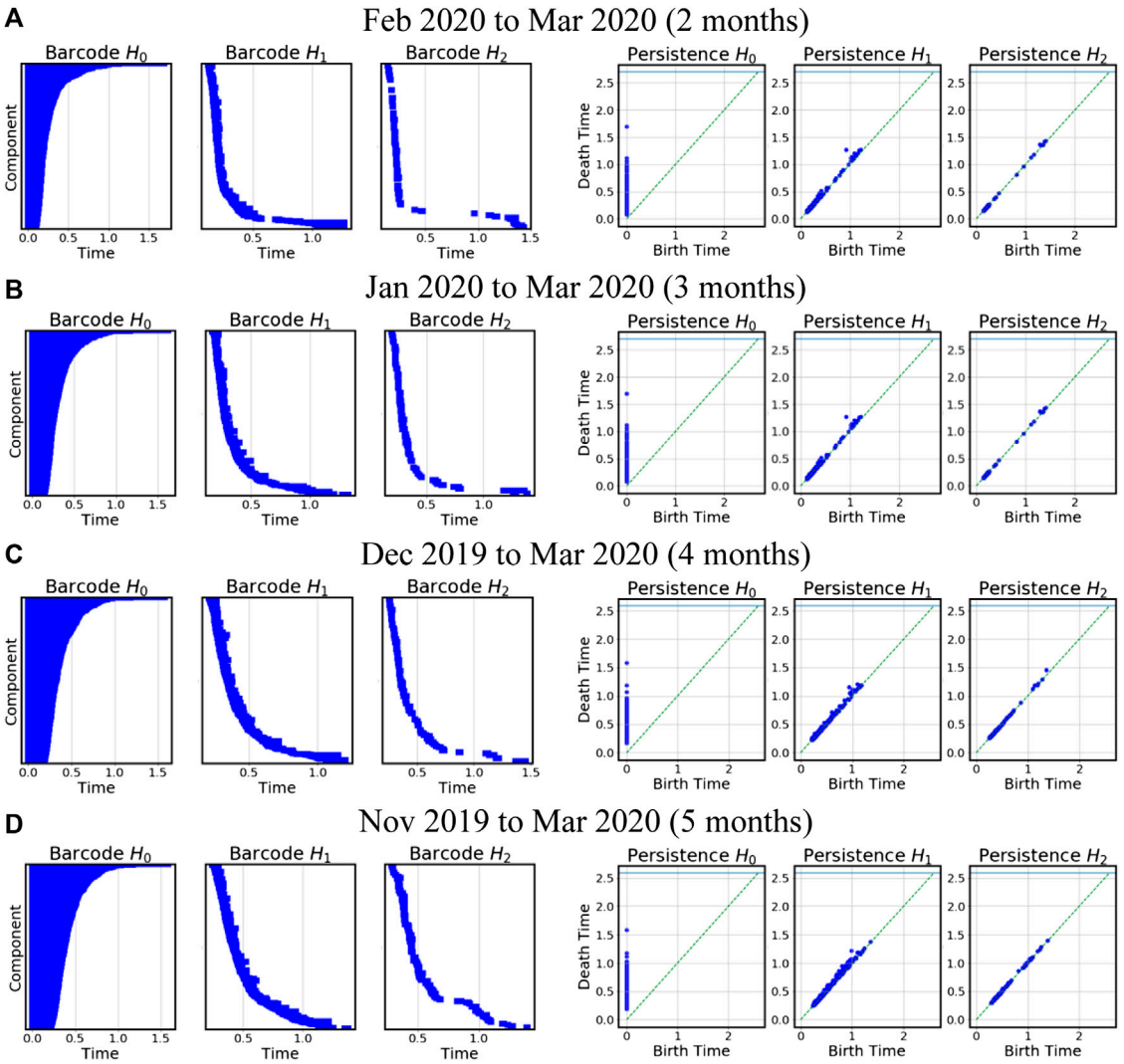


FIGURE 12 | The barcodes and persistence diagrams covering the pandemic COVID-19 crash in March 2020. The first column is a two-month window result, and the subsequent columns correspond to 3, 4, 5, and 6-months time windows, respectively.

find the barcodes falling off rapidly between $\epsilon = 0$ and $\epsilon = 0.2$, and more slowly thereafter. This suggests that $\epsilon = 0.2$ ($\rho = 0.97$) is a characteristic scale that emerged only during the COVID-19 crash. To quantify the persistence-weakening phenomena in TWSE, we selected three periods and analyzed the H_1 and H_2 persistence diagrams (see **Supplementary Figure S3**). In the normal market state during the six periods (Jan 2017, Jun 2017) to (Jun 2017 to Nov 2017), the two principal variances were found to be $\sigma_1^2 = 0.361 \pm 0.036$, and $\sigma_2^2 = 0.066 \pm 0.003$ for H_1. For H_2, we found that $\sigma_1^2 = 0.371 \pm 0.07$, and $\sigma_2^2 = 0.036 \pm 0.001$. For the period (Sep 2018, Feb 2019), which covers the mini-crash, we measured $\sigma_1^2 = 0.763$, and $\sigma_2^2 = 0.046$ for H_1; for H_2, we measured $\sigma_1^2 = 0.860$, and $\sigma_2^2 = 0.024$. Finally, for the period (Oct 2019, Mar 2020), we obtained $\sigma_1^2 = 0.770$ and $\sigma_2^2 = 0.041$ in H_1; for H_2, σ_1^2 increased to 0.931, while σ_2^2 become 0.020. To conclude, during the two market crashes in TWSE, the second principal variance was reduced, implying a shortened persistence lifetime, a manifestation of the persistence-weakening phenomena that come along with crashes.

Another way to quantify the persistence weakening is through the persistent entropy, $E(F) = -\sum_{i=1}^{n} p_i \log(p_i)$ where F is the distribution of lifetimes $l_i = y_i - x_i$ (x_i, y_i are the birth time and the death time of homology group i in the barcode with n segments), $S_L = \sum_{i}^{n} l_i$ is the sum of all lifetimes, and $p_i = l_i/S_L$ can be thought of as the 'weight' of homology group i in the barcode [44, 45]. The persistent entropy $E(F)$ is maximum when all homology groups have the same lifetimes, and is minimum when the lifetimes of homology groups are all different. $E(F)$ thus allows us to distinguish between narrow and broad distributions of lifetimes, as well as smoothly varying and multimodal distributions of lifetimes. We chose to compute $E(F)$ for the same three periods used to calculate the covariance matrix and principal variances. For the normal market state during

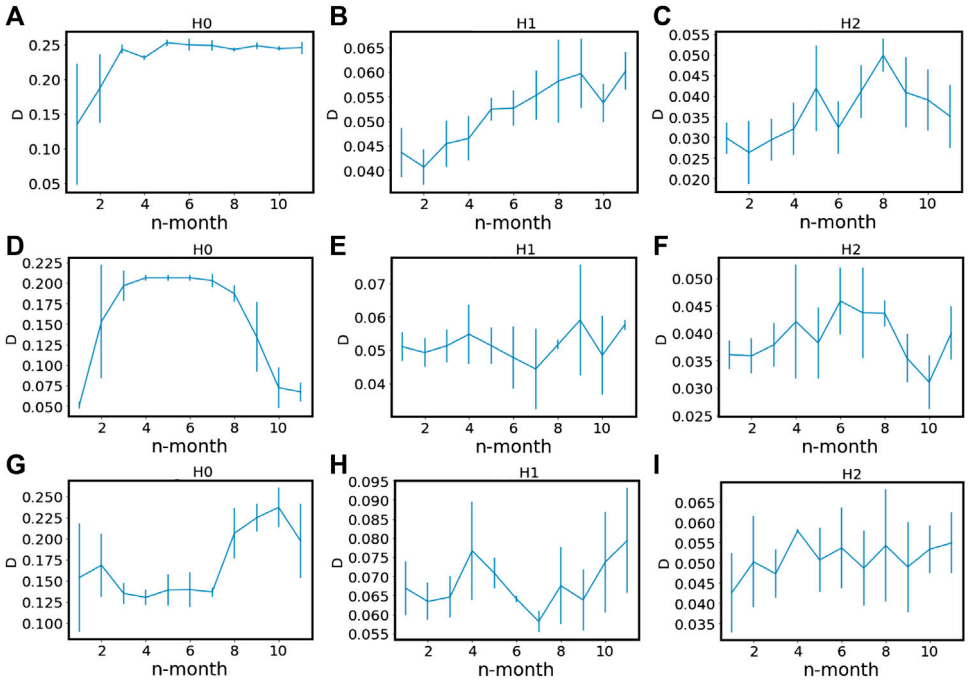


FIGURE 13 | Bottleneck distances D calculated for **(A)** H_0, **(B)** H_1, and **(C)** H_2 for the origin month (we pick three points in Jan 2017 as origins, and calculated D with subsequent n = 1 months). The solid lines are the mean values, and the vertical bars are the standard deviation for each data point. **(D–F)** are the same as **(A–C)** but for normal market states of TAIEX, and **(G–I)** we select three points in Sep 2018 as origins, and calculated D with subsequent n = 11 months, covering the mini market crash of TAIEX.

the six periods (Jan 2017, Jun 2017) to (Jun 2017 to Nov 2017), we found that $E(F) = 2.8185 \pm 0.004$, 2.741 ± 0.028, 2.733 ± 0.072 in 0–2 dim respectively; for the period (Sep 2018, Feb 2019), $E(F)$ remained roughly the same at 2.80 for 0-dim, but decrease to 2.55 for 1-dim, and more significantly to 2.1 for 2-dim. For the COVID-19 crash, $E(F)$ for 0-dim remained at 2.80, while for the other two dims, they became 2.40 and 1.94. This suggests that $E(F)$ for 2-dim changes most dramatically across market crashes.

For SGX (**Supplementary Appendix Figure A2**), the persistence-weakening phenomena are less significant, except for (Sep 2017, Feb 2018), (Nov 2017, Apr 2018), and (Dec 2017, May 2018). In these periods, persistence-weakening only occurs in the 2-dim persistence diagrams but not in their 1-dim counterparts. This is rather different from those observed in the two TAIEX crashes. We believe this is because over the period Jan 2017 to Apr 2019, the largest downward movement of the STI is still smaller than the smaller TAIEX crash, as such the persistence weakening is less prominent. Referring to **Figure 11B**, we believe these downward movements were market corrections in the STI, and the analysis we introduced thus far cannot help to classify them. Consequently, we introduce the Betti numbers and Euler characteristics in the next subsection to resolve this issue.

Other Works Addressing Persistence

In 2015, Teh and Cheong [38] studied dynamics in the SGX during the Global Financial Crisis using a cluster fusion-fission approach. They found that before the crisis, a giant cluster of stocks emerged in the SGX. This later broke up into small clusters after Lehman Brothers went bankrupt. Also, they found that the probability that a pair of stock remain in the same cluster decays exponentially with two time scales i.e. 3 weeks, and 7 weeks. They called these temporal correlations the 'persistence' of stocks. In our work, since our sliding window size is one month, we can also measure the persistence in both time scales, in terms of n-holes that emerge in the two-time windows.

We show the mean value of bottleneck distance D (we pick three points in Jan 2017 as origins, and calculated D with subsequent n = 11 months for the origin), and its standard deviation in SGX over the whole of 2017 in **Figure 13**. We discovered that D increased steadily over the next four windows for H_0, and then saturated around 0.25, whereas for H_1 and H_2, D also increased but less significantly over the n-windows. A larger D implies that the homology groups are less persistent, whereas the converse means the persistence is stronger. As for the case of TWSE, we investigated two periods. The first period is (Jan 2017, Dec 2017), the same as the first period studied for the SGX, and the second is (Sep 2017, Aug 2018), which is in the middle of the mini-crash. For these two periods, we observe dissimilar features for H_0. In the first period, D grew from an initial value of 0.05, and saturated around 0.2, before dropping steeply to 0.07. For the second period, H_0 stayed between 0.13 and 0.15, before jumping to a larger value of roughly 0.22 seven months later. As expected, the bottleneck distance increases and then decreases over the course of a market crash.

TABLE 1 | (A) Euler characteristics, and *k*th Betti numbers for $k = 0 – 3$. Data collected in STI from Jan 2017 to Nov 2018. **(B)** Euler characteristics, and *k*th Betti numbers for $k = 0 – 3$. Data collected in TAIEX from Jun 2018 to Dec 2018, May 2019 to Oct 2019. Two periods cover the two crashes. We have calculated β_k specifically for the period in TAIEX (Jan 2017 to Jun 2017) 10 times to test if β_k fluctuates; our results confirmed that all arrived at the same β_k and χ. We therefore will not include the error bars for the Betti numbers and Euler characteristics.

Intervals	χ	β_0	β_1	β_2	β_3
(A) Euler characteristics, and *k*th Betti numbers for $k = 0 – 3$. Data collected in STI from Jan 2017 to Nov 2018					
01/17 – 06/17	14	3	7	42	24
02/17 – 07/17	16	4	3	64	49
03/17 – 08/17	30	3	7	69	35
04/17 – 09/17	8	4	5	65	56
05/17 – 10/17	25	4	14	54	19
06/17 – 11/17	10	2	20	47	19
07/17 – 12/17	35	3	22	69	15
08/17 – 01/18	–12	5	22	14	9
09/17 – 02/18	–25	25	53	6	3
10/17 – 03/18	–1	23	35	16	5
11/17 – 04/18	–4	28	38	11	5
12/17 – 05/18	–5	24	41	18	6
01/18 – 06/18	–11	23	58	34	10
02/18 – 07/18	45	5	17	70	13
03/18 – 08/18	63	3	23	100	17
04/18 – 09/18	40	3	30	75	8
05/18 – 10/18	28	3	23	59	11
06/18 – 11/18	15	4	31	64	22
07/18 – 12/18	–7	2	30	69	48
08/18 – 01/19	–9	7	46	43	13
09/18 – 02/19	–19	28	70	28	5
10/18 – 03/19	34	2	29	71	10
11/18 – 04/19	11	3	29	60	23
(B) Euler characteristics, and *k*th Betti numbers for $k = 0 – 3$. Data collected in TAIEX from Jun 2018 to Dec 2018, May 2019 to Oct 2019					
06/18 – 11/18	23	2	5	62	36
07/18 – 12/18	27	45	21	3	
08/18 – 01/19	28	40	16	4	
09/18 – 02/19	54	74	18	1	
10/18 – 03/19	26	36	21	11	0
11/18 – 04/19	–3	6	34	44	19

(Continued on following page)

TABLE 1 | (*Continued*) **(A)** Euler characteristics, and *k*th Betti numbers for $k = 0 - 3$. Data collected in STI from Jan 2017 to Nov 2018. **(B)** Euler characteristics, and *k*th Betti numbers for $k = 0 - 3$. Data collected in TAIEX from Jun 2018 to Dec 2018, May 2019 to Oct 2019. Two periods cover the two crashes. We have calculated β_k specifically for the period in TAIEX (Jan 2017 to Jun 2017) 10 times to test if β_k fluctuates; our results confirmed that all arrived at the same β_k and χ. We therefore will not include the error bars for the Betti numbers and Euler characteristics.

Intervals	χ	β_0	β_1	β_2	β_3
12/18 – 05/19	−10	23	44	11	0
05/19 – 10/19	−36	19	79	24	
06/19 – 11/19	−2	6	22	28	14
07/19 – 12/19	−38	1	6	10	43
08/19 – 01/20	−33	1	6	23	51
09/19 – 02/20	−7	11	46	31	3
10/19 – 03/20	98	106	10	2	

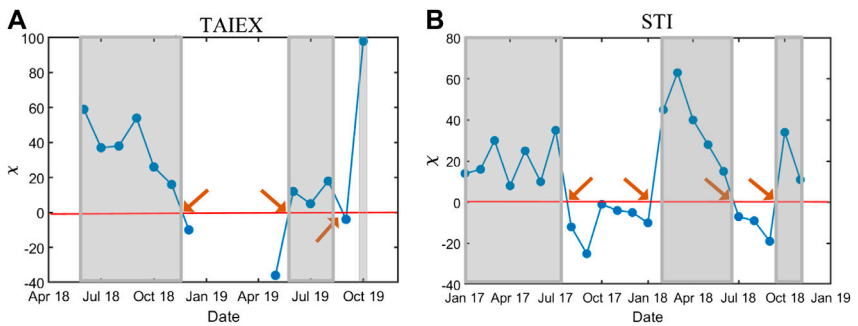


FIGURE 14 | The Euler characteristic χ in **(A)** TAIEX and **(B)** STI against dates. Each date represents a six-month period in which the correlation matrix was constructed. Besides, different dates are associated with different values of ϵ. The gray bands are over the same periods shown in **Figure 11** and cover Euler characteristics that are positive.

*k*th Betti Numbers and Euler Characteristics

In the literature, econophysicists use cross correlations to distinguish between different market states, like the bull and bear market states, as well as the market correction state. In a market correction state, the market condition resembles a random walk process, and thus the cross correlations between stocks are random matrix-like, and the distribution of their eigenvalues resembles a Marčenku–Pastur distribution (MPD). Following Reimann et al. and Santos et al. [46, 47], we can also use the *k*th Betti numbers and Euler characteristics as fingerprints to classify market states in the STI and TAIEX. According to Reimann et al. and Santos et al., different correlation matrices can have higher similar fingerprints and thus represent the same topologies.

In [46, 48], the authors also proposed to use the Euler entropy $S_x = \ln(|\chi|)$ as an alternative entropy construct, instead of the conventional Boltzmann entropy. They used the Euler entropy to inform whether there are topological phase transitions at any specific time or correlation values. According to their findings, a negative χ can be geometrically connected to a sheet of hyperboloid with negative curvature, at $\chi = 0$ the hyperboloid become cone-like, on the edge of breaking into two hyperboloids, and finally a positive χ, where the hyperboloid breaks into two hyperboloids. Hence, when χ changes from a positive value to a negative one, we can identify a critical point. At these points, the Euler entropy explodes ($\ln|0| \to -\infty$) and become singular. In statistical mechanics, when the system approaches a critical point, we expect to see the susceptibility function become non-analytic. In view of this, we can also use the Euler entropy to analyze and classify different market states.

Here, we show the *k*th Betti number and Euler characteristics for different periods in the SGX (TWSE) in **Table 1**. For TWSE, we chose two periods of time, i.e. (Jun 2017, Dec 2018), and (May 2019, Oct 2019), to calculate χ. These periods correspond to the two TAIEX crashes. Also, we calculated up to 2-dim Betti numbers, because for TWSE, we were not always able to compute the 3-dim Betti numbers. From July 2018 to Nov 2018, we found that χ was positive, and become negative in Dec 2018. From June 2019 to Sep 2019, χ stayed close to zero, and then suddenly jumped to 98 in Oct 2019, whose time window

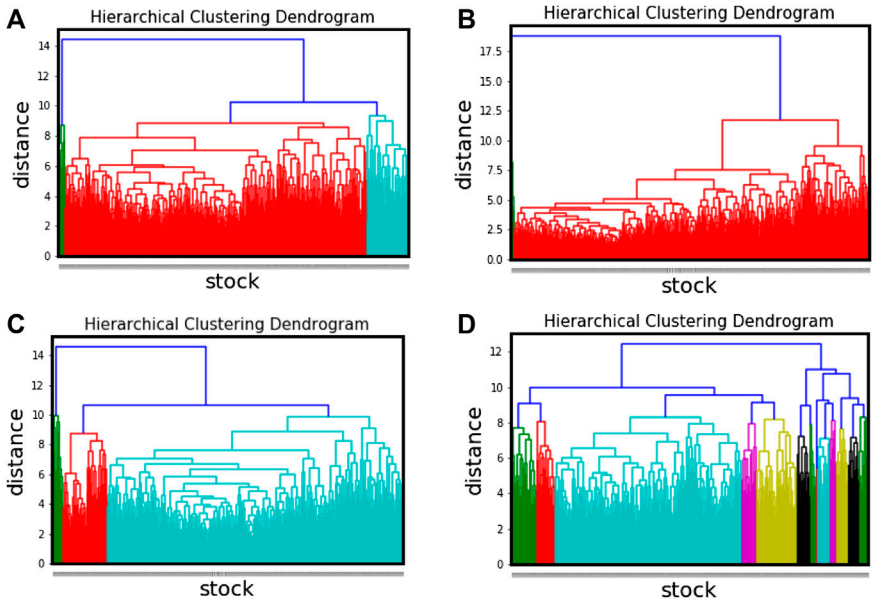


FIGURE 15 | The hierarchical clustering dendrogram for four periods in TAIEX, i.e. **(A)** from Jun 2018 to Nov 2018, **(B)** Sep 2018 to Feb 2019, **(C)** Oct 2018 to Mar 2019, and **(D)** Dec 2018 to May 2019.

included the COVID-19 crash in Mar 2020. Based on our results, both crashes seem to be associated with large positive χ values instead of negative ones. We can understand a positive χ as the result of many isolated hyperspheres, while a negative χ comes from averaging the curvature over hyperbolic bridges after some hyperspheres merged. This conclusion is also supported by the behavior of β_0, whose average values over the two periods are 31.6 and 98 respectively, suggesting that the stock components are fragmented rather than agglomerated. On the other hand, while χ is 23 and 27 respectively for (June 2018, Nov 2018) and (Jul 2018, Dec 2018), the values for β_0, β_1, and β_2 are different, implying that their topologies are dissimilar. Our findings agree with our hypothesis that during crashes, stock components tend to break up into fragments, even though the overall cross correlations are high.

Going on to the SGX, where the Euler characteristic was computed up to the 3-dim Betti numbers, we see from **Figure 14B** four topological transitions (marked by brown arrows). These imply that from Jan 2017 to Apr 2019, even though the signatures were weak in the cross correlations, SGX switched between different topological phases. We classified the market period from Jan 2017 to July 2017 as the first market state, where χ has an average value of around 20. The second market state was from Aug 2017 to Jan 2018, when χ became negative. The third market state was from Feb 2018 to Jun 2018, where χ became positive again. Finally, the fourth market state started from Jul 2018 and ended in Sep 2018, during which χ turned negative a second time. Thereafter, χ was positive for the last two months. The Betti numbers in **Table 1** show more subtle behaviors that the χ alone cannot reveal. For example, in the first period, we see that $\beta_0 \approx 3$ and $\beta_2 \approx 58.5$, whereas β_1 was separated into two groups, one averaging 5.5, while the other averaging 18.6. β_3 was also separated into two groups, one having an average of 41, while the other averaging 17.6. These are in line with the insights we developed in **TDA of Toy Models and Hypothesis on Real Markets** Section, that we cannot deduce the topology of the data by simply looking at χ, but must also check the details of β_n. We found similar situations for other periods (Aug 2017 to Jan 2018, Jul 2018 to Sep 2018) in SGX.

To show that we indeed observe in the real market data topological changes described in our hypothesis in **TDA of Toy Models and Hypothesis on Real Markets** Section, we investigated specifically the mini-crash of TAIEX over four time periods. One is just before the crash (Jun 2018 to Nov 2018), two is during the crash (Sep 2018 to Feb 2019) and (Oct 2018 to Mar 2019), and the last is just after the crash (Dec 2018 to May 2019). Here let us point out an important limitation of the Betti numbers, i.e. they do not tell us how big the clusters are. For example, the same set of Betti numbers can describe a collection of clusters that are roughly the same size, some with holes, some without; this market is not close to a crash. Or it can describe a collection of clusters, one of which is a giant cluster containing most of the holes; a market like this is close to a crash. This means that β_n must be supplemented by traditional clustering analysis, where it is easier to see giant clusters, but difficult to understand topological changes.

To this end, we show in **Figure 15** the results of average-linkage hierarchical clustering based on the cross-correlation matrices of the four periods. In the first period, we found one giant cluster co-existing with two small clusters. $\beta_0 = 2$ for this period is close to the number of clusters we found, confirming our hypothesis that before the crash, we have a growing giant cluster. By tracking which clusters the 671 stocks belong to, we found that in the second period, one of the smaller clusters was absorbed by

the giant cluster. In the third period, this membership information revealed the initial stages of the giant cluster breaking up, as it ejected two smaller clusters. For the last period, we now found 12 different clusters, suggesting that we were near the end of the market crash. This fusion and fission phenomenon is in line with the hypothesis we made (see **Figure 7D**). To tease out subtle topological changes to the giant cluster as the market crash progressed, we looked to changes in the Betti numbers. This is possible because, unlike the use of a single criterion to group stocks in hierarchical clustering, TDA uses multiple criteria to accomplish this task. For example, it is possible for the giant cluster to continue absorbing stocks while it is ejecting others. The overall change in the number of clusters is reflected in β_0, however with β_0 alone we cannot distinguish between $1 - 2 = -1$ and $5 - 6 = -1$. These two scenarios may be differentiated by the other Betti numbers. Specifically, we found that β_2 changed from $62 \to 3 \to 4 \to 1 \to 11 \to 44 \to 11$ starting from the period (Jun 2018 to Nov 2018) to the period (Dec 2018 to May 2019), implying an initial decrease in the number of 2D holes just before the crash (**Figure 7A**), becoming 1 in the middle of the crash (**Figure 7B**), before increasing again just after the crash (**Figure 7D**).

We also wanted to check if there are any handle-breaking events during real market crashes. For example, if we start with a torus, the Betti numbers would be $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$. If the handle of the torus breaks, the object remaining would be homomorphic to a sphere, which has $\beta_0 = 1$, $\beta_1 = 0$, and $\beta_2 = 1$, indicating that handle breaking is a topological change whose signature is $(\Delta\beta_0, \Delta\beta_1, \Delta\beta_2) = (0, -2, 0)$. To track these topological changes, we kept only the giant cluster in each of the period, and recomputed the Betti numbers within these components. During the TAIEX mini-crash, we found the sequence $1 \to 4 \to 1 \to 2$ for β_0, which tells us that the giant cluster for the second period is the least homogeneous. Going beyond β_0, we found the sequence $23 \to 61 \to 1 \to 0$ for β_1. This tells us that the initial giant cluster already contained many irreducible loops, and this number of irreducible loops increased further in the second period as the giant cluster increased in size. By the third period, most of these irreducible loops have disappeared, and by the fourth period, the giant cluster remaining has a simple topological structure. Finally, for β_2, we found the sequence $78 \to 29 \to 44 \to 2$. Specifically, β_2 (the number of enclosed volumes) and β_1 (the number of irreducible loops) can together tell us more about the topology of the simplicial complex. For example, $\beta_1 = 0, \beta_2 = 1$ for a spherical shell, whereas for a torus, $\beta_1 = 2, \beta_2 = 1$. In the first period, we found that $\beta_1 < \beta_2$. This tells us that the giant cluster contains many enclosed volumes that are not holes (because every hole in the simplicial complex must be accompanied by irreducible loops). In the second period, we found instead that $\beta_1 > \beta_2$, and in fact $\beta_1 \approx 2\beta_2$, suggesting that all the enclosed volumes have become holes. The number of handles thus increased from the first period to the second period (although we cannot exclude the possibility that a few of them might have broken, although it is unlikely for many to have broken). From **Table 1**, we see that the giant cluster broke up most vigorously during the second and third period. Here we see that beyond this fragmentation, the topological changes associated with the second and third periods are very different: in the second period, enclosed volumes became holes, whereas in the third period, the handles of these holes broke and more enclosed volumes emerged. Furthermore, because β_2 was large in the fourth period, the fragmentation products are closer to being spherical shells than they are to solid spheres.

MST, PMFG, and TMFG

In the econophysics literature, we celebrate insights on stock markets obtained using correlation filtering methods. From Mantegna's work [8], we learned to project an arbitrary correlation matrix onto a minimal spanning tree, requiring only $N - 1$ links when there are N nodes, to visualize the correlational structure of stock markets. However, there is no reason why we should admit only $N - 1$ links. According to Tumminello et al. [9], the number of non-intersecting links in a graph G with genus g is at most $3(n - 2 + 2g)$, and therefore we may project the correlation matrix onto manifolds with different genus g to keep more links or fewer links. The simplest such projection is onto a sphere ($g = 0$), or other manifolds with a small genus. The graph that results from projection onto a sphere is planar and is therefore called a planar maximally filtered graph (PMFG). A related method, the triangular maximally filtered graph (TMFG) [10], that checks local planarity but not globally that the genus is zero. This is computationally more efficient and can be parallelized for very large datasets. However, there is no reason to believe that $g = 0$ is the optimum genus for all correlation matrices computed from stock markets. We believe that genus g implied in **Table 1** is optimum because they are computed in an unbiased fashion through the TDA filtration procedure. We can use this optimum genus to systematically improve the efficacy of such information filtering methods.

MST methods have been used to track topological changes during market crashes. To name a few, Onnela et al. [49, 50] investigated the US stocks during the 1987 Black Monday and found that the diameter of the MST decreased during the market crash, so this feature can be used as a universal indicator of market crashes. We ourselves also used the MST of the 10 US Dow Jones economic sectors [51] and the 36 Nikkei industry indices [52] in conjunction with time-series segmentation, to find a core-fringe structure during crizes. In the same spirit, Wilinski et al. [53] and Sienkiewicz et al. [54] investigated market crashes in the Frankfurt Stock exchange (FSE), and the Warsaw Stock Exchange (WSE), and concluded that a two-transition process characterizes market crashes universally. The first transition is from a hierarchical scale-free MST to a superstar-like MST, followed by a second transition to a power-law MST decorated with star-like trees or hubs. In using the MST, they have assumed that loops (β_1) in the networks can be ignored. In this sense, the present is a natural extension to what they have done, where we take a more detailed look into the topological transitions.

Ultimately, informational filtering methods such as MST or PMFG are designed to produce connected graphs and are thus not the best choice for identifying fragmented clusters. To identify these, we can of course use the minimal spanning

forest (MSF) [55] or the directed bubble hierarchical tree (DBHT) [56, 57] methods modified from the MST or PMFG. Here, we would like to stress that most clustering approaches are limited to β_0 and β_1, and cannot differentiate topological changes beyond β_1. TDA is promising because it is an elegant extension beyond *0*-simplices and *1*-simplices, allowing us to unravel subtle topological changes during market crashes.

Outlook and Perspective

Several future directions are possible based on this work. First, in **MST, PMFG, and TMFG** Section, we mention that the upper bound of the number of links for information filtering methods involving projection to manifolds with genus g is given by $3(n-2+2g)$. The Euler characteristic listed in **Tables 1(A, B)** can be also be used to calculate the genus g via $\chi = 2(1-g)$. With this we are not required to assume *a priori* that the optimal manifold to project has $g = 0$ or close to 0. Second, Bubenik [58] had proposed to use persistence landscapes, which is a *Banach space* that can be converted from persistence diagrams. We then can do statistical averaging of the persistence landscapes, and use the result to design persistence weighted kernels, see for example this very recent work [59]. Persistence weighted kernels can fully maximize the strength of ML algorithms in making stock price predictions. Third, we identified two market crashes in TWSE and several topological phase transitions in SGX.

A pearl of common wisdom that can be gleaned from [47] is that the number of simplices (kth Betti numbers), in general, will peak at $k = 6$ to $k = 8$, before dropping to zero at $k = 11$. More computing resources are required to carry out future works in this direction to test at which k the number of simplices actually peak, and at which k it finally dropped to zero. Also, in Reimann's work, they investigated directed simplices instead of undirected ones. The former finds applications in educational science, for example [60]. Also, a recent work that applied persistent homology in investigating co-occurrence networks had shown promising results [61].

CONCLUSION

In this work, we collected daily price data from SGX and TWSE and analyzed them using persistent homology and TDA toolkits. We then made a case for TDA to be employed alongside the other state-of-the-art network embedding techniques including the MST, PMFG, TMFG, in analyzing the topological structures. We were drawn to the application of Persistent Homology (PH) and TDA in complex systems for three reasons:) PH and TDA are unbiased;) they scan through a full range of correlation values instead of using only one or two specific values; and) it is less susceptible to random noises.

