FINANCIAL INTERMEDIATION VERSUS DISINTERMEDIATION: OPPORTUNITIES AND CHALLENGES IN THE FINTECH ERA

EDITED BY: Meryem Duygun, Shatha Qamhieh Hashem and Alessandra Tanda
PUBLISHED IN: Frontiers in Artificial Intelligence

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-475-7
DOI 10.3389/978-2-88966-475-7

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

FINANCIAL INTERMEDIATION VERSUS DISINTERMEDIATION: OPPORTUNITIES AND CHALLENGES IN THE FINTECH ERA

Topic Editors:
Meryem Duygun, University of Nottingham, United Kingdom
Shatha Qamhieh Hashem, An-Najah National University, Palestine
Alessandra Tanda, University of Pavia, Italy

Citation: Duygun, M., Hashem, S. Q., Tanda, A., eds. (2021). Financial intermediation versus disintermediation: Opportunities and challenges in the FinTech era. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-475-7

Table of Contents

04 ***Editorial: Financial Intermediation Versus Disintermediation: Opportunities and Challenges in the FinTech Era***
Meryem Duygun, Shatha Qamhieh Hashem and Alessandra Tanda

06 ***Where Did FinTechs Come From, and Where Do They Go? The Transformation of the Financial Industry in Germany After Digitalization***
Barbara Brandl and Lars Hornuf

18 ***On the Improvement of Default Forecast Through Textual Analysis***
Paola Cerchiello and Roberta Scaramozzino

25 ***Initial Coin Offerings: Risk or Opportunity?***
Anca Mirela Toma and Paola Cerchiello

33 ***Explainable AI in Fintech Risk Management***
Niklas Bussmann, Paolo Giudici, Dimitri Marinelli and Jochen Papenbrock

38 ***Network Models to Enhance Automated Cryptocurrency Portfolio Management***
Paolo Giudici, Paolo Pagnottoni and Gloria Polinesi

49 ***The Bitcoin as a Virtual Commodity: Empirical Evidence and Implications***
Cinzia Baldan and Francesco Zen

62 ***How Risk Profiles of Investors Affect Robo-Advised Portfolios***
Dmitri Boreiko and Francesca Massarotti

71 ***FinTech: A New Hedge for a Financial Re-intermediation. Strategy and Risk Perspectives***
Anna Omarini

EDITORIAL
published: 14 January 2021
doi: 10.3389/frai.2020.629105

Editorial: Financial Intermediation Versus Disintermediation: Opportunities and Challenges in the FinTech Era

Meryem Duygun[1], *Shatha Qamhieh Hashem*[2] and *Alessandra Tanda*[3]*

[1]Nottingham University Business School, University of Nottingham, Nottingham, United Kingdom, [2]Department of Banking and Financial Sciences, An-Najah National University, Nablus, Palestine, [3]Department of Economics and Management, University of Pavia, Pavia, Italy

Keywords: FinTech, disintermediation, finance and credit sector, peer to peer, ICOS, cryptocurrencies, robo advisors, risk management

Editorial on the Research Topic

Financial Intermediation Versus Disintermediation: Opportunities and Challenges in the FinTech Era

OPEN ACCESS

Edited and reviewed by:
Paolo Giudici,
University of Pavia, Italy

***Correspondence:**
Alessandra Tanda
alessandra.tanda@unipv.it

Specialty section:
This article was submitted to Artificial Intelligence in Finance, a section of the journal Frontiers in Artificial Intelligence

Received: *13 November 2020*
Accepted: *18 November 2020*
Published: *14 January 2021*

Citation:
Duygun M, Hashem SQ and Tanda A (2021) Editorial: Financial Intermediation Versus Disintermediation: Opportunities and Challenges in the FinTech Era. Front. Artif. Intell. 3:629105. doi: 10.3389/frai.2020.629105

Financial Technology (FinTech) emerged in the 21st century as a significant and innovative force that profoundly disrupts the traditional financial intermediation channels. FinTech has changed both the way financial products are produced and how they are distributed. Technology extensively used by FinTech has been a driver for improving and automating the business model for banks and other financial institutions. The combination of modern technological advancements and financial products and services challenged the traditional financial industry to provide better financial solutions to their clients. Equally FinTech set further pressure by allowing non-financial businesses to provide tailored digitalized financial solutions. Thus, the traditional workflow of the financial industry is currently being disrupted by FinTech and by its incumbents.

This special issue "Financial intermediation vs. disintermediation: Opportunities and challenges in the FinTech era" targets the improvements and challenges that emerged with FinTech. It collects valuable contributions for the interested audience in understanding and analysing the FinTech phenomenon from multiple perspectives.

In this issue, the study contributed by Omarini ("FinTech: a new hedge for a financial re-intermediation. Strategy and risks perspectives") emphasizes that new entrants and technologies are changing dramatically the banking process from an isolated silo approach to an open banking approach. Here the traditional banking business model is questioned as new ways of innovation, cooperation are arising in a manner that is altering the conceptualization of conventional banking business models. New entrants are putting pressure on the traditional banking processes by offering innovative cost efficient financial services and products through the use of lean structures and an appealing seamless customer experience.

Another contribution by Brandl and Hornuf titled "Where Did FinTechs Come From, and Where Do They Go? The Transformation of the Financial Industry in Germany After Digitalization" investigates how the financial industry is being transformed by FinTech companies in Germany. The study points out that traditional participants in the financial system are trying to approach the digitalization of their business with various strategies and different degrees of success. The authors indicate that there is an increase in the number of strategic partnerships established between traditional financial agents and FinTech. Nevertheless banks and financial institutions remain reluctant in opening up their structures to FinTech, often because of the difficulties that they face due to their different IT standards and infrastructure. Thus they are partially participating in the digitalization surge. The authors pointed out the major attention dedicated by academics and observers to the emerging trend of cryptoassets, i.e. Bitcoin. Cryptocurrencies and cryptoassets were proposed initially as an alternative to the legal tender (currencies) which created some worries at the supervisory level. Worth to note that they showed an important development, but remained not so widely spread. This is because of the possibilities introduced by these assets on the market. Overtime, this contributed to a decrease in the threat to the normal currencies, but an increase of cryptoassets in the world of investment.

An additional contribution to the special issue is provided by Baldan and Zen "The Bitcoin as a Virtual Commodity: Empirical Evidence and Implications". The focus of this study is on the role of distributed ledger technology (DLT) and of Bitcoin, and particularly on its pricing dynamics. The virtual coin shows an unpredictable pattern in most of the cases with high volatility. The evidence provided by the authors signals the need of further research on this issue. They also suggest to explain the Bitcoin prices not only with the profit and cost function, but also with additional explanatory parameters including technical features, i.e. the internet, and financial variables, i.e. financial indexes.

Using data drawn from the same asset class, a contribution is presented by Giudici et al. ("Network Models to Enhance Automated Cryptocurrency Portfolio Management"). This study provides an original approach to build efficient portfolio allocation strategies including cryptoassets and other volatile financial instruments. By means of a network model, they reach enhanced portfolio results in terms of performance and risk compared to what was obtained through more traditional models.

Cryptocurrencies can also be used to raise money similarly to what happens during an IPO. The Initial Coin Offerings (ICO) are a way for companies to raise funds through a cryptocurrency linked to the company. The surge of ICO has been accompanied by fraudulent initiatives and this has determined important losses. Toma and Cerchiello investigate this matter in their contribution titled "Initial Coin Offerings: Risk or Opportunity?". By analysing the Telegram chats, white papers and websites through sentiment analysis, the authors show that several elements convey a positive sentiment that helps in detecting genuine ICOs from those that are not so.

Another area that has been transformed by FinTech is the area of lending. The technological development brought new entrants to offer peer-to-peer lending services or lending by crowdfunding platforms. Nevertheless, the financial and banking sectors can employ such technological developments as financing and lending alternatives, not only in terms of capitalizing on the new forms of intermediation in the credit industry, but also in terms of benefiting from such improvements in the process of lending. Furthermore, the FinTech new technologies can boost the predictive power of traditional models in the estimation of credit risk. The paper presented by Cerchiello and Scaramozzino "On the Improvement of Default Forecast Through Textual Analysis" provides an interesting example. Through the use of text mining on a sample of transactions carried out by borrowers, the authors propose an augmented model for credit risk scoring and show that the approach produces interesting results in terms of improved accuracy with respect to the baseline model.

More sophisticated estimation techniques and big data models might not necessarily call for process and results "unexplainability" as maintained by Bussmann et al. Their paper "Explainable AI in FinTech Risk Management" provides an application of sophisticated estimation techniques that employ artificial intelligence, but still have the characteristic of being explainable. This is particularly appreciable in the financial industry, where many stakeholders need to understand not only the result of a process (e.g., credit scoring, investment recommendation, etc.) but also the process itself and its implications. These stakeholders include investors, consumers, regulators and supervisory authorities.

Asset management is another segment of financial services that has benefited from the technological changes over time such as the developments in electronic trading. In the new wave of FinTech transformation, asset management is being electronically integrated within a number of important applications. Among these is the tool of automated advisory services which is becoming more widespread, despite that it has received - so far - less attention than other services, as in the case of lending or payments.

The paper provided by Boreiko and Massarotti on "How Risk Profiles of Investors Affect Robo-Advised Portfolios" investigates the investment suggestions provided to investors by 53 different digital advisors. They find that robo-advisory services, although formally complying with regulatory provisions on investments, show high variability in the investment recommendations even for the same risk-type model investor. Their results underline the need for a harmonised approach by regulators on this innovation that is introduced by FinTech.

The FinTech phenomenon has many facets and every facet presents its novelty, its advantages and risks that investors, customers, industries, policy makers and regulators have to consider. This special issue provides insights on some aspects of FinTech innovation, ranging from the most successful ones (such as peer-to-peer lending and cryptoassets) to more specific technological innovations and modelling approaches that move forward the knowledge and understanding on the FinTech phenomenon. We remark that while this special issue represents a significant contribution to the existing FinTech literature, still more research on the themes of the special issue will be needed to develop new methodologies and tools to enhance the understanding of the FinTech phenomenon.