We then utilize three toy models to illustrate our hypothesis in **Introduction** Section, that is "in different market states, their topological features are also changing accordingly, and TDA can be effective in scrutinizing these changes." We showed in these toy models, including spheres, toruses, and ellipsoids, how χ, the Betti numbers, the barcodes, and persistence diagrams change with topological changes. Also, we use schematic diagrams to illustrate different market states, what the topologies could be like, and argue what their possible Betti numbers and χ's could be.

Our results revealed unexpected and promising findings in the stock markets. In TWSE, we found a small crash from Sep 2018 to Jan 2019, followed by a larger crash in March 2020, which is due to the COVID-19 pandemic. For these two crashes, we performed three tests using TDA methods. The first test was to quantify a persistence-weakening phenomenon in the barcodes and persistence diagrams. This persistence-weakening phenomenon was also discovered in the SGX, suggesting that it might be universal. However, there were no reported crashes in the SGX for the period studied. To understand this apparent inconsistency, in the second test we calculated the Betti numbers and the Euler characteristic of different 6-months windows in both markets. Our results suggest that market crashes in TAIEX and STI are associated with $\chi > 0$, but the market crash signatures are stronger and have cleaner interpretations in β_0. When we scrutinized the changes to β_0, β_1, and β_2 of the giant cluster over four time periods before, during, and after the TAIEX mini-crash, we found that at the beginning of the crash, the giant cluster has many holes and many more enclosed volumes. As the market crash progressed, these enclosed volumes first became holes, before the handles of these holes broke, to give rise to fragmentation products that were closer to spherical shells than they are to solid spheres. Finally, in the last test, we found the dim-2 persistent entropy decreasing significantly across market crashes. To conclude we found that TDA confirmed most parts of our hypothesis, but also suggested that the topological changes surrounding a market crash are more complex than what we had imagined.

DATA AVAILABILITY STATEMENT

All Python and Matlab scripts are provided, along with instructions on how to use them. These will download the raw data from Yahoo! Finance, perform the necessary computations to give the final results. The files are provided in this link. https://researchdata.ntu.edu.sg/dataset.xhtml?persistentId=doi:10.21979/N9/8XMZGF.

AUTHOR CONTRIBUTIONS

SC and PT-WY conceived the study, PT-WY collected the data, SC and PT-WY analyzed the data and interpreted the results, SC and PT-WY wrote and reviewed the manuscript.

FUNDING

This research is supported by a startup grant from the Nanyang Technological University.

ACKNOWLEDGMENTS

We thank our NTU colleagues Melvin Soh Hwee Jin and Jeric Ho Yew Kee for their assistance in providing advice and

solutions to our inquiries on high-performance computing-related issues. P T.W.Y. acknowledges support from ERI@N. We also thank the two anonymous reviewers for their insightful and constructive comments that helped improved our revised manuscript greatly.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2021.572216/full#supplementary-material.

REFERENCES

1. Laloux L, Cizeau P, Bouchaud J-P, Potters M. Noise dressing of financial correlation matrices. *Phys Rev Lett* (1999) 83(7):1467–70. doi:10.1103/PhysRevLett.83.1467
2. Plerou V, Gopikrishnan P, Rosenow B, Nunes Amaral LA, Stanley HE. Universal and nonuniversal properties of cross correlations in financial time series. *Phys Rev Lett* (1999) 83(7):1471–4. doi:10.1103/PhysRevLett.83.1471
3. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Stanley HE. A random matrix theory approach to financial cross-correlations. *Phys Stat Mech Appl* (2000) 287(3-4):374–82. doi:10.1016/s0378-4371(00)00376-9
4. Sandoval L, Franca IDP. Correlation of financial markets in times of crisis. *Phys Stat Mech Appl* (2012) 391(1-2):187–208. doi:10.1016/j.physa.2011.07.023
5. Mantegna RN, Stanley HE. Stochastic process with ultraslow convergence to a Gaussian: the truncated Levy flight. Epub 1994/11/28. *Phys Rev Lett* (1994) 73(22):2946–9. PubMed PMID: 10057243. doi:10.1103/PhysRevLett.73.2946
6. Mandelbrot BB, Van Ness JW. Fractional brownian motions, fractional noises and applications. *SIAM Rev* (1968) 10(4):422–37. doi:10.1137/1010093
7. Mandelbrot BB, Fisher AJ, Calvet LE. *A multifractal model of asset returns* New Haven: Yale University (1997).
8. Mantegna RN. Hierarchical structure in financial markets. *The European Physical Journal B* (1999) 11(1):193–7. doi:10.1007/s100510050929
9. Tumminello M, Aste T, Di Matteo T, Mantegna RN. A tool for filtering information in complex systems. *Proc. Natl. Acad. Sci. U.S.A* (2005) 102(30):10421–6. doi:10.1073/pnas.0500298102
10. Massara GP, Di Matteo T, Aste T. Network filtering for big data: triangulated maximally filtered graph. *Journal of Complex Networks* (2016) 5(2):161–78. doi:10.1093/comnet/cnw015
11. Zomorodian A, Carlsson G. Computing persistent homology. *Discrete Comput Geom* (2004) 33(2):249–74. doi:10.1007/s00454-004-1146-y
12. G Singh, F Mémoli, GE Carlsson, editors. *Topological methods for the analysis of high dimensional data sets and 3d object recognition*. Geneva: SPBG (2007).
13. H Edelsbrunner, D Letscher, A Zomorodian, editors. Topological persistence and simplification. Proceedings 41st annual symposium on foundations of computer science. Redondo Beach, CA, USA, 2000, IEEE (2000). doi:10.1109/SFCS.2000.892133
14. Carlsson G. Topology and data. *Bull Am Math Soc* (2009) 46(2):255–308. doi:10.1090/s0273-0979-09-01249-x
15. Hatcher A. *Algebraic topology*. Cambridge University Press (2002).
16. Edelsbrunner H, Harer J. *Computational topology: an introduction*. American Mathematical Soc. (2010). doi:10.1007/978-3-540-33259-6_7
17. Ghrist RW. *Elementary applied topology*. Seattle: CreateSpace Independent Publishing Platform (2014).
18. Eilenberg S, Steenrod N. *Foundations of algebraic topology*. Princeton University Press (2015).
19. Munkres JR. *Elements of algebraic topology*. New York: CRC Press (2018).
20. Cotton FA. *Chemical applications of group theory*. John Wiley & Sons (2003).
21. Dresselhaus MS, Dresselhaus G, Jorio A. *Group theory: application to the physics of condensed matter*. Springer Berlin Heidelberg (2007).
22. Strang G. Linear Algebra and its applications*: thomson*. Boston: Cengage Learning (2006).
23. Lay DC. *Linear algebra and its applications*. Addison-Wesley (2012).
24. Barabási A-L. *Network science*. Cambridge University Press (2016).
25. West DB. *Introduction to graph theory*. Upper Saddle River: Prentice-Hall (2001).
26. Newman M. *Networks: an introduction*. Oxford: OUP Oxford (2010).
27. Hausmann J-C. On the Vietoris-Rips complexes and a cohomology theory for metric spaces. *Ann Math Stud* (1995) 138:175–88. doi:10.1515/9781400882588-013
28. Toth CD, O'Rourke J, Goodman JE. *Handbook of discrete and computational geometry*. Chapman and Hall/CRC (2017).
29. H Edelsbrunner, editor. Smooth surfaces for multi-scale shape representation. *International conference on foundations of software technology and theoretical computer science*. Springer (1995).
30. De Silva V, Carlsson GE. Topological estimation using witness complexes. *SPBG* (2004) 4:157–66. doi:10.2312/SPBG/SPBG04/157-166
31. Otter N, Porter MA, Tillmann U, Grindrod P, Harrington HA. A roadmap for the computation of persistent homology. Epub 2017/01/01. *EPJ Data Sci* (2017) 6(1):17, 2017 . PubMed PMID: 32025466; PubMed Central PMCID: PMCPMC6979512. doi:10.1140/epjds/s13688-017-0109-5
32. Edelsbrunner H, Harer J. Persistent homology-a survey. *Contemp Math* (2008) 453:257–82.
33. Bauer U. Ripser: a lean C++ code for the computation of Vietoris-Rips persistence barcodes. *Software* (2017). available at. https://github.com/Ripser/ripser (Accessed September 04, 2020).
34. Carrara N. *TDA toolkit* (2018). [updated September09/04/2020]. doi:10.5281/zenodo.1436034
35. Adams H, Tausz A, Vejdemo-Johansson M. Java{P}lex: {A} research software package for persistent (co)homology. In: Hong H, Yap, CK, editors. *Proceedings of ICMS*. 2014 (2014).
36. Sornette D. *Why stock markets crash: critical events in complex financial systems*. New Jersey: Princeton University Press (2017).
37. Lux T. Herd behaviour, bubbles and crashes. *Econ J* (1995) 105(431):881–96. doi:10.2307/2235156
38. Teh BK, Cheong SA. Cluster fusion-fission dynamics in the Singapore stock exchange. *The European Physical Journal B* (2015) 88(10):263. doi:10.1140/epjb/e2015-60456-y
39. Teh BK, Goo YW, Lian TW, Ong WG, Choi WT, Damodaran M, et al. The Chinese Correction of February 2007: how financial hierarchies change in a market crash. *Phys Stat Mech Appl* (2015) 424:225–41. doi:10.1016/j.physa.2015.01.024
40. Kleinen T, Held H, Petschel-Held G. The potential role of spectral properties in detecting thresholds in the Earth system: application to the thermohaline circulation. *Ocean Dynam* (2003) 53(2):53–63. doi:10.1007/s10236-002-0023-6
41. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, et al. Early-warning signals for critical transitions. Epub 2009/09/04. *Nature* (2009) 461(7260):53–9. PubMed PMID: 19727193. doi:10.1038/nature08227
42. Folke C, Carpenter S, Walker B, Scheffer M, Elmqvist T, Gunderson L, et al. Regime shifts, resilience, and biodiversity in ecosystem management. *Annu Rev Ecol Evol Syst* (2004) 35(1):557–81. doi:10.1146/annurev.ecolsys.35.021103.105711
43. Scheffer M, Carpenter S, Foley JA, Folke C, Walker B. Catastrophic shifts in ecosystems. Epub 2001/10/12. *Nature* (2001) 413(6856):591–6. PubMed PMID: 11595939. doi:10.1038/35098000
44. Rucco M, Castiglione F, Merelli E, Pettini M. Characterisation of the idiotypic immune network through persistent entropy. Proceedings of ECCS Springer proceedings in complexity. Springer (2014). p. 117–28. 2016.
45. Chintakunta H, Gentimis T, Gonzalez-Diaz R, Jimenez M-J, Krim H. An entropy-based persistence barcodes. *Pattern Recogn* (2015) 48(2):391–401. doi:10.1016/j.patcog.2014.06.023
46. Fan S, Raposo EP, Coutinho-Filho MD, Copelli M, Stam CJ, Douw L. Topological phase transitions in functional brain networks. Epub 2019/10/24. *Phys Rev E* (2019) 100(3-1):032414, 2019. PubMed PMID: 31640025. doi:10.1103/PhysRevE.100.032414

47. Reimann MW, Nolte M, Scolamiero M, Turner K, Perin R, Chindemi G, et al. Cliques of neurons bound into cavities provide a missing link between structure and function. Epub 2017/07/01. *Front Comput Neurosci* (2017) 11:48, 2017. PubMed PMID: 28659782; PubMed Central PMCID: PMCPMC5467434. doi:10.3389/fncom.2017.00048
48. Fan S, da Silva LCB, Coutinho-Filho MD. Topological approach to microcanonical thermodynamics and phase transition of interacting classical spins. *J Stat Mech Theor Exp* (2017) 2017(1):013202. doi:10.1088/1742-5468/2017/1/013202
49. Onnela J-P, Chakraborti A, Kaski K, Kertiész J. Dynamic asset trees and portfolio analysis. *The European Physical Journal B-Condensed Matter and Complex Systems* (2002) 30(3):285–8.
50. Onnela J-P, Chakraborti A, Kaski K, Kertesz J. Dynamic asset trees and Black Monday. *Phys Stat Mech Appl* (2003) 324(1-2):247–52.
51. Zhang Y, Lee GHT, Wong JC, Kok JL, Prusty M, Cheong SA. Will the US economy recover in 2010? A minimal spanning tree study. *Phys Stat Mech Appl* (2011) 390(11):2020–50.
52. Cheong SA, Fornia RP, Lee GHT, Kok JL, Yim WS, Xu DY, et al. The Japanese economy in crises: a time series segmentation study. *Economics: the Open-Access. Open-Assessment E-Journal* (2012) 6(2012-5):1–81.
53. Wiliński M, Sienkiewicz A, Gubiec T, Kutner R, Struzik Z. Structural and topological phase transitions on the German Stock Exchange. *Phys Stat Mech Appl* (2013) 392(23):5963–73.
54. Sienkiewicz A, Gubiec T, Kutner R, Struzik ZR. *Dynamic structural and topological phase transitions on the Warsaw Stock Exchange: A phenomenological approach*. arXiv preprint arXiv:13016506 (2013).
55. Di Gesù V, Sacco B. Some statistical properties of the minimum spanning forest. *Pattern Recognition* (1983) 16(5):525–31.
56. Song W-M, Di Matteo T, Aste T. Hierarchical information clustering by means of topologically embedded graphs. *PloS One* (2012) 7(3):e31929.
57. Musmeci N, Aste T, Di Matteo T. Relation between financial market structure and the real economy: comparison between clustering methods. *PloS One* (2015) 10(3):e0116201.
58. Bubenik P. Statistical topological data analysis using persistence landscapes. *J Mach Learn Res* (2015) 16(1):77–102. doi:10.5555/2789272.2789275
59. G Kusano, Y Hiraoka, K Fukumizu, editors. Persistence weighted Gaussian kernel for topological data analysis. New York: International Conference on Machine Learning (2016).
60. Goh WP, Kwek D, Hogan D, Cheong SA. Complex network analysis of teaching practices. *EPJ Data Science* (2014) 3(1):1–16. doi:10.1140/epjds/s13688-014-0034-9
61. Salnikov V, Cassese D, Lambiotte R, Jones NS. Co-occurrence simplicial complexes in mathematics: identifying the holes of knowledge. Epub 2018/01/01. *Appl Netw Sci* (2018) 3(1):37, 2018 . PubMed PMID: 30839828; PubMed Central PMCID: PMCPMC6214324. doi:10.1007/s41109-018-0074-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yen and Cheong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 12 April 2021
doi: 10.3389/fphy.2021.602944

Characteristics of Principal Components in Stock Price Correlation

Wataru Souma *

College of Science and Technology, Nihon University, Funabashi, Japan

The following methods are used to analyze correlations among stock returns. 1) The meaningful part of the correlation is obtained by applying random matrix theory to the equal-time cross-correlation matrix of assets returns. 2) Null-model randomness is implemented *via* rotational random shuffling. 3) Principal component analysis and Helmholtz-Hodge decomposition are used to extract leading and lagging relationships among assets from the complex correlation matrix constructed from the Hilbert-transformed data set of asset returns. These methods are applied to price data for 445 assets from the S&P 500 from 2010 to 2019 (2,510 business days). Additional analysis and discussion clarify key aspects of leading and lagging relationships among business sectors in the market. Numerical investigation of these dataset reveals the possibility that leading and lagging relationships among business sectors may depend on gross market conditions.

Keywords: S&P 500, stock return, cross-correlation matrix, random matrix theory, principal component, complex correlation matrix, complex hilbert principal component analysys, helmholtz-Hodge decomposition

OPEN ACCESS

Edited by:
Wei-Xing Zhou,
East China University of Science and Technology, China

Reviewed by:
Gholamreza Jafari,
Shahid Beheshti University, Iran
Xiao Han,
University of California, Davis, United States

***Correspondence:**
Wataru Souma
wataru.soma@gmail.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *04 September 2020*
Accepted: *10 February 2021*
Published: *12 April 2021*

Citation:
Souma W (2021) Characteristics of Principal Components in Stock Price Correlation. Front. Phys. 9:602944. doi: 10.3389/fphy.2021.602944

1 INTRODUCTION

The analysis of big data can reveal novel aspects of nature and society. However, data often contain noise, making it necessary to distinguish the signal from the noise. Principal component analysis (PCA), independent component analysis, machine learning, and other techniques have been applied to extract the meaningful components of various datasets. About 20 years ago, random matrix theory (RMT) was introduced to distinguish the components of a dataset from the noise. [1, 2] developed a "null-hypothesis" test based on RMT. In paticular, they compared the properties of empirical equal-time cross-correlation matrix to those of a random matrix and considered deviations from the random matrix case to suggest the presence of meaningful information. They compared the distribution of eigenvalues of this empirical cross-correlation matrix with the Marčenko-Pastur distribution [3], which is theoretically derived from so-called random Wishart matrices. They considered the eigenvector corresponding to the largest eigenvalue to represent the "market" itself. They also compared the distributions of the components of eigenvectors with the Porter-Thomas distribution [4], finding that the eigenvector corresponding to the largest eigenvalue differed remarkably from the Porter-Thomas distribution.

[5] confirmed the findings by [1, 2]; the meaningful part represents a market mode and group structures, such as industry categories and stocks with large market capitalization. [6] applied RMT to the equal-time cross-correlation matrix of assets listed on the first division of the Tokyo Stock Exchange (TSE). [7] clarified the structure of the meaningful part of the equal-time cross-correlation matrix of assets listed on the New York Stock Exchange (NYSE). [8] investigated the empirical

equal-time cross-correlation matrix of stock price fluctuations on the National Stock Exchange of India, finding that this emerging market exhibited strong correlations in the movements of stock prices compared to developed markets such as the NYSE. [9] analyzed the empirical equal-time cross-correlation matrix of stock price fluctuations on the Tehran stock exchange and in the Dow Jones Industrial Average (DJIA), showing that the DJIA is more sensitive to global perturbations. [10] investigated the structures of networks constructed from principal components of the empirical equal-time cross-correlation matrices of stock price fluctuations on the Tehran stock exchange and in the DJIA. [11] constructed an autocorrelation matrix of a time series and analyzed it based on the random-matrix theory approach and fractional Gaussian noises.

[5] constructed a "filtered" cross-correlation matrix, from eigenvalues and eigenvectors outside the random matrix bound and applied this cross-correlation matrix to portfolio optimization [12]. The result they obtained shows that predicted risk was much closer to the realized risk than the traditional portofolio optimaization. [13] applied the portfolio optimization method to the stocks listed on the first division of the TSE and showed that the performance of the portfolio constructed by this method was usually better than that of market index such as TOPIX. [14] extended this portfolio optimization method to a case involving a short sale of stocks.

RMT is a powerful method for distinguishing meaningful components and noise in financial time-series data. The null hypothesis of randomness in this method assumes randomness in cross-correlation and autocorrelation. However, the autocorrelation of stock returns cannot be considered random (for example, see [15]. Thus, a new method is needed that preserves autocorrelation but randomizes cross-correlation. [16, 17] developed a method referred to as rotational random shuffling (RRS). In RRS, empirical time-series data are shuffled rotationally in the time direction with a periodic boundary condition imposed. Therefore, equal-time cross-correlation matrices constructed from RRS time series preserve almost all the autocorrelation information of each time series while randomizing cross-correlation. By comparing the distribution of eigenvalues of this RSS cross-correlation matrix with that of the empirical cross-correlation matrix, meaningful components and noise can be successfully distinguished.

It is natural to consider the application of RMT to different-time cross-correlation matrix. [18] introduced so-called complex Hilbert principal component analysis (CHPCA), in which the cross-correlation matrix is defined in the complex space. The components of eigenvectors of the complex cross-correlation matrix distribute in the complex plane, allowing the recognition of lead-lag relationships between components based on the difference in angle between them. [19] applied CHPCA to time-series data set for 483 assets representing the S&P 500 from 2008 to 2011 (1,009 business days) and constructed a correlation network in which pairs of assets with phase differences below a certain threshold were weighted based on correlation strength. [20] explored data from 1990 to 2012 for foreign exchanges and stock markets in 48 countries using CHPCA and extracted a significant lead-lag relationship between the markets. [21] applied CHPCA to a time-series data for assets listed on the NYSE from 2005 to 2014 and clarified lead-lag relationships among stocks, investment trusts, real estate investment trusts (REITs), and exchange traded funds (ETFs). [22, 23] applied CHPCA to the early warning indicators of financial crizes proposed by the Bank of Japan and explored changes in lead-lag relationships between indices before and after financial crizes.

When applying CHPCA to time series data, we need to explicitly extract the lead-lag relationship between the time series. [24, 25]; and [26] applied the Helmholtz-Hodge decomposition (HHD) to extract circular and gradient flows in a complex network. [27] applied CHPCA and HHD to monthly time series of 57 US macroeconomic indicators and five trade/money indexes, confirming statistically significant co-movements among these time series and identifying noteworthy economic events. [28] summarized CHPCA, RRS, and HHD and applied these methods to economic time-series data.

The purpose of the present paper is twofold. The first is to introduce a recently developed method to analyze stock return correlations. The second is to highlight a novel aspect of leading and lagging relations of business sectors in the market. In **Section 2**, log returns of stock prices are defined, and an empirical equal-time cross-correlation matrix is constructed for 445 assets from the S&P 500 from 2010 to 2019 (2,510 business days). A method is also presented for calculating the eigenvalues and eigenvectors of this cross-correlation matrix and applies RMT and RRS to distinguish the meaningful part from the noise. Furthermore, it is shown that the eigenvector corresponding to the largest eigenvalue represents the market mode and meaning components without the principal component represent group mode. In **Section 3**, the dataset is investigated using CHPCA, RRS, and HHD and lead-lag relationships among assets are discussed. In **Section 4**, an application of CHPCA to portfolio theory is sketched. **Section 5** is devoted to summary and discussion.

2 APPLICATION OF RMT AND RRS

In this section, the equal-time cross-correlation matrix is defined. RMT is then applied to distinguish the meaningful components from the noise components. After that, RRS is introduced to distinguish the meaning components from the noise components.

2.1 Equal-Time Cross-Correlation Matrix

This paper investigates data for 445 assets from the S&P 500 for dates obtained 2010–2019 (2,510 business days). By denoting an opening price of stock n on day t as $o_n(t)$ and a closing price of stock n on day t as $c_n(t)$, the daily log return of stock n on day t is defined as

$$r_n(t) = \ln\left[\frac{c_n(t)}{o_n(t)}\right] \tag{1}$$

where ln represents the natural logarithm. Here, $n = 1, 2, \ldots, N = 445$, and $t = 1, 2, \ldots, T = 2510$. For each stock n, the time-average of $r_n(t)$ is denoted as $\langle r_n \rangle$, and the standard deviation of $r_n(t)$ is denoted as σ_n. These are defined by

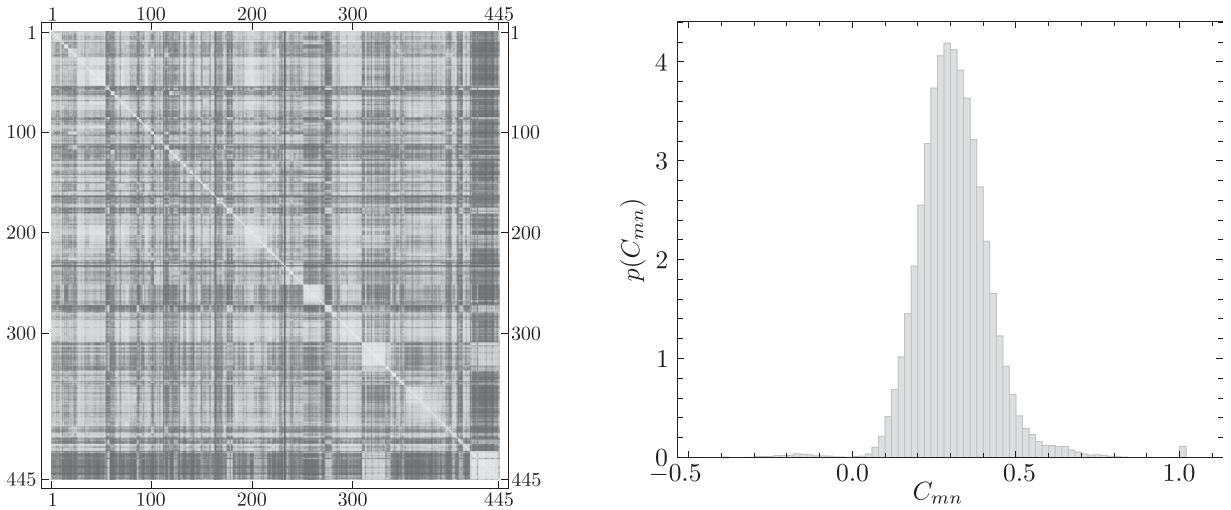


FIGURE 1 | (Left) Visualization of the equal-time cross-correlation matrix for the data for 445 assets from S&P 500 from 2010 to 2019 (2,510 business days). The shade is proportional to correlation strength, with white color corresponding to $C_{nn} = 1$ and the color becoming dark as the correlation becomes large-magnitude negative. **(Right)** The distribution $p(C_{mn})$ of components of the equal-time cross-correlation matrix. The components are almost all positive. The right tail of the distribution is thicker than the left tail.

$$\langle r_n \rangle = \frac{1}{T}\sum_{t=1}^{T} r_n(t), \qquad \sigma_n = \sqrt{\frac{1}{T}\sum_{t=1}^{T}[r_n(t) - \langle r_n \rangle]^2} \tag{2}$$

A normalized log return of asset n is denoted as $w_n(t)$, and define it by

$$w_n(t) = \frac{r_n(t) - \langle r_n \rangle}{\sigma_n} \tag{3}$$

Thus, a component of equal-time cross-correlation matrix is defined by

$$C_{mn} = \frac{1}{T}\sum_{t=1}^{T} w_m(t) w_n(t) \tag{4}$$

The left panel of **Figure 1** depicts an equal-time cross-correlation matrix. In this figure, shade indicates the strength of the positive correlation. White color corresponds to $C_{nn} = 1$, with darker shades representing weaker correlations, and yet darker shades representing negative correlations. The darkest shade corresponds to $C_{mn} = -0.515641$. Because the stocks are arranged in industry codes orders, the block pattern seen in the figure roughly corresponds to a grouping by industry. The right panel of **Figure 1** shows the distribution of components of the equal time cross-correlation matrix. This figure shows that nearly all correlations are positive. Furthermore, the right tail of the distribution is thicker than the left tail.

2.2 Application of RMT

Calculation of eigenvalues λ_R for this cross-correlation matrix produces **Figure 2**. Here, subscript R represents the eigenvalue rankings. The left panel of **Figure 2** shows the distribution of eigenvalues. The largest eigenvalue is $\lambda_1 = 143.516$, and the smallest eigenvalue is $\lambda_{445} = 0.0638128$. The right panel of **Figure 2** shows the distribution in the range of small eigenvalues. The solid line is the probability distribution function of the so-called Marčenko-Pastur distribution, which is derived from RMT in the limit $N \to \infty$ and $T \to \infty$ by fixing $Q = N/T$:

$$p(\lambda) = \left(1 - \frac{1}{Q}\right)^{+} \delta(\lambda) + \frac{1}{2\pi Q}\frac{\sqrt{(\lambda - \lambda_-)^{+}(\lambda_+ - \lambda)^{+}}}{\lambda} \tag{5}$$

where $(x)^{+} = \max(0, x)$; $\delta(x)$ denotes Dirac's delta function; and $\lambda_{\pm}$ is defined by

$$\lambda_{\pm} = \left(1 \pm \sqrt{Q}\right)^2 \tag{6}$$

In this paper, $\lambda_+ = 2.01941$ denotes the upper bound of eigenvalue λ, and $\lambda_- = 0.335172$ denotes the lower bound of λ.

In RMT extraction of the meaningful part of the correlation structure, empirical eigenvalues larger than λ_+ signify the meaningful part. In particular, in the cross-correlation matrix of stock returns, the largest eigenvalue corresponds to the market mode, and the remaining meaningful part correspond to group modes, such as, industry sectors. In this analysis, it was found that $\lambda_1 > \lambda_2 > \ldots > \lambda_{17} > \lambda_+$, so, 17 meaningful components were retained.

In traditional PCA, Monte Carlo simulations and so-called scree graphs are used to extract meaningful components. In the present method, the time series of each stock is randomly shuffled to generate an equal-time cross-correlation matrix. This manipulation breaks both the autocorrelation and the cross-correlation. It is derived from a similar concept as the application of RMT. If we construct the equal-time

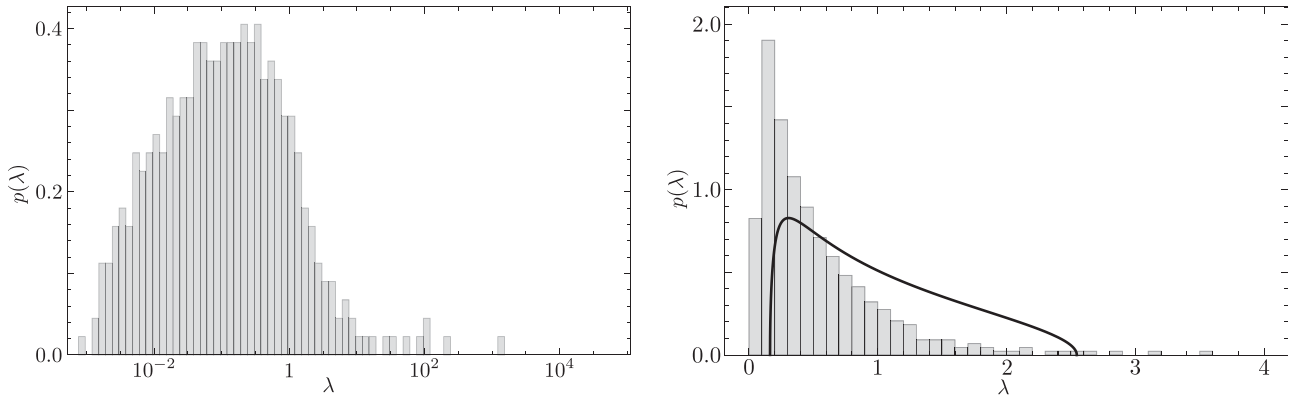


FIGURE 2 | (Left) Distribution $p(\lambda)$ of eigenvalues λ of the empirical equal-time cross-correlation matrix. **(Right)** Empirically obtained distribution $p(\lambda)$ of eigenvalues λ in the range of small eigenvalues. The solid line is the Marčenko-Pastur distribution under RMT as the theoretical curve given by **Eq. 5**.

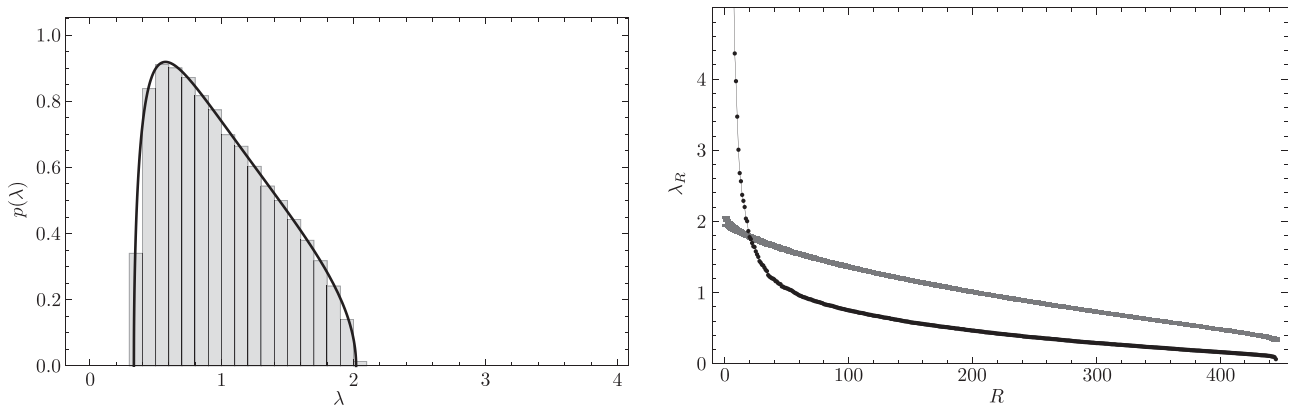


FIGURE 3 | (Left) Distribution $p(\lambda)$ of eigenvalues λ of the equal-time cross-correlation matrix constructed from randomly shuffled time series. The solid line represents the Marčenko-Pastur distribution derived under RMT as the theoretical curve given by **Eq. 5**. **(Right)** Scree graph of eigenvalues. The abscissa represents eigenvalue rankings R, and the ordinate represents empirically obtained eigenvalues λ_R. The curve with error bars depicts the simulated distribution of eigenvalues using random shuffling (RS). To obtain this curve, we repeated this manipulation 20 times and calculated the mean value and standard deviation. Each error bar represents three times the standard deviation. The thin line with filled circles depicts the distribution of eigenvalues of the empirical equal-time cross-correlation matrix. The meaningful part can be obtained by comparing these two distributions. If the upper bound for eigenvalues derived from the randomly shuffled cross-correlation matrix is denoted as $\lambda_{\max}$, then $\lambda_1 > \lambda_2 > \ldots > \lambda_{19} > \lambda_{\max} = 1.7947$. Hence, 19 meaningful components should be retained for this data set.

cross-correlation matrix from those randomly shuffled time series, we can obtain the histogram shown in the left panel of **Figure 3**. The solid line in this figure corresponds to the Marčenko-Pastur distribution given by **Eq. 5**. From this figure, we can recognize the equivalence between the traditional PCA and the application of RMT.

The right panel of **Figure 3** shows the scree graph. In this figure, the abscissa corresponds to the eigenvalue rankings and the ordinate corresponds to the magnitude of eigenvalues. The curve with error bars in this figure depicts the eigenvalue distribution of the randomly shuffled cross-correlation matrix. The thin line with filled circles in this figure depicts the distribution of eigenvalues of the empirical equal-time cross-correlation matrix. If we denote the upper bound of eigenvalue derived from the randomly shuffled cross-correlation matrix as $\lambda_{\max}$, we obtain $\lambda_1 > \lambda_2 > \ldots > \lambda_{19} > \lambda_{\max} = 1.7947$. Hence, there are 19 meaningful components in the dataset.

2.3 Application of the RRS

As stated above, when we make a randomly shuffled cross-correlation matrix, we break both the autocorrelation and the cross-correlation conditions. However, it has been reported that the stock price has an autocorrelation tendency. Thus, we need to develop a method that preserves autocorrelation but randomizes the crosscorrelation. [16, 17] developed a method referred to as RRS. In RRS, we shuffle the empirical time-series data rotationally in the time direction and impose the periodic boundary condition:

$$w_n(t) \to w_n(Mod[t + \tau, T]) \tag{7}$$

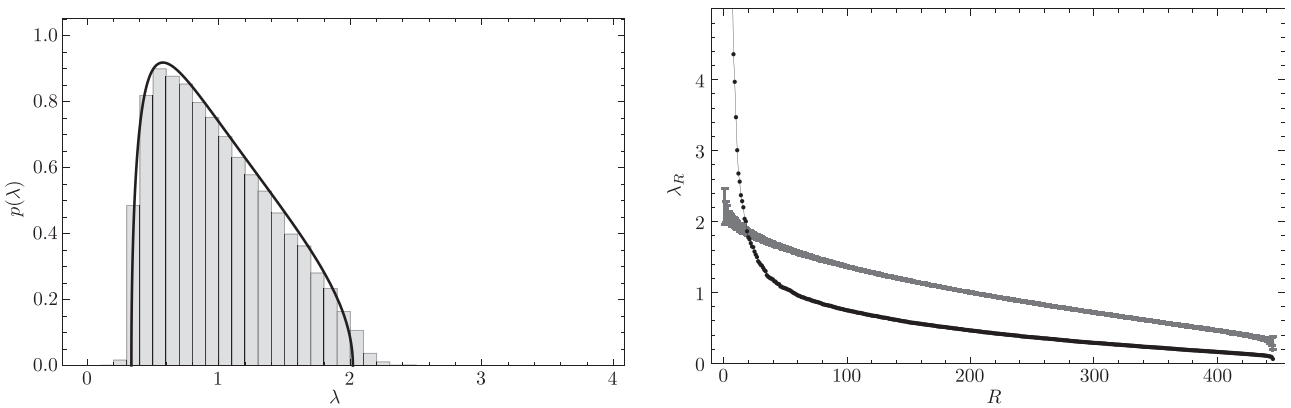


FIGURE 4 | (Left) Distribution $p(\lambda)$ of eigenvalues λ of the equal-time cross-correlation matrix constructed by rotational random shuffling (RRS). The solid line represents the Marčenko-Pastur distribution derived under RMT as the theoretical curve given by **Eq. 5**. **(Right)** Scree graph of eigenvalues. The abscissa represents eigenvalue rankings R, and the ordinate represents empirically obtained eigenvalues λ_R. The thin line with filled circles depicts the empirically obtained distribution of eigenvalues of the empirical equal-time cross-correlation matrix. The curve with error bars depicts the simulated distribution of eigenvalues using RRS. To obtain this curve, this manipulation was repeated 20 times, after which the mean value and standard deviation were calculated. Each error bar represents three times the standard deviation. The thin line with filled circles in this figure depicts the distribution of eigenvalues of the empirical equal-time cross-correlation matrix. The meaningful part can be obtained by comparing these two distributions. If the upper bound for eigenvalues derived from the RRS cross-correlation matrix is denoted as $\lambda_{\max}$, then $\lambda_1 > \lambda_2 > \ldots > \lambda_{19} > \lambda_{\max} = 1.7947$. Hence, 19 meaningful components should be retained for this data set.

Here, $\tau \in [0, T-1]$ is a (pseudo-) random integer that is different for each n. For example, if $\tau = 1537$ for stock 1, $\tau = 2128$ for stock 2, . . ., $\tau = 138$ for stock N, the time series of normalized log returns is given by

$$
\begin{aligned}
w_1 &= \{w_1(1538), w_1(1539), \ldots, w_1(2510), w_1(1), w_1(2), \ldots, w_1(1537)\} \\
w_2 &= \{w_2(2129), w_2(2130), \ldots, w_2(2510), w_2(1), w_2(2), \ldots, w_2(2128)\} \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
w_N &= \{w_N(139), w_N(140), \ldots, w_N(2510), w_N(1), w_N(2), \ldots, w_N(138)\}
\end{aligned}
$$

Such a rotationally randomly shuffled time series allows the cross-correlation matrix to be constructed and eigenvalues to be calculated. An example is shown in the histogram in the left panel of **Figure 4**. The solid line in this figure corresponds to the Marčenko-Pastur distribution given by **Eq. 5**. This figure shows that the distribution of eigenvalues is almost the same as the Marčenko-Pastur distribution based on RMT except for the large eigenvalue range.

The right panel of **Figure 4** shows the scree graph. In this figure, the abscissa corresponds to eigenvalue rankings, and the ordinate corresponds to eigenvalue magnitude. The curve with error bars in this figure depicts the eigenvalue distribution of the RRS cross-correlation matrix. The thin line with filled circles in this figure depicts the distribution of eigenvalues of the empirical equal-time cross-correlation matrix. Again, if the upper bound of eigenvalues derived from the RRS cross-correlation matrix is denoted as $\lambda_{\max}$, then $\lambda_1 > \lambda_2 > \ldots > \lambda_{19} > \lambda_{\max} = 1.7947$ is obtained. Hence, 19 meaningful components are retained. Although the numbers of meaningful components in RMT and RRS are equal, this result is a coincidence specific to the data set at hand.

Figure 5 shows the distribution of components of the top 20 eigenvectors, $v_1, \ldots, v_{20}$. The thin vertical lines in these figures separate business sectors. RMT suggests that the distribution of the components of each eigenvector is given by the Poter-Thomas distribution:

$$
p(v) = \frac{N}{2\pi} \exp\left(-\frac{Nv^2}{2}\right) \tag{8}
$$

The first eigenvector v_1 consists of components of similar magnitude and is referred to as the market mode. In the second eigenvector, there is a negative peak in the rightmost sector, which corresponds to the utility sector. In the third eigenvector, there is a negative peak in the left sector, which corresponds to the bank sector. In the fourth eigenvector, there is a positive peak in the middle sector, which corresponds to the oil and gas equipment and service sector. In the fifth eigenvector, there is a negative peak in the right middle sector, which corresponds to the REIT sector. The panels from the sixth eigenvector to the 20th eigenvector have peaks in some sectors containing a small number of assets. However, sometimes it is difficult to extract the meaning of each principal component. Thus, the correlation matrix was split into three parts:

$$
\begin{aligned}
C &= \sum_{R=1}^{N} \lambda_R v_R v_R^T \\
&= \lambda_1 v_1 v_1^T + \sum_{R=2}^{19} \lambda_R v_R v_R^T + \sum_{R=20}^{N} \lambda_R v_R v_R^T \\
&= C_{Market} + C_{Group} + C_{Noise}
\end{aligned} \tag{9}
$$

It is important to understand why the largest eigenvalue and the corresponding eigenvector are referred to as representing the market mode. The market index on day t is denoted as $w_M(t)$ and defines it by the scalar product of $w(t)$ and the first eigenvector v_1:

$$
w_M(t) = w(t) \cdot v_1 \tag{10}
$$

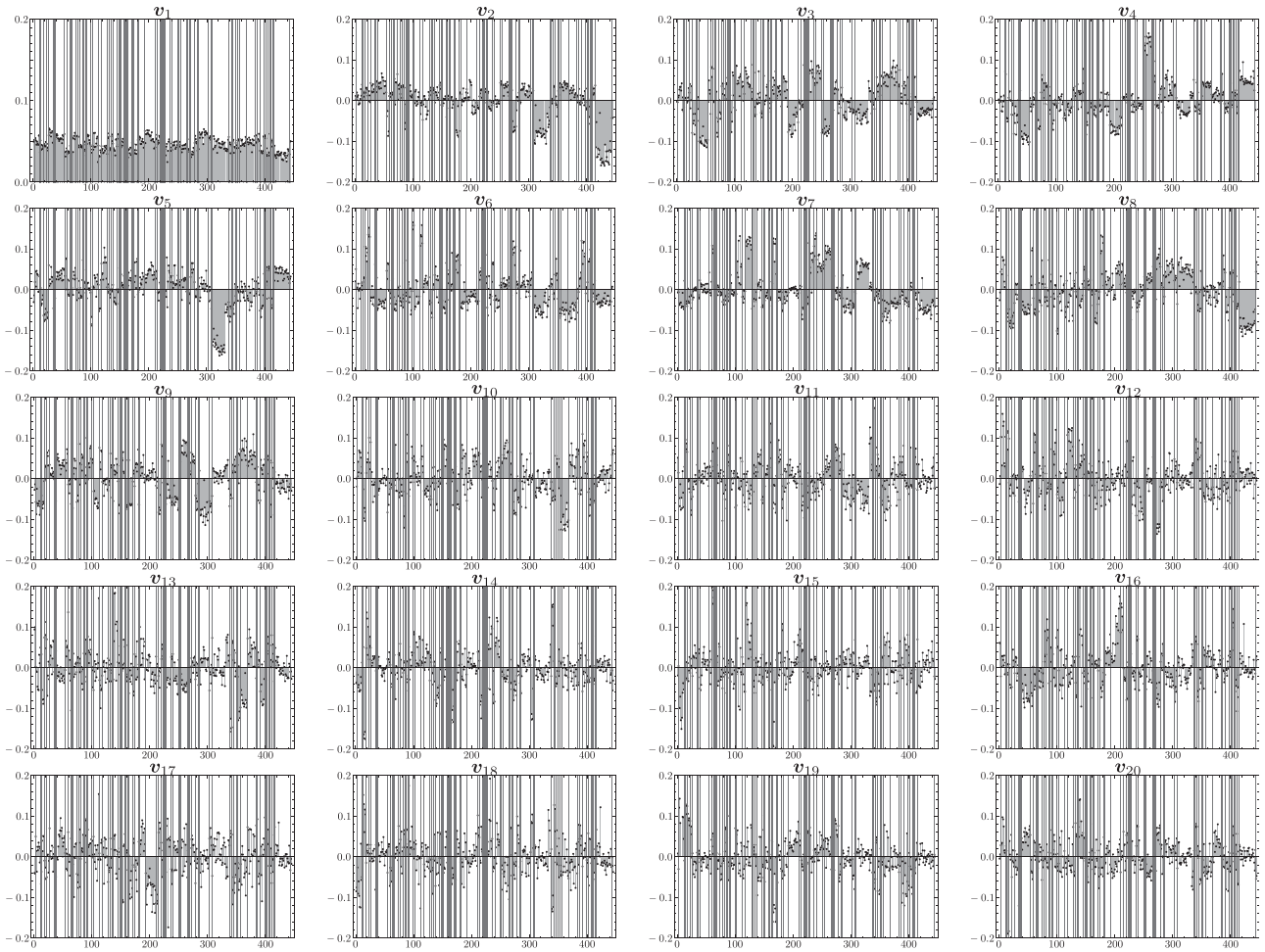


FIGURE 5 | Distribution of components of the top 20 eigenvectors, $v_1, \ldots, v_{20}$. The abscissa represents n, and the ordinate represents to the components $v_{R,n}$ of eigenvector v_R. Here, R is the eigenvalue rankings. The thin vertical lines in these figures separate business sectors.

i.e., weighting the average return with the weight given by the first eigenvector. On the other hand, the S&P 500 is used to characterize the entire market. The normalized log return on day t from open to close of the S&P 500 is denoted as $w_{SP}(t)$. **Figure 6** shows the scatter plot of $w_M(t)$ vs. $w_{SP}(t)$. This figure shows that $w_M(t)$ and $w_{SP}(t)$ exhibit a strong, positive correlation. The dashed line in this figure shows a linear function with the slope given by Pearson's correlation index $\rho = 0.852$ and with the intercept equal to 0. This correlation coefficient is almost the same as that obtained by [5].

3 APPLICATION OF CHPCA AND HHD

In this section, the complex correlation matrix is defined. RRT is then applied to distinguish the meaning components from the noise components, and CHPCA is introduced. After that, HHD is presented in order to clarify the lead-lag relationships among assets.

3.1 Complex Correlation Matrix

A simple definition of different-time correlation is given by $Corr[w_m(t), w_n(t+\Delta t)]$, $(\Delta t = 1, \ldots, T-1)$. However, if N and T are extremely large, a huge number of combinations must be investigated. Therefore, a complex correlation matrix is introduced to overcome this problem.

We consider the Fourier transform of the daily log returns of asset n as represented by

$$r_n(t) = \sum_{k=0}^{T} [a_n(\omega_k)\cos(\omega_k t) + b_n(\omega_k)\sin(\omega_k t)] \tag{11}$$

where $\omega_k = 2\pi k/T \geq 0$. The Hilbert transform of $r_n(t)$ is given by

$$\widehat{r}_n(t) = \sum_{k=0}^{T} [b_n(\omega_k)\cos(\omega_k t) - a_n(\omega_k)\sin(\omega_k t)] \tag{12}$$

We define a complex log return $\tilde{r}_n(t)$ as

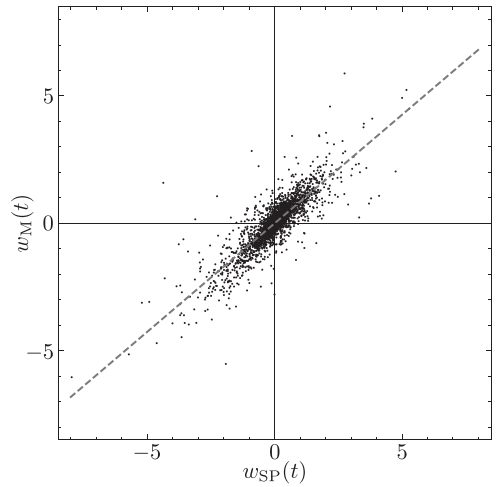


FIGURE 6 | Scatter plot of the normalized S&P 500 index $w_{SP}(t)$ and the normalized market index $w_M(t)$. The dashed line is linear function with slope ρ, which is Pearson's correlation index, equal to 0.852, and intercept equal to 0.

$$\tilde{r}_n(t) = r_n(t) + i\hat{r}_n(t) = \sum_{k=0}^{T} c_n(\omega_k)e^{-i\omega_k t} \tag{13}$$

where i denotes an imaginary unit defined by $i^2 = -1$. For each asset n, we define a time average $\langle\tilde{r}_n\rangle$ and a standard deviation $\tilde{\sigma}_n$ as follows.

$$\langle\tilde{r}_n\rangle = \frac{1}{T}\sum_{t}^{T}\tilde{r}_n(t), \qquad \tilde{\sigma}_n = \sqrt{\frac{1}{T}\sum_{t=1}^{T}|\tilde{r}_n(t) - \langle\tilde{r}_n\rangle|^2} \tag{14}$$

We define the normalized complex log return $\tilde{w}_n(t)$ as

$$\tilde{w}_n(t) = \frac{\tilde{r}_n(t) - \langle\tilde{r}_n\rangle}{\tilde{\sigma}_n} \tag{15}$$

Thus, the time-average of $\tilde{w}_n(t)$ is zero, and its standard deviation is one. Each component of the complex correlation matrix is defined by

$$\tilde{C}_{mn} = \frac{1}{T}\sum_{t=1}^{T}\tilde{w}_m(t)\tilde{w}_n^{\dagger}(t) \tag{16}$$

Herein, † represents the transposed complex conjugate.

The elements of the complex correlation matrix distribute on the complex plane, as shown in the upper left panel of **Figure 7**. The lower left panel of **Figure 7** shows the distribution of the real parts of the elements of the complex correlation matrix. This distribution is almost the same as for the case of the equal-time cross-correlation matrix shown in the right panel of **Figure 1**. The upper right panel of **Figure 7** shows the distribution of the imaginary parts of the elements of the complex correlation matrix. This panel shows a symmetrical distribution.

3.2 Complex Hilbert Principal Component Analysis

Figure 8 is obtained by calculating the eigenvalues λ_R for the cross-correlation matrix. As in **Section 2.2**, here the subscript R again represents the eigenvalue rankings. The left panel of **Figure 8** shows the distribution of the logarithms of eigenvalues. The largest eigenvalue is $\lambda_1 = 143.71$, and the smallest eigenvalue is $\lambda_{445} = 0.0442842$. The right panel of **Figure 8** shows the distribution in the small eigenvalue region. The solid line is the Marčenko-Pastur distribution given by **Eq. 5** with $Q = 2N/T$.

Figure 9 shows the scree graph. In this figure, the abscissa corresponds to the eigenvalue rankings and the ordinate corresponds to eigenvalue magnitudes. The curve with error bars in this figure shows the eigenvalue distribution of the RRS complex correlation matrix. The thin line with filled circles in this figure depicts the distribution of eigenvalues of the empirical complex cross-correlation matrix. If we again denote the upper bound for eigenvalues derived from the RRS cross-correlation matrix as $\lambda_{\max}$ we again obtain $\lambda_1 > \lambda_2 > \ldots > \lambda_{16} > \lambda_{\max} = 2.18894$. Hence, 16 meaningful components are retained for this dataset.

Figure 10 shows the distribution of each component for the top 16 eigenvectors $v_1, \ldots, v_{16}$ in the complex plane. In this case, the Poter-Thomas distribution, which is the null hypothesis of randomness, is given by

$$p(v) = \frac{N}{\pi}\exp\left(-N|v|^2\right) \tag{17}$$

In the complex plane, we regard the clockwise direction from the positive real axis as corresponding to leading components, whereas the counterclockwise direction from the positive real axis corresponds to the lagging components. Components of the first eigenvector v_1 distribute along the positive real axis. This means that the phase difference, i.e., the difference between leading and lagging, is small for the first eigenvector. Thus, we refer to the first eigenmode as the market mode. On the other hand, components of the 2nd to 16th eigenvectors distribute over a wide region in the complex plane. This behavior suggests group structure.

3.3 Helmholtz-Hodge Decomposition

We decompose the complex correlation matrix into the meaningful part and the noise part as

$$\begin{aligned}\tilde{C} &= \sum_{R=1}^{N}\lambda_R v_R v_R^{\dagger} \\ &= \sum_{R=1}^{16}\lambda_R v_R v_R^{\dagger} + \sum_{R=17}^{N}\lambda_R v_R v_R^{\dagger} \\ &= \tilde{C}_{Principal} + \tilde{C}_{Noise}\end{aligned} \tag{18}$$

where † represents taking the complex conjugate of a vector. The left panel of **Figure 11** shows the meaningful part of the complex correlation matrix. The introduction of a lower bound for the magnitudes of elements of the principal part of the complex correlation matrix produces, the right panel of **Figure 11**. The components of the real matrix F are the absolute values of the components of this constrained meaningful correlation matrix. Here, F is considered the weighted adjacency matrix. The components of this matrix can then be written as

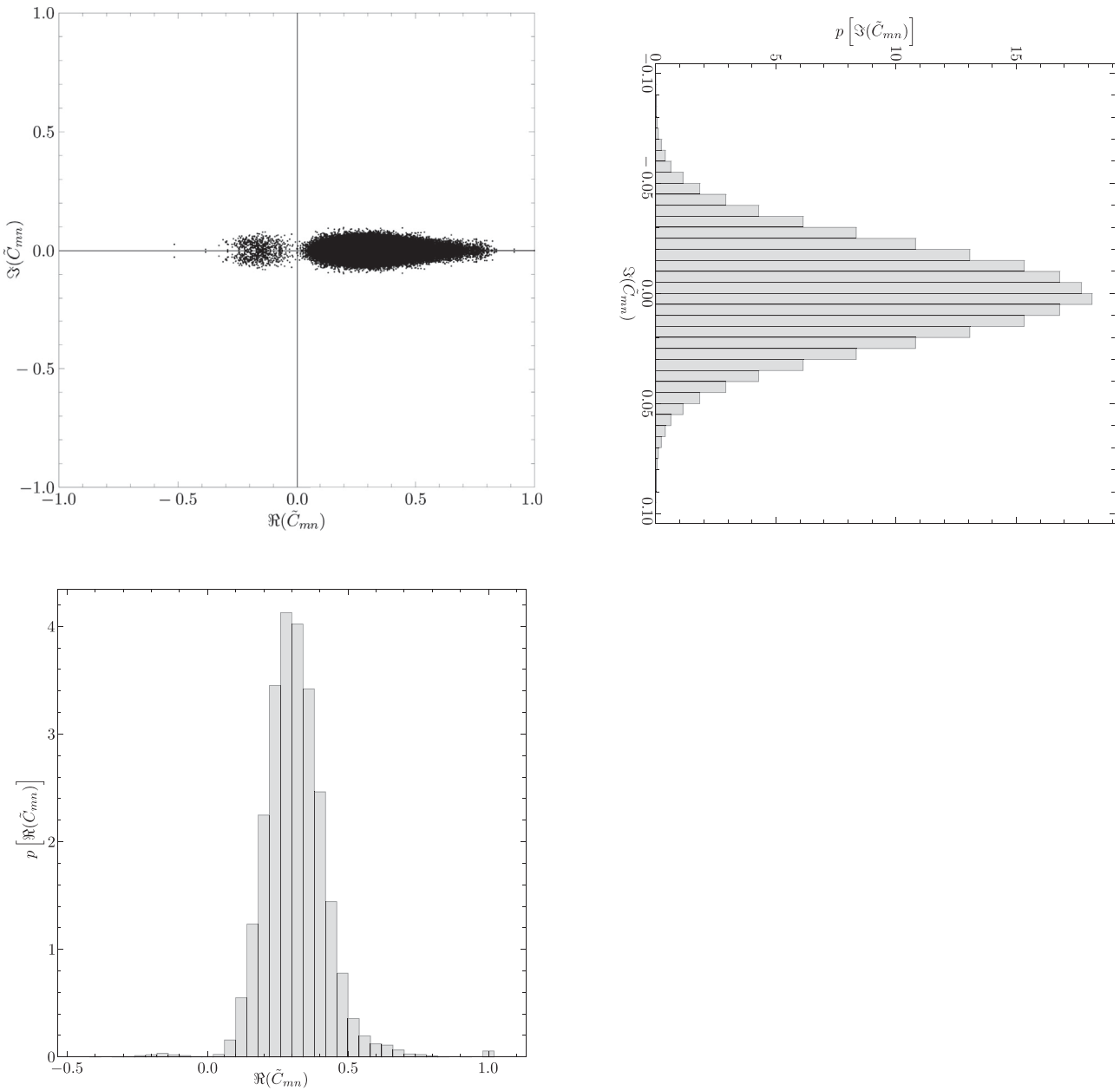


FIGURE 7 | (Upper left) Distribution of the elements $\tilde{C}_{mn}$ of the complex correlation matrix $\tilde{C}$ in the complex plane $[\mathfrak{R}(\tilde{C}_{mn}), \mathfrak{I}(\tilde{C}_{mn})]$. **(Upper right)** Distribution $p[\mathfrak{I}(\tilde{C}_{mn})]$ of the elements of the imaginal part $\mathfrak{I}(\tilde{C}_{mn})$ of the complex correlation matrix. **(Lower left)** Distribution $p[\mathfrak{R}(\tilde{C}_{mn})]$ of the elements of the real part $\mathfrak{R}(\tilde{C}_{mn})$ of the complex correlation matrix.

$$F_{mn} = F_{mn}^{(c)} + F_{mn}^{(p)} \tag{19}$$

where $F_{mn}^{(c)}$ corresponds to the circular flow in the network defined by

$$\sum_{n=1}^{N} F_{mn}^{(c)} = 0 \tag{20}$$

On the other hand, $F_{mn}^{(p)}$ corresponds to the gradient flow in the network defined by

$$F_{mn}^{(p)} = \gamma_{mn}(\phi_m - \phi_n) \tag{21}$$

Here, ϕ_m is the Helmholtz-Hodge potential. By using **Eqs 16**, **17** can be rewritten as

$$\sum_{n=1}^{N} [F_{mn} - \gamma_{mn}(\phi_m - \phi_n)] = 0 \tag{22}$$

By solving **Eq. 18**, we obtain the Helmholtz-Hodge potential shown in **Figure 12**. In this figure, the leading components show a

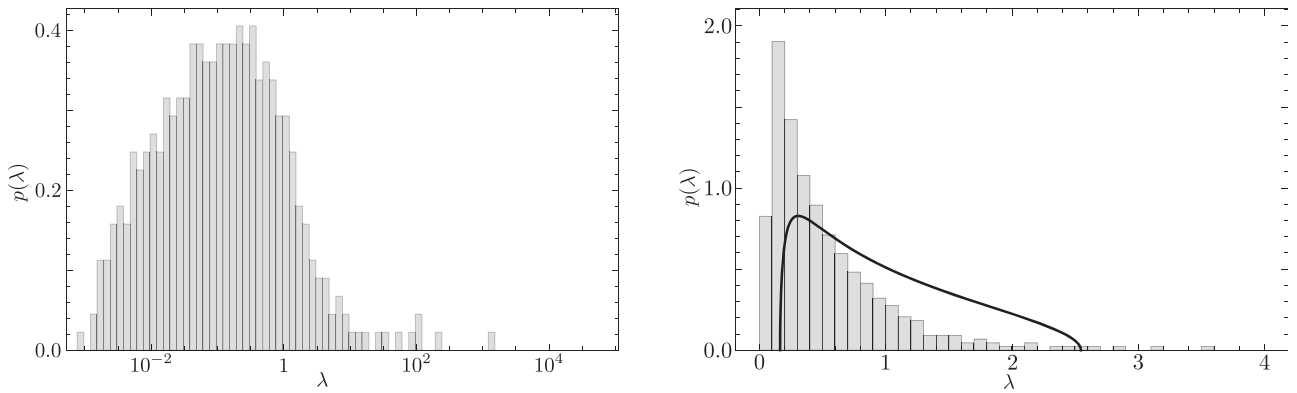


FIGURE 8 | (Left) Distribution $p(\lambda)$ of eigenvalues λ of the empirical equal-time cross-correlation matrix. **(Right)** Empirically obtained distribution $p(\lambda)$ of eigenvalues λ in the range of small eigenvalues. The solid line is the Marčenko-Pastur distribution derived under RMT as the theoretical curve given by **Eq. 5** with $Q = 2N/T$.

small value of the Helmholtz-Hodge potential, while the lagging components show a large value.

The average values $\langle\phi\rangle$ of the Helmholtz-Hodge potential for some major sectors are shown in **Table 1**. This table shows that the semiconductors industry is the most strongly leading, while the drug manufacturing industry is the most strongly lagging. On the other hand, [28] explored 483 assets from the S&P 500 for 4-years from 2008 to 2011 (1,009 business days). He obtained the result that the financial sector is the most strongly leading, while the telecommunications and service sector is the most strongly lagging. Therefore, we suspect that the lead-lag structure depends on the gross market conditions of the period investigated. However, clarifying this suspicion is a problem for future study.

4 APPLICATION OF CHPCA TO THE PORTFOLIO THEORY: A SKETCH

As a problem for future study, we consider the application of CHPCA to construct a portfolio by following Markowitz's portfolio theory [12]. We represent the fraction of wealth invested in asset n as ξ_n. If we denote the number of assets as K, ξ_n is normalized by

$$\sum_{n=1}^{K} \xi_n = 1 \tag{23}$$

By using the complex log return of each asset $\tilde{r}_n$ defined by **Eq. 9**, we define the complex log return of the portfolio $\tilde{r}_P$ as

$$\tilde{r}_P = \sum_{n=1}^{K} \xi_n \tilde{r}_n = \sum_{n=1}^{K} \xi_n r_n + i \sum_{n=1}^{K} \xi_n \hat{r}_n \tag{24}$$

However, the portfolio return must be a real number, so we need to impose the following constraint:

$$\sum_{n=1}^{K} \xi_n \hat{r}_n = 0 \tag{25}$$

The risk of the portfolio is defined by the variance:

$$\begin{aligned} \tilde{\sigma}_P^2 &= \sum_{m=1}^{K}\sum_{n=1}^{K} \xi_m \xi_n \tilde{C}_{mn} \tilde{\sigma}_m \tilde{\sigma}_n \\ &= \sum_{m=1}^{K}\sum_{n=1}^{K} \xi_m \xi_n \Re\left(\tilde{C}_{mn}\right) \tilde{\sigma}_m \tilde{\sigma}_n + \sum_{m=1}^{K}\sum_{n=1}^{K} \xi_m \xi_n \Im\left(\tilde{C}_{mn}\right) \tilde{\sigma}_m \tilde{\sigma}_n \end{aligned} \tag{26}$$

Here again, the risk must be a real number, so we need to impose the following constraint:

$$\sum_{m=1}^{K}\sum_{n=1}^{K} \xi_m \xi_n \Im\left(\tilde{C}_{mn}\right) \tilde{\sigma}_m \tilde{\sigma}_n = 0 \tag{27}$$

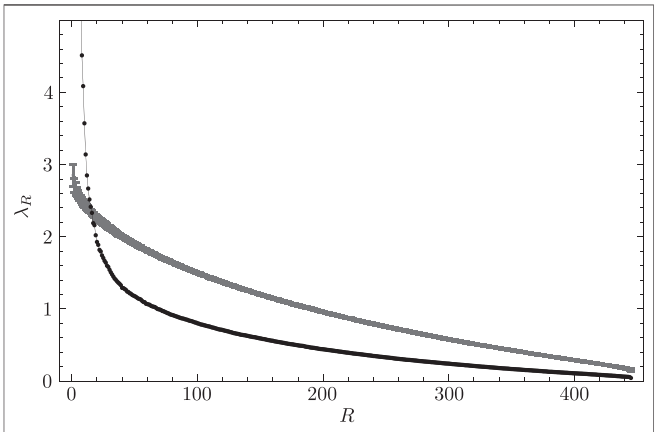


FIGURE 9 | Scree graph of eigenvalues. The abscissa represents eigenvalue rankings R, while the ordinate represents empirically obtained eigenvalues λ_R. The curve with error bars in this figure shows the eigenvalue distribution of the RRS complex correlation matrix. To obtain this curve, this manipulation was repeated 20 times, after which the mean value and standard deviation were calculated. Each error bar represents three times the standard deviation. The thin line with filled circles in this figure depicts the distribution of eigenvalues of the empirical equal-time cross-correlation matrix. The meaningful part can be obtained by comparing these two distributions. If the upper bound for eigenvalues derived from the RRS cross-correlation matrix is denoted as $\lambda_{\max}$, then $\lambda_1 > \lambda_2 > \ldots > \lambda_{16} > \lambda_{\max} = 2.18894$. Hence, 16 meaningful components should be retained for this data.