AUTHOR CONTRIBUTIONS

MD, SH, and AT contributed to the design of the editorial and to the writing of the manuscript.

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 shared affiliation with one of the authors AT at time of review.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 10 March 2020
doi: 10.3389/frai.2020.00008

Where Did FinTechs Come From, and Where Do They Go? The Transformation of the Financial Industry in Germany After Digitalization

*Barbara Brandl[1] and Lars Hornuf[2,3,4,5]**

[1] Faculty of Social Science, Institute of Sociology, Goethe University Frankfurt, Frankfurt, Germany, [2] Faculty of Business Studies and Economics, University of Bremen, Bremen, Germany, [3] Max Planck Institute for Innovation and Competition, Munich, Germany, [4] Center of Finance, University of Regensburg, Regensburg, Germany, [5] CESifo, Faculty of Economics, University of Munich (LMU), Munich, Germany

The digitalization of financial services opened a window for new players in the financial industry. These start-ups take on tasks and functions previously reserved for banks, such as financing, asset management, and payments. In this article, we trace the transformation of the industry after digitalization. By using data on FinTech formations in Germany, we provide first evidence that entrepreneurial dynamics in the FinTech sector are not so much driven by technology as by the educational and business background of the founders. Furthermore, we investigate the reactions of traditional banks to the emergence of these start-ups. In contrast with other emerging industries such as biotechnology, a network analysis shows that FinTechs have mostly engaged in strategic partnerships and only a few banks have acquired or obtained a financial interest in a FinTech. We explain the restraint of traditional banks to fully endorse the new possibilities of digitalized financial services with the characteristics of the technology itself and with the postponed fundamental decisions of banks to modernize their IT infrastructure.

Keywords: application programming interface, API, crowdfunding, financial technology, FinTech, financial industry, IT infrastructure, robo-advice

JEL classification: L86, M13, O16, O32

OPEN ACCESS

Edited by:
Dror Y. Kenett,
Johns Hopkins University,
United States

Reviewed by:
Alessandra Tanda,
University of Pavia, Italy
Emanuela Raffinetti,
University of Milan, Italy

****Correspondence:***
Lars Hornuf
hornuf@uni-bremen.de

Specialty section:
This article was submitted to Artificial Intelligence in Finance, a section of the journal Frontiers in Artificial Intelligence

Received: *11 November 2019*
Accepted: *21 February 2020*
Published: *10 March 2020*

Citation:
Brandl B and Hornuf L (2020) Where Did FinTechs Come From, and Where Do They Go? The Transformation of the Financial Industry in Germany After Digitalization. Front. Artif. Intell. 3:8. doi: 10.3389/frai.2020.00008

INTRODUCTION

A key question in the economics of innovation literature is how industries change as they absorb technological breakthroughs. Are the existing companies able to incorporate the new technologies in their business routines? Does the emergence of new technologies open a window for new types of companies and thus reshape the structure of the industry? How does this change process affect the central institutions and organizations? Mapping the genesis of a new or transformed industry is a challenge because it is a multi-layered process that is both shaped and determined by the emerging technologies and their interplay with the existing institutions in the industry. Therefore, most research works backward from contemporary cases to develop a story about how institutions and organizations were purposefully created and rationally chosen to meet the upcoming challenges (Powell et al., 2012). In this article, we analyze the current transformation process of the financial industry. Who are the new players? When and where did they emerge? How did the type of technology shape this process? How do incumbents react to the challengers?

Digitalization has significantly challenged many traditional industries. This is especially true for the communication industry, the entertainment and media sector, and, more recently, the financial industry. The innovation that has the potential to turn the financial industry upside down is the digitalization of banking business segments such as financing, asset management, and payments. In the past, banks have been able to integrate digital financial innovations, such as Internet banking, and have established new digital technological infrastructures, such as SWIFT or TARGET2-Securities. However, most of the financial innovations were absorbed in the digital back end of banks where customers only indirectly benefited from them. The digitalization of front-end services has created opportunities for new companies. The emerging players in the financial sector are called FinTechs, an acronym for start-ups that commercialize technological financial innovations. Although these new start-ups are a heterogeneous group with diverse interests and business plans, they all have one thing in common: they take on tasks and functions that were traditionally reserved for banks (Puschmann, 2017).

FinTechs can roughly be grouped into four categories: financing (e.g., crowdfunding, crowdlending, crowdinvesting), asset management (e.g., robo advice, social trading, factoring), payments (e.g., crypto currencies, alternative payment systems), and other (e.g., search engines, infrastructure providers)[1]. In the past decade, the number of FinTech start-up formations and market volume in all four segments were steadily growing (Dorfleitner et al., 2017). In 2016, the total volume of all FinTechs in the segments of financing and asset management active in the German market was 7.9 billion EUR (Dorfleitner et al., 2019). The total transaction volume processed through FinTechs in the payment segment was estimated to be 17 billion EUR in 2015 (Dorfleitner et al., 2017).

The traditional players in the financial sector, however, have only reluctantly participated in these new technological possibilities especially in the areas of financing and asset management. This is only partly true for the payment segment. The two main initiatives to digitalize payment systems are the online payment system paydirekt, which is available to customers of around 1,400 German banks and savings banks[2], and the instant payment system RT1, which was launched in January 2017 by 40 European Bank and provides a real-time processing facility for pan-European payments. However, traditional players in the financial sector have only reluctantly participated in the new technological possibilities of digitalized financial services and their market penetration is still small as compared, for example, to the mobile payment incumbent PayPal. Although recent years have witnessed some acquisitions of FinTechs by banks, most FinTech start-ups are not yet controlled by banks.

Despite the rapidly changing environment in the financial industry, almost no studies have investigated the FinTech–bank relationship and how the emergence of FinTechs affects the traditional banking sector. A notable exception is the study of Cumming and Schwienbacher (2018), who investigate the pattern of venture capital investments in FinTechs around the world. They find that venture capital investments in FinTechs can be attributed to differences in the enforcement of financial regulation among start-ups and banks after the financial crisis. Haddad and Hornuf (2019) evidence that countries witness more FinTech start-up formations when the economy is well-developed, venture capital is readily available, and people have more mobile telephone subscriptions. The available labor force and the number of secure Internet servers increase the number of FinTech start-ups in a country as well. Puschmann (2017) defines the term FinTech and presents a categorization of the phenomenon. More recently, Hornuf et al. (2018) have investigated the factors that drive banks to form alliances with FinTechs in Canada, France, Germany, and the United Kingdom. They find that banks are significantly more likely to form alliances with FinTechs when they pursue a well-defined digital strategy and/or employ a Chief Digital Officer. Furthermore, they evidence that markets react more strongly if digital banks rather than traditional banks announce a bank-fintech alliance.

However, to the best of our knowledge, no study has examined the genesis of the transformation of the industry and its interplay with the dominate players in the financial sector. To close this gap in the literature, we chose an explorative approach to map the emergence of FinTechs in the German financial industry. To gain a deeper understanding of these processes, we combine insights from transaction cost theory and concepts of economic sociology. Empirically, we trace this process by using data on the FinTech founders and their professional biographies. Furthermore, we collect data on investments and strategic cooperations of banks with FinTechs in Germany. We conduct a simple network analysis based on this dataset. Theoretically, we base our analysis on insights from transaction cost theory (Coase, 1937; Williamson, 1981) and its further development by organizational theorists, who provide a specific focus on technology development (Teece, 1986, 1998).

This paper proceeds as follows: section Innovation in the Financial Industry provides an overview of the development and the current state of digitalization in the financial sector. We also address the difficulties and the potential of digitalization in the financial service industry. Section Methods and Data presents our data and method. In section Results, we present our findings and argue that the limitation of banks to fully integrate FinTechs can be explained by the characteristics of the technology and postponed fundamental decisions of banks regarding their IT infrastructure. In section Conclusion, we conclude that the future of digital financial innovations will not be decided by technological superiority but by institutional factors. Thus, the future diffusion of digital financial innovations rests on a coordination problem, the solution of which depends on the establishment of novel, effective institutions and organizations.

[1] We apply the typology that was developed by Dorfleitner et al. (2017), which focuses on the German FinTech market. Another typology of FinTech business models was suggested by Eickhoff et al. (2017).

[2] See https://www.paydirekt.de/ueberuns/.

INNOVATION IN THE FINANCIAL INDUSTRY

Merton (1992) identifies four core functions of financial services that innovation needs to address: (1) the moving of funds across time and space (e.g., saving accounts, credit cards), (2) the pooling of funds (e.g., stock markets), (3) managing risk (e.g., derivative products), and (4) extracting information to support decision making and to address asymmetric information problems (e.g., markets for products that deal with default probabilities such as swaps). Given these specific requirements, innovations in finance differ in many respects from the innovations in other fields. Because of the specific features of innovations in the financial industry, financial innovations were rarely the subject of traditional innovation studies and their inquiries. A noteworthy exemption is Awrey (2013), who argues that innovation in the financial sector can only insufficiently be understood by the neoclassical concept of innovation, which describes innovation as a rational answer to market frictions. Instead, he suggests a theoretical perspective in which law in the form of public regulation and private contractual agreements is regarded as a catalyst for innovation in the financial sector. Lerner and Tufano (2011) suggest that innovations in finance contain dynamics that differ from innovation processes in other fields, because the technology behind financial innovation is rather trivial. Digitalization enabled the definition of atomic small business-to-business and business-to-consumer services, which has changed the structural conditions of the financial industry and the possibilities for innovation. As a result, technological innovations are no longer excluded from the financial sector but deeply interwoven in the creation of new firms and financial products.