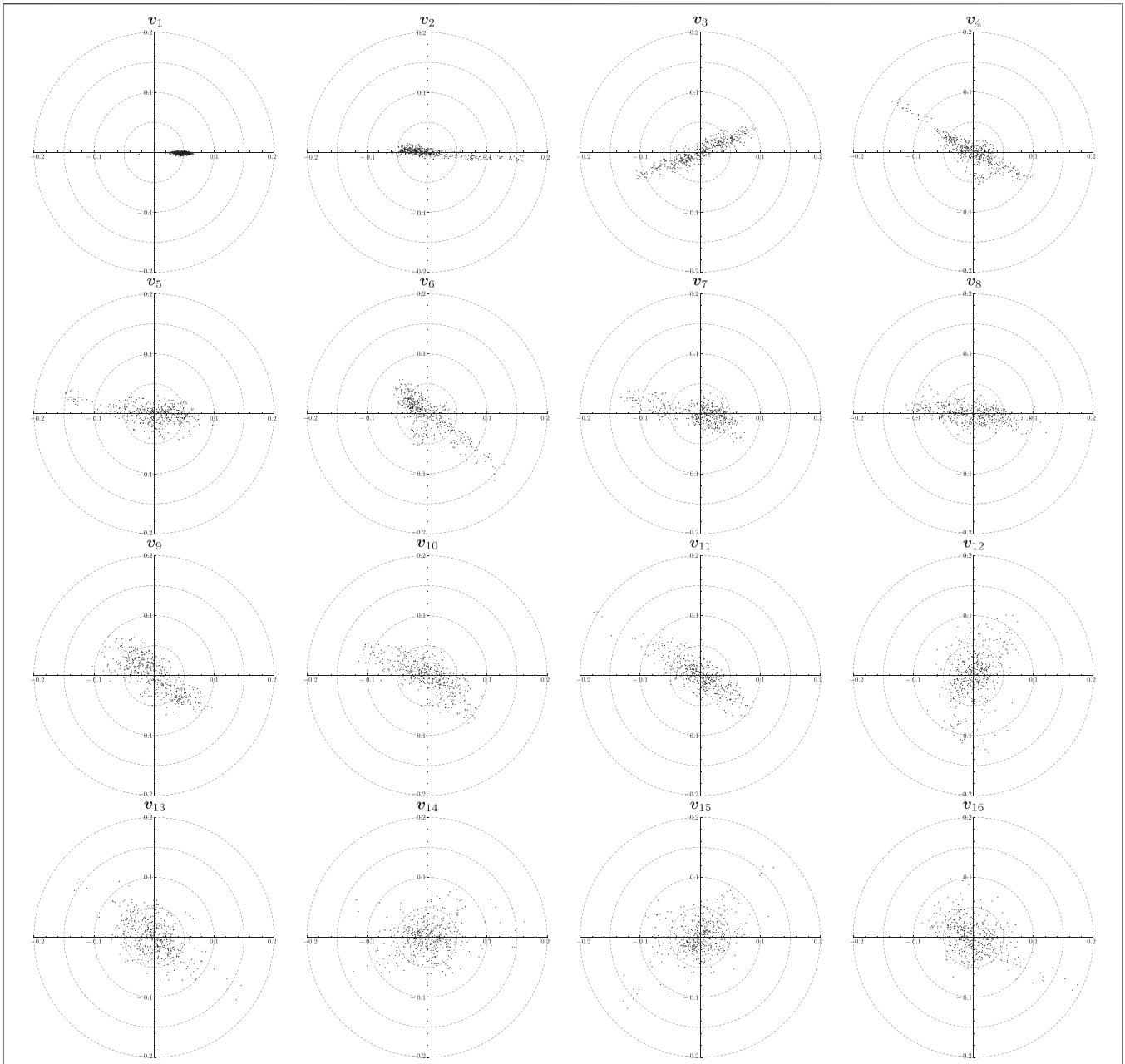


FIGURE 10 | Distribution of components of the top 16 eigenvectors, $v_1, \ldots, v_{16}$. In each figure, the abscissa represents $\Re(v_{R,n})$, and the ordinate represents $\Im(v_{R,n})$. Here, R is the eigenvalue rankings.

Therefore, under the conditions given in **Eqs 19**, **21**, **23**, a portfolio can be created that minimizes risk under the assumed returns.

5 CONCLUSION

An analysis of price data for 445 assets from the S&P 500 from 2010 to 2019 (2,510 business days) provided the basis for an exploration of recent developments in distinguishing the meaningful part from the noise part in correlation structures in big data. Application of RMT to the equal-time cross-correlation matrix was found to be a useful method for obtaining the meaningful components of the correlation structure. However, the null hypothesis of randomness underlying RMT destroyed both real autocorrelation and real cross-correlation in the data. In order to preserve autocorrelation, we introduce RRS. In

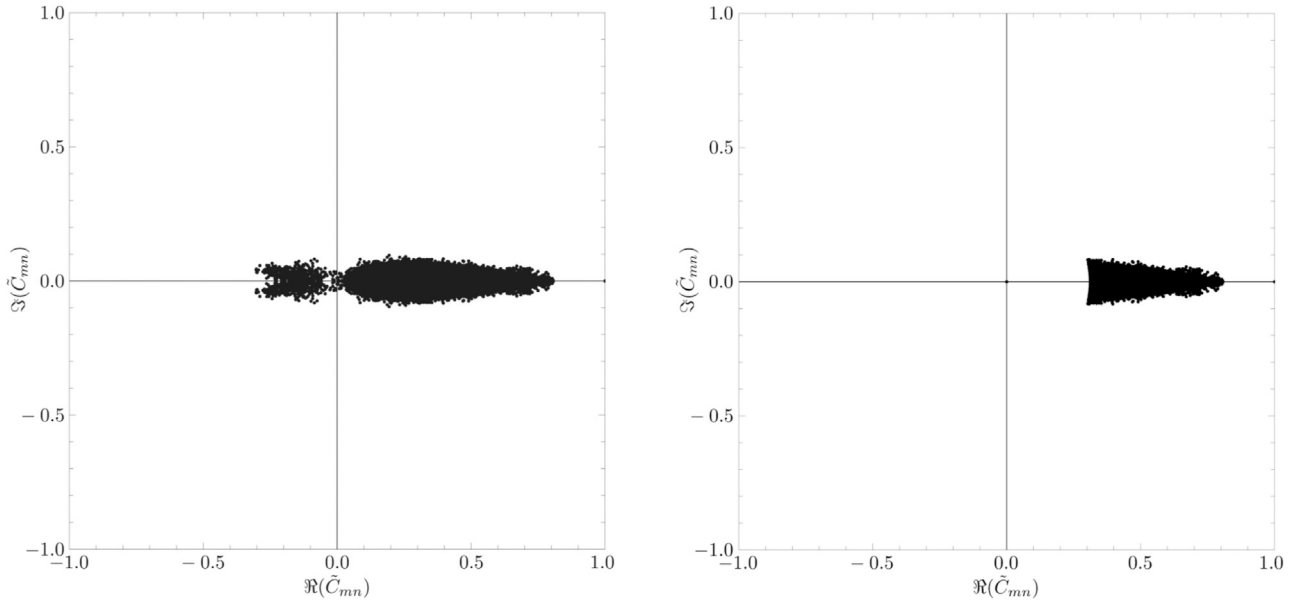


FIGURE 11 | (Left) Distribution of the components of the meaningful part of the complex correlation matrix. **(Right)** Distribution of the components of the constrained meaningful part of the complex correlation matrix.

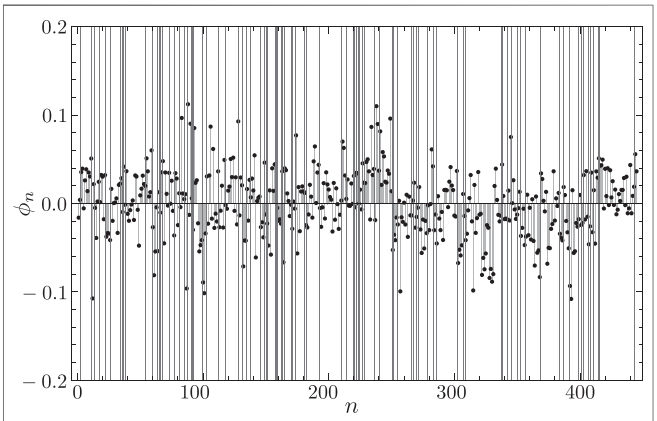


FIGURE 12 | Distribution of Helmholtz-Hodge potential for each asset. The abscissa represents n, while the ordinate represents Helmholtz-Hodge potential ϕ_n. The thin vertical lines in this figure separate business sectors.

TABLE 1 | Helmholtz-Hodge potentials $\langle\phi\rangle$ for some major business sectors.

Sector	# Assets	$\langle\phi\rangle$
Semiconductors	12	–0.03334
REIT	29	–0.02810
Software	15	–0.00789
Insurance	18	0.00042
Pharmaceutical retai	17	0.00106
Banks	16	0.00562
Diagnostics and research	10	0.01490
Utilities	28	0.01529
Information technology services	10	0.01661
Drug manufacture	11	0.02270

the case of this paper, the number of meaningful components for RMT and for RRS happened to be. We also introduced CHPCA for investigating the various different-time cross-correlations. By using both CHPCA and HHD, we clarified the lead-lag relationships for some major business sectors.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://en.wikipedia.org/wiki/List_of_S%26P_500_companies, https://github.com/datasets/s-and-p-500-companies.

AUTHOR CONTRIBUTIONS

WS wrote this paper by himself.

FUNDING

This work was supported by JSPS KAKENHI Grant No. JP20228860 and National Bank Academic Research Promotion Foundation in 2019.

ACKNOWLEDGMENTS

The author would like to thank Hiroshi Iyetomi, Hideaki Aoyama, Yoshi Fujiwara, Yuichi Ikeda, Hiroshi Yoshikawa, and Irena Vodenska for useful discussions.

REFERENCES

1. Laloux L, Cizeau P, Bouchaud J-P, Potters M. Noise dressing of financial correlation matrices. *Phys Rev Lett* (1999) 83:1467–70. doi:10.1103/physrevlett.83.1467
2. Plerou V, Gopikrishnan P, Rosenow B, Nunes Amaral LA, Stanley HE. Universal and nonuniversal properties of cross correlations in financial time series. *Phys Rev Lett* (1999) 83:1471–4. doi:10.1103/physrevlett.83.1471
3. Marčenko VA, Pastur LA. Distribution of eigenvalues for some sets of random matrices. *Mathematics USSR-Sbornik* (1967) 1:457–83. 10.1070.SM1967v001n04ABEH001994
4. Porter CE, Thomas RG. Fluctuations of nuclear reaction widths. *Phys Rev* (1956) 104:483–91. doi:10.1103/physrev.104.483
5. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, Stanley HE. Random matrix approach to cross correlations in financial data. *Phys Rev E, Stat Nonlinear, Soft Matter Phys* (2002) 65:066126. doi:10.1103/physreve.65.066126
6. Utsugi A, Ino K, Oshikawa M. Random matrix theory analysis of cross correlations in financial markets. *Phys Rev E, Stat Nonlinear, Soft Matter Phys* (2004) 70:026110. doi:10.1103/physreve.70.026110
7. Kim D-H, Jeong H. Systematic analysis of group identification in stock markets. *Phys Rev E Stat Nonlinear Soft Matter Phys* (2005) 72:046133. doi:10.1103/physreve.72.046133
8. Pan RK, Sinha S. Collective behavior of stock price movements in an emerging market. *Phys Rev E Stat Nonlinear Soft Matter Phys* (2007) 76:046116. doi:10.1103/physreve.76.046116
9. Namaki A, Jafari GR, Raei R. Comparing the structure of an emerging market with a mature one under global perturbation. *Physica A: Stat Mech its Appl* (2011) 390:3020–5. doi:10.1016/j.physa.2011.04.004
10. Namaki A, Shirazi AH, Raei R, Jafari GR. Network analysis of a financial market based on genuine correlation and threshold method. *Physica A: Stat Mech its Appl* (2011) 390:3835–41. doi:10.1016/j.physa.2011.06.033
11. Jamali T, Jafari GR. Spectra of empirical autocorrelation matrices: a random-matrix-theory-inspired perspective. *EPL* (2015) 111:10001. doi:10.1209/0295-5075/111/10001
12. Markowitz H. Portfolio selection*. *J Finance* (1952) 7:77–91. doi:10.1111/j.1540-6261.1952.tb01525.x
13. Fujiwara Y, Souma W, Murasato H, Yoon H. Application of PCA and random matrix theory to passive fund management. in: H Takayasua, editors. *Practical fruits of econophysics*. Berlin, Germany: Springer (2006). p. 226–30.
14. Souma W. Toward a practical application of econophysics: an approach from random matrix theory (written in Japanese). *Appl Math* (2005) 15:45–59. doi:10.11540/bjsiam.15.3239
15. Lo AW, Craig MacKinlay A. An econometric analysis of nonsynchronous trading. *J Econom* (1990) 45:181–211. doi:10.1016/0304-4076(90)90098-e
16. Iyetomi H, Nakayama Y, Aoyama H, Fujiwara Y, Ikeda Y, Souma W. Fluctuation-dissipation theory of input-output interindustrial relations. *Phys Rev E Stat Nonlinear Soft Matter Phys* (2011a) 83:016103. doi:10.1103/physreve.83.016103
17. Iyetomi H, Nakayama Y, Yoshikawa H, Aoyama H, Fujiwara Y, Ikeda Y, et al. What causes business cycles? analysis of the Japanese industrial production data. *J Jpn Int Economies* (2011) 25:246–72. doi:10.1016/j.jjie.2011.06.002
18. Arai Y, Yoshikawa T, Iyetomi H. Complex principal component analysis of dynamic correlations in financial markets. *Front Artif Intelligence Appl* (2013) 255:111–9. doi:10.3233/978-1-61499-264-6-111
19. Arai Y, Yoshikawa T, Iyetomi H. Dynamic stock correlation network. *Proced Comp Sci* (2015) 60:1826–35. doi:10.1016/j.procs.2015.08.293
20. Vodenska I, Aoyama H, Fujiwara Y, Iyetomi H, Arai Y. Interdependencies and causalities in coupled financial networks. *PLoS One* (2016) 11:e0150994. doi:10.1371/journal.pone.0150994
21. Souma W, Aoyama H, Iyetomi H, Fujiwara Y, Vodenska I. Construction and application of new analytical methods for stock correlations: toward the construction of prediction model of the financial crisis (written in Japanese). *JWEIN* (2016) 1–8.
22. Souma W, Iyetomi H, Yoshikawa H. Application of complex Hilbert principal component analysis to financial data. in IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); 2017 July 4–8; Turin, Italy. New York, NY: IEEE (2017) 2:391–4.
23. Souma W, Iyetomi H, Yoshikawa H. The leading and lagging structure of early warning indicators for detecting financial crises (written in Japanese). *RIETI Pol Discussion Paper Ser 18-P-005* (2018). p. 1–26.
24. Kichikawa Y, Iyetomi H, Iino T, Inoue H. Hierarchical and circulating flow structure in an interfirm transaction network. *Book of abstracts* (2017) 12. Available from: https://core.ac.uk/download/pdf/148338502.pdf#page=27.
25. Iyetomi H, Ikeda Y, Mizuno T, Ohnishi T, Watanabe T. International trade relationship from a multilateral. *Book of abstracts* (2017) 253. Available from: https://core.ac.uk/download/pdf/148338502.pdf#page=27.
26. Kichikawa Y, Iyetomi H, Iino T, Inoue H. Community structure based on circular flow in a large-scale transaction network. *Appl Netw Sci* (2019) 4:92. doi:10.1007/s41109-019-0202-8
27. Iyetomi H, Aoyama H, Fujiwara Y, Souma W, Vodenska I, Yoshikawa H. Relationship between macroeconomic indicators and economic cycles in United States. *Scientific Rep* (2020) 10:1–12. doi:10.1038/s41598-020-70100-3
28. Iyetomi H. Collective phenomena in economic systems. in: H Aoyama, Y Aruka, H Yoshikawa, editors *Complexity, heterogeneity, and the methods of statistical Physics in economics*. Berlin, Germany: Springer (2020). p. 177–201.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Souma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Physics

ORIGINAL RESEARCH
published: 23 April 2021
doi: 10.3389/fphy.2021.525162

Regional Economic Convergence in China: A Comparative Study of Nighttime Light and GDP

Qiao-Li Xiao[1], *Yue Wang*[1,2] *and Wei-Xing Zhou*[1,3]*

[1]School of Business, East China University of Science and Technology, Shanghai, China, [2]Research Institute of Sports Economics, East China University of Science and Technology, Shanghai, China, [3]Department of Mathematics, East China University of Science and Technology, Shanghai, China

Nighttime lights (NTLs) have been used as a proxy for economic growth in recent years. To verify the effectiveness of NTL in measuring regional economies, this article studies the regional economic convergence phenomenon in China's provinces by a comparative analysis of NTL data and GDP data from 1992 to 2013. It is found that there is a significant difference between the results of club convergence between NTL and GDP; GDP high-growth clubs are mainly concentrated in the east and central areas, while NTL's high-growth clubs are mostly concentrated in the central and west areas. Besides, the growth rate gaps between GDP clubs are relatively flat, while the growth rate gaps between NTL clubs are large. From the perspective of influencing, factors of the regional convergence, technological innovation, and industrial structure have a significant impact on GDP and NTL, and industrial structure has opposite effects on GDP clubs and NTL clubs. Besides the above factors, for NTL convergence clubs, population growth rate, economic openness, and resource consumption are also significant.

Keywords: nighttime lights, GDP, regional economies, club convergence, log t test

OPEN ACCESS

Edited by:
Valerio Capraro,
Middlesex University, United Kingdom

Reviewed by:
Gholamreza Jafari,
Shahid Beheshti University, Iran
Xu Tian,
Nanjing Agricultural University, China

***Correspondence:**
Wei-Xing Zhou
wxzhou@ecust.edu.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *08 January 2020*
Accepted: *04 February 2021*
Published: *23 April 2021*

Citation:
Xiao Q-L, Wang Y and Zhou W-X (2021) Regional Economic Convergence in China: A Comparative Study of Nighttime Light and GDP. Front. Phys. 9:525162. doi: 10.3389/fphy.2021.525162

1 INTRODUCTION

1.1 Outlining the Problem

The nighttime lights (this study is referred to as NTLs) data released by the National Oceanic and Atmospheric Administration (NOAA) show that NTL is a unique surface landscape phenomenon; it is representing the laws of human activities on the surface of the Earth and closely related to the socio-economic development of human beings. The advancement of technology and the development of NTL data resources have led to the more widespread use of NTL data for characterizing levels of economic growth and exploring the temporal and spatial characteristics and heterogeneity of economic development. This has become a trend in several related fields, and China is one of the most commonly researched regions in the literature on applying NTL to economic and social issues [1]. On the one hand, China's high-speed economic growth over the past 40 years has been met with great interest by global economists; on the other hand, although China's economy has maintained rapid growth as a whole, large gaps still exist in regional economic development. As China faces a period of critical strategic opportunity for development and potential social tensions, more attention has been given to balancing economic development between different regions. Also, China's official GDP statistics have received a question for statistical difficulties, structural incomparable, and human manipulation [2].

For the above reasons, the practice of measuring economic development with the easily obtainable and more objective NTL data has been widely adopted in China, especially in the investigation of regional economic development issues. However, these studies have not thoroughly investigated the reliability of using NTL as an indicator of regional economic gaps. Questions remain over the difference between NTL data and traditional GDP statistics in measuring regional economic disparities, the difference between the two in reflecting objective economic facts and laws, and what kind of GDP economic information can and cannot be characterized by NTL. At the same time, previous studies also ignore China's economy's characteristics in a transition period, including the balanced growth of multiple regional economies and the heterogeneity of economic convergence, both of which directly affect the establishment of research perspectives and selection of research methods.

Given the above problems, as well as in consideration of the regional heterogeneity of development in China and the time-varying convergence rate, this article takes China's provincial NTL data and GDP data as the research object, breaks the traditional eastern, central, and western division of economic zones, first using the $\log t$ test method based on the nonlinear time-varying factor model to find the convergence clubs of both types of data, analyze the club convergence characteristics of the two, and then use ordered logit model to explore the influencing factors of club membership. The two types of data are then compared and discussed in terms of club membership, regional distribution of members, differences in growth rate between clubs, and club membership factors. The two types of data are also differentiated according to their characteristics as economic indicators. It is hoped that these comparisons and discussions can answer the reliability of using NTL data to measure the regional economic disparities.

This work contributes to the literature in three aspects. First, the $\log t$ test method based on the nonlinear time-varying factor model is introduced into the regional economic convergence study of NTL and GDP, enriching the research perspective of spatial correlation analysis between the two kinds of data. Second, using the ordered logit model to explore the factors affecting convergence is a process that clarifies the similarities and differences between the NTL and GDP data in characterizing regional economic development. It also verifies the advantages and disadvantages of analysis of economic growth performed with NTL data. Third, this study will broaden the source channels of economic growth substitution variables, and is of great significance for exploring the characteristics and reliability of NTL data as economic indicators. It will also help to promote a better coordinated and higher quality development of China's different regions.

1.2 Literature Review

Since NTL has been shown to have a significantly high correlation with regional economic activities (GDP, GRP, etc.) and social activities (population, energy, etc.) [3–9] and NTL can be used as a proxy variable for social and economic activities. An increasing number of studies have begun to use this data to characterize the economy and study the spatial and temporal patterns of regional economies and industries [10–14], urban development [15–21], human activities and their effects [22–24], and energy consumption [22, 25]. Of special interest is the research conducted in Refs. [26–29]. These studies propose a model, based on the proof of the correlation between NTL and GDP that can use the NTL data to correct GDP. This model provides estimates of economic growth for countries whose statistical agencies lack reliability and carry great significance for studying the economics of NTL. The above research enriches the surrogate indicators of economic variables, expands the research methods available for analyzing economic and social problems, opens up research areas in other related issues, and illustrates the great potential of using NTL data in economics [30].

Before introducing NTL, many scholars used GDP as the primary source of economic research data despite some skepticism toward China's GDP. With the widespread use of NTL in economics research, many scholars have begun using NTL data to investigate China's economic and social problems, particularly in research related to the correlation between NTL and economic and social indicators [31–33], and in studies that use NTL data to estimate economic and social indicators [34–37]. In a study on regional differences in economic development, Liu et al. conducted an empirical investigation on the spatial differences and spillover effects of urban economic growth in China and found that the economic development of urban agglomerations shows significant spatial nonequilibrium characteristics [38]. Liu and Du use NTL data to re-examine the regional disparity and stochastic convergence characteristics of China's economic development and helping identify further the economic development of convergence clubs in China's various regions [39]. Based on the convergence analysis framework of the growth theory, Wang et al. utilize a dynamic panel data model to assess the dynamic trend of China's regional economic gap and find that between 1992 and 2012, China's initial nighttime light intensity, whether at the provincial level or the prefecture-city level, regions that initially have lower nighttime light brightness levels exhibit faster growth rates [40]. This level of growth deviates somewhat from per capita GDP trends. Ding and Zhou found that the high concentration of regional NTL on the southeast coast is consistently stable from 1992 to 2013 and is self-reinforcing. They also discovered that population growth, regional surface area, and distance from the coastline are all negatively correlated with NTL [41]. Each of these studies either directly investigates the overall convergence trend of NTL or uses NTL to characterize economic development and investigate spatial differences, convergence characteristics, and development trends of China's regional economies, and help to shed light on the nature and heterogeneity of China's regional economic growth model.

We believe that there are still some shortcomings in the perspectives and methods that use NTL to characterize regional economic development in research. From a research perspective, they ignore the similarities and differences between the economy represented by NTL and the economy as defined by GDP. Most studies are based on a positive correlation between

NTL and the economy, directly using NTL data as a substitute for GDP data, without giving proper attention to the information that is provided by GDP but absent in NTL data, and the new information provided by NTL data but absent in GDP data. Although a few scholars have mentioned this concern, none have carried out an in-depth investigation or argumentation on the topic. In terms of research methodology, the current mainstream convergence research methods are combing with statistical indicators such as the δ convergence index, using the conditional β convergence or club convergence methods to examine the regional convergence phenomenon, and to look for the influencing factors for the regional convergence. However, these methods ignore the fact that China is in a "transition period." When the economy must converge, and there is heterogeneity in the rate of convergence, if the income level and convergence rate of the underdeveloped regions are lower than the developed regions, there is a period in which developed regions grow faster than underdeveloped regions, but this does not affect their ultimate convergence. This period is called the "transition period." In 1986, Deng Xiaoping put forward the Reform and Opening Up strategy of "Let a part of the population get rich first, they will carry more people into wealth, and eventually everyone will enjoy common prosperity." Since then, China has been in the transitional period of economic transformation. Compared with the traditional tools for measuring and analyzing economic trends, the nonlinear time-varying factor model proposed by Phillips and Sul based on the convergence club division method can more accurately evaluate the transitional period of the Chinese economy [42]. This method came to be widely used in the economic convergence analysis worldwide after its proposal [43–49].

2 DATA AND METHOD

2.1 Data Sets

The most widely used NTL in the academic community is acquired by the Optical Imaging Linear Scanning Service sensor of the United States Digital Defense Meteorological Satellite Program (DMSP/OLS). The device effectively monitors radiation information such as city lights, fishing fires, and forest fires. It provides the world's most extended time series (1992–2013) of nighttime light-to-earth observation data provided by the NOAA's National Geophysical Data Center, including average visible lights, stable lights, and cloud-free coverage. In the stable lights images, unstable light sources such as aurora and wildfire and the interference of moonlight and clouds have been removed. The final data value is the annual average gray value of cloudless stable light, which provides a unique research perspective for human activity detection. Although the data has a low resolution (approximately 1 km) and lacks calibration and urban center saturation [50], the corrected DMSP/OLS data have become the most commonly used data source for NTL remote sensing research.

This article uses DMSP/OLS stable lights data from 34 provincial-level administrative regions in China from 1992 to 2013. It combines the NTL data with the GDP data to explore differences between the use of NTL and GDP in regional economic convergence research. The NTL data have been prepared and corrected by Cao et al. [51], based on well-established methods [25, 52, 53].

TABLE 1 | Descriptive statistics of variables.

Variables	N	Mean (%)	SD	Min (%)	Max (%)
GDP growth rate	714	11.258	3.62	-5.88	26.76
NTL mean growth rate	714	7.137	7.685	0.000	72.459

The average growth rates of NTL and GDP are then taken as research variables. The NTL growth rate of each region is calculated from the average value of the NTL from 1992 to 2013, so the growth rate data for a total of 21 years from 1993 to 2013. In order to remove the inflation effect, the real GDP growth rates from 1993 to 2013 are adopted in our analysis. The real GDP growth rates are retrieved from the National Bureau of Statistics and the Compilation of 60 Years of Statistics of the People's Republic of China. The descriptive statistics for each variable can be found in **Table 1**.

Between 1993 and 2013, both GDP and NTL growth rates experienced several fluctuations, and the magnitude and extent of the changes did not show significant consistency. The regions with high GDP growth rates are Inner Mongolia, Guangdong, and Tianjin. The high-growth areas with NTL are Tibet, Yunnan, and Xinjiang. Whether it is GDP or NTL, the regions with lower growth rates are concentrated in Macao, Hong Kong, and Taiwan.

2.2 Club Convergence Test

The convergence club division method proposed by Phillips and Sul is to judge the convergence of economic growth and price index changes in a transitional economy at a given level of significance and to identify and divide the convergence clubs 42. The advantage of using this method lies in that it can fully consider the heterogeneity of observed entities without the need for more stringent economic assumptions about model parameters or convergence clubs, enabling the analytic framework to be applied to a broader range of complete data that characterizes economic development.

The club convergence test method is for a nonlinear time-varying factor model:

$$y_i(t) = \delta_i(t)\mu(t), \tag{1}$$

where $y_i(t)$ is a time series, $\mu(t)$ is a single common component, and $\delta_i(t)$ is a time-varying parameter containing the perturbation term, expressing change in individual heterogeneity over time. First, the relative transfer coefficient is defined to eliminate the common factors between individuals. Only the heterogeneous time-varying features of the individual are retained, which is called the relative transition path:

$$h_i(t) = \frac{y_i(t)}{\frac{1}{N}\sum_{j=1}^{N} y_j(t)} = \frac{\delta_i(t)}{\frac{1}{N}\sum_{j=1}^{N} \delta_j(t)} \tag{2}$$

A semiparametric model is then constructed as follows:

$$\delta_i(t) = \delta_i + \frac{\sigma_i \xi_i(t)}{L(t)t^{\alpha}}, \tag{3}$$

where $L(t)$ is a slowlychanging function that increases and diverges at infinity and $\xi_i(t)$ is weakly dependent upon t but is *i.i.d.* $(0, 1)$ over i. This article uses $L(t) = \log(t)$ and through regression of the equation:

$$\log\left(\frac{H_1}{H_t}\right) - 2\log L(t) = \widehat{c} + \widehat{b}\log t + \widehat{\mu}_t, \tag{4}$$

which verifies the original hypothesis of convergence:

$$\begin{cases} \mathcal{H}_0 : \delta_i = \delta \quad \text{and} \quad \alpha \geq 0 \\ \mathcal{H}_1 : \delta_i \neq \delta \quad \text{or} \quad \alpha < 0 \end{cases} \tag{5}$$

In **Eq. 4**, $t = [rT], [rT] + 1, \ldots, T$, $0 < r < 1$, and T is the original length of the time series, where $[rT]$ represents the integer part of rT. During the regression process, a small portion of r of the time series needs to be discarded, which will help focus the test's attention on the later trend of the time series. In addition, $H_t = (1/N)\sum_{i=1}^{N}(h_{it} - 1)^2$ where h_{it} is the relative transition path in **Eq. 2** and $\widehat{b} = 2\widehat{\alpha}$ where $\widehat{\alpha}$ is the least squares estimate of α in $\mathcal{H}_0$. The one-sided t test is used to test $\widehat{\alpha} \geq 0$ in the original null hypothesis, specifically, at the 5% significance level, if the t statistic of $\widehat{b}$ is less than −1.65, then the original hypothesis of convergence is rejected.

The above test method is called the $\log t$ test, which is the essential condition for judging whether there is a convergence club. The specific convergence club division is determined by sorting the sample area, selecting the core group, and adding the group member. The $\log t$ test is used in both the core group selection, and the addition of group members, whereby identifying club convergence is completed.

2.3 The Ordered Logit Regression

In order to search out the various factors affecting club membership and evaluate their importance in determining club membership, this article employs the ordered logit model, based on the club convergence results, to identify the influencing factors [45]. This method can estimate how the explanatory variable variation changes the probability that a given region belongs to the affiliated club while fixing all other variables to its sample mean. In addition to the indicators of population growth rate, technological innovation covered by the neoclassical growth theory model, and human capital emphasized by the endogenous growth model, other factors that are mainly involved in the convergence of regional economic growth include the degree of opening up, marketization, employment rate, geographical factors, policy factors, and industries structures [54–60]. Considering the characteristics of China's economic development and the results of club convergence, combined with factors that may affect economic activity and NTL, this article examines the influencing factors of club formation from six dimensions: capital investment, technological innovation, industrial structure, market vitality, economic openness, and resource consumption. Capital investment is divided into physical capital and human capital, where the physical capital is measured by fixed capital investment and the human capital is divided into the number of employees and population growth rate. The technological innovation is measured by the number of patents granted. The industrial structure is divided into industrial proportion and service industry proportion, which are respectively measured by the secondary industry output and the tertiary industry output for GDP proportion. Market vitality is measured by the total retail sales of social consumer goods. Economic openness is measured by the amount of foreign direct investment. Resource consumption is measured by the consumption of coal. In order to eliminate inflation, all nominal data are deflated based on 1993. We use the fixed asset investment price index to deflate fixed capital investment, the consumer price index to deflate consumer goods' total retail sales, and the U.S. GDP deflation index to deflate foreign direct investment.

Considering that the club is divided according to the growth rate, all variables with the exceptions of population growth rate, industrial proportion, and service industry proportion use the increment data.

The data comes from the China Statistical Yearbook, the regional statistical yearbooks, and the China Energy Statistics Yearbook. Due to the lack of relevant data in Hong Kong, Macao, and Taiwan, convergence factor testing excludes these three regions, and only uses relevant data from the mainland's 31 areas for the impact factor analysis.

3 RESULTS

3.1 Club Convergence in GDP Growth Rates

According to Phillips and Sul's recommendation 42, this article takes a value of $r = 0.3$ for the $\log t$ test of the GDP growth rate of China's provinces. First, the estimation equation for the $\log t$ test for China's 34 regions is as follows:

$$\log\left(\frac{H_1}{H_t}\right) - 2\log[\log(t)] = \underset{(-11.81)}{-1.464} + \underset{(2.904)}{0.14}\log t \tag{6}$$

It can be seen that $\widehat{b} = 0.14 \pm 0.048 > 0$ and $t_k = 2.904 > -1.65$. The regression result accepts the null hypothesis of convergence, which shows that the GDP growth rates of Chinese provinces reflect a nationwide convergence trend.

Next, clubs are divided among Chinese provinces to identify economic convergence in the country. In accordance with the recommendations of Phillips and Sul, during the initial sorting process, the last 1/2 data of the entire time span of the observation values were selected 42, the top-ranking region (Inner Mongolia) was used as the reference region, and Tianjin, Chongqing, were then added to perform $\log t$ regression and calculate the t statistic until the t statistic is less than −1.65. We found that for $k = \{1, 2\}$ or $\{1, 2, 3\}$ and $t_k = -1.292$ or -6.479; when Chongqing was added, $t_k < -1.65$. At this point the adding of regions was paused. For the group $k = \{1, 2\}$, the value of t_k was the largest, so the core group was confirmed as Inner Mongolia and Tianjin. Then, one region was added at a time to the core

TABLE 2 | GDP convergence clubs.

Club	Club members	Club size	Core group t_k	Whole group t_k	$\hat{b}$	$\hat{\alpha}$
1	IM, TJ, MO, GD	4	-1.292	10.015	2.007 ± 1.073	1.004
2	CQ, SN, JS, SD, QH, GZ, FJ, HB, AH, XZ, HA, ZJ, HE, SH	14	6.226	7.622	1.682 ± 0.977	0.841
3	JL, SC, GX, JX, HN, LN, SX, NX, HI, GS, YN, HL, XJ, BJ	14	14.523	7.355	0.608 ± 0.039	0.304
4	HK, TW	2	4.866	4.866	2.206 ± 0.453	1.103

TABLE 3 | GDP club identification process.

Qrder	t_k	t_k	Club	t_k	t_k	Club	t_k	t_k	Club
IM	SR	CR	Club 1				HK	SR	Club 4
TJ	–1.292	CR	Club 1				TW	4.866	Club 4
CQ	–6.479	–11.705		SR	CR	Club 2			
SN	–13.226	–13.731		6.226	CR	Club 2			
MO	11.303	9.360	Club 1						
GD	1.192	2.431	Club 1						
JL	–15.034	–23.201		0.688	–0.132		SR	CR	Club 3
SC	–13.139	–18.208		–2.614	–2.540		–0.705	CR	Club 3
JS	–0.713	–0.720		3.504	4.121	Club 2			
SD	–1.020	–1.138		2.396	2.940	Club 2			
GX	–5.713	–20.832		–3.971	–1.298		1.609	CR	Club 3
JX	–17.750	–30.360		–1.111	–3.415		–13.649	6.634	Club 3
QH	–6.653	–1.396		17.132	22.288	Club 2			
GZ	–3.474	–0.268		14.477	10.889	Club 2			
FJ	–0.570	–0.468		2.945	3.595	Club 2			
HN	–14.141	–27.632		–13.637	–4.417		0.183	3.431	Club 3
HB	–3.328	–5.921		0.410	1.221	Club 2			
LN	–18.349	–36.889		–4.169	–6.121		–42.562	5.230	Club 3
AH	–1.276	–1.180		2.182	2.715	Club 2			
XZ	–0.820	–1.083		3.026	3.467	Club 2			
HA	–4.123	–5.715		0.019	0.537	Club 2			
SX	–11.457	–32.357		–2.774	–0.950		1.800	2.657	Club 3
NX	–22.311	–41.493		0.638	–1.893		0.640	9.108	Club 3
HI	–14.986	–21.170		–29.715	–82.210		–14.108	3.438	Club 3
GS	–19.418	–42.349		–16.118	–10.310		–6.469	0.251	Club 3
YN	–14.062	–17.920		–82.567	–31.452		–12.285	1.925	Club 3
ZJ	–0.431	–0.357		2.132	2.525	Club 2			
HE	–2.916	–2.953		0.213	0.687	Club 2			
HL	–28.266	–65.982		–1.632	–5.331		–5.544	6.954	Club 3
XJ	–19.100	–27.277		–18.248	–49.762		–36.256	0.067	Club 3
BJ	–6.982	–10.648		–1.818	–0.863		0.675	0.943	Club 3
SH	–2.642	–2.644		0.058	0.547	Club 2			
HK	–47.154	–60.820		–108.934	–99.370		–84.7323	–114.307	
TW	–78.993	–159.575		–86.974	–286.319		–962.4422	–83.117	

In view of the limitation of the width of the table, we put the logt test process after Club 3 on the upper right of the table. SR refers to standard region.CR refers to core region.

group, with 50% as the significance level, 0 was used as the critical value to check the size of the t statistic; when $t_k > 0$, it was classified as a club member. It was ultimately determined that the Club 1 members are Inner Mongolia, Tianjin, Macao, and Guangdong.

After excluding Inner Mongolia, Tianjin, Macao, and Guangdong, the $\log t$ test was continued as described above. The members of the second convergence club (Chongqing, Shaanxi, Jiangsu, Shandong, Qinghai, Guizhou, Fujian, Hubei, Anhui, Tibet, Henan, Zhejiang, Hebei, and Shanghai), the third convergence club (Jilin, Sichuan, Guangxi, Jiangxi, Hunan, Liaoning, Shanxi, Ningxia, Hainan, Gansu, Yunnan, Heilongjiang, Xinjiang, and Beijing) and the fourth convergence club (Hong Kong and Taiwan) were identified. See **Tables 2,3** for details. All clubs have a whole group t_k greater than 2, indicating that these clubs are stronger clubs. The speed of convergence $\hat{\alpha}$ is not the same among the four clubs. Club 4 has the fastest convergence rate of 110.3%, Club 1's convergence speed is 100.4%, and the Club 2 and 3's convergence speed is 84.1% and 30.4%, respectively.

The trend of each region's transition path is shown in the top panel of **Figure 1**. The club members have prominent convergence characteristics. In the first few years, Club 1 members' transition paths are quite different, but in the subsequent years, it shows a characteristic of convergence. Club 2 members' transition paths gradually narrowed in 2004

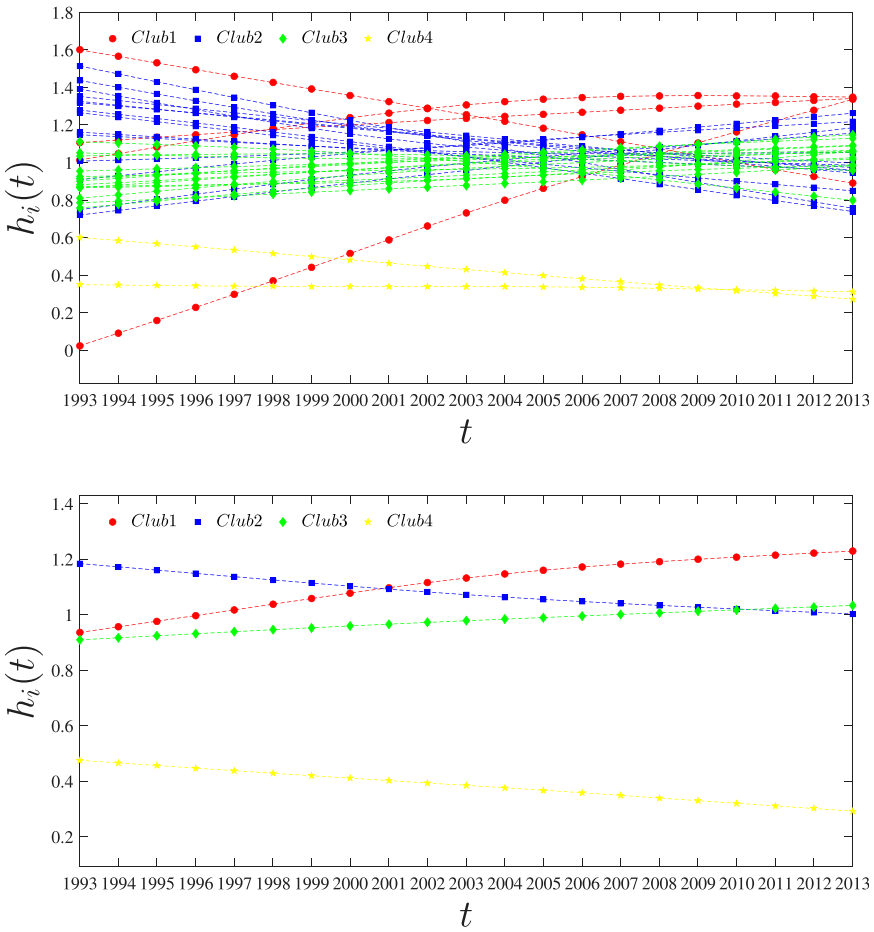


FIGURE 1 | Top: transition path of GDP growth rate. **Bottom**: GDP convergence club average transition path.

and 2005, and then the gap gradually widened, but it still converged to one club. The transition paths of Club 3 members are relatively consistent. The transition paths of Hong Kong and Taiwan included in Club 4 show a significant trend of convergence. The initial level of their transition paths is low, and the level of transition paths in 2013 is also low, which is significantly lower than the previous three clubs. It is worth noting that Macao, which has the lowest initial level, surpassed the transition paths of most regions and became a member of Club 1, reflecting the rapid growth of Macao's GDP.

At the same time, the club convergence results also showed obvious geographic effects. Both Club 1 and Club 2 are mainly concentrated in central and eastern China, except Tibet and Qinghai. While Club 3 is mostly concentrated in the western and northeast areas of China, the convergence result of the log$t$ test captures the spatial effects of GDP growth.

Each club's transition paths were averaged, and the transition trend of each club's overall transition path is shown in the bottom panel of **Figure 1**. It can be seen from the figure that in the initial years of the time series, the transition paths of Club 1 and Club 3 is relatively concentrated, but then the gap gradually widened. Club 2's transition path was the highest at first, but eventually lags behind Club 1, especially in 2001; it was a key point in reversing the transition paths of Club 1 and 2. Overall, clubs 1, 3, and 4's transition paths show an upward trend, while the transition path of Club 2 shows a downward trend. That is to say, the GDP transition paths of the northeast and part of the central and western regions have an upward trend. Although the eastern region took the lead at first, it has subsequently shown a downward trend. One reason for the existence of this catching-up effect may be the "Western Development" and the "Rise of the Central China" strategies introduced by the Chinese government in the early 20th century; as investment in the central and western regions continued to increase, the economic growth rate also continued to grow and even surpassed that of the eastern area.

The transition path of Club 4 has continued to decline since 1992. It also markedly differs from the other three clubs and is the club with the lowest growth rate; this is due to the social system of Hong Kong and Taiwan differing from the mainland and can be regarded as exceptional cases.

TABLE 4 | Nighttime light convergence clubs.

Club	Club members	Club size	Core group t_k	Whole group t_k	$\hat{b}$	$\hat{a}$
1	XZ, JX, CQ, QH, IM, SC, GZ, GS, NX, SN, HN, YN, AN, XJ	14	3.382	3.349	0.308 ± 0.375	0.154
2	ZJ, JS, GX, HI, HL, HB, JL, FJ, HA,SD, LN, TJ	12	2.372	−1.0751	−0.073 ± 0.167	−0.037
3	SX, GD, SH	3	4.73	3.355	0.303 ± 0.686	0.152
4	HK, MO	2	−1.057	−1.057	−1.44 ± 1.36	−0.72
Nonconvergent	HE, BJ, TW	3				

TABLE 5 | NTL club identification process.

Qrder	t_k	t_k	Club	t_k	t_k	Club	t_k	t_k	t_k	Club
XZ	SR	CR	1							
JX	2.958	CR	1							
CQ	3.382	CR	1			BJ	SR	NC		
QH	0.321	2.871	1			TW	−7.442	SR	NC	
IM	0.531	2.623	1			HK	−9.157	−3.576	SR	4
SC	0.017	3.136	1			MO	−8.176	−7.632	−1.057	4
GZ	1.901	3.277	1							
GS	−0.416	1.898	1							
NX	0.554	2.306	1							
SN	0.156	2.452	1							
HN	1.494	3.127	1							
YN	−31.470	3.239	1							
ZJ	−42.568	−2.693		SR	CR	2				
AH	−6.502	0.854	1							
XJ	−17.588	0.860	1							
JS	−5.100	−0.392		2.372	CR	2				
GX	−32.778	−2.315		2.175	2.876	2				
HI	−4.707	−2.047		0.709	3.239	2				
HL	−34.798	−21.571		−2.979	2.684	2				
HB	−5.859	−1.757		1.929	2.894	2				
JL	−9.358	−5.337		0.979	0.984	2				
FJ	−124.893	−8.461		−5.658	2.604	2				
HA	−61.528	−18.426		−57.047	1.666	2				
SD	−70.587	−25.392		−73.276	0.409	2				
LN	−12.045	−7.962		−0.599	1.392	2				
TJ	−67.572	−42.692		−13.222	0.392	2				
HE	−60.384	−22.722		−49.956	−15.198		SR	UC		
SX	−50.626	−23.052		−38.353	−15.987		−12.749	SR	CR	3
GD	−40.442	−25.990		−27.833	−16.977		−2.943	4.730	CR	3
SH	−35.280	−29.323		−23.428	−17.420		−3.481	4.543	3.355	3
BJ	−43.330	−41.499		−28.490	−23.206		−13.187	−17.802	−15.994	
TW	−24.647	−29.921		−15.675	−17.387		−8.200	−7.608	−8.611	
HK	−43.784	−80.153		−24.139	−32.027		−12.024	−11.054	−13.344	
MO	−29.294	−44.897		−17.662	−22.060		−9.425	−8.660	−10.374	

In view of the limitation of the width of the table, we put the logt test process after Club 3 on the upper right of the table. SR refers to standard region.CR refers to core region.NC refers to nonconvergence.

In summary, there is evidence that the GDP growth rates of China's 34 regions exhibited club convergence from 1992 to 2013. The club identification results reflect significant geographical features and are consistent with China's regional economic development characteristics.

3.2 Club Convergence in Nighttime Light Growth Rates

In the convergence test of nighttime lights growth rate, we also take $r = 0.3$. First, the overall estimation equation for the $\log t$ test for China's 34 regions is:

$$\log\left(\frac{H_1}{H_t}\right) - 2\log\left[\log(t+1)\right] = \underset{(6.219)}{1.619} - \underset{(-14.319)}{1.454}\log t \qquad (7)$$

It can be seen that $\hat{b} = -1.454 \pm 0.093 < 0$ and $t_k = -14.319 < -1.65$, rejecting the null hypothesis of convergence. It indicates that the NTL growth rate of the Chinese provinces does not converge as a whole. The clubs were then divided into regions to identify NTL convergence in China.

For the initial sorting process, the last 2/3 of the entire period's data was selected for ranking. Based on this ranking, Tibet ranked

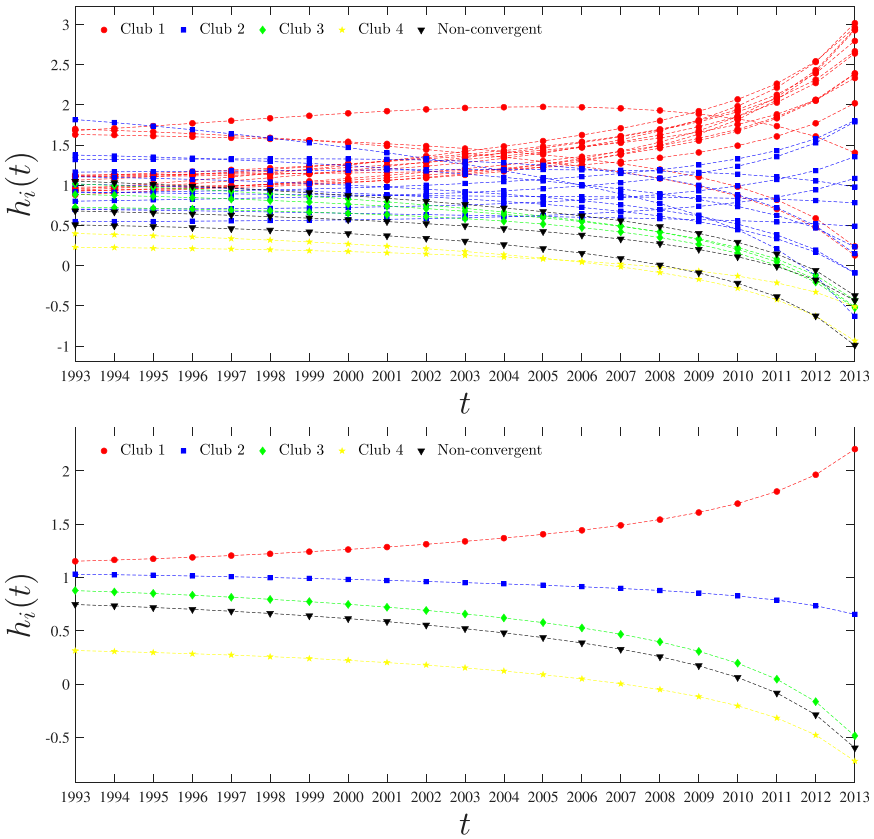


FIGURE 2 | Top: NTL transition path. **Bottom**: NTL convergence club average transition path.

first and was used as the reference region. Tibet, Jiangxi, Chongqing, Qinghai, Inner Mongolia, Sichuan, Guizhou, Gansu, Ningxia, Shaanxi, Hunan, and Yunnan were then added to carry out the $\log t$ regression and calculate the t statistic until the t statistic is less than −1.65. We found that when Yunnan was added, $t_k = -31.47 < -1.65$. Hence, the addition of regions was paused. For group $k = \{1, 2, 3\}$, the value of t_k was the largest, confirming the core group as comprising Tibet, Jiangxi, and Chongqing. Each area was then added one by one into the core group. According to the t statistic, the members of the first club were finally determined to be Tibet, Jiangxi, Chongqing, Qinghai, Inner Mongolia, Sichuan, Guizhou, Gansu, Ningxia, Shaanxi, Hunan, Yunnan, Anhui, and Xinjiang.

After excluding the members of the first club, the above test method was repeated to determine the second convergence club (Zhejiang, Jiangsu, Guangxi, Hainan, Heilongjiang, Hubei, Jilin, Fujian, Henan, Shandong, Liaoning, and Tianjin), the third convergence club (Shanxi, Guangdong, and Shanghai), the fourth convergence club (Hong Kong and Macao) and separate regions not belonging to any convergence clubs (Hebei, Beijing, and Taiwan). See **Tables 4**,**5** for details. The t_k value of the entire group of Clubs 1 and 3 is greater than Club 2, indicating that the convergence is relatively strong. The t_k value for the whole group of Clubs 2 and 4 is less than 0, indicating that their convergence is relatively weak. In terms of convergence speed, Clubs 1 and 3 have a faster convergence rate of 15.4% and 15.2%. The convergence speeds of Clubs 2 and 4 are not statistically significant.

The trend of each region's transition path is shown in the top panel of **Figure 2**. Except for Tibet, the growth rate of nighttime lights in all regions began to stabilize. From 2004, the growth rate gap between them has gradually opened, but there are still four clubs that are converging. Geographical effects still exist. Club 1 is mainly concentrated in the central and western China. Club 2 is mainly concentrated in the eastern and northeast areas, indicating that the convergence result also captures the spatial effect of NTL growth to a certain extent.

The average transition path of each club is averaged, and the transition trend of each club's overall transition path is shown in the bottom panel of **Figure 2**. It can be seen from the graph that in the initial years of the time series, the transition paths of Clubs 1, 2, and 3 are relatively concentrated, and the initial growth rate of Club 4 (Hong Kongand Macao) is the lowest, even lower than the areas that do not converge. Subsequently, the transition path of Club 1 gradually increased, Club 2 had a slight downward trend, and Clubs 3 and 4 had significant downward trends. The difference in NTL growth rates between clubs has gradually increased.

In summary, there is evidence that the NTL growth rates of China's 34 regions did not have an overall convergence during

TABLE 6 | Robustness test of GDP convergence.

Clubs	Club size	Members	*t*-value
Panel A: $L(t) = \log(t)$			
1	4	IM, TJ, MO, GD	–1.292
2	14	CQ, SN, JS, SD, QH, GZ, FJ, HB, AH, XZ, HA, ZJ, HE, SH	6.226
3	14	JL, SC, GX, JX, HN, LN, SX, NX, HI, GS, YN, HL, XJ, BJ	14.523
4	2	HK, TW	4.866
Panel **B**: $L(t) = \log\log(t)$			
1	4	IM, TJ, MO, GD	–1.45
2	14	CQ, SN, JS, SD, QH, GZ, FJ, HB, AH, XZ, HA, ZJ, HE, SH	6.08
3	14	JL, SC, GX, JX, HN, LN, SX, NX, HI, GS, YN, HL, XJ, BJ	8.198
4	2	HK, TW	4.34
Panel C: $L(t) = \log\log(t+1)$			
1	4	IM, TJ, MO, GD	–1.283
2	14	CQ, SN, JS, SD, QH, GZ, FJ, HB, AH, XZ, HA, ZJ, HE, SH	6.248
3	14	JL, SC, GX, JX, HN, LN, SX, NX, HI, GS, YN, HL, XJ, BJ	13.871
4	2	HK, TW	4.813

TABLE 7 | Robustness test of NTL convergence.

Clubs	Club size	Members	*t*-value
Panel A: $L(t) = \log(t)$			
1	14	XZ, JX, CQ, QH, IM, SC, GZ, GS, NX, SN, HN, YN, AN, XJ	3.382
2	12	ZJ, JS, GX, HI, HL, HB, JL, FJ, HA,SD, LN, TJ	2.372
3	3	SX, GD, SH	4.73
4	2	HK, MO	–1.057
Nonconvergent	3	HE, BJ, TW	
Panel **B**: $L(t) = \log\log(t)$			
1	12	XZ, JX, CQ, QH, IM, SC, GZ, GS, NX, SN, HN, YN	2.948
2	10	ZJ, AH, XJ, JS, GX, HI, HL, HB, JL, FJ	3.683
3	6	SD, LN, TJ, HE, SX, GD	3.236
4	2	HK, MO	–1.167
Nonconvergent	4	HA, SH, BJ, TW	
Panel C: $L(t) = \log\log(t+1)$			
1	14	XZ, JX, CQ, QH, IM, SC, GZ, GS, NX, SN, HN, YN, AN, XJ	3.478
2	12	ZJ, JS, GX, HI, HL, HB, JL, FJ, HA,SD, LN, TJ	4.611
3	3	SX, GD, SH	4.7
4	2	HK, MO	–1.045
Nonconvergent	3	HE, BJ, TW	

1992–2013 but exhibited club convergence. The growth rates gap between different clubs increases with the year.

3.3 Robustness

In order to avoid the influence of the slow change function $L(t)$ on the results, this article uses three different situations $L(t) = \log(t)$, $L(t) = \log[\log(t)]$, and $L(t) = \log[\log(t+1)]$, respectively, in the robustness tests, mainly examining the robustness of club size and club member [43]. The test results of GDP growth rates are shown in **Table 6**. The club size and members are exactly the same for $L(t) = \log(t)$, $L(t) = \log[\log(t)]$, and $L(t) = \log[\log(t+1)]$, with only a slight difference in the *t*-values. The robustness results show that the effect of the slow-changing function on the convergence of China's GDP club is ignorable and the GDP club convergence results are very robust.

The robustness tests for nighttime lights are shown in **Table 7**. The club size and members are exactly the same for $L(t) = \log(t)$ and $L(t) = \log[\log(t+1)]$, with only a slight difference in the *t*-values. When $L(t) = \log[\log(t)]$, the club size and members are different from the first two cases, but the club members still have a remarkable overlap with the first two cases. The robustness results show that the effect of the slow-changing function on the convergence of China's nighttime lights club is minor and the NTL club convergence results are relatively robust.

3.4 Convergence Factor Testing

Due to the lack of relevant data in Hong Kong, Macao, and Taiwan, Club 4 is not included in discussing the club convergence factors for GDP and NTL. According to the principles of club classification, as the club group rank increases, the growth rate of clubs in the corresponding region is lower; on the contrary, the lower the club group ranking, the higher the growth rate of clubs in the corresponding region. The overall fit of the GDP clubs gives $N = 651$ with $R^2 = 0.046$ and $p = 0.000$, indicating that the equation is relatively significant. In terms of the direction of

TABLE 8 | Marginal effect of GDP convergence clubs and NTL convergence clubs.

	GDP			Nighttime light		
Explanatory variables	**Club 1**	**Club 2**	**Club 3**	**Club 1**	**Club 2**	**Club 3**
Fixed capital	0.000	0.000	−0.000	0.001	−0.000	−0.000
Investment increment	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Employees increment	0.143	0.024	−0.038	−0.188	0.111	0.077
	(0.408)	(0.068)	(0.109)	(0.169)	(0.1)	(0.07)
Population	0.003	0.006	−0.009	0.023[c]	−0.014[c]	−0.009[c]
Growth rate	(0.002)	(0.004)	(0.006)	(0.006)	(0.004)	(0.003)
Patents increment	0.046[b]	0.077[b]	−0.124[b]	0.185[c]	−0.109[c]	−0.076[c]
	(0.023)	(0.038)	(0.06)	(0.065)	(0.039)	(0.028)
Secondary industry	0.007[c]	0.012[c]	−0.019[c]	−0.017[c]	0.01[c]	0.007[c]
Proportion	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)
Tertiary industry	0.003[b]	0.004[b]	−0.007[b]	−0.017[c]	0.01[c]	0.007[c]
Proportion	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
Social consumer goods	0.001	0.002	−0.003	−0.027	0.016	0.011
Retail increment	(0.014)	(0.024)	(0.038)	(0.034)	(0.02)	(0.014)
Foreign direct	0.005	0.008	−0.012	0.062[b]	−0.037[b]	−0.026[b]
Investment increment	(0.003)	(0.005)	(0.008)	(0.031)	(0.018)	(0.013)
Coal consumption	0.041	0.068	−0.109	0.327[b]	−0.193[b]	−0.134[b]
Increment	(0.062)	(0.104)	(0.166)	(0.159)	(0.095)	(0.066)

[a]*Indicates significant at the 10% level.*
[b]*Indicates significant at the 5% level.*
[c]*Indicates significant at the 1% level.*

the parameter estimation results, all variables have a negative impact on club grading, that is, as a region's mean of these explanatory variables increases by one unit, the probability that the region will move toward a club with a higher economic growth rate will increase.

Excluding Hong Kong, Macao, Taiwan and nonconvergent regions, the overall fitting result of the NTL club gives $N = 609$ with $R^2 = 0.134$ and $p = 0.000$. The equation is relatively significant, and the direction of the parameter estimation results is different from the GDP estimation results. Fixed asset investment increment, population growth rate, patents increment, foreign direct investment increment, and coal consumption increment have a negative impact on club grading, that is, as a region's mean of these explanatory variables increases, the probability that the region will move toward a club with a higher NTL growth rate will increase. Laborer increment, industrial share, service industry share, and total retail sales of social consumer goods increment have a positive impact on club grading, that is, as a region's mean of these explanatory variables increases, the probability that the region will move toward a club with a lower NTL growth rate will increase.

We further calculated the marginal effects of explanatory variables on GDP club division. The results are shown in columns 2, 3, and 4 of **Table 8**. Clubs 1 and 2 have the same direction for each variable, which is precisely the opposite of Club 3. Among these variables, the coefficients of patents increment, secondary industry proportion, and tertiary industry proportion are significant. If the patents increment increases by a mean of 1%, the likelihood that the region belongs to Clubs 1 and 2 will increase 0.046% and 0.077%, respectively, while the likelihood of belonging to Club 3 will decrease 0.124%. If the secondary industry proportion increases by a mean of 1%, the likelihood that the region belongs to Clubs 1 and 2 will increase 0.007% and 0.012%, respectively, while the likelihood of belonging to Club 3 will decrease 0.019%. If the mean tertiary industry proportion increases by 1%, the likelihood of the region belonging to Clubs 1 and 2 will increase 0.003% and 0.004%, respectively, while the likelihood of belonging to Club 3 will decrease 0.007%. These results indicate that technological innovation and industrial structure played an essential role in promoting economic growth. In fact, the results of marginal effects are consistent with the realities of economic development. The high GDP growth clubs are mainly concentrated in the eastern regions, which have an economic development level faster than the western regions. High-tech industries are mainly concentrated in the east developed areas, and the industrial and service industries have developed rapidly since China's economic transformation, which was initially put forward in 2001. This is an important reason why the eastern regions' economic growth is higher than that of the central and western regions.