Significant technological innovations in the financial industry began in the 1960s with the installation of ATM machines and continued with the computerizing of core banking operations (Millo et al., 2005). Today, digitalization has enabled start-ups to extract profitable parts of banking operations in market segments that were previously not often catered to by banks. For example, crowdfunding is the practice in which entrepreneurs raise capital for a project or product from the larger public, often without a securities prospectus. Crowdfunding can be either reward based (the investor pre-purchases a product or service) or equity based (investors pool money to support a project or company). Two functions of financial innovation—that is, moving of funds in time and the pooling of funds from non-sophisticated investors—became possible through new technologies: online platforms that provide an infrastructure to connect individuals who are willing to invest in artists, start-ups, or non-governmental organizations that want to raise money for their projects. The digitization of financial services, however, not only implies a new way of providing financial services but also questions the traditional relationships between lenders and borrowers and between entrepreneurs and customers and thereby challenges the dominant position of banks.

Although the digitalization of *financial services* has so far been portrayed as novel, it would be wrong to consider the digitalization of the *financial industry* a recent phenomenon. The expenditures of the financial sector for IT devices and services have traditionally been rather high. By 1979, the financial industry had already dedicated 32% of all expenses to IT, which was the highest share of all sectors, a number that even increased to 38% in 1992 (Scott et al., 2017). The high share of IT expenses can be explained by the financial sector being the first industry to employ computers on a large scale in its work processes. The first wave of adaptation to the early telecommunication and information technologies had already begun in the late 1950s and peaked in the 1980s. Franke (1987) states that in 1980, half of banks' fixed capital expenditure was for computers or in some form computer related. As a result, the digital architecture of the financial system as well as the internal business routines of banks date back to that time.

This early adoption of computers by the financial industry was possible through Common Business-Oriented Language (COBOL), a problem-oriented programming language that was developed in 1959 as one of the first languages to program business applications. While the early programming languages were predominately used for scientific purposes, COBOL is a hardware-independent software that has the capacity to access and manipulate masses of data (Beyer, 2012). Although in the meantime other more manageable and speedier programming languages such as Java or Python have become available, the clear majority of software applications of banks and credit card companies are still based on COBOL and mainframes.

The outdated IT infrastructure is at least partly responsible for the difficulties banks are facing today in the digitalization of financial services. Although the current infrastructure is highly resilient and robust, it is also very costly to maintain and update. COBOL performs well in the traditional core activities of banks, such as the daily settlement of payments, but is monstrously complex and not well-suited to integrate fast and flexible applications. While the growth rates of IT expenses in the financial industry are still above average today, it appears that traditional banks must invest much more to replace the existing IT infrastructure.

METHODS AND DATA

To examine where FinTechs come from and how the financial industry has changed since it began absorbing digital innovations, we use a mixed-methods approach. To learn more about why FinTech have emerged, we first investigate the educational and professional background of the founders. If FinTech is a technology-driven activity or, in line with Awrey (2013), is a result of legal arbitrage opportunities, this should to some extent be reflected in the founder backgrounds. To describe the current state of collaboration and consolidation, we then conduct a social network analysis of the cooperation and investment activities of FinTechs and the financial industry.

Method and Data

To investigate how FinTechs interact with banks, we conduct a network analysis, which enables us to gain a better understanding of the current market structure through visualization techniques (Powell and Grodal, 2005; Powell et al., 2005; Scott and Carrington, 2011). Our network of banks and FinTechs in the German market is represented as a graph constructed of *nodes* (companies) and *links* (type of connection between companies). In particular, we differentiate between three types of nodes—banks, FinTechs, and FinTech banks—and three types of links—investments, strategic partnerships, and spin-offs. We represent the nodes as dots and the links as lines in a graphical illustration. The more links one node holds, the bulkier is the respective dot. Light blue dots represent FinTechs, dark blue dots banks, and intermediate blue dots FinTech-banks.

Our initial dataset consists of 436 FinTechs that operate in the German market and which Dorfleitner et al. (2017) identified. We excluded the category Search Engines and Comparison Portals as well as other FinTechs, because these firms might be more similar to comparison portals such as Check24 than start-ups that seek to transform financial services. To create a dataset that mirrors all ties between banks and German FinTechs, we supplemented our FinTech data with information on 62 national and international banks. The majority of banks (84%) and FinTechs (78%) that are active in the Germen market originate from Germany. Foreign companies mostly originate from other European countries, predominantly the United Kingdom, Switzerland, and France.

Four FinTechs also possess a banking license. We subsequently refer to these companies as "FinTech-banks." We define "FinTech-banks" as start-ups that provide banking services to others and were founded after the year 2002. Given this relatively recent foundation of companies such as solarisBank, biw Bank für Investments und Wertpapiere, and N26 Bank, these banks are more similar to start-ups then traditional German banks such as Deutsche Bank and Commerzbank. A key source of data was the website www.paymentandbanking.com[3], which continually maps the connections of FinTechs with other companies. The database was previously used in an analysis by Gimpel et al. (2018) to develop a taxonomy of FinTechs start-ups. For 171 FinTechs in our dataset we were able to identify a connection with at least one bank. To categorize the different types of collaborations, we hand-collected additional data from company press releases, annual reports, websites, and trade magazines. We did not identify investments, strategic partnerships, or spin-offs of banks with German FinTechs before 2010 and therefore limited our analyses to the period from 2010 to 2017.

In addition to firm-specific information, such as the founding date and place, we compiled a unique dataset on the professional biographies of 542 FinTech founders, most notably their field of study and former employers. We hand-collected the data from company websites and social media profiles and supplemented them with a survey among the founders. It should be noted that the dataset on German FinTech founders and the dataset on the M&A activities of banks and FinTechs does not perfectly overlap. This is because not all founders could be identified and the dataset on cooperations includes foreign firms as well. For some FinTechs we were not able to find any information about their founders. This is because these FinTechs were often founded by other companies as spin offs ($n = 35$). Spin offs were most often set up in the category crowdfunding (12 spin offs out of 65 crowdfunding FinTechs). Frequently, small firms or non-profit organizations, such as sports clubs or artist associations, founded crowdfunding platforms to raise money related to their specific activities.

Variables

The *education* of the FinTechs founder is an indicator of the type of innovation a FinTech has developed or aims to develop. In the entrepreneurship literature, the educational background of founders is one of the most widely studied variables. In human capital theory, the variable is often used to understand the transition of individuals to entrepreneurs (Brüderl et al., 1992; Lazear, 2005; Kim et al., 2006). Although we are not aware of any literature that uses the founders' field of study as an indicator of the emergence of a sector and the innovations arising from it, we would expect that founders with a background in science are more prone to science-based innovations while founders with a business administration background have more competences in the implementation of business model innovations.

The *former employer* of a FinTech founder can be considered a proxy for the degree to which the technology must be adapted to a specific context. In the sociology literature and management research on entrepreneurship, the professional background and education are often included under the term "imprinting" (Hambrick and Mason, 1984; Wiersema and Bantel, 1992; Ding, 2011). Here, the assumption is that individuals are subjected to a socialization process during their professional education. This process, in turn, deeply shapes their vision of the firm, values, and information-processing patterns. Research further argues that the professional experience influences founders and their performance because it shapes their networks (Haunschild et al., 1999; Rider, 2012). Moreover, we assume that individuals acquire specific human capital during their professional experience (Becker, 1962; Nonaka, 1994), which can be specific intellectual assets that they later use to found a company. We conjecture that founders who have previously worked for a bank or a management consultancy have specific knowledge about the IT infrastructure of banks and the needs and potential of bank customers. Founders who worked in non-bank-related industries or come from universities are less likely to have such specific knowledge.

The way a company or bank secures *access to a certain technology* also provides insights into the type of technology itself. We define a collaborative tie or alliance as any contractual arrangement to exchange or pool resources between banks and FinTechs or between FinTechs and FinTechs. We differentiate among three types of access to a technology. First, we define investments as the financial interest of one firm in another firm. This category includes the full integration of another company as well as the purchase of shares. The investment of a firm indicates the desire to limit access to a certain technology for rival companies. Second, the establishment of a strategic partnership between firms indicates the non-exclusive access to a technology and suggests that a company wants to participate in the knowledge or the customer base of another company, without

[3]http://www.paymentandbanking.com kindly provided their permission for the usage of this data.

taking the risk of a full acquisition. Third, the setup of spin-offs by an already-established company or bank indicates the ability or at least the intention of traditional players to take part in the innovation process.

RESULTS

The Origins of FinTech in Germany

The first FinTechs in Germany emerged in the late 2000s. Research has argued that one reason for the emergence of FinTech companies was the recent financial crises (Haddad and Hornuf, 2019). One the one hand, trust in traditional banks was lost, and customers were seeking alternative ways to handle their banking activities. On the other hand, the financial crises made obtaining capital more difficult for firms (Lopez de Silanes et al., 2015), as banks restricted their lending activities. For many bank customers who did not receive capital from traditional banks, crowdlending and crowdfunding platforms provided a much-appreciated alternative. It might further be argued that FinTechs emerged because a bevy of bankers became unemployed due to layoffs at traditional banks. A first hint that the financial crises was indeed a trigger for many FinTech activities is that many FinTech sub-segments started up around the time of the financial crises (e.g., crowdfunding, crowdlending) or in the years that followed (Haddad and Hornuf, 2019).

Moreover, the emergence of many FinTechs around the same time indicates that there was no technological evolution over time, in which one innovation came first and other companies later built on it. A reason for this finding might be that many FinTechs are based on online services and algorithms. As with any software, FinTech innovations are thus a highly context-dependent technology. This means that innovation is not so much driven by scientific findings as by a constant process of adaptation. We explain this observation in detail in the next section.

Furthermore, Dorfleitner et al. (2017) show that the formation of FinTechs is concentrated in specific local areas; more than half of all FinTech formations took place in only four German cities. The uncontested center of the entrepreneurial activity is Berlin, which represents around one-quarter of all FinTech formations in Germany. Berlin is followed by Munich, Frankfurt, and Hamburg. While we also found distinct local centers of the entrepreneurial activity, we could not identify specific intellectual centers from which the formation of FinTechs emanated. In our dataset, we found neither specific universities nor previous employers from which founders of FinTechs originated. The 542 FinTech founders are spread over 169 universities. LMU Munich is the university where most FinTech founders came from (15 founders), followed by the European Business School (11 founders) and the WHU—Otto Beisheim School of Management (10 founders). The former employers of FinTech founders are even more diverse, with founders originating from 268 different employers. By contrast, the majority (92%) of FinTech founders are male, and thus gender is largely homogeneous. The share of female FinTech founders is even smaller than the already low share of women who establish start-ups in the German economy (15%) (Bundesverband Deutsche Startups, 2019), consistent with the historically low participation of women in finance and the science, technology, engineering, and mathematics fields.

Our interpretation of these findings is that the innovation behind FinTechs is not so much science and technology driven as based on learning and doing. This analytical distinction stems from traditional innovation studies that claim that the dynamics of innovation differ by industry (Jensen et al., 2007; Binz and Truffer, 2017). One way to analyze these differences is to determine whether a technology is universal or context dependent. Teece (1998) defines knowledge assets, in contrast with material assets, as assets that, by their nature, cannot readily be sold and bought. We argue, however, that technologies differ in the degree to which they can be exchanged in one interaction. More precisely, technologies can be grouped on a continuum between being universally applicable and highly context dependent. Technologies that tend more toward universality work regardless of their area of application, while more context-dependent technologies must be adapted to specialized conditions (Dasgupta and David, 1994). While in science and technology-driven industries, such as biotechnology, the emergence of intellectual centers such as university or industrial complexes are typical (Powell et al., 2012), the lack of such innovation centers in the financial industry indicates an innovative dynamic that is more strongly driven by factors such as the adaptation to the specific needs of customers.

Founders of FinTechs: Not Tech Geeks but Businesspeople

FinTechs founders have higher formal degrees than average. Whereas, Metzger (2017) reports that 50% of digital founders in Germany have vocational training as their highest level of education, 92% of the FinTech founders have a degree from higher education institutions. Furthermore, 14% of the FinTech founders hold a doctoral degree, which is far above the average founder education level in Germany. Given that academics generally have better job opportunities, FinTech founders are more likely to be opportunity rather than necessity entrepreneurs. When considering the specific educational background of the FinTech founders, it becomes clear that the overwhelming majority have a business background. **Figure 1** shows that 55% of the 348 FinTech founders have a degree in business administration or a related field, such as management, finance, or accounting. Another 19% have a background in science or engineering, and only 9% have a degree in computer science. The remaining 18% have a background in law (6%), media (5%), or other fields (6%). These numbers differ slightly in the various FinTech sub-segments. For example, in the sub-segment of crowdfunding, many founders (14%) have a media background. This can be explained by the specific purpose of *crowdfunding*. In crowdfunding, entrepreneurs intend to raise capital for projects or products from the larger public. This FinTech sub-segment is populated particularly by artists, who develop cultural products that reflect the underlying cultural ideas of their geographic region (Mollick, 2014). In other segments such as *robo advice*, founders more frequently (28%) have a science or IT degree, which is likely due to the challenge

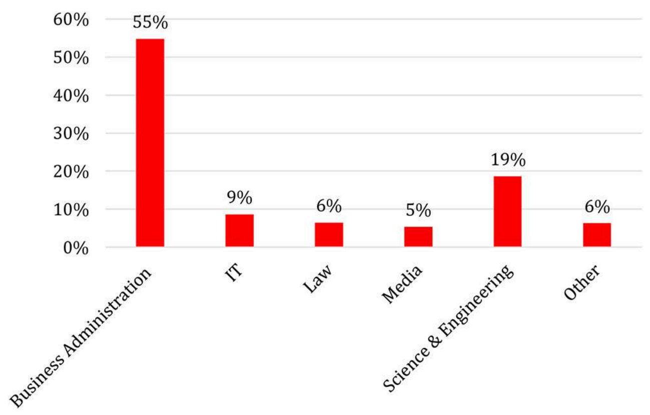

FIGURE 1 | Educational background of FinTech founders (in %), $N = 422$.

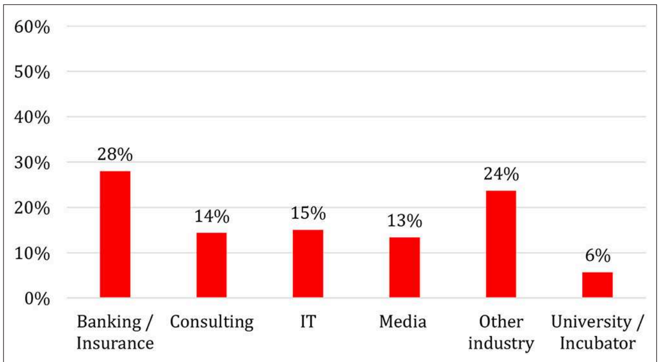

FIGURE 2 | Former employers of FinTech founders (in %), $N = 450$.

to create algorithms that identify an investment strategy; thus, the technological part of the innovation is stronger. In line with human capital theory, our data indicate that the educational background of the FinTech founders relates well to the work content of the respective FinTech sub-segment.

Despite this variation in the sub-segments, the overall trend is clear. Although the digitalization of financial services strengthens the linkage between finance and technology, the entrepreneurial activity is driven by founders from a business background. One interpretation of this empirical finding is that the technological innovation behind FinTechs is rather trivial, while the implementation of innovations, such as the acquisition of customers or the establishment of new standards, is more challenging. This argument is supported by Lerner et al. (2015), who find that financial patents show lower performance in common proxies for the quality of patents, such as the number of citations of scientific publications within the patent or the number of litigations associated with a patent (Lerner, 2002; Tufano, 2003).

An analysis of the former employers of FinTech founders confirms our previous findings. As **Figure 2** illustrates, most FinTech founders (28%) previously worked for banks or insurance companies. The share of founders who come from consulting firms is also high (14%). Many management consultants who founded a FinTech company presumably also had a focus on the financial industry. Conversely, we observe a relatively small share of FinTechs that were founded directly out of universities (6%). The number of founders who were previously employed in the IT sector is also rather low (15%), which is surprising given that most FinTech business models are based on software solutions.

We conjecture that the large percentage of FinTech founders with a banking or insurance background is due to the specific requirements of technological innovation in the financial industry. Software in general is a highly context-dependent technology because it requires adaptation to specific contexts. It is virtually impossible to develop software in a context-independent environment and sell a pre-arranged product to customers. Rather, the real work of software developers begins after the first users have begun using a digital product. Programmers of software companies constantly need to fix bugs, adapt their software to a constantly changing hardware environment, and specify their products to the requirements of their customers. The superiority of a software lies not in an initial advantage of a better innovation but in the constant work of contextualization. In the case of banking, the requirement to adapt innovations to a specific context is even stronger, because application programming interfaces (APIs) are not standardized but differ from one bank to another. Often, the nature of an API is only known by the former employees of a specific bank. Although banks are generally willing to share this information, individuals with such specific knowledge and personal connections with bank employees have an advantage.

The necessity to adapt products to existing conditions creates different innovation dynamics than in traditional industries, which are often more science and technology driven. Products that are more science and technology based, such as in the pharmaceutical industry, can be developed in a clean laboratory environment. Although the transformation of these innovations in a real-life context is sometimes more difficult and costlier than expected, the major share of costs in pharma or biotechnology for innovations emerges *ex ante* in the R&D process. In contrast, companies in more context-dependent industries such as software development often must invest more heavily in their service departments after product creation. As we show in the next section, the type of technology also influences the regime of appropriation. While a competitor in a science and technology-driven industry such as biotechnology would have a significant advantage by neglecting the patent of a rival company, someone who illegally uses an algorithm does not automatically have access to the desired product or software, because the software is constantly being improved and developed. In the case of biotechnology, a large share of the technology is expressed in the intellectual property right itself. In the case of software development, a larger share of innovation occurs through the contextualization within a technological architecture.

The contextualization of software and technology is especially important in the financial industry. The current IT infrastructure in the financial industry was created decades ago and evolved incrementally, without a consistent architectural design. Because each bank developed its IT infrastructure to a large extent for itself, contextualization led to a lack of standards (e.g., for APIs). Start-ups therefore cannot create one universal solution for all banks but often must adapt their innovation to the specific technological context of each bank.

M&A Activities in the Financial Industry: Little Investments, More Strategic Partnerships

To understand the dynamics unleashed by technological innovation, it is insightful to analyze the incentive structure of companies to integrate or license the new technology. In stark contrast with other emerging industries such as biotechnology, we find evidence that banks do not predominately use the direct integration of start-ups to gain access to the desired technology, but rather employ another form of coordinating intellectual assets: the strategic partnership. **Figure 3** maps the contractual links between banks and FinTechs and between FinTechs and FinTechs, while **Figure 4** shows only the investment of one company in another. We find that only 19% of all contractual links are actual investments, while the overwhelming majority (74%) are strategic partnerships, and 7% are spin-offs. A strategic partnership is a contractually fixed relationship between two firms. In general, a strategic partnership between a bank and a FinTech or between a FinTech and another FinTech means that one company uses the software of the other company, usually by paying a transaction-based fee. For example, many banks use the video identification tools of FinTech companies, to verify the identity of potential customers in a legally admissible way that is convenient for the customer. In contrast, many FinTechs do not possess a banking license, because their business focus is mainly on front-end operations. For example, many crowdlending portals transfer credit requests to a bank, which consequently originates the loan. Most FinTech-banks such as Fidor Bank, solarisBank, and FinTech AG—biw Bank für