The marginal effect of explanatory variables on NTL club division can be seen in **Table 8**. Clubs 2 and 3 have the same direction of change for each variable, which is precisely the opposite of Club 1. When these data are combined with the NTL transition paths (bottom panel of **Figure 2**), it can be found that the transition path of Club 1 is gradually increasing, that is, the NTL growth rates increase with time and regarded as a high growth rates club. The transition paths of Clubs 2 and 3 have a downward trend, that is, the NTL growth rates decline with time and regarded as a low growth rates clubs. From the results presented in **Table 8**, it can be found that, with the increase of fixed capital investment, population growth rate, patents

TABLE 9 | Comparison of GDP and NTL club members

Panel A: GDP growth rate					
Club	**Club 1**	**Club 2**	**Club 3**	**Club 4**	**—**
Inclusion groups		A, C	B, D		
Members outside the groups	IM, TJ, MO, GD	HE, SH	SX, BJ	HK, TW	
Panel **B**: NTL growth rate					
Club	Club 1	Club 2	Club 3	Club 4	Nonconvergent
Inclusion groups	A, B	C, D			
Members outside the groups	IM	TJ	SX, GD, SH	HK, MO	HE, BJ, TW

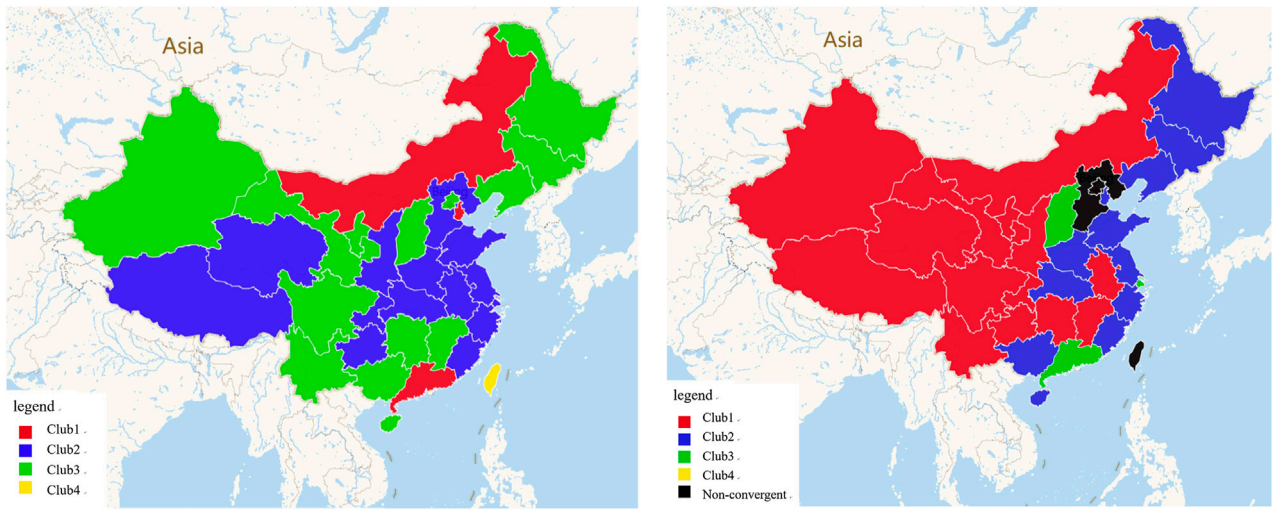


FIGURE 3 | Club division maps.

increment, foreign direct investment, and coal consumption increment, the probability that a region has a high growth rate of NTL will increase. Increases in employee increment, the secondary industrial proportion, the tertiary industrial proportion, and retail sales of social consumer goods will increase the probability that a region will have a low growth rate of NTL. The marginal effect of the population growth rate, patents increment, secondary industrial proportion, tertiary industrial proportion, foreign direct investment, and coal consumption increment is significant. If the population growth rate mean increases by 1%, the probability of region belonging to Club 1 will increase 0.023%, while the probability of belonging to Clubs 2 and 3 will decrease 0.014% and 0.009%, respectively. If the patent increment mean increases by 1%, the probability of belonging to Club 1 will increase 0.185%, while the probability of belonging to Clubs 2 and 3 will decrease 0.109% and 0.076%, respectively. If the secondary industrial proportion mean increases by 1%, the probability of belonging to Club 1 will decrease 0.017%, while the probability of belonging to Clubs 2 and 3 will increase 0.01% and 0.007%. If the mean tertiary industrial proportion is increased by 1%, the probability of belonging to Club 1 will decrease 0.017%, while the probability of belonging to Clubs 2 and 3 will increase by 0.01% and 0.007%, respectively. The influence of foreign direct investment and coal consumption increment can also be observed from **Table 8**.

3.5 Comparison of GDP and NTL Growth Rate Convergence Characteristics

From the perspective of club membership, we identified the common regions where the club divisions' results coexist. We refer to these common areas as groups. Group A includes six regions (Tibet, Chongqing, Qinghai, Guizhou, Shaanxi, and Anhui), Group B contains seven regions (Jiangxi, Sichuan, Gansu, Ningxia, Hunan, Yunnan, and Xinjiang), Group C includes six regions (Jiangsu, Shandong, Fujian, Hubei, Henan, and Zhejiang), and Group D includes five regions (Guangxi, Hainan, Heilongjiang, Jilin, and Liaoning). These four groups and their corresponding clubs all appear in the two different club division results, as shown in **Table 9**. It can be clearly seen that in terms of the order of regional growth rate, the results of the two clubs show certain degree of difference.

From the perspective of regional distribution, the maps in **Figure 3** show that in GDP convergence clubs, the regions in higher ranked clubs are mainly concentrated in the central and eastern areas, that is, these regions have higher GDP growth rates. The western and northeast regions belong to clubs with a relatively

low ranking and these regions have relatively low GDP growth rates. Simultaneously, Tibet and Qinghai are located in the west but are divided into the same club as the eastern coastal areas, indicating that Tibet and Qinghai's GDP growth rates are higher than those of other western regions. In the NTL convergence clubs, the regions in higher ranked clubs are mainly concentrated in the west and central areas, which means these regions have higher NTL growth rates. The eastern and northeast regions belong to clubs with a relatively low ranking, which means these regions have relatively low NTL growth rates.

Judging from the gap in transition paths between clubs, the difference between the two types of data is significant. Due to the particularity of the data in Hong Kong, Macao, and Taiwan, only the 31 regions in the mainland are considered. The transition paths of GDP Clubs 1, 2, and 3 show a catching-up effect. The transition paths between clubs first converge, then diverge, with a relatively flat path of divergence. The transition paths of each club are in line with the patterns of linear development. The transition paths of Clubs 1 and 3 are on rise, while Club 2 is on decline. The transition paths of NTL Clubs 1, 2, and 3 do not show a catching-up effect. With the annual increase, the transition path of each club grew intenser, showing a nonlinear development pattern. The transition path of Club 1 is on the rise, and the transition paths of Clubs 2, 3, and 4 show a downward trend.

From the perspective of influencing factors of club division, fixed capital investment increment, population growth rate, patents increment, foreign direct investment increment, and coal consumption increment have the same direction of influence on the GDP growth rate and NTL growth rate, but there are subtle differences in the significance. For the GDP growth rate, the most significant factor is patents increment, followed by industrial proportion and industrial proportion. For the NTL growth rate, in addition to the above three variables, population growth rate, foreign direct investment increment, and coal consumption increment are also significant.

4 CONCLUSION

With the development of science and technology, the easily accessible and more objective NTL data is increasingly valued and used as a surrogate variable for economic growth. Besides, China is undergoing a transitional period characterized by strategic developmental opportunities and social tensions. The consideration of regional economic convergence where convergence speed is heterogeneous is crucial for the coordinated development of regional economies. To explore the reliability of economic variables in NTL data, from the perspective of regional economic convergence, the log t test based on the nonlinear time-varying factor model was used for club convergence analysis on the inter-provincial DMSP/OLS NTL data and GDP data from 1992 to 2013. It was found that between 1992 and 2013, China's provincial NTL and GDP growth rates exhibited club convergence. The high-GDP-growth clubs are mainly concentrated in the eastern and central regions, while the high-NTL-growth clubs are mainly concentrated in the central and west regions. The GDP club's catching-up effect between areas is consistent with the neoclassical growth model's general conclusion. Also, the growth rate gap among clubs differs between the two data types. The growth rate gap among GDP clubs is relatively flat, while the growth rate gap among NTL clubs is large. Therefore, there is a significant difference between NTL and GDP club convergence.

This article further explored the club convergence mechanisms. We have found that patents increment, industrial proportion, and service industry proportion have a significant impact on both GDP and NTL. Among them, the increase in the proportion of industry and service industry has opposite effects on GDP clubs and NTL clubs, while the increase in patent increment has the same impact on GDP clubs and NTL clubs. This confirms that the industrial structure of eastern coastal areas has been upgraded faster than central and western regions. In addition, we found that, for NTL clubs, population growth rate, foreign direct investment, and coal consumption are also significant factors. An increase in population growth rate, foreign direct investment, and coal consumption will increase the probability that a region belongs to a high-growth club. It explains that there are also certain differences between GDP clubs and NTL clubs in terms of influencing factors.

On the whole, from 1992 to 2013, there were certain differences in the results of club convergence between China's provincial NTL and GDP growth. The difference between NTL and GDP growth is mainly reflected in regions and influencing factors. High GDP growth regions correspond to low NTL growth regions, while low GDP growth regions correspond to high NTL growth regions. GDP growth is mainly influenced by technological innovation and industrial structure, while NTL growth is mainly affected by the population growth rate, technological innovation, industrial structure, opening to the international world, and resource consumption. The impact of industrial structure on GDP and NTL is the opposite. At present, when GDP statistics are difficult to obtain or are of low quality, a large number of studies use NTL brightness as a representative of the level of economic development. But judging from the results of this work, NTL is not a good substitute for GDP. Our results suggest that by distinguishing the connotations and differences between GDP and NTL and combining the characteristics of the two, a more reasonable alternative indicator of economic development can be constructed to make it more in line with the specific facts and laws of economic activities.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://ngdc.noaa.gov/eog/download.html.

AUTHOR CONTRIBUTIONS

WZ conceived the research. QX analyzed the data and prepared the figures. WZ and QX analyzed the results and wrote the paper. All authors discussed and reviewed the manuscript.

FUNDING

This work was partly supported by the Fundamental Research Funds for the Central Universities.

REFERENCES

1. Bennett MM, and Smith LC Advances in using multi-temporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics. *Remote Sens Environ* (2017) 192:176–97. doi:10.1016/j.rse.2017.01.005
2. Rawski TG Chinese industrial production, 1952-1971. *Rev Econ Stat* (1973) 55:169–81. doi:10.2307/1926991
3. Elvidge CD, Baugh KE, Kihn EA, Kroehl HW, Davis ER, and Davis CW Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *Int J Remote Sens* (1997) 18:1373–9. doi:10.1080/014311697218485
4. Doll CNH, Muller JP, and Elvidge CD Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. *Ambio: J Human Environ* (2000) 29:157–62. doi:10.1579/0044-7447-29.3.157
5. Sutton PC, and Costanza R Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecol Econ* (2002) 41:509–27. doi:10.1016/S0921-8009(02)00097-6
6. Sutton PC, Elvidge CD, and Ghosh T Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. *Int J Ecol Econ Stat* (2007) 8:5–21.
7. Ghosh T, Anderson S, Powell R, Sutton PC, and Elvidge CD Estimation of Mexico's informal economy and remittances using nighttime imagery. *Remote Sens* (2009) 1:418–44. doi:10.3390/rs1030418
8. Ghosh T, Powell RL, Elvidge CD, Baugh KE, Sutton PC, and Anderson S Shedding light on the global distribution of economic activity. *Open Geogr J* (2010) 3:147–60. doi:10.2174/1874923201003010147
9. Elvidge CD, Baugh KE, Zhizhin M, and Hsu FC Why VIIRS data are superior to DMSP for mapping nighttime lights. *Proc Asia-Pacific Adv Netw* (2013) 35:62–9. doi:10.7125/APAN.35.7
10. Ebener S, Murray C, Tandon A, and Elvidge CC From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *Int J Health Geogr* (2005) 4:5. doi:10.1186/1476-072X-4-5
11. Doll CNH, Muller JP, and Morley JG Mapping regional economic activity from night-time light satellite imagery. *Ecol Econ* (2006) 57:75–92. doi:10.1016/j.ecolecon.2005.03.007
12. Blumenstock JE Fighting poverty with data. *Science* (2016) 353:753–4. doi:10.1126/science.aah5217
13. Li D, and Li X Applications of night-time light remote sensing in eveluation of socioeconomic development. *J Macro-quality Res (in Chinese)* (2015) 3:1–8. doi:10.13948/j.cnki.hgzlyj.2015.04.001
14. Mellander C, Lobo J, Stolarick K, and Matheson Z Night-time light data: a good proxy measure for economic activity?. *PLoS ONE* (2015) 10:e0139779. doi:10.1371/journal.pone.0139779
15. Imhoff ML, Lawrence WT, Elvidge CD, Paul T, Levine E, Privalsky MV, et al. Using nighttime DMSP/OLS images of city lights to estimate the impact of urban land use on soil resources in the United States. *Remote Sens Environ* (1997) 59:105–17. doi:10.1016/s0034-4257(96)00110-1
16. Chen J, Zhuo L, Shi PJ, and Ichinose T The study on urbanization process in China based on DMSP/OLS data :Development of a light index for urbanization level estimation. *J Remote Sens (In Chinese)* (2003) 7:168–75. doi:10.11834/jrs.20030302
17. Small C, Pozzi F, and Elvidge CD Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sens Environ* (2005) 96:277–91. doi:10.1016/j.rse.2005.02.002
18. Noor AM, Alegana VA, Gething PW, Tatem AJ, and Snow RW Using remotely sensed night-time light as a proxy for poverty in Africa. *Popul Health Metrics* (2008) 6:1–13. doi:10.1186/1478-7954-6-5
19. Zhang QL, and Seto KC Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. *Remote Sens Environ* (2011) 115:2320–9. doi:10.1016/j.rse.2011.04.032
20. Stathakis D, Tselios V, and Faraslis I Urbanization in European regions based on night lights. *Remote Sensing Appl Soc Environ* (2015) 2:26–34. doi:10.1016/j.rsase.2015.10.001
21. Ma T, Zhou YK, Zhou CH, Haynie S, Pei T, and Xu T Night-time light derived estimation of spatio-temporal characteristics of urbanization dynamics using DMSP/OLS satellite data. *Remote Sens Environ* (2015) 158:453–64. doi:10.1016/j.rse.2014.11.022
22. Amaral S, Camara G, Monteiro AMV, Quintanilha JA, and Elvidge CD Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data. *Comput Environ Urban Syst* (2005) 29:179–95. doi:10.1016/j.compenvurbsys.2003.09.004
23. Amaral S, Monteiro AMV, Camara G, and Quintanilha JA DMSP/OLS night-time light imagery for urban population estimates in the Brazilian Amazon. *Int J Remote Sens* (2006) 27:855–70. doi:10.1080/01431160500181861
24. Axbard S Income opportunities and sea piracy in Indonesia: evidence from satellite data. *Appl Econ* (2016) 8:154–94. doi:10.1257/app.20140404
25. Elvidge CD, Ziskin D, Baugh KE, Tuttle BT, Ghosh T, Pack DW, et al. A fifteen year record of global natural gas flaring derived from satellite data. *Energies* (2009) 2:595–622. doi:10.3390/en20300595
26. Chen X, and Nordhaus WD Using luminosity data as a proxy for economic statistics. *Proc Natl Acad Sci U.S.A* (2011) 108:8589–94. doi:10.1073/pnas.1017031108
27. Henderson V, Storeygard A, and Weil DN A bright idea for measuring economic growth. *Amer Econ Rev* (2011) 101:194–9. doi:10.1257/aer.101.3.194
28. Henderson JV, Storeygard A, and Weil DV Measuring economic growth from outer space. *Amer Econ Rev* (2012) 102:994–1028. doi:10.1257/aer.102.2.994
29. Chen X, and Nordhaus W A test of the new VIIRS lights data set: population and economic output in Africa. *Remote Sens* (2015) 7:4937–47. doi:10.3390/rs70404937
30. Donaldson D, and Storeygard A The view from above: applications of satellite data in economics. *J Econ Persp* (2016) 30:171–98. doi:10.1257/jep.30.4.171
31. Ma T, Zhou CH, Pei T, Haynie S, and Fan JF Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: a comparative case study from China's cities. *Remote Sens Environ* (2012) 124:99–107. doi:10.1016/j.rse.2012.04.018
32. Yu B, Shi K, Hu Y, Huang C, Chen Z, and Wu J Poverty evaluation using NPP-VIIRS nighttime light composite data at the county level in China. *IEEE J Sel Top Appl Earth* (2015) 8:1217–29. doi:10.1109/JSTARS.2015.2399416
33. Shi K, Yu B, Huang Y, Hu Y, Yin B, Chen Z, et al. Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: a comparison with DMSP-OLS data. *Remote Sens* (2014) 6:1705–24. doi:10.3390/rs6021705
34. Xu K, Chen F, and Liu X The truth of China economic growth: evidence from global night-time light data. *Econ Res J* (2015) 50:17–29. [in Chinese].
35. Li X, Xu HM, Chen XL, and Li C Potential of NPP-VIIRS nighttime light imagery for modeling the regional economy of China. *Remote Sens* (2013) 5:3057–81. doi:10.3390/rs5063057
36. Huang X, Schneider A, and Friedl MA Mapping sub-pixel urban expansion in China using MODIS and DMSP/OLS nighttime lights. *Remote Sens Environ* (2016) 175:92–108. doi:10.1016/j.rse.2015.12.042
37. He C, Ma Q, Li T, Yang Y, and Liu Z Spatiotemporal dynamics of electric power consumption in Chinese Mainland from 1995 to 2008 modeled using DMSP/OLS stable nighttime lights data. *J Geogr Sci* (2012) 22:125–36. doi:10.1007/s11442-012-0916-3
38. Liu HJ, Pei YF, and Jia WX Spatial differences and spillover effects of economic development of urban agglomerations in China: on DMSP/OLS nighttime light data from 1992 to 2013. *Finance Trade Res (in Chinese)* (2017) 28:1–12. doi:10.19337/j.cnki.34-093/f.2017.11.001
39. Liu HJ, and Du GJ Regional inequality and stochastic convergence in China. *J Quant Tech Econ* (2017) 2017:43–59. [in Chinese]. doi:10.13653/j.cnki.jqte.2017.10.003
40. Wang XB, Huang LX, Xu XX, and Li X Revisit China's regional economic convergence: evidence from the DMSP/OLS night light data. *Chin Econ Q (in Chinese)* (2017) 16:877–96.
41. Ding HF, and Zhou YX Viewing the spatial and temporal pattern of regional economic development in China from night light. *Macroeconomics* (2017) 2017:130–8.[in Chinese]. doi:10.16304/j.cnki.11-3952/f.2017.03.011
42. Phillips PCB, and Sul D Transition modeling and econometric convergence tests. *Econometrica* (2007) 75:1771–855. doi:10.1111/j.1468-0262.2007.00811.x
43. He Y Economic convergence in transition economy of China: an empirical study with a nonlinear time-varying factor model. *Econ Res J* (2008) 2008:39–51.[in Chinese].

44. Fischer C Price convergence in the EMU? Evidence from micro data. *Eur Econ Rev* (2012) 56:757–76. doi:10.1016/j.euroecorev.2012.01.008
45. Bartkowska M, and Riedl A Regional convergence clubs in Europe: identification and conditioning factors. *Econ Model* (2012) 29:22–31. doi:10.1016/j.econmod.2011.01.013
46. Meng H, Xie WJ, and Zhou WX Club convergence of house prices: evidence from China's ten key cities. *Int J Mod Phys B* (2015) 29:1550181. doi:10.1142/S0217979215501817
47. Tian X, Zhang XH, Zhou YH, and Yu XH Regional income inequality in China revisited: a perspective from club convergence. *Econ Model* (2016) 56:50–8. doi:10.1016/j.econmod.2016.02.028
48. Lyncker KV, and Thoennessen R Regional club convergence in the EU: evidence from a panel data analysis. *Empir Econ* (2017) 52:525–53. doi:10.1007/s00181-016-1096-2
49. Sun XH, and Cao Y Convergence of clubs for urban economic growth: Identification methods and convergence mechanisms–An empirical test from 347 administrative regions in China. *Mod Econ Sci (in Chinese)* (2018) 40:14–25.
50. Ma L, Wu J, Li W, Peng J, and Liu H Evaluating saturation correction methods for DMSP/OLS nighttime light data: a case study from China's cities. *Remote Sens* (2014) 6:9853–72. doi:10.3390/rs6109853
51. Cao ZY, Wu ZF, Kuang YQ, and Huang NS Correction of DMSP/OLS night-time light images and it's application in China. *J Geo-Information Sci (in Chinese)* (2015) 17:1092–102. doi:10.3724/SP.J.1047.2015.01092
52. Wu J, He S, Peng J, Li W, and Zhong X Intercalibration of DMSP-OLS night-time light data by the invariant region method. *Int J Remote Sens* (2013) 34:7356–68. doi:10.1080/01431161.2013.820365
53. Liu Z, He C, Zhang Q, Huang Q, and Yang Y Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. *Landsc Urban Plan* (2012) 106:62–72. doi:10.1016/j.landurbplan.2012.02.013
54. Cai F, and Du Y Convergence and divergence of regional economic growth in China. *Econ Res J* (2000) 2000:30–7. [in Chinese].
55. Shen KR, and Ma J The characteristics of "Club Convergence" of China's economic growth and it's cause. *Econ Res J* (2002) 2002:33–9. [in Chinese].
56. Lin YF, and Liu PL Chinese development strategy and economic convergence. *Econ Res J* (2003) 2003:19–25. [in Chinese].
57. Tang XB, and Chen XS Analysis of economic convergence and its influence factors in eight major regions of China. *J Renmin Univ China* (2007) 2007:106–13. [in Chinese].
58. Pan XF, and Yang Y Rsearch on regional innovation club convergence and influencing factors in China. *Stud Sci Sci (in Chinese)* (2014) 32:130–8. doi:10.16192/j.cnki.1003-2053.2014.02.015
59. Zhu GZ, Qiao KY, and Yu JH Is provincial economic growth convergence in China?. *Chin Econ Q (in Chinese)* (2014) 13:1171–94. doi:10.13821/j.cnki.ceq.2014.03.016
60. Dai M, and Mao R The industrial heterogeneity, the industrial structure, and the provincial growth convergence in China. *Manage World* (2015) 2015:34–46+62+187. [in Chinese]. doi:10.19744/j.cnki.11-1235/f.2015.06.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Xiao, Wang and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Physics

ORIGINAL RESEARCH
published: 13 April 2022
doi: 10.3389/fphy.2022.593493

Network Structures for Asset Return Co-Movement: Evidence From the Chinese Stock Market

Huai-Long Shi[1]* *and Huayi Chen*[2]

[1]School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, China, [2]College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China

This article focuses on the detailed network structure of the co-movement for asset returns. Based on the Chinese sector indices and Fama-French five factors, we conducted return decomposition and constructed a minimum spanning tree (MST) in terms of the rank correlation among raw return, idiosyncratic return, and factor premium. With the adoption of a rolling window analysis, we examined the static and time-varying characteristics associated with the MST(s). We obtained the following findings: 1) A star-like structure is presented for the whole sample period, in which market factor *MKT* acts as the hub node; 2) the star-like structure changes during the periods for major market cycles. The idiosyncratic returns for some sector indices would be disjointed from *MKT* and connected with their counterparts and other pricing factors; and 3) the effectiveness of pricing factors are time-varying, and investment factor *CMA* seems redundant in the Chinese market. Our work provides a new perspective for the research of asset co-movement, and the test of the effectiveness of empirical pricing factors.

Keywords: Chinese stock market, asset pricing, sector index, minimum spanning tree, co-movement

OPEN ACCESS

Edited by:
Gabjin Oh,
Chosun University, South Korea

Reviewed by:
Michele Bellingeri,
University of Parma, Italy
Dehua Shen,
Tianjin University, China

***Correspondence:**
Huai-Long Shi
hlshi@nuist.edu.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *10 August 2020*
Accepted: *14 March 2022*
Published: *13 April 2022*

Citation:
Shi H-L and Chen H (2022) Network Structures for Asset Return Co-Movement: Evidence From the Chinese Stock Market. Front. Phys. 10:593493. doi: 10.3389/fphy.2022.593493

1 INTRODUCTION

Co-movement refers to the correlated or similar movement of two or more entities. In finance, the co-movement of asset returns is crucial to the cross-sectional diversification and management of systematic risk. It also affects the way shocks are transmitted and thus the level of systemic risk [1]. Understanding the co-movement for asset returns is essential and beneficial in the academic field and practical investment. According to the seminal work by Ref. [2], asset co-movement could be decomposed into two parts. One part is "reasonable" co-movement, which is related to the fundamentals of assets. And the other part is "excess" co-movement, or "friction-based" co-movement, which is linked to investor sentiment and is beyond the explanation of fundamental change with respect to the assets. Motivated by Ref. [2], a strand of ensuing studies have been carried out in the literature, with the adoption of the indicators in terms of R^2 or its variants derived from the regression on the market performance [1, 3–5].

However, most literature has concentrated more on the co-movement between individual assets and market performance rather than that among the assets themselves. In practice, people may be more concerned with the co-dependency or cross-correlation for various assets. It was virtually realized *via* the indirect evaluation of the co-movement between each asset and market performance [6], because of computational complexity in early years. The overall market performance serves as the "intermediary" during the process.

Along with the sharp drop in computing costs, an increasing number of researchers have turned their attention to more direct and accurate metrics of the co-movement among the assets. In

particular, based upon the network theory, a surge of literature is coming out to unveil the co-movement structures within or across asset classes. The network-based approach typically consists of the following two steps: 1) constructing the matrix in terms of cross-correlation coefficients for given assets, and thereby 2) mapping it to a graph, in which the nodes refer to the assets with links or edges connecting them.[1] Therefore, the topological properties associated with the graph or network can be further obtained to describe the structure of the co-movement.

For the first step, Pearson correlation is one of the most popular methods, especially in the early literature, in which it is simple to capture the linear correlations between the assets. Some similar alternatives include partial correlation [7], co-integration [8], and lead–lag relation [9–11]. Although these approaches function well in featuring the linear correlation among assets, they fail to capture the non-linear correlation that is more prevalent in financial markets. In response, more approaches have been employed, such as copula [12], mutual information [13], and rank correlation [14]. For the second step, it is easy to build a complete graph or network based on the raw correlation matrix, containing rich and even redundant information. An information filter is necessary, and a number of network-based methods have been developed to this end. For instance, minimum spanning tree (MST, hereafter) is one of the most popular methods [15–17]. Assuming that there are a total of N nodes in the original network, $N - 1$ most important edges would be retained in the MST. Planar maximally filtered graph (PMFG, hereafter) has also been frequently adopted in the literature [18]. The PMFG retains more information and exhibits more robustness than the MST. It is mainly realized by the fact that the nodes have no cross-linkage on the planar. In comparison, the most intuitive way is to conduct the threshold method [19, 20], which is able to filter out all the nodes with linkage weights greater than a specific critical value.

In this work, we concentrate on the detailed co-movement of the assets proxied by industrial sector indices in China. China has witnessed rapid development in its financial market in recent decades, while some phenomena still characterize the Chinese market, including the less transparent information environment at the market and firm levels, and a more significant proportion of irrational individual investors [21]. These result in remarkably distinct performance in the stock market from other mature markets [22], which necessitates further study. Furthermore, despite rich research on the co-movement for the assets in China, insufficient attention has been paid to a more detailed co-movement structure, which plays an increasingly important role in portfolio selection and risk management. On the other hand, with the rapid development of "factor zoo" [23], the performance of assets can be attributed to the compensation for various kinds of risks or the premium for investing styles. In this vein, it pays to perform an anatomy of the co-movement structure for the asset, which is the focus of our work.

Given that the network-based approach is more informative about detailed co-movement structures than those traditional methods with respect to the linear regression model (i.e., R^2 and its variants), we constructed the MST with layered structures in terms of correlations among systematic, idiosyncratic, and raw returns of assets. Based on the network theory, topological properties associated with the MST can be derived to describe the co-movement structure. For the whole sample period, we found that the MST presented a star-like connection, with the hub node being market factor *MKT*, and all sector nodes were connected to market factor *MKT*. Size factor *SMB* was directly linked to the sector composite (*Cps*), while the other three factors were connected through *SMB*. Our plot of the MST also implied the effectiveness of Fama-French pricing factors in China: market factor *MKT* played an important role, while investment factor *CMA* seemed redundant. These findings are consistent with the prior literature [24, 25].

With the adoption of the rolling window analysis, we observed that the topological properties associated with the network structure are time-varying. In particular, in major market cycles, the star-like structure would be accordingly changing. This is featured by the fact that a few idiosyncratic returns for sector indices would be disjointed from market factor *MKT* and then connected to their counterparts, which is consistent with the findings from Ref. [14]. Meanwhile, pricing factors would generally exhibit more favorable pricing efficacy.

The contributions of our work are as follows:

Our first contribution is to probe into the detailed co-movement structures of asset returns in China, to which scarce attention has been paid. In fact, to the best of our knowledge, our work is the first which is aimed for this purpose. A mounting of the literature has focused on the individual stock co-movement in China as well as the driving factors of change in the co-movement structure [11, 26], while another strand of the literature has examined the co-movement structure across the asset classes, such as that for the individual stocks in the Chinese market and other mature markets [27], and that between global oil prices and China's commodity sectors [28]. In comparison, our focus on detailed co-movement structures of asset returns differentiates our work from the literature mentioned previously. Specifically, we conduct the return decomposition based on the Fama-French five-factor model for China and further construct the MST with layered structures. Based on this, we probe into the inter-structure for the co-movement between asset raw return and price premium, and the intra-structures for the factor premiums and idiosyncratic returns.

Our second contribution is to extend Ref. [14] by considering the weighted schemes for the co-movement structure in asset returns. The authors of Ref. [14] adopt the network-based approach to examine the asset returns' co-movement in the US market. Based on the Fama-French three-factor model, they decompose the returns of industrial sector indices into systematic and idiosyncratic parts. They further examine the properties associated with time-varying MSTs with unweighted

[1]The edge exists between the nodes for assets if they have a higher pairwise correlation, which will be detailed in the following sections. Note that "edge" and "link" are interchangeable in our work.

edges between the returns. Their results suggest that unexpected industries connect idiosyncratically through the dot-com bust. Different from their work, we focus on unweighted as well as weighted schemes in our work. Our results based on weighted schemes reveal more explicit and remarkable fluctuations in the measures for the co-movement structure for the period of major cycles in China than the results based on unweighted schemes, including the period around 2007, 2015, and 2018, corresponding to three major bursts of stock bubbles in China. In comparison, the results based on unweighted schemes convey less information, especially highlighted by our results for the intra-structure of returns' co-movement.[2]

Last but more importantly, our work adds to the literature on empirical asset pricing by providing evidence of the time-varying effectiveness of pricing factors in the Chinese stock market. By conducting the rolling window analysis, our work lends support to the importance of market factor *MKT* but the redundancy of the investment factor *CMA*, which is also consistent with empirical findings in previous literature. In fact, the majority of research works focusing on asset co-movement are conducted based on cross-sectional and time-series regressions. In contrast, we employ the network-based approaches to fulfill the goal. In this regard, our research also provides a new perspective on the effectiveness evaluation of empirical pricing factors.

The article proceeds as follows: **Section 2** and **Section 3** introduce the method and data employed in the study; **Section 4** presents empirical findings; **Section 5** introduces the results of robustness check; and **Section 6** concludes.

2 METHODOLOGY

2.1 Decomposition of Asset Return

Our interest is on the detailed co-movement structures for asset prices. We examine it by focusing on the inter- and intra-structures of the co-movement among raw, systematic, and idiosyncratic asset returns. Therefore, our first step is to conduct the decomposition of raw asset returns. According to the theoretical backgrounds of the macro-economy and empirical asset pricing, asset return can be explained by several common risk factors regarding fundamentals, plus the compensation for the idiosyncratic shocks whose expected value is zero [29]. In light of this, we can base on a specific popular pricing model to do the return decomposition [21, 30]. We resort to one of the most popular pricing models, that is, Fama-French five-factor model (FF5, hereafter) for the Chinese stock market [31]. More importantly, five factors from FF5 can function more favorably to describe the systematic risks and deliver better explanation power to the performance regarding individual assets and asset portfolios than Fama-French three-factor [32] and CAPM models [33–35] for China [24, 25]. Specifically, we regress the return time-series for each of the 28 sector indices against the Fama-French five factors. In this way, the idiosyncratic return for each sector indices can be captured by the residuals plus the intercept.[3] The process mentioned previously can be described as follows:

$$ER_i = \alpha_i + \beta_{\mathrm{MKT},i}R_{\mathrm{MKT}} + \beta_{\mathrm{SMB},i}R_{\mathrm{SMB}} + \beta_{\mathrm{HML},i}R_{\mathrm{HML}} + \beta_{\mathrm{RMW},i}R_{\mathrm{RMW}} + \beta_{\mathrm{CMA},i}R_{\mathrm{CMA}} + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right), \tag{1}$$

where ER_t denotes daily log return for stock index i, and R_{MKT} denotes value-weighted market portfolio returns, both in excess of the risk-free rate; R_{SMB}, R_{HML}, R_{RMW}, and R_{CMA} constructed following Refs. [31, 36, 37] represent size, value, profitability, and investment factors, respectively. ε_i denotes the idiosyncratic return of asset i. Accordingly, the raw asset return can be split into systematic component that is captured by $\beta_{\mathrm{MKT},i}R_{\mathrm{MKT}} + \beta_{\mathrm{SMB},i}R_{\mathrm{SMB}} + \beta_{\mathrm{HML},i}R_{\mathrm{HML}} + \beta_{\mathrm{RMW},i}R_{\mathrm{RMW}} + \beta_{\mathrm{CMA},i}R_{\mathrm{CMA}}$, and the idiosyncratic return that is captured by $\alpha_i + \varepsilon_i$.

2.2 Measurement of Co-Movement

As alternatives, we gauge the co-movement between each pair of return series with rank correlation, namely, Spearman's ρ and Kendall's τ, which are able to describe the non-linear relationship among various components based on return decomposition.[4]

Spearman's ρ is equivalent to Pearson's linear correlation applied to the rankings of each return series. Assume two return series $A = \{A_i\}_1^n$ and $B = \{B_i\}_1^n$, if all the ranks are distinct; then the equation could be simplified to

$$\rho_{A,B}^{Spearman} = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \tag{2}$$

where $d_i = rg(A_i) - rg(B_i)$ is the difference between the two ranks of each observation in A and B, and n is the length of each series [38].

Kendall's τ coefficient can be defined as

$$\tau_{A,B} = \frac{2K}{n(n-1)}, \tag{3}$$

where $K = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} f(A_i, A_j, B_i, B_j)$, and

$$f\left(A_i, A_j, B_i, B_j\right) = \begin{cases} 1, & if \ \left(A_i - A_j\right)\left(B_i - B_j\right) > 0 \\ 0, & if \ \left(A_i - A_j\right)\left(B_i - B_j\right) = 0 \\ -1, & if \ \left(A_i - A_j\right)\left(B_i - B_j\right) < 0 \end{cases}, \tag{4}$$

According to **Eq. 4**, the Kendall correlation between two variables will be high when observations have similar ranks and low when observations have dissimilar ranks. Specifically, the value of τ

[2]Similarly, results from Ref. [14] suggest that large fluctuations of measures of co-movement structures coincide with the burst of the dot-com bubble, while they remain relatively stable for other periods with major risk events (from Figure 4 through Figure 5 in Ref. [14]). We argue that this may be, to a large extent, due to the unweighted scheme adopted by the authors.

[3]The efficacy of the Fama-French five-factor models for the Chinese stock market is not our focus in this work. We consider pricing factors possibly regarding fundamentals in the study to serve more favorably for our purpose of return decomposition.

[4]It is noted that although we report main results based on rank correlation, our findings still hold and are even more significant when Pearson's ρ is adopted.

ranges from −1 to +1. A value of −1 indicates that one's ranking is the reverse of the other, while a value of +1 indicates that the two rankings are the same. A value of 0 indicates no relationship between them [39].

Both of the aforementioned correlation coefficients measure monotonicity relationships. Strictly speaking, the Kendall correlation is preferred over the Spearman correlation because of a smaller gross error sensitivity and a smaller asymptotic variance [40], which nevertheless makes no difference in our study. Unless noted otherwise, Spearman's ρ is adopted as the measurement of correlation, and we leave out the superscript *Spearman* in the left hand of **Eq. 2** in the remainder of the article.

In light of Ref. [14], we introduce a special correlation structure for the purpose of this work. In particular, we set the correlation coefficient between raw and idiosyncratic returns as one. In other words, raw return Y and its idiosyncratic component I is perfectly correlated, and together, they can be further viewed as a new type of node labeled as Z. The correlation coefficient of special nodes pair $Z \sim Z$ is, in fact, the correlation coefficient of their idiosyncratic components $I \sim I$, and the correlation coefficient of special node and factor premium $Z \sim F$ equals that of its raw return and factor premium $Y \sim F$. Assuming there are a total of N assets and M risk factors, that is, the size of the network is $N + M$, we thus obtain the correlation matrix C as follows:

$$C = \begin{pmatrix} \rho_{I_1,I_1} & \cdots & \rho_{I_1,I_N} & \rho_{Y_1,F_1} & \cdots & \rho_{Y_1,F_M} \\ \cdots & \ddots & \cdots & \cdots & \cdots & \cdots \\ \rho_{I_N,I_1} & \cdots & \rho_{I_N,I_N} & \rho_{Y_N,F_1} & \cdots & \rho_{Y_N,F_M} \\ \rho_{F_1,Y_1} & \cdots & \rho_{F_1,Y_N} & \rho_{F_1,F_1} & \cdots & \rho_{F_1,F_M} \\ \cdots & \cdots & \cdots & \cdots & \ddots & \cdots \\ \rho_{F_M,Y_1} & \cdots & \rho_{F_M,Y_N} & \rho_{F_M,F_1} & \cdots & \rho_{F_M,F_M} \end{pmatrix}, \tag{5}$$

2.3 Construction of the Network

We can construct an undirected network as follows: First, correlation matrix C formed according to **Eq. 5** needs to be transformed to a distance matrix before constructing the network. We follow Ref. [41] to construct the distance measurement for nodes pair (A, B) in the network:

$$D_{A,B} = \sqrt{1 - |\rho_{A,B}|}, \tag{6}$$

which is also proven to satisfy the properties of distance norm: 1) $D^{(.)}_{A,B} = 0$ if and only if $A = B$, 2) $D^{(.)}_{A,B} = D^{(.)}_{B,A}$, and 3) $D^{(.)}_{A,B} <= D^{(.)}_{A,K} + D^{(.)}_{K,B}$.[5]

Second, distance matrix D constructed previously still contains rich information regarding co-movement structures. It can map to a complete graph, in which each pair of vertices is connected by an edge. However, our purpose is to conduct the study based on a more concise structure containing the most important links or edges. Accordingly, the MST is employed to filter the original network. We follow the process similar to that under Kruskal's algorithm [42]. Specifically, the lower-diagonal elements of C are sorted in order. We then rely on the sorted values to screen out the most significant links, so as to construct an MST. The links with lower values are preferred, and all nodes are connected sequentially through the shortest distance. As a result, $N + M - 1$ links are retained, through which all of the $N + M$ nodes are connected in the network, compared to a total of $\frac{(N+M)(N+M-1)}{2}$ edges in its complete graph.

2.4 Topological Properties of the Network

We investigate both static and dynamic co-movement structures by examining the network's topological properties. In order to capture the detailed structure of the co-movement, the properties are required to characterize the tightness as well as the shape of the network structure. It is stressed that the results of all metrics based on both unweighted and distance-weighted (see **Eq. 6**) networks are presented in our work, and the comparison of the results would be conducive to detecting the effectiveness of pricing factors.

For the tightness of the entire network, we consider node distance as a proxy, $dis_{A,B}$, for node pair (A, B) [43–45], which acts as an intuitive but favorable measurement for the level of co-movement.[6] The unweighted distance between A and B is defined as the shortest path length between them. According to **Eq. 6**, the weighted distance between A and B can be defined as the length of the shortest D-weighted path between them. It is noted that the weighted node distance functions better in capturing the dynamics of network tightness, given that the unweighted node distance may remain unchanged over time. To gauge the tightness of the network more accurately, we employ the indicator in terms of nodes' distance, which is similar with network *efficiency* [46]:

$$NT = \frac{\sum_{A \neq B} \frac{1}{dis_{A,B}}}{(N + M)(N + M - 1)}, \tag{7}$$

where $dis_{A,B}$ represents the unweighted or weighted distance between A and B, and $N + M$ denotes the network size. Specifically, the network structure is tight (loose), when NT is larger (smaller), thus indicating the higher (lower) level of co-movement.

In addition, we attempt to devote more attention to the sub-networks, or detailed structures of co-movement, including the inter-structure of co-movement across raw return and factor premium, and the intra-structures for idiosyncratic returns and factor premiums. As for the sub-network based on given node types, NT fails to measure the network tightness, since the nodes may be connected by other nodes outside the sub-network

[5]It is noted that conventional method of mapping the correlation to the distance relies on $D_{A,B} = \sqrt{2(1 - \rho_{A,B})}$ [15, 16]. In this vein, positive (negative) correlation gives rise to the short (long) distance. However, this is not what is required in our work. High magnitudes of pairwise correlations, irrespective of their directions, would imply the close relationship between the assets, on which the high level of co-movement is based. Thus, **Eq. 6** serves the purposes of our work better.

[6]Thanks for the comment offered by one of the anonymous reviewers, and we notice that tightness of the network in our study is similar to the conception of network robustness [46]. Despite this, our work concentrates on evaluating network connectivity for assets and pricing factors. Therefore, we adopt "tightness" rather than "robustness" to highlight this focus.

of our interest. We thus consider the average link size of a given sub-network as an alternative. It is defined by the following:

$$\bar{l}_S = \frac{\sum_{e\in S} l_e}{\sum_{e\in G} l_e}. \tag{8}$$

where l_e represents the length of link e in the network. S refers to the sub-network of our interest, while the entire network is denoted as G. It is noted that the change of link size for the entire network is controlled for in **Eq. 8**. By doing so, we can obtain more accurate results for the co-movement structure.[7] In the case of an unweighted network, the length of link e equals one, if it exists in network S, and zero otherwise. In this regard, $\sum_{e\in S} l_e$ is given by the total link size for all the links of network S, and $\sum_{e\in G} l_e$ equals $N + M - 1$. In this vein, a larger (smaller) value of $\bar{l}_S$ implies the higher (lower) level of tightness, and thereby the lower (higher) level of co-movement for a given network S. In the case of a weighted network, l_e refers to distance measurement d depicted by **Eq. 6**. Thus, $\sum_{e\in S} l_e$ and $\sum_{e\in G} l_e$ equal the sum of d in the network S and G, respectively. When the absolute link size is fixed, a larger (smaller) value of $\bar{l}_S$ implies the lower (higher) level of the co-movement across the returns in given network S.

As for the shape of the network structure, we consider the node degree, node strength, and their variants as the major measurements. Node degree is defined as the number of links connecting a given node to the rest of the network. Its weighted version, or node strength, is defined as the sum of weights (or distances) of links connected to the node. Here we focus on the maximum and the standard deviation of the node degree (strength). The maximum degree of the network is defined as the degree of the node with the greatest number of edges incident to it. We define the maximum strength of the network as the strength of the node with the greatest weight of edges incident to it. Intuitively, the evolution of maximum degree (strength) indicates a time-varying network shape. The standard deviation of node degree (strength) measures the amount of dispersion of nodes' degree (strength), which plays a more critical role in characterizing the shape of the network. Specifically, a larger (smaller) standard deviation of degree (strength) implies that the shape of the network is more inclined to be the star-like (chain-like) connection.

3 DATA

Obviously, individual stocks within the same industry or sector, or more broadly, some style of "portfolio," share the co-movement to a large extent [2]. Therefore, we adopt the sector indices as our data sample, instead of individual stocks. By doing this, the industry effects are controlled for in our work so that a more precise co-movement structure can be further evaluated. Specifically, our work is conducted based on 28 sector indices, issued by Shenyin and Wanguo Securities Co., Ltd (http://www.swsresearch.com), and the period spans from 5 January, 2000 to 29 March, 2019. With the closing price of sector indices on a daily basis, we can further obtain their daily log returns, that is, $r_t = \log(p_t) - \log(p_{t-1})$. Summary statistics associated with daily log returns for different indices are reported in Panel A of **Table 1**. It is noted that price limit trading rules implemented currently in China became effective since December 1996, requiring that the maximum daily price fluctuation in terms of the last closing price is ±10% for risk stocks, as highlighted by the results in the columns of Max and Min. It is also evident that each sector produces a time-series average positive profit, and the most outperformed is F&D with the return of 0.05% per day, or around 12.6% ($\approx 0.05\% \times 252$) per year. The standard deviation (std) of daily returns ranges from 1.7% to 2.3%; thus, the annualized standard deviation ranges from about 27% to 37%, which is remarkably higher than those of major assets around the world.[8] In addition to these, we can observe that a majority of sector returns are negatively skewed, indicating the relatively higher possibility of a flash crash. The kurtosis is beyond five for all sector indices, accompanied with the evidence of negatively skewed distribution, which suggests that the sector indices are not normally distributed. Accordingly, this can be further confirmed by the significant statistics of Jarbe–Barque tests at the 1% level.

Additionally, for the decomposition of the asset return, we also retrieve a risk-free interest rate as well as Fama-French five factors for China on a daily basis from the China Stock Market and Accounting Research (CSMAR) database, which is the comprehensive database for Chinese business research, covering data on the Chinese stock market.[9] Specifically, the risk-free interest rate is proxied by the 3 month fixed deposit benchmark interest rate published by the China Central Bank. Pricing factors are constructed strictly following Ref. [31]. Summary statistics associated with Fama-French pricing factors are also presented in Panel B of **Table 1**. Apparently, the market factor, *MKT*, outperforms other four factors and even 28 sector indices, with an annual return of 15% ($\approx 0.06\% \times 252$). *MKT* has larger magnitudes of maximum (11%) and minimum (9%), thus implying the higher volatility, which can also be confirmed by its daily standard deviation of 1.7%, or around 27% per year, at least twice than those for other four factors. It is noted that although *MKT* is the most volatile among the factors, it is still located within the lower range of the standard deviation for sector indices. Accompanied with its highest profits, *MKT* produces the highest Sharpe ratio and acts as

[7]We argue that it is necessary to consider the weight changes for the edges in the network. For the inter-structure for raw return and factor premium, we observe a significant increase in the weighted link size, but a decrease in the unweighted link size. This is mainly due to the increase in distance measurement for the entire network. Unsurprisingly, the results are consistent after allowing for weight changes.

[8]For the recent decade, global commodities have achieved the highest annualized standard deviation of about 17%, which is obviously lower than those for sector indices in China. In the Chinese market, the majority of investors are retail investors, causing larger irrational and speculative behaviors and thereby higher volatility [22]. For more information regarding the historical volatility of worldwide assets, refer to https://advisor.visualcapitalist.com/asset-class-risk-and-return/.

[9]As for the detailed construction of Fama-French five factors, we refer interested readers to Ref. [31] as well as the documents by the CSMAR (https://www.gtadata.com/), which is not the focus in our study.

TABLE 1 | Descriptive statistics of log returns for 28 sector indices and the premiums for Fama-French five factors.

Full name	Abbr	Code	Max	Min	Mean	Std	Skew	Kurt	JB-Stats
Panel A: Sector indices									
Agriculture and forestry	A&F	801010	0.09	–0.10	0.0002	0.019	–0.48	6.25	2,232**
Automobile	Auto	801880	0.09	–0.10	0.0003	0.019	–0.45	6.30	2,277**
Bank	Bank	801780	0.10	–0.11	0.0002	0.019	0.19	7.21	3,474**
Biotechnology	Biotech	801150	0.09	–0.09	0.0004	0.018	–0.46	6.38	2,388**
Building and decoration	B&D	801720	0.09	–0.10	0.0001	0.018	–0.33	6.53	2,508**
Building materials	BM	801710	0.09	–0.10	0.0003	0.020	–0.48	6.06	1998**
Chemical	Chem	801030	0.09	–0.09	0.0002	0.018	–0.52	6.36	2,405**
Commercial trade	ComT	801200	0.09	–0.10	0.0002	0.018	–0.55	6.49	2,604**
Communications	Comm	801770	0.10	–0.10	0.0001	0.020	–0.23	6.03	1826**
Composite	Cps	801230	0.09	–0.10	0.0001	0.019	–0.66	6.04	2,134**
Computer	Cpt	801750	0.10	–0.10	0.0003	0.021	–0.34	5.28	1,098**
Electrical equipment	ElecE	801730	0.09	–0.09	0.0003	0.019	–0.43	5.96	1844**
Electronic	Elec	801080	0.09	–0.10	0.0001	0.021	–0.53	5.43	1,373**
Food and drink	F&D	801120	0.09	–0.09	0.0005	0.017	–0.15	6.24	2058**
Household appliances	Happ	801110	0.09	–0.09	0.0003	0.019	–0.23	5.91	1,687**
Leisure and services	L&S	801210	0.09	–0.10	0.0003	0.020	–0.43	6.12	2041**
Light manufacturing	LMF	801140	0.09	–0.10	0.0001	0.018	–0.68	6.94	3,389**
Mechanical equipment	MechE	801890	0.09	–0.09	0.0003	0.019	–0.53	6.34	2,388**
Media	Media	801760	0.10	–0.11	0.0002	0.021	–0.36	5.35	1,177**
Mining	Mining	801020	0.10	–0.10	0.0002	0.021	–0.13	5.90	1,651**
National defense	ND	801740	0.10	–0.10	0.0002	0.022	–0.34	5.89	1718**
Non-bank financial	NBF	801790	0.10	–0.10	0.0003	0.023	0.05	5.59	1,310**
Non-ferrous metals	NFMet	801050	0.09	–0.10	0.0002	0.022	–0.33	5.64	1,443**
Real estate	Rest	801180	0.09	–0.10	0.0003	0.020	–0.37	5.90	1744**
Steel	Steel	801040	0.09	–0.10	0.0001	0.019	–0.30	6.35	2,249**
Textile and apparel	T&A	801130	0.09	–0.10	0.0001	0.019	–0.70	7.02	3,532**
Transportation	Trans	801170	0.10	–0.10	0.0001	0.017	–0.48	7.34	3,846**
Utilities	Util	801160	0.09	–0.10	0.0001	0.017	–0.53	7.30	3,804**
Panel B: Fama-French factors									
Market factor	MKT		0.11	-0.09	0.0006	0.017	-0.19	8.01	4,912**
Size factor	SMB		0.05	-0.07	0.0004	0.008	-1.01	9.89	10,009**
Value factor	HML		0.05	-0.04	0.0001	0.006	0.67	9.54	8,659**
Profitability factor	RMW		0.05	-0.03	0.0000	0.006	0.80	9.00	7,495**
Investment factor	CMA		0.03	-0.03	0.0000	0.004	-0.27	6.23	2081**

*This table presents the full name, the abbreviation of the name (Abbr.), as well as the code (only for sector index), maximum (Max), minimum (Min), mean (Mean), standard deviation (Std), skewness (Skew), kurtosis (Kurt), and Jarque–Bera test statistic (JB-stats) of daily log return for each index (Panel A) and pricing factor (Panel B). The superscripts * and ** denote the statistical significance at 5% and 1%, respectively.*

the best "asset" within the framework of portfolio theory [47]. Except for *MKT*, the maximum and minimum magnitudes for the other four factors are lower than 5% and 7% per day, respectively. Relatively speaking, *SMB* and *HML* are more profitable than the others, with the annual return of 10% ($\approx 0.04\% \times 252$) and 2.5% ($\approx 0.01\% \times 252$), respectively, while *RMW* and *CMA* produce the profits approaching zero per day. *SMB* and *HML* are more volatile, with the annual standard deviation of 12.7% and 9.5%, respectively, than 6.3% for *CMA*. Similarly, five factors are not normally distributed, as suggested by the results of their skewness, kurtosis, and the statistics under the Jarbe–Barque test.

4 EMPIRICAL RESULTS

4.1 Network Connectedness Over the Whole Sample Period

To begin with, we examine the results over the whole sample period. With a daily log return for 28 stock indices, we first retrieve their idiosyncratic returns by performing the regression of their raw returns on Fama-French five-factor premiums, as described by **Eq. 1**. Factor premium, raw return, and its corresponding idiosyncratic return are employed to calculate the correlation matrix C in terms of Spearman's ρ. With the adoption of the distance measurement in **Eq. 6**, C is transformed to D, and its MST can further be obtained, as presented in **Figure 1**. The network comprises 33 nodes with 32 most significant linkages connecting them. The white circles in the figure represent different sector indices, while the gray ones represent pricing factors. The width of edges between nodes denotes the reciprocal of the distance between them. The edges in bold imply that the shorter the distance the closer the relationship between the connecting nodes.

It is evident that all sector indices are connected to factor *MKT*, as shown in the plot of the MST, which suggests that sector indices exhibit a high degree of co-movement with market factor *MKT* compared to other risk factors. Apparently, market factor *MKT* has the maximum degree, while the other four factors are

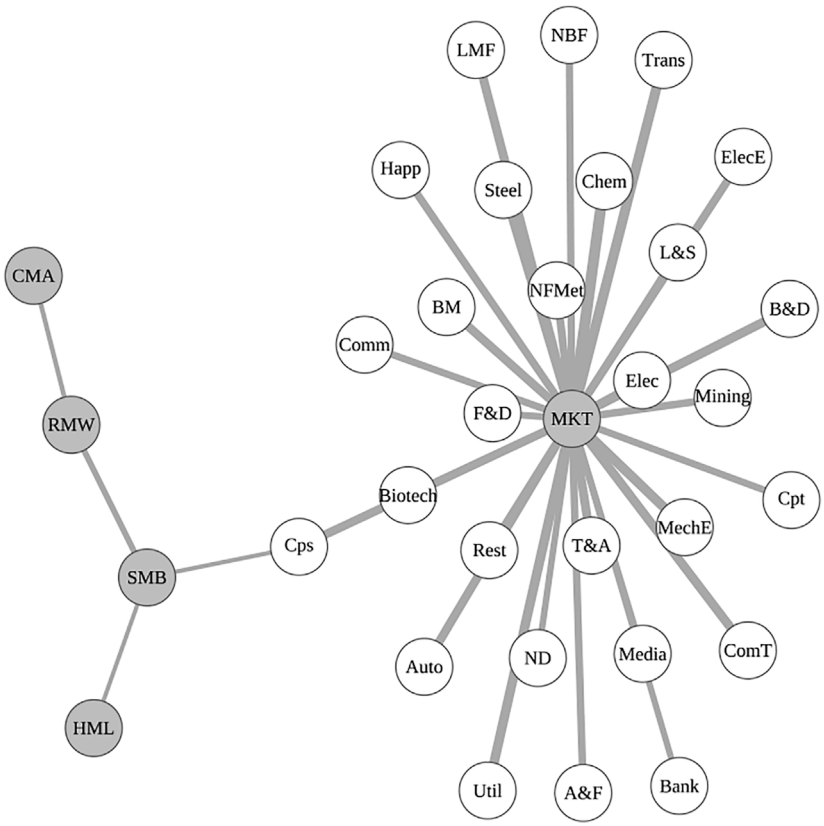


FIGURE 1 | MST for the whole sample period from 1 July 2000 to 29 March 2019. White circles denote 28 sector indices, while gray circles denote five pricing factors.

connected with other sector nodes through sector *Cps*. Specifically, size factor *SMB* is directly linked to sector *Cps*, while the other three factors are connected through *SMB*. In comparison, investment factor *CMA* lies in the most periphery of the MST structure.

We also notice that it is more distant between pricing factors than that between *MKT* and sector indices. That is, the links connecting *MKT* with sector indices are more preferred during the construction of the MST. In some sense, the results aforementioned also unveil the relative strength of the pricing effectiveness of risk factors in the Chinese stock market, which are in accordance with empirical findings from the literature on asset pricing. As documented in previous research works, Fama-French five factors in emerging markets behave differently from other mature markets. It is widely acknowledged that market factor *MKT* is not enough to capture the common risks associated with assets, and thus, "factor zoo" has been developed to fulfill this end [23]. Furthermore, new pricing factors that have been proven in mature markets are applied to emerging markets, including China. As for the Chinese market, size factor *SMB* and profitability factor *RMW* are found to function well [24], while investment factor *CMA* seems redundant [24, 25], and the results regarding value factor *HML* are mixed [24, 48, 49], which is consistent with our findings in **Figure 1**.

4.2 Dynamics of Network Connectedness

In order to probe into the dynamics for the co-movement among raw returns, factor premiums, and their idiosyncratic returns, rolling or moving window analysis is conducted [50–53]. It is noted that the rolling window analysis has been frequently adopted in the literature. Researchers record the behavior regarding the variable of interest in each moving window, thereby obtaining its dynamics over time. In the fields of finance, the authors of Ref. [54] put forward the famous Fama-MacBeth regression to process the panel data by evaluating the time-series estimates of slopes based on cross-sectional regression in each moving window, which has been widely employed in financial studies; recent research works include Refs. [55–58] among others. The authors of Ref. [59] propose variance decomposition based on the VAR model to detect the spillover effect of variables of interest, whose dynamics are also obtained by re-estimating the model for each moving window [60, 61]. There is a body of literature proposing the econophysics-based methods to process financial data, with the adoption of the rolling window analysis as well [62]. It is noted that the rolling window analysis has also been applied to other fields, including particle-in-cell simulation [63], recurrence plot-based complexity measurement [64], mutual information estimation [65], and the study of the dynamical response of a population [66].

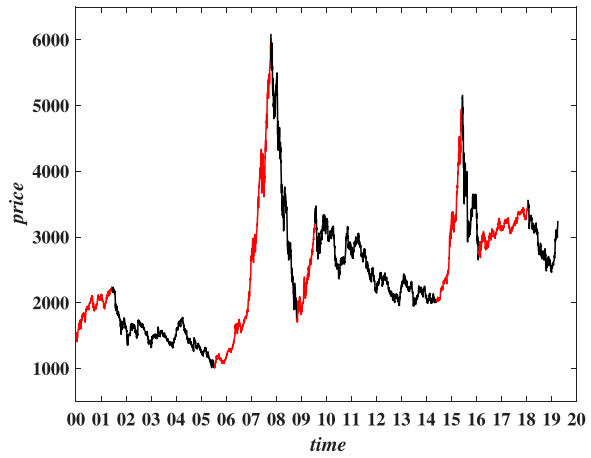


FIGURE 2 | (Color line) Rough partition of market cycles in terms of the daily closing price of the SHCI.

As for our work, the most intuitive and efficient way is to evaluate the properties associated with the MST in each moving window and further examine their dynamic behaviors. The observations in each time window should be sufficient to guarantee a precise estimation. Following Ref. [14], the sizes of the rolling window and the step are set as 120 trading days and 1 day, respectively. As such, we follow the process similar to what is illustrated in the previous subsection, except that at each trading day t during the sample period, we construct the MST based on the data derived from the decomposition of the index return over the past period from $t - 1$ to $t - 120$. To capture the changing structure of co-movement, we examine the evolution of the topological properties of the MSTs.

We also wonder whether the topological characteristics of the network are related to market performance. To this end, we perform a rough division in terms of market performance proxied by the Shanghai Component Index (SHCI, hereafter). As shown in **Figure 2**, market trends over the period from January 2000 to March 2019 are roughly divided into five market cycles, which are typically composed of the phases of accumulation, mark-up, distribution, and mark-down. In the figure, the red curve represents the rising phase of the market, while the black curve represents the downward or adjusting phase. We focus on the network structures over three time periods corresponding to the three most important market cycles, namely, the periods around 2007, 2015, and 2018.

The results are presented in **Figures 3**, **4**. On the whole, our results unveil the strong correlation between the evolution of topological properties of the network structure and market cycles, regardless of which topological properties are considered.

According to **Eq. 7**, *NT* plays an important role in characterizing the level of network tightness. In the top panel of **Figure 3**, we can readily observe that both unweighted and weighted results vary over time and are closely related to the major market cycles, represented by the shaded parts in the plot. *NT* experiences a sharp decrease during all three major periods, thus indicating the lower level of co-movement. It is more pronounced in the unweighted case, since *NT* remains relatively stable with slight fluctuations most of the time, except for the three periods of our interest. It is also worth noting that the evolution of the average distance is not exactly coherent with the market trend. The Chinese market has witnessed the most significant bubble formation and bursting during the period around 2007, while the market cycle around 2018 is relatively insignificant. However, **Figure 3** presents a seemingly more pronounced increase of the average distance realized for the period around 2018, rather than 2007. This finding can also be observed in the following results with respect to other properties. In addition, we also measure the network tightness by the link size for the entire network, as described by $\sum_{e\in G} l_e$ according to **Eq. 8**. Here we focus on the results for the weighted case, as shown in the bottom panel of **Figure 3**. It is noted that the unweighted link size for the entire network remains the value of 32 all the time, while the result for the weighted case is obviously time-varying. Also, consistent with findings based upon the results of *NT*,

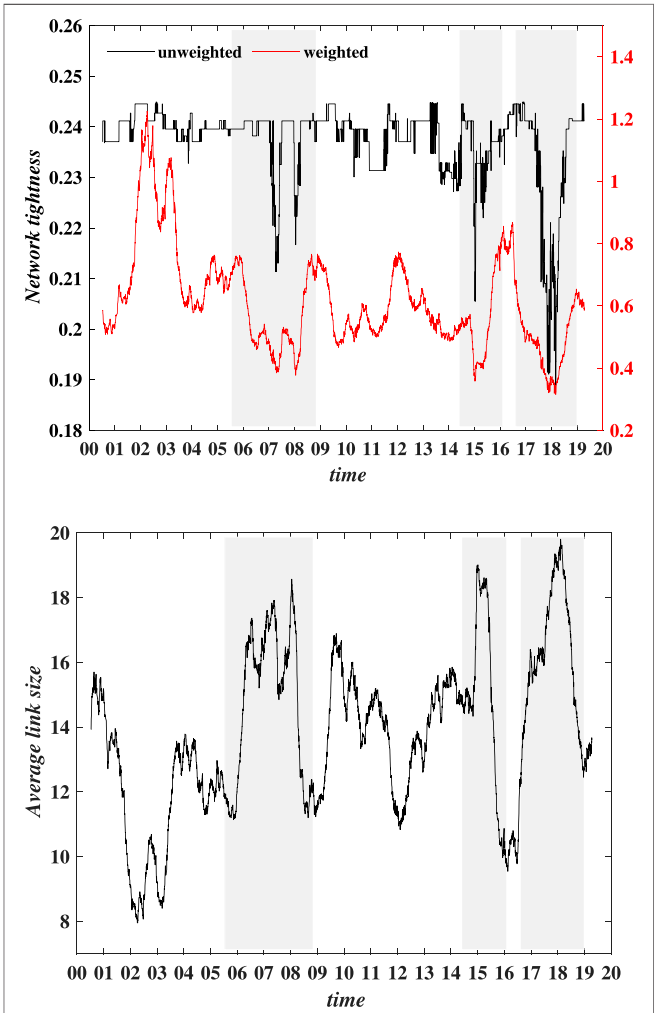


FIGURE 3 | (Color line) Evolution of network tightness indicator *NT* and weighted link size for the entire network, presented in the top and bottom panels, respectively. The sample period is from 1 July 2001 to 29 March 2019. The shaded parts in each plot correspond to the three major market cycles.