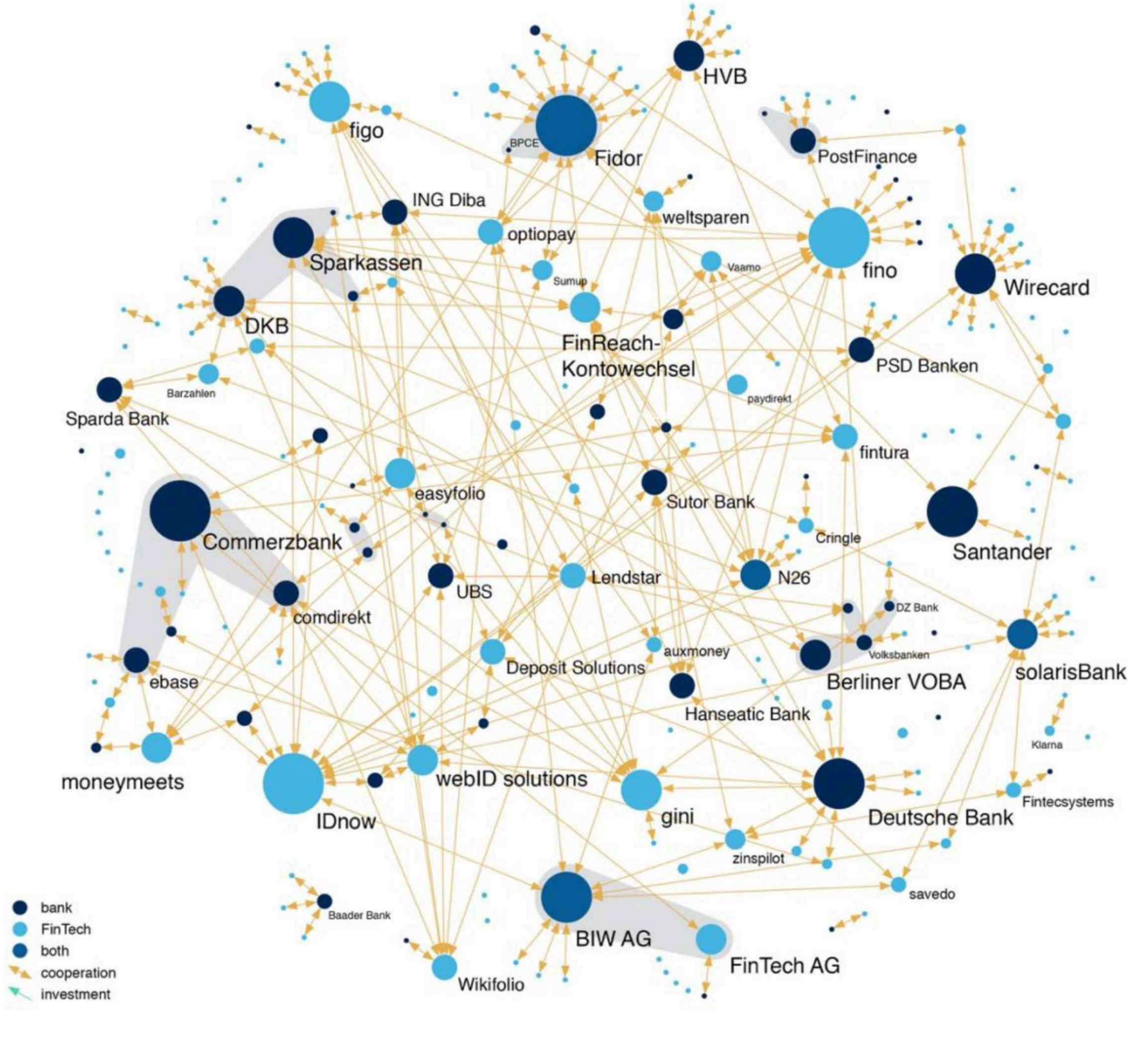

FIGURE 3 | Bank and FinTech cooperation.

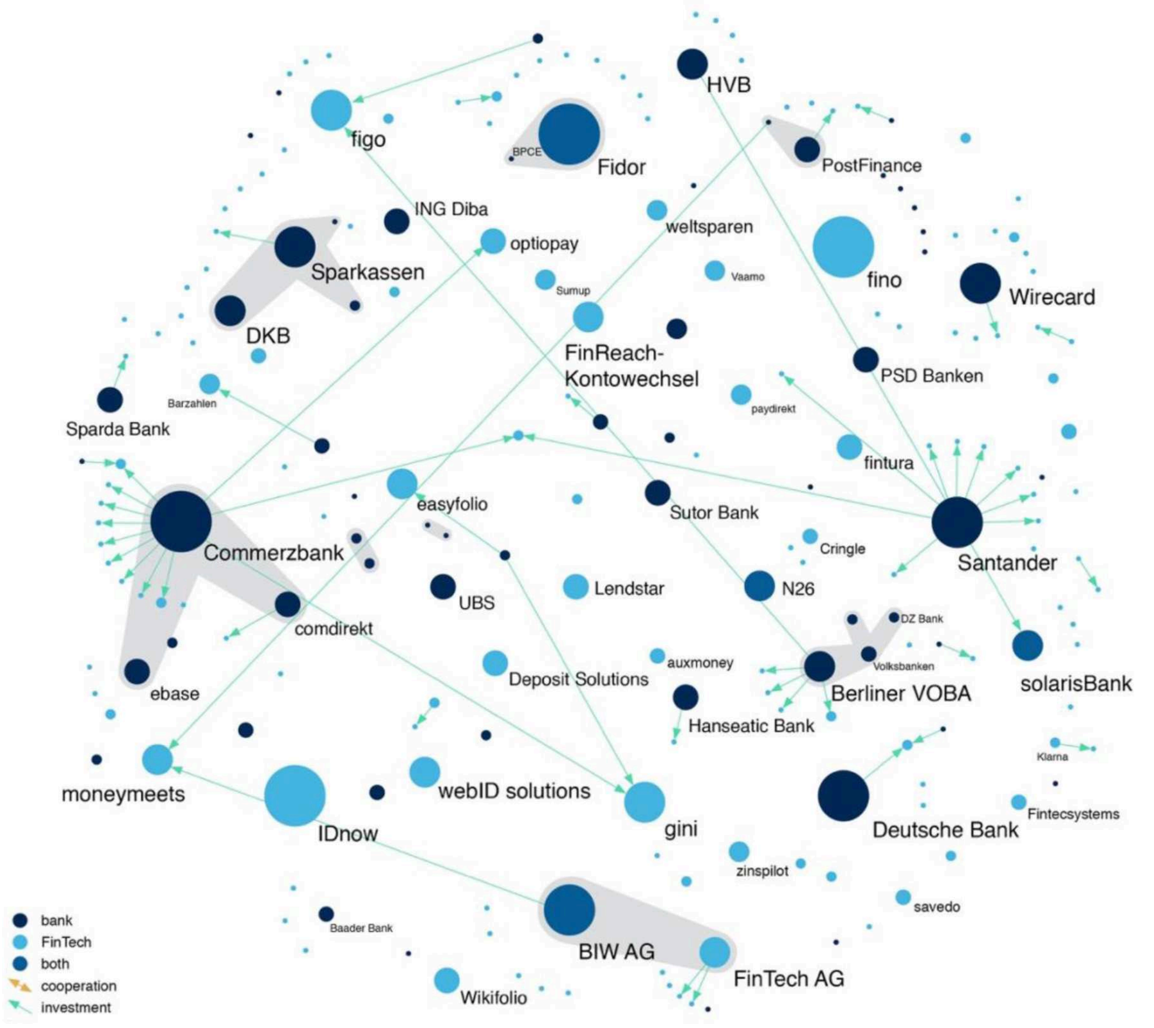

FIGURE 4 | Bank and FinTech investments.

Investments und Wertpapiere have realized the potential that stems from the FinTech sector and have specialized in what is called "banking as a service" (BaaS) or "banking as a platform" (BaaP). These banks provide end-to-end processes that ensure the execution of atomic small financial services on demand. BaaP works through APIs provided by the respective FinTech-bank. In other cases, strategic partnerships simply consist of banks making the products of FinTechs available to their customers, such as Consorsbank, which offers crowdinvesting products from Seedmatch and the possibility of social trading through wikifolio. While the bank might only receive a revenue share from the FinTech, the strategic partnership can help the bank maintain customers who might otherwise also switch their core banking activities to a new FinTech competitor. The low percentage of spin-offs (7% of all contractual links) indicates the inability of the established players to actively take part in the innovation process.

To gain a better understanding of the calculus of companies aiming to obtain access to a desired technology, the theories of Teece (1986, 1998) are a useful reference. He shows that the optimal strategy of a company to gain access to a technology depends on the *regime of appropriation*. Referring to Teece's work, Graff et al. (2003) suggest a spectrum of channels to coordinate complementary intellectual assets. The most extreme strategy of a firm is the complete internalization of external assets through integration. On the other end of the continuum stands the purchase of a non-exclusive license for a technology from an independent firm.

The regime of appropriability refers to environmental factors, such as the design of intellectual property rights or the features of the technology itself, that govern an innovator's ability to capture the profits generated by an innovation (Teece, 1986). Moreover, Teece (1986) differentiates between tight and weak regimes of appropriability. A tight regime reflects a situation in which the technology is relatively easy to protect from imitation; correspondingly, a weak appropriability regime describes a situation in which it is almost impossible to protect a new technology from unwanted usage. The characteristics of a regime of appropriability emerge through the interplay between the design of the intellectual property rights (e.g., patents, trademarks, copyrights) and the nature of the knowledge that requires protection[4].

The context-dependent nature of software affects not only the innovative process *per se* but also the regime of appropriability. We illustrate the interplay of the regime of appropriability and the structure of the industry with another nascent industry with similar starting conditions: early plant biotechnology. In the 1980s, traditional companies secured access to the new technologies of the biotechnology industry by simply acquiring the start-ups. One reason for this drastic strategy was the regime of appropriability. In biotechnology, the weak regime of appropriability impeded the contracting over technologies between different companies and created incentives to fully integrate the start-ups (Graff et al., 2003; Marco and Rausser, 2008; Schneider, 2010). The difficulties in defining intellectual property rights in biotechnology stemmed from the novelty of the field, which led to the paradox situation that the patent offices had to decide on an issue, which was largely unclear even from a scientific perspective. Moreover, in contrast with chemistry or other fields of science, the patent offices could not rely on existing jurisprudence, which makes the results of a court ruling usually more unpredictable. Charles (2002) argues that Monsanto managers realized early on that gens could not be licensed like other technologies such as software. The European counterpart and direct competitor in the early days of agro-biotechnology AgrEvo (now Bayer) did not draw the same conclusions and tried a model of gaining access to various gens and biotechnological tools via license agreements, which did not work out and ended—at least temporarily—in the great defeat of European agrochemical companies in the global seed market (Bijman, 2001; Charles, 2002).

Another observation that differs from adaptation processes in other industries is that FinTech start-ups are active players in reshaping the industry. In other nascent industries (e.g., biotechnology), the start-ups were largely passive, while the dominant players appropriated crucial parts of the new technology (Owen-Smith and Powell, 2003; Powell et al., 2012). In our dataset of German FinTechs, we observe that start-ups themselves take an active role in M&A activities. Some FinTechs began as regular start-ups but eventually received a bank license (e.g., solarisBank) or founded a bank as a spin-off of their FinTech (e.g., FinTech AG—biw Bank für Investments und Wertpapiere). These FinTech-banks coordinated their intellectual assets almost exclusively through BaaS or BaaP. However, we also observe FinTechs that directly integrated other FinTechs by obtaining a minority interest in or acquiring them. In 2013, the Swedish FinTech Klarna, a digital payment service provider, acquired its direct competitor Billpay and, in 2016, Coockies.

The strategies of traditional banks regarding the coordination of their intellectual assets differs widely. While banks such as Santander and Commerzbank largely acquired start-ups from the financial industry, other traditional banks, including HypoVereinsbank and Deutsche Bank, are more cautious about investing in FinTechs. These banks more often engage in strategic partnerships (see **Figure 3**).

As the emergence of FinTechs is a recent phenomenon, we expect further consolidation in the coming years. However, we argue that the dominance of strategic partnerships as well as the currently missing consolidation is not only a transitional phenomenon that will vanish in the foreseeable future but also has structural reasons that are rooted in the technology and industry itself.

The first reason consolidation of the German FinTech industry is missing lies in the regime of appropriability. In contrast with the biotechnological industry, the appropriability regime of software is tight. This may seem surprising, as the intellectual property protection of algorithms and software is weak in general, especially in Germany and other European countries, where computer programs cannot be patented (Eimer, 2011). Developers of proprietary software must protect their innovations with a weaker intellectual property right—namely, the copyright. In many cases, therefore, the software developers do not depend on intellectual property rights to appropriate their innovations. Software firms often sell software licenses (end-user license agreements) and keep the source code secret. In the case of server-side software, the source code is not protected by a copyright but is kept by a trade secret. Thus, the regime of appropriation is rather tight for software innovations, not only because of the ease of contracting over software but also because explicit knowledge is not an advantage *per se*, as all software products must be adapted to a specific firm and context over time.

Although the license agreements of software companies are not always effective in excluding free-riding customers that do not pay for the service, especially at a business-to-business level, private appropriation of innovations works well in general in the software industry. One reason is that the documentation of software usage takes place automatically through networks and protocols. From the perspective of firms that must contract over technologies, the automatized documentation of usage implies a reduction in transaction costs. While in biotechnology the monitoring and enforcement problems associated with the technology resulted in a structural advantage for large companies that could afford to engage in patent disputes (Haedicke, 2008; Schubert et al., 2011; Gill et al., 2012), the low costs of licensing and monitoring software have led to opportunities for small firms in the financial industry.

A weak regime of appropriability results in a structural advantage for large companies, as is evident in the biotechnological industry. Cost-efficient monitoring and litigating intellectual property right infringements can only be achieved by globally operating companies that maintain

[4]The knowledge economics and organizational sociology literature streams describe several classes of problems related to the tasks of coordinating knowledge between different units or companies and determining whether an appropriability regime is tight or weak (Pisano, 1990; Nonaka, 1994; Heller and Eisenberg, 1998; Graff et al., 2003; Brandl and Glenna, 2016): (1) monitoring and enforcement problems due to the difficulty in determining who is using a certain technology and whether the users pay for it, (2) entitlement problems due to poorly defined and/or silent blocking of intellectual property rights, and (3) monitoring problems due to the difficulty of measuring and therefore contracting over certain types of technology or knowledge.

branches in different jurisdictions (Haedicke, 2008). The same is true at the business-to-consumer level. The ability to exclude free-riding customers from using the technology often makes a monitoring system necessary (Schubert et al., 2011). The installation and maintenance of a monitoring system is costly, and large companies can more easily realize economies of scale to implement such a system. As a result, a week regime of appropriability increases transaction costs and therefore creates a structural incentive to integrate the company of interest. In contrast, a tight regime of appropriability opens opportunities for small companies (Hall and Zidonis, 2001). Because the contracting over knowledge is reliable, companies can license their technological assets to other firms. In other words, in tight regimes of appropriability transactions costs are lower. Therefore, the incentive for companies to fully integrate a start-up is also lower, as an acquisition is also associated with higher costs and risks.

A second reason that prevents banks from fully integrating FinTechs is the current design of the market infrastructure in the financial industry. The software that underpins the infrastructure was designed for digital requirements that were defined decades ago. The lack of coordination among banks led to siloed data stores maintained by individual participants. Most experts agree that players in the financial industry must reconcile the current system sooner or later, especially regarding the status of transactional data. The current lack of coordination in common technological standards and banking functions leads to a hesitation among banks to fully integrate FinTechs. The coordination of intellectual assets of strategic partnerships, however, enables banks to appropriate the knowledge of FinTechs without needing to make fundamental decisions about the future of their IT infrastructure. Finally, strategic partnerships are an efficient way for banks to overcome their cultural legacy, extensive regulatory provisions, and compliance issues, which also allows them to approach different technologies without having to commit to a specific one.

CONCLUSION

We began this article by stating that the digitalization of financial services could potentially turn the industry upside down. Not only do FinTechs have a streamlined cost structure, but they also provide novel services to customers, such as fully digital financing or investment solutions. In contrast, most of the traditional banks rely on a historic and expensive IT infrastructure. Our analysis indicates that the current wave of digitalization does not unleash the same groundbreaking dynamics as other innovations such as biotechnology. We explain that the hesitation of many banks to fully endorse the new possibilities of digitalized financial services is due to the context dependency of the software and the tight appropriation regime. In other words, the characteristics of the technology allow banks to participate only partly in the new wave of digitalization through strategic partnerships, without needing to change their own outdated IT infrastructure.

Our results might be only one part of the story. Our analysis is restricted to the German market, and different regions of the world might show another pattern. Moreover, we only focus on a specific type of FinTech. Although the current FinTechs challenge the traditional banks through their digitally optimized cost structure and their affinity to new technology, they are not a self-sustaining alternative to the banking system. Almost all currently active FinTechs rely on banks, mostly because they do not possess a banking license, which is required to conduct core banking operations (taking deposits and extending loans).

In the future, another type of FinTech might become more important. FinTechs do not just offer services that work in parallel to the current system but also provide technologies and services that aim to fully replace the current structures and organizations of the financial industry. These FinTechs either possess a banking license or rely on self-sufficient systems such as blockchain, the technology that stands behind Bitcoin and other crypto currencies, and thus could supersede traditional banks (Tapscott and Tapscott, 2016). Blockchain is an open, distributed ledger that can verifiably and permanently record transactions between two parties. As contracts and records of transactions are the defining institutions of the economy, the technological transformation of these institutions has the potential to cause a deeper social change process than the current FinTechs have. Blockchain start-ups offer technological solutions for various problems. The first product based on blockchain, Bitcoin, for example, promises to be a currency that is protected from the access of central banks during economic crises (Valkenburgh, 2016). Digital Asset Holdings, a software firm located in New York City, offers software solutions based on the distributed ledgers technology for the entire architecture of the financial industry, such as central clearing houses and central securities depositories (Digital Asset, 2016). Thus, unlike all other FinTech innovations, blockchain has the potential to replace the financial industry as such. For this to happen, the new technology needs to be trusted by all market participants. Trust, however, is a critical asset that is built up only in the long run.

Because the financial industry was until now not capable of agreeing on common standards for a new IT infrastructure and because of the political inability to solve this coordination problem, the strategic development of firms and the future of the blockchain technology as such are highly unclear. Iansiti and Lakhani (2017) compare the current situation of blockchain to the early days of e-mail. Before the adoption of the transmission control protocol and Internet protocol (TCP/IP), the telecommunications architecture was based on *circuit switching*, a method in which the connection between two parties had to be physically pre-established with an electrical circuit. The adoption of TCP/IP not only created a new communication architecture but also paved the way for entirely new technological applications such as the World Wide Web.

Technologies that aim to replace the current ones create an enormous coordinative challenge for the economic agents involved, as success depends entirely on the establishment of novel, effective institutions, and organizations. This challenge can be met by a company, an industrial pressure group, or the state by implementing legal regulations. Many studies in the field of comparative political economy have shown that different types of economies have established varying ways of dealing with coordination problems in relation to technology and platform services. Hall and Soskice (2001) argue that in liberal economies (e.g., United States), firms compete over standard settings, while in coordinated economies (e.g., Germany), industry-specific networks coordinate a collaborative process of standard settings. To enable complex technological adaptation, in some industries the state must step in, if firms or other organizations fail to coordinate on technological standards (Casper and Soskice, 2004).

The infrastructure that underpins the financial industry was developed decades ago. Although it incrementally evolved, the current system is largely unable to meet the current regulatory and market needs of the financial system. It is an open secret that international banks such as Goldman Sachs, J.P. Morgen, and Citigroup as well as companies such as Deutsche Börse and PricewaterhouseCoopers are actively acquiring blockchain start-ups and applying for patents on parts of the blockchain technology on their own business models. The makeover of the digital architecture of financial markets, the reconciliation of the siloed data stores, and the agreement on standards is an enormous coordinative challenge. It is still an open question, however, whether the financial industry will be able to deal with this challenge. The danger of leaving this challenge to market players is that banks either are not capable of agreeing on a common standard or, even worse, will decide on the wrong standard. The task of designing a novel IT architecture for the financial markets is therefore also a political question, which can only be answered democratically.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available. The data was partly obtained from https://paymentandbanking.com. Publishing the curricular information of fintech founders might violate their privacy.

AUTHOR CONTRIBUTIONS

Both authors have made substantial, direct, and intellectual contribution to the work, and approved it for publication. Both authors were involved in data collection and theoretical framing. Data analysis was conducted by BB.

REFERENCES

Awrey, D. (2013). Toward a supply-side theory of financial innovation. *J. Comp. Econ.* 41, 401–419. doi: 10.1016/j.jce.2013.03.011

Becker, G. S. (1962). Investment in human capital: a theoretical analysis. *J. Political Econ.* 70, 9–49. doi: 10.1086/258724

Beyer, K. W. (2012). *Grace Hopper and the Invention of the Information Age (Reprint)*. Cambridge, MA: Lemelson Center Studies in Invention and Innovation.

Bijman, J. (2001). AgrEvo: from crop protection to crop production. *AgBioForum J. AgroBiotechnol. Manage. Econ.* 4, 20–25.

Binz, C., and Truffer, B. (2017). Global innovation systems—a conceptual framework for innovation dynamics in transnational contexts. *Res. Policy* 46, 1284–1298. doi: 10.1016/j.respol.2017.05.012

Brandl, B., and Glenna, L. (2016). Intellectual property and agricultural science and innovation in Germany and the United States. *Sci. Technol.Hum. Values* 42, 622–656. doi: 10.1177/0162243916675954

Brüderl, J., Preisendörfer, P., and Ziegler, R. (1992). Survival chances of newly founded business organizations. *Am. Sociol. Rev.* 57, 227–242. doi: 10.2307/2096207

Bundesverband Deutsche Startups (2019). *Female Founders Monitor*. Available online at: https://deutscherstartupmonitor.de/fileadmin/ffm/ffm_2019/studie_ffm_2019.pdf (accessed January 3, 2020).

Casper, S., and Soskice, D. W. (2004). "Sectoral systems of innovation and varieties of capitalism: explaining the development of high-technology entrepreneurship in Europe," in *Sectoral Systems of Innovation. Concepts, Issues and Analyses of Six Major Sectors in Europe,* ed F. Malerba (Cambridge: Cambridge University Press), 348–387.

Charles, D. (2002). *Lords of the Harvest: Biotech, Big Money, and the Future of Food (Softcover Ed edition)*. Cambridge, MA: Basic Books.

Coase, R. H. (1937). The nature of the firm. *Economica* 4, 386–405. doi: 10.1111/j.1468-0335.1937.tb00002.x

Cumming, D., and Schwienbacher, A. (2018). Fintech venture capital. *Corp. Gover. Int. Rev.* 26, 374–389. doi: 10.1111/corg.12256

Dasgupta, P., and David, P. A. (1994). Toward a new economics of science. *Res. Policy* 23, 487–521. doi: 10.1016/0048-7333(94)01002-1

Digital Asset (2016). "The digital asset platform," in *Non-Technical White Paper* (New York, NY).

Ding, W. W. (2011). The impact of founders' professional-education background on the adoption of open science by for-profit biotechnology firms. *Manage. Sci.* 57, 257–273. doi: 10.1287/mnsc.1100.1278

Dorfleitner, G., Hornuf, L., Schmitt, M., and Weber, M. (2017). *The German FinTech market*. Heidelberg: Springer.

Dorfleitner, G., Hornuf, L., Schmitt, M., and Weber, M. (2019). "Allgemeiner marktüberblick," in *FinTech Handbuch - Digitalisierung, Recht,* eds F. Möslein and S. Omlor (Munich: C.H. Beck; Finanzen), 21–38.

Eickhoff, M., Muntermann, J., and Weinrich, T. (2017). "What do fintechs actually do? a taxonomy of fintech business models," in *ICIS 2017 Proceedings,* 22. Available online at: http://aisel.aisnet.org/icis2017/EBusiness/Presentations/22 (accessed January 3, 2020).

Eimer, T. R. (2011). *Arenen und Monopole: Softwarepatente in den USA und in Europa*. Wiesbaden: Springer-Verlag.

Franke, R. H. (1987). Technological revolution and productivity decline: computer introduction in the financial industry. *Technol. Forecast. Soc. Change* 31, 143–154. doi: 10.1016/0040-1625(87)90046-1

Gill, B., Brandl, B., Böschen, S., and Schneider, M. (2012). Autorisierung. Eine wissenschafts- und wirtschaftssoziologische perspektive auf geistiges eigentum. *Berliner J. Soziol.* 22, 407–440. doi: 10.1007/s11609-012-0197-9

Gimpel, H., Rau, D., and Röglinger, M. (2018). Understanding FinTech start-ups – a taxonomy of consumer-oriented service offerings. *Electron. Mark.* 28, 245–264 doi: 10.1007/s12525-017-0275-0

Graff, G. D., Rausser, G. C., and Small, A. A. (2003). Agricultural biotechnology's complementary intellectual assets. *Rev. Econ. Stat.* 85, 349–363. doi: 10.1162/003465303765299864

Haddad, C., and Hornuf, L. (2019). The Emergence of the global fintech market: economic and technological determinants. *Small Bus. Econ.* 53, 81–105. doi: 10.1007/s11187-018-9991-x

Haedicke, M. (2008). "Innovationssteuerung durch patente im bereich der biotechnologie," in *Geistiges Eigentum: Schutzrecht oder Ausbeutungstitel. Zustand und Entwicklungen im Zeitalter von Digitalisierung und Globalisierung,*ed O. Depenheuer (Berlin: Springer), 111–120.

Hall, B., and Zidonis, R. (2001). The patent paradox revisited: an empirical study of patenting in the U.S. semiconductor Industry 1979-1995. *RAND J. Econ.* 32, 101–128. doi: 10.2307/2696400

Hall, P. A., and Soskice, D. W. (2001). *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford: Oxford University Press.

Hambrick, D. C., and Mason, P. A. (1984). Upper echelons: the organization as a reflection of its top managers. *Acad. Manage. Rev.* 9, 193–206. doi: 10.5465/amr.1984.4277628

Haunschild, P. R., Henderson, A. D., and Davis-Blake, A. (1999). "CEO demographics and acquisitions: network effects of educational and functional background," in *Corporate Social Capital and Liability,* eds R. T. A. J. Leenders and S. M. Gabbay (New York, NY: Springer), 266–283.

Heller, M. A., and Eisenberg, R. S. (1998). Can patents deter innovation? *Anticommons Biomed. Res. Sci.* 280, 698–701. doi: 10.1126/science.280.5364.698

Hornuf, L., Klus, M., Lohwasser, T., and Schwienbacher, A. (2018). How do banks interact with fintechs? Forms of alliances and their impact on bank value. *CESifo Working Paper* (Amsterdam).

Iansiti, M., and Lakhani, K. (2017, January–February). The truth about blockchain. *Harvard Business Review*. 95, 118–127.

Jensen, M. B., Johnson, B., Lorenz, E., and Lundvall, B. Å. (2007). Forms of knowledge and modes of innovation. *Res. Policy* 36, 680–693. doi: 10.1016/j.respol.2007.01.006

Kim, P. H., Aldrich, H. E., and Keister, L. A. (2006). Access (not) denied: the impact of financial, human, and cultural capital on entrepreneurial entry in the United States. *Small Bus. Econ.* 27, 5–22. doi: 10.1007/s11187-006-0007-x

Lazear, E. (2005). Entrepreneurship. *J. Labor Econ.* 23, 649–680. doi: 10.1086/491605

Lerner, J. (2002). Where does state street lead? a first look at finance patents, 1971 to 2000. *J. Finance* 57, 901–930. doi: 10.1111/1540-6261.00446

Lerner, J., Baker, M., Speen, A., and Leamon, A. (2015). *Financial Patent Quality: Finance Patents After State Street (Harvard Business School Working Paper No. 16-068)*. Cambridge, MA: Harvard Business School.

Lerner, J., and Tufano, P. (2011). *The Consequences of Financial Innovation: A Counterfactual Research Agenda (NBER Working Paper Series No. 16780)*. Cambridge, MA: National Bureau of Economic Research.

Lopez de Silanes, F., McCahery, J., Schoenmaker, D., and Stanisic, D. (2015). *The European Capital Markets Study Estimating the Financing Gaps of SMEs*. Available onlien at: http://www.dsf.nl/wp-content/uploads/2015/09/European-Capital-Markets-Study_2015_FINAL-15-7.pdf (accessed December 27, 2016).

Marco, A. C., and Rausser, G. (2008). The role of patent rights in mergers: consolidation in plant biotechnology. *Am. J. Agric. Econ.* 90, 133–151. doi: 10.1111/j.1467-8276.2007.01046.x

Merton, R. C. (1992). Financial innovation and economic performance. *J. Appl. Corp. Finance* 4, 12–22. doi: 10.1111/j.1745-6622.1992.tb00214.x

Metzger, G. (2017). *Digitale Gründer werden ihrer Vorreiterrolle gerecht (Digital Founders Live Up to Their Pioneering Role)*. KfW Research, Nr. 157 (available in German language). Available online at: https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-Fokus-Volkswirtschaft/Fokus-Nr.-157-Januar-2017-Digitale-Gr%C3%BCnder.pdf

Millo, Y., Muniesa, F., Panourgias, N. S., and Scott, S. V. (2005). Organised detachment: clearinghouse mechanisms in financial markets. *Inform. Organ.* 15, 229–246. doi: 10.1016/j.infoandorg.2005.02.003

Mollick, E. (2014). The dynamics of crowdfunding: an exploratory study. *J. Bus. Ventur.* 29, 1–16. doi: 10.1016/j.jbusvent.2013.06.005

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5, 14–37. doi: 10.1287/orsc.5.1.14

Owen-Smith, J., and Powell, W. W. (2003). The expanding role of university patenting in the life sciences: assessing the importance of experience and connectivity. *Res. Policy* 32, 1695–1711. doi: 10.1016/S0048-7333(03)00045-3

Pisano, G. P. (1990). The R&D boundaries of the firm: an empirical analysis. *Adm. Sci. Q.* 35, 153–176. doi: 10.2307/2393554

Powell, W. W., and Grodal, S. (2005). "Networks of innovators," in *The Oxford Handbook of Innovation,* eds J. Fagerberg, D. C. Mowery, and R. R. Nelson (Oxford: Oxford University Press), 56–85.

Powell, W. W., Packalen, K., and Whittington, K. (2012). "Organizational and institutional genesis. The emergence of high-tech clusters in the life sciences," in *Emergence of Organizations and Markets*, eds J. F. Padgett and W. Powell (Princeton, NJ: Princeton University Press), 434–465.

Powell, W. W., White, D. R., Koput, K. W., and Owen-Smith, J. (2005). Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences. *Am. J. Sociol.* 110, 1132–1205. doi: 10.1086/421508

Puschmann, T. (2017). Fintech. *Bus. Inform. Syst. Eng.* 59, 69–76. doi: 10.1007/s12599-017-0464-6

Rider, C. I. (2012). How employees' prior affiliations constrain organizational network change: a study of U.S. venture capital and private equity. *Admin. Sci. Q.* 57, 453–483. doi: 10.1177/0001839212461633

Schneider, C. (2010). The battle for patent rights in plant biotechnology: evidence from opposition fillings. *J. Technol. Trans.* 36, 565–579. doi: 10.1007/s10961-010-9200-9

Schubert, J., Böschen, S., and Gill, B. (2011). Having or doing intellectual property rights? Transgenic seed on the edge between refeudalisation and napsterisation. *Eur. J. Sociol.* 22, 407–440. doi: 10.1017/S0003975611000014

Scott, J., and Carrington, P. (2011). *The SAGE Handbook of Social Network Analysis*. Los Angeles, CA: Sage Publications.

Scott, S. V., Van Reenen, J., and Zachariadis, M. (2017). The long-term effect of digital innovation on bank performance: an empirical study of SWIFT adoption in financial services. *Res. Policy* 46, 984–1004. doi: 10.1016/j.respol.2017.03.010

Tapscott, D., and Tapscott, A. (2016). *Blockchain Revolution: How the Technology Behind Bitcoin is Changing Money, Business, and the World*. New York, NY: Portfolio.

Teece, D. J. (1986). Profiting from technological innovation: implications for integration, collaboration, licensing and public policy. *Res. Policy* 15, 285–305. doi: 10.1016/0048-7333(86)90027-2

Teece, D. J. (1998). Capturing value from knowledge assets: the new economy, markets for know-how, and intangible assets. *Calif. Manage. Rev.* 40, 55–79. doi: 10.2307/41165943

Tufano, P. (2003). "Financial innovation," in *Handbook of the Economics of Finance, Vol. 1*, eds G. M. Constantinides, M. Harris, and R. M. Stulz (Amsterdam: Elsevier), 307–335.

Valkenburgh, P. (2016). *Framework for Securities Regulation of Cryptocurrencies.* Coin Center Report. Washington, DC: Coin Center.

Wiersema, M. F., and Bantel, K. A. (1992). Top management team demography and corporate strategic change. *Acad. Manage. J.* 35, 91–121. doi: 10.2307/256474

Williamson, O. (1981). The economics of organization: the transaction cost approach. *Am. J. Sociol.* 87, 548–577. doi: 10.1086/227496

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.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 07 April 2020
doi: 10.3389/frai.2020.00016

Check for updates

On the Improvement of Default Forecast Through Textual Analysis

*Paola Cerchiello and Roberta Scaramozzino**

Department of Economics and Management, University of Pavia, Pavia, Italy

Textual analysis is a widely used methodology in several research areas. In this paper we apply textual analysis to augment the conventional set of account defaults drivers with new text based variables. Through the employment of *ad hoc* dictionaries and distance measures we are able to classify each account transaction into qualitative macro-categories. The aim is to classify bank account users into different client profiles and verify whether they can act as effective predictors of default through supervised classification models.

Keywords: text analysis, credit scoring, default, classification models, finance

1. INTRODUCTION

The change in all sectors of the economy that we are witnessing in recent years is so rapid that it speaks of the fourth industrial revolution. In the era of big data, across all sectors companies' main asset has become data. There is an increasing use of data, as the use of digital technologies increases, the amount of information collected increases exponentially. As a result, firms sit upon swathes of data, but the key is being able to derive value from them. In the financial sector, data are used for multiple purposes, one of which is credit scoring. This refers to the techniques used to assign creditworthiness to a customer to be able to distinguish "good" from "bad" customers i.e., clients who will repay their financing and those that will be insolvent. The probability that an applicant is insolvent is determined by analyzing the information on the latter (Hand and Henley, 1997). Credit scoring models are fundamental for banks to guarantee a correct forecast of default risk for financed loans, which translates into a reduction in losses and an increase in profits. There are numerous techniques for this purpose (Efron, 1979; Baesens et al., 2003; Jayasree and Balan, 2013; Emekter et al., 2015). Although nowadays most of the models in question use quantitative information, typically financial data, the latter is no longer sufficient to properly profile customers in a world that is now increasingly digital. This type of information, called hard information, is contrasted by another category of so-called soft information. Soft information is the term used to indicate information obtained through textual analysis, in this case, we talk about unstructured data. Text mining arises in this context. Text mining is the mechanism for extracting relevant information from unstructured documents to discover unknown patterns (Gupta and Lehal, 2009; Aggarwal and Zhai, 2012).

Even in today's standards, the traditional approach, which uses only hard information, is that which is widely used by firms but there is lack of studies that analyze textual information (Fei et al., 2015; Allahyari et al., 2017). Jiang et al. (2018) demonstrate how the use of textual data can increase the predictive power of a model, combining soft information with typically financial information analyzing the main p2p platforms in China. Groth and Muntermann (2011) state that exposure to intraday market risk management can be discovered through the use of text mining. Chan and Franklin (2011) show, through the use of textual data, that the forecast accuracy of their model improves similar traditional models by 7%. The advantages and disadvantages of Hard Information are analyzed by Liberti and Petersen (2018) who examine how information influences financial markets. Cornée (2019) demonstrates

Edited by:
Shatha Qamhieh Hashem,
An-Najah National University, Palestine

Reviewed by:
Yuze Yuan,
Minjiang University, China
Roberto Savona,
University of Brescia, Italy
Benjamin Miranda Tabak,
Fundação Getúlio Vargas, Brazil

****Correspondence:***
Roberta Scaramozzino
roberta.scaramozzino01@universitadipavia.it

Specialty section:
This article was submitted to Artificial Intelligence in Finance, a section of the journal Frontiers in Artificial Intelligence

Received: *30 December 2019*
Accepted: *13 March 2020*
Published: *07 April 2020*

Citation:
Cerchiello P and Scaramozzino R (2020) On the Improvement of Default Forecast Through Textual Analysis. Front. Artif. Intell. 3:16. doi: 10.3389/frai.2020.00016

the importance of textual information for credit scoring by analyzing 389 presences of a French bank. Grunert et al. (2005) analyzing German SMEs, compare a model based on financial data, one on textual data and one mixed demonstrates how the latter is the best in terms of predicting loan defaults.

In this paper, we propose a credit scoring model that utilizes text mining. The variables extracted through textual analysis are used as predictors in the model. To extract this information we have classified the bank transactions into macro-categories and then considered the frequencies of each macro category and the total amounts. We then compared the classical model based on financial information and the one with the addition of variables derived from textual analysis. The rest of the paper is organized as follows: in section Methodology the methodology used is shown, in section Data we analyze the data, in section Results we report the results obtained and finally the paper is discussed and future research presented.

2. METHODOLOGY

In this section we explain the methodology used. We developed a default risk prediction model by combining financial information and textual information. The first step of the analysis was the extraction of relevant variables from the texts provided in the transactions. The method consists of 2 parts: pre-processing and knowledge extraction. The first one needed a lot of work and demanded most of the time spent on the analysis. Preprocessing plays a significant role in Text mining. Any task of text mining fully depends on the preprocessing step. High-quality of preprocessing always yielded superior results (Kumar and Ravi, 2016). In the preprocessing, we have cleaned up the texts of the transactions. We have created the textual corpus starting from the original text through the removal of stop words (typical and specific for the context), tokenization, removal of errors, and stemming. We then created the document term-matrix. The matrix describes the frequency of the terms in the documents. In our case, the rows represent the transactions and the columns correspond to the terms. The columns are extracted through the analysis of the corpus regarding the dictionaries, these were created manually in terms of key-value: by inserting the subject of the service as a key and the category of the transaction as value. The final goal of text mining was to obtain topics (macro-areas of interest relating to transactions) to create new variables for each one created and to insert these variables in the credit scoring model. The first problem encountered is due to spelling errors, due to this problem many transactions were not found in dictionaries. This obstacle is because the transaction descriptions are handwritten by the bank operators, who often abbreviate the words or write absently without paying attention. To overcome such issue we, therefore, decided to use a distance measure, the Levenshtein one, to attribute the word to the closest transaction under consideration (Levenshtein, 1966). This is a measure accounting for the difference between two strings, which is the minimum number of changes necessary to transform one word into another. Mathematically:

$$lev_{a,b}(i,