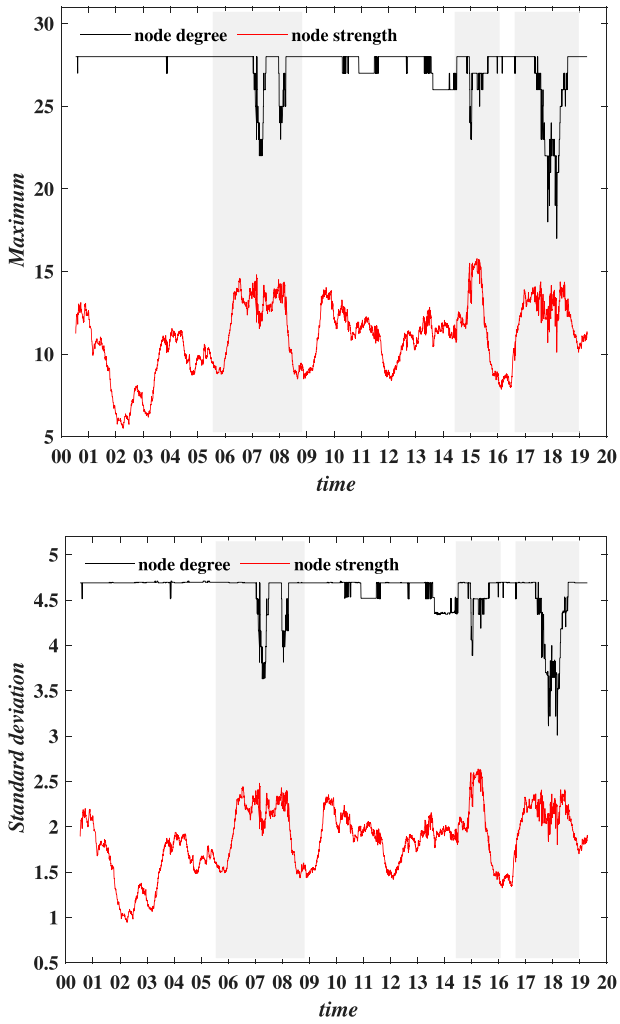


FIGURE 4 | (Color line) Evolution of maximum and standard deviation of node degree (strength), presented in the top and bottom panels, respectively. The sample period is from 1 July 2001 to 29 March 2019. The shaded parts in each plot correspond to the three major market cycles.

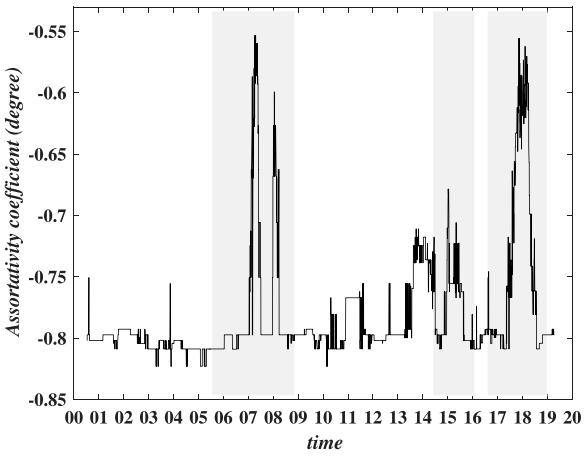


FIGURE 5 | Evolution of assortativity coefficient in terms of node strength. The sample period is from 1 July 2001 to 29 March 2019. The shaded parts in each plot correspond to the three major market cycles.

link size exhibits a significant increase during three major periods, indicating the lower level of co-movement.

On the other hand, the node degree (strength) of the network can be employed to depict the shape of the network. The results for maximum and standard deviation of the node degree (strength) are presented in the top and bottom panels of **Figure 4**, respectively. The maximum degree (strength) shows the importance of the most connected node in the network, while the standard deviation of the node degree (strength) denotes the dispersion of node degree (strength) across the network. As shown in the figure, node strength is more volatile over time, while node degree remains stable, with the value being around 28 at most times.[10] This implies that the degree of the remaining nodes fluctuates within a small range; accompanied with the results in the previous section, market factor *MKT* is supposed to be the hub node over the rolling time windows, which needs further validation. In addition, we can observe that maximum degree experiences a sharp decrease in all three market cycles, which indicates that some edges of the hub node are disjointed. Meanwhile, a pronounced increase in node strength suggests that the level of co-movement between the hub node and its connected nodes is weakening. As for the standard deviation of degree (strength), the results are basically the same as that for the maximum degree (strength). The results above imply a star-like connection in the network structure, which is also time-varying, especially over the period corresponding to major market cycles.

In order to provide more evidence about the existence of star-like connection, we further resort to the assortativity coefficient [14], and the results are shown in **Figure 5**. Obviously, the assortativity coefficient is changing over time and exhibits more significant fluctuations in three major cycles. It is noted that all results over the rolling windows are negative, which confirms that nodes with an extremely high degree tend to connect with nodes with very low degrees. This is also consistent with our previous findings as well as those from the prior literature [14].

It is noted that our results are similar to those from Ref. [14] that adopts a network-based approach to examine the asset returns' co-movement in the US market. Based on the Fama-French three-factor model, the authors decompose the returns of industrial sector indices into systematic and idiosyncratic parts. Their results also suggest the time-varying co-movement structures over the period from 1970 to 2015 (see Figures 2, 3 from Ref. [14]). However, they observe large fluctuations of measures for co-movement structures over the period around the dot-com bubble only. It is somewhat out of surprise that no large fluctuations are observed for the period around 2008, corresponding to the Global Financial Crisis (GFC). This is

[10]It is noted that the theoretical maximum is 32, which is the number of edges in the network.

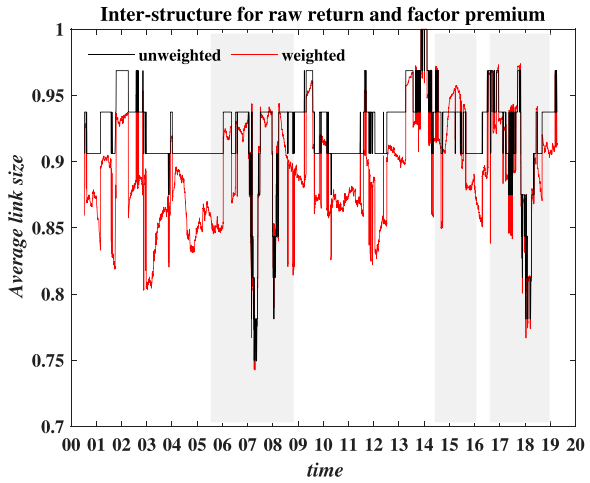


FIGURE 6 | (Color line) Evolution of the average link size of intra-structure for idiosyncratic returns of sector indices. The sample period is from 1 July 2001 to 29 March 2019. The shaded parts in each plot correspond to the three major market cycles.

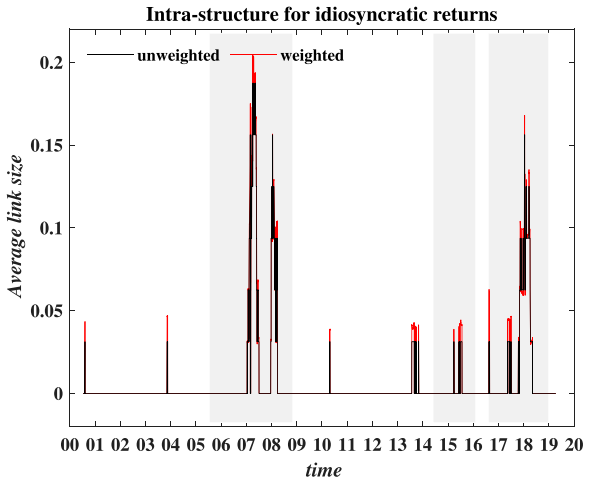


FIGURE 7 | (Color line) Evolution of the average link size of intra-structure for the idiosyncratic returns of sector indices. The sample period is from 1 July 2001 to 29 March 2019. The shaded parts in each plot correspond to the three major market cycles.

partly due to their adoption of unweighted schemes, which may be insensitive to small changes in co-movement structures. Our results in the following text would further highlight this.

In addition, to further understand the MST structures in the three market cycles of our interest, we plot the MSTs for three moments that are randomly chosen out in each of the periods for three market cycles. The results are shown in **Supplementary Figure S1** through **Supplementary Figure S3**. It is evident that compared with the results for the whole sample period in **Figure 1**, **Supplementary Figure S1** presents a perfect star-like network, though some sector indices are disjointed from market factor *MKT*, and instead, they are connected to their counterparts. A similar finding is obtained by Ref. [14]. The authors also observe that *MKT* acts as a hub node that connects the majority of idiosyncratic returns most of the time, implying the critical role in pricing the assets, while some of the industrial sectors would be disjointed from *MKT* and connect with themselves through the risky periods (see Figure 6 from Ref. [14]).

The structure for pricing factors is also slightly different. Among them, value factor *HML* is connected with sector *Steel*, through which it is connected with market factor *MKT*. The other three factors are connected in sequence and connected with market factor *MKT* through sector *Bank*. In comparison, in the plots for 2015 and 2018, as shown in **Supplementary Figures S2,3**, respectively, the MST structures are remarkably different from what is suggested in **Figure 1**, and the nodes of sector indices are more inclined to connect with their counterparts and also with other pricing factors. Additionally, we observe that value factor *HML* and profitability factor *RMW* take up more important positions in the MST, and size factor *SMB* acts as the bridge for other factors. However, investment factor *CMA* seems less important than the other factors in the plots.

4.3 Detailed Structure of Co-Movement

With the rolling window analysis, we further attempt to acquire more information regarding the detailed structure for different node pairs. In doing so, we divide the entire network structure into three parts, that is, intra-structures for both factor premiums and idiosyncratic returns, and the inter-structure for raw return and factor premium. We will evaluate the characteristics with respect to each sub-network. Here, we employ the average link size as the main measurement of the network structure, as described in **Eq. 8**, instead of average distance that only captures the characteristics associated with the entire network structure.

4.3.1 Inter-Structure for Raw Return and Factor Premium

Figure 6 presents the results with respect to the inter-structure of co-movement between raw return and factor premium. Similar to previous results, the unweighted result of the link size is less volatile. Specifically, in the result of the unweighted case, its fluctuation is around 0.95 over most periods, which suggests that about 95% of connected edges in the network exist between the sector indices and pricing factors. In comparison, the weighted results fluctuate more frequently and range between 0.7 and 0.96. As to the results for the three market cycles, we can observe much more pronounced fluctuations in both cases, indicating a sudden change in the MST structure.

It is noted that the results for three major market cycles are mixed. For the periods around 2007 and 2018, both unweighted and weighted values exhibited a sharp decline with wild fluctuation, implying a loose network structure. In comparison, the results over the period around 2015 were different. The unweighted value remains stable, while the weighted value peaks instead of slumping down, suggesting the lower level of co-movement.

4.3.2 Intra-Structure for Idiosyncratic Returns

The results in **Figure 7** reveal important changes in the network structures for major market cycles. The average

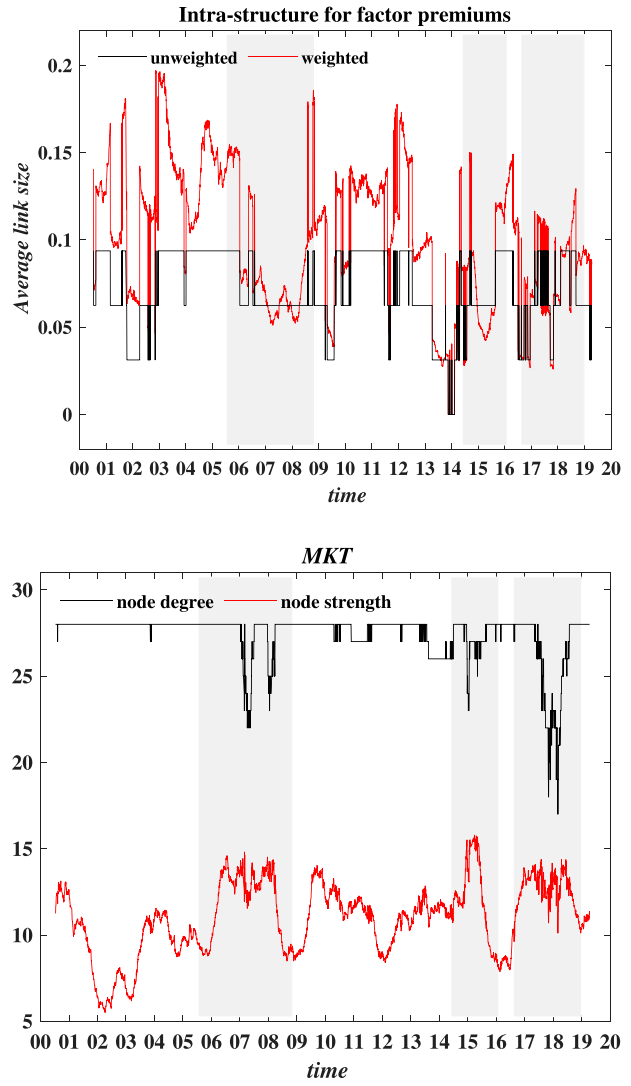


FIGURE 8 | (Color line) Evolution of the average link size of intra-structure for factor premiums and the node degree (strength) of market factor *MKT*, presented in top and bottom plots, respectively. The sample period is from 1 July 2001 to 29 March 2019. The shaded parts in each plot correspond to the three major market cycles.

link size in the figure remains zero most of the time, which indicates no connection exists between idiosyncratic returns in the network. We have already obtained the results for the whole sample period in **Figure 1** that sector indices are all connected through market factor *MKT*. However, in some time windows, such as three major cycles, the connectivity for idiosyncratic returns is significantly enhanced, regardless of whether it is in the weighted or unweighted case. This may be caused by the characteristics embedded in the MST. As the correlation between sector nodes and common risk factors is weakened, they have to connect to other nodes instead. On the other hand, this also implies the possibility of potential pricing factors hiding in idiosyncratic returns, which requires further investigation.

4.3.3 Intra-Structure for Factor Premiums

Finally, we move to the results for intra-structures of co-movement for factor premiums. The results are presented in **Figure 8**. Unweighted results fluctuate slightly, ranging from 0% to 10%, while weighted results fluctuate frequently, but they are all within a range from 0% to 20%, and weighted average link size has a downward trend in fluctuations over time. For three major market cycles, both the weighted and unweighted results exhibit different decrease degrees. In fact, the link size for pricing factors is expected to be as small as possible. This is in line with the intention of asset pricing theory, which is to explore irrelevant but economically meaningful risk factors. As such, the co-movement between factors and assets is expected to be as strong as possible, demonstrating factor pricing effectiveness. Our results provide evidence that Fama-French five factors are not suitable for the Chinese stock market on the ground that some risk factors play their roles only in major market cycles, which is consistent with our previous finding.

We attempt to explore further how Fama-French five factors are connected. We examine the evolution of node degree (strength) for each factor. Here, we report the results on the market factor *MKT* only, whose degree remains at a high level for most time windows. As shown in the bottom panel of **Figure 8**, node degree of *MKT* only fluctuates slightly, with the value being around 28 in the plot. This also provides evidence of a close relationship between the *MKT* and other sector indices. The dynamic changes of node strength over time seem to be irregular. However, for major market cycles, node strength of *MKT* factor would be increasing, which implies the weakening of the relationship with its connected nodes. In comparison, node degree (strength) for the other four factors fluctuates more frequently and exhibited a significant upward trend only in the major cycle around 2018, as shown in **Supplementary Figure S4** through **Supplementary Figure S7**.

5 ROBUSTNESS CHECK

We employ alternative correlation metrics to conduct the robustness check, including Pearson's ρ and Kendall's τ. It is noted that similar results are obtained based on different measurements, and our main findings still hold. In addition, we also examine the sensitivity of the results when the size of the rolling step and window varies. The results also indicate that our findings are not sensitive to the settings associated with the rolling window analysis.[11]

6 DISCUSSION AND SUMMARY

In this work, we probe into the detailed structure of co-movement for the asset price. Specifically, using the data of Chinese sector indices and Fama-French five factors, we perform the return decomposition and construct a minimum spanning tree (MST) in

[11]The results mentioned previously could be provided upon the requests.

terms of the rank correlation coefficients among raw return, idiosyncratic return, and factor premium.

For the whole sample period, the MST presents a star-like connection with the hub node being market factor *MKT*, and all sector nodes are connected to market factor *MKT*. We find that the topological properties associated with the network structure are time-varying in the rolling window analysis. In particular, in major market cycles, the star-like structure would be changing. This is featured by the fact that a few idiosyncratic returns for sector indices would be disjointed from market factor *MKT* and then connected to their counterparts. Also, pricing factors will generally exhibit more favorable effectiveness of pricing assets.

Our results imply the importance of monitoring assets' co-movement in practice. Regulators should be alert to the structural changes in assets' co-movement and take measures to avoid further risk spillovers. Specifically, given that strengthened co-movement of idiosyncratic returns may lead to continued amplification of risk, much more attention should be paid to industrial sectors that are disjointed from crucial risk factors. In addition, our work also provides a new perspective for the research on the co-movement structure for the asset price. The effectiveness of new pricing factors can also be evaluated within this framework. It is also noted that this work's focus is to describe characteristics associated with the time-varying co-movement structure, whose driving forces will be left in our future work.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are subject to the following licenses/restrictions: The data for Chinese stock indices used in this study are openly available. Access to the data can be obtained from SWS research. The data for pricing factors and interest rate are owned by CSMAR, and there are fees associated with the database. Requests to access these datasets should be directed to SWS research: http://www.swsresearch.com/EN/default.aspx; CSMAR: http://www.gtarsc.com/#/index.

AUTHOR CONTRIBUTIONS

HLS: data curation, conceptualization, investigation, methodology, funding acquisition, and project administration; HC: software, validation, visualization, and writing—original draft.

FUNDING

This work was partly supported by Humanities and Social Sciences Fund of Ministry of Education of China (20YJC790113), the Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (2019SJA0156), and the Startup Foundation for Introducing Talent of NUIST.

ACKNOWLEDGMENTS

We would like to acknowledge the reviewers for their constructive comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2022.593493/full#supplementary-material

REFERENCES

1. Malceniece L, Malcenieks K, Putniņš TJ. High Frequency Trading and Comovement in Financial Markets. *J Financial Econ* (2019) 134:381–99. doi:10.1016/j.jfineco.2018.02.015
2. Barberis N, Shleifer A, Wurgler J. Comovement. *J Financial Econ* (2005) 75: 283–317. doi:10.1016/j.jfineco.2004.04.003
3. Morck R, Yeung B, Yu W. The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock price Movements? *J Financial Econ* (2000) 58:215–60. doi:10.1016/S0304-405X(00)00071-4
4. Hameed A, Kang W, Viswanathan S. Stock Market Declines and Liquidity. *J Financ* (2010) 65:257–93. doi:10.1111/j.1540-6261.2009.01529.x
5. Karolyi GA, Lee K-H, Van Dijk MA. Understanding Commonality in Liquidity Around the World. *J Financial Econ* (2012) 105:82–112. doi:10.1016/j.jfineco.2011.12.008
6. Markowitz H. Portfolio Selection. *J Finance* (1952) 7:77–91. doi:10.2307/2975974
7. Kenett DY, Huang X, Vodenska I, Havlin S, Stanley HE. Partial Correlation Analysis: Applications for Financial Markets. *Quantitative Finance* (2015) 15: 569–78. doi:10.1080/14697688.2014.946660
8. Tu C. Cointegration-based Financial Networks Study in Chinese Stock Market. *Physica A: Stat Mech its Appl* (2014) 402:245–54. doi:10.1016/j.physa.2014.01.071
9. Billio M, Getmansky M, Lo AW, Pelizzon L. Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors. *J Financial Econ* (2012) 104:535–59. doi:10.1016/j.jfineco.2011.12.010
10. Curme C, Tumminello M, Mantegna RN, Stanley HE, Kenett DY. Emergence of Statistically Validated Financial Intraday lead-lag Relationships. *Quantitative Finance* (2015) 15:1375–86. doi:10.1080/14697688.2015.1032545
11. Bu H, Tang W, Wu J. Time-varying Comovement and Changes of Comovement Structure in the Chinese Stock Market: A Causal Network Method. *Econ Model* (2019) 81:181–204. doi:10.1016/j.econmod.2019.03.002
12. Liu B-Y, Ji Q, Fan Y. A New Time-Varying Optimal Copula Model Identifying the Dependence across Markets. *Quantitative Finance* (2017) 17:437–53. doi:10.1080/14697688.2016.1205208
13. Guo X, Zhang H, Tian T. Development of Stock Correlation Networks Using Mutual Information and Financial Big Data. *PloS one* (2018) 13:e0195941. doi:10.1371/journal.pone.0195941
14. de Carvalho PJC, Gupta A. A Network Approach to Unravel Asset price Comovement Using Minimal Dependence Structure. *J Banking Finance* (2018) 91:119–32. doi:10.1016/j.jbankfin.2018.04.012
15. Mantegna RN. Hierarchical Structure in Financial Markets. *Eur Phys J B* (1999) 11:193–7. doi:10.1007/s100510050929
16. Bonanno G, Caldarelli G, Lillo F, Mantegna RN. Topology of Correlation-Based Minimal Spanning Trees in Real and Model Markets. *Phys Rev E* (2003) 68:046130. doi:10.1103/PhysRevE.68.046130

17. Oh G. Grouping Characteristics of Industry Sectors in Financial Markets. *Physica A: Stat Mech its Appl* (2014) 395:261–8. doi:10.1016/j.physa.2013.09.031
18. Tumminello M, Aste T, Di Matteo T, Mantegna RN. A Tool for Filtering Information in Complex Systems. *Proc Natl Acad Sci U.S.A* (2005) 102:10421–6. doi:10.1073/pnas.0500298102
19. Onnela J-P, Kaski K, Kertész J. Clustering and Information in Correlation Based Financial Networks. *The Eur Phys J B - Condensed Matter* (2004) 38:353–62. doi:10.1140/epjb/e2004-00128-7
20. Tse CK, Liu J, Lau FCM. A Network Perspective of the Stock Market. *J Empirical Finance* (2010) 17:659–67. doi:10.1016/j.jempfin.2010.04.008
21. Shi H-L, Zhou W-X. Horse Race of Weekly Idiosyncratic Momentum Strategies with Respect to Various Risk Metrics: Evidence from the Chinese Stock Market. *North Am J Econ Finance* (2021) 58:101478. doi:10.1016/j.najef.2021.101478
22. Shi H-L, Jiang Z-Q, Zhou W-X. Profitability of Contrarian Strategies in the Chinese Stock Market. *PLOS ONE* (2015) 10:e0137892. doi:10.1371/journal.pone.0137892
23. Cochrane JH. Presidential Address: Discount Rates. *J Financ* (2011) 66:1047–108. doi:10.1111/j.1540-6261.2011.01671.x
24. Guo B, Zhang W, Zhang Y, Zhang H. The Five-Factor Asset Pricing Model Tests for the Chinese Stock Market. *Pacific-Basin Finance J* (2017) 43:84–106. doi:10.1016/j.pacfin.2017.02.001
25. Lin Q. Noisy Prices and the Fama-French Five-Factor Asset Pricing Model in China. *Emerging Markets Rev* (2017) 31:141–63. doi:10.1016/j.ememar.2017.04.002
26. Huang C, Deng Y, Yang X, Cao J, Yang X. A Network Perspective of Comovement and Structural Change: Evidence from the Chinese Stock Market. *Int Rev Financial Anal* (2021) 76:101782. doi:10.1016/j.irfa.2021.101782
27. Chiang TC, Lao L, Xue Q. Comovements between Chinese and Global Stock Markets: Evidence from Aggregate and Sectoral Data. *Rev Quant Finan Acc* (2016) 47:1003–42. doi:10.1007/s11156-015-0529-x
28. Chen P. Global Oil Prices, Macroeconomic Fundamentals and China's Commodity Sector Comovements. *Energy Policy* (2015) 87:284–94. doi:10.1016/j.enpol.2015.09.024
29. Ross SA. The Arbitrage Theory of Capital Asset Pricing. *J Econ Theor* (1976) 13:341–60. doi:10.1016/0022-0531(76)90046-6
30. Blitz D, Hanauer MX, Vidojevic M. The Idiosyncratic Momentum Anomaly. *Int Rev Econ Finance* (2020) 69:932–57. doi:10.1016/j.iref.2020.05.008
31. Fama EF, French KR. A Five-Factor Asset Pricing Model. *J Financial Econ* (2015) 116:1–22. doi:10.1016/j.jfineco.2014.10.010
32. Fama EF, French KR. Common Risk Factors in the Returns on Stocks and Bonds. *J Financial Econ* (1993) 33:3–56. doi:10.1016/0304-405X(93)90023-5
33. Sharpe WF. Capital Asset Prices: a Theory of Market Equilibrium under Conditions of Risk*. *J Finance* (1964) 19:425–42. doi:10.1111/j.1540-6261.1964.tb02865.x
34. Lintner J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Rev Econ Stat* (1965) 47:13–37. doi:10.2307/1924119
35. Mossin J. Equilibrium in a Capital Asset Market. *Econometrica* (1966) 34:768–83. doi:10.2307/1910098
36. Fama EF, French KR. Dissecting Anomalies with a Five-Factor Model. *Rev Financ Stud* (2016) 29:69–103. doi:10.1093/rfs/hhv043
37. Fama EF, French KR. International Tests of a Five-Factor Asset Pricing Model. *J Financial Econ* (2017) 123:441–63. doi:10.1016/j.jfineco.2016.11.004
38. Spearman C. The Proof and Measurement of Association between Two Things. *Am J Psychol* (1904) 15:72–101. doi:10.2307/1412159
39. Kendall MG. A New Measure of Rank Correlation. *Biometrika* (1938) 30:81–93. doi:10.1093/biomet/30.1-2.81
40. Croux C, Dehon C. Influence Functions of the Spearman and Kendall Correlation Measures. *Stat Methods Appl* (2010) 19:497–515. doi:10.1007/s10260-010-0142-z
41. Zheng Z, Yamasaki K, Tenenbaum JN, Stanley HE. Carbon-dioxide Emissions Trading and Hierarchical Structure in Worldwide Finance and Commodities Markets. *Phys Rev E* (2013) 87:012814. doi:10.1103/PhysRevE.87.012814
42. Kruskal JB. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proc Amer Math Soc* (1956) 7:48–50. doi:10.1090/s0002-9939-1956-0078686-7
43. Newman MEJ. Scientific Collaboration Networks. I. Network Construction and Fundamental Results. *Phys Rev E* (2001) 64:016131. doi:10.1103/PhysRevE.64.016131
44. Newman MEJ. Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality. *Phys Rev E* (2001) 64:016132. doi:10.1103/PhysRevE.64.016132
45. Opsahl T, Agneessens F, Skvoretz J. Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths. *Social Networks* (2010) 32:245–51. doi:10.1016/j.socnet.2010.03.006
46. Bellingeri M, Bevacqua D, Scotognella F, Alfieri R, Nguyen Q, Montepietra D, et al. Link and Node Removal in Real Social Networks: A Review. *Front Phys* (2020) 8:228. doi:10.3389/fphy.2020.00228
47. Markowitz H. Portfolio Selection*. *J Finance* (1952) 7:77–91. doi:10.2307/297597410.1111/j.1540-6261.1952.tb01525.x
48. Hu GX, Chen C, Shao Y, Wang J. Fama-French in China: Size and Value Factors in Chinese Stock Returns. *Int Rev Finance* (2019) 19:3–44. doi:10.1111/irfi.12177
49. Liu J, Stambaugh RF, Yuan Y. Size and Value in China. *J Financial Econ* (2019) 134:48–69. doi:10.1016/j.jfineco.2019.03.008
50. Fenn DJ, Porter MA, Williams S, McDonald M, Johnson NF, Jones NS. Temporal Evolution of Financial-Market Correlations. *Phys Rev E* (2011) 84:026109. doi:10.1103/PhysRevE.84.026109
51. Ross GJ. Dynamic Multifactor Clustering of Financial Networks. *Phys Rev E* (2014) 89:022809. doi:10.1103/PhysRevE.89.022809
52. Shi H-L, Jiang Z-Q, Zhou W-X. Time-varying Return Predictability in the Chinese Stock Market. *Rep Adv Phys Sci* (2017) 01:1740002. doi:10.1142/S2424942417400023
53. Shi H-L, Zhou W-X. Wax and Wane of the Cross-Sectional Momentum and Contrarian Effects: Evidence from the Chinese Stock Markets. *Physica A: Stat Mech its Appl* (2017) 486:397–407. doi:10.1016/j.physa.2017.05.078
54. Fama EF, MacBeth JD. Risk, Return, and Equilibrium: Empirical Tests. *J Polit Economy* (1973) 81:607–36. doi:10.1086/260061
55. Atilgan Y, Bali TG, Demirtas KO, Gunaydin AD. Left-tail Momentum: Underreaction to Bad News, Costly Arbitrage and Equity Returns. *J Financial Econ* (2020) 135:725–53. doi:10.1016/j.jfineco.2019.07.006
56. Hou K, Xue C, Zhang L. Replicating Anomalies. *Rev Financ Stud* (2020) 33:2019–133. doi:10.1093/rfs/hhy131
57. Ali U, Hirshleifer D. Shared Analyst Coverage: Unifying Momentum Spillover Effects. *J Financial Econ* (2020) 136:649–75. doi:10.1016/j.jfineco.2019.10.007
58. Parsons CA, Sabbatucci R, Titman S. Geographic lead-lag Effects. *Rev Financ Stud* (2020) 33:4721–70. doi:10.1093/rfs/hhz145
59. Diebold FX, Yılmaz K. On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms. *J Econom* (2014) 182:119–34. doi:10.1016/j.jeconom.2014.04.012
60. Diebold FX, Yilmaz K. Trans-Atlantic Equity Volatility Connectedness: U.S. And European Financial Institutions, 2004-2014. *J Financial Econom* (2015) 14:nbv021–127. doi:10.1093/jjfinec/nbv021
61. Bostanci G, Yilmaz K. How Connected Is the Global Sovereign Credit Risk Network? *J Banking Finance* (2020) 113:105761. doi:10.1016/j.jbankfin.2020.105761
62. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, Stanley HE. Random Matrix Approach to Cross Correlations in Financial Data. *Phys Rev E* (2002) 65:066126. doi:10.1103/PhysRevE.65.066126
63. Yoffe SR, Lehe R, Ersfeld B, Brunetti E, Vieux G, Noble A, et al. Particle-in-cell Simulation of Plasma-Based Amplification Using a Moving Window. *Phys Rev Res* (2020) 2:013227. doi:10.1103/PhysRevResearch.2.013227
64. Marwan N, Wessel N, Meyerfeldt U, Schirdewan A, Kurths J. Recurrence-plot-based Measures of Complexity and Their Application to Heart-Rate-Variability Data. *Phys Rev E* (2002) 66:026702. doi:10.1103/PhysRevE.66.026702

65. Moon Y-I, Rajagopalan B, Lall U. Estimation of Mutual Information Using Kernel Density Estimators. *Phys Rev E* (1995) 52:2318–21. doi:10.1103/PhysRevE.52.2318
66. McDonald M, Suleman O, Williams S, Howison S, Johnson NF. Impact of Unexpected Events, Shocking News, and Rumors on Foreign Exchange Market Dynamics. *Phys Rev E* (2008) 77:046110. doi:10.1103/PhysRevE.77.046110

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Shi and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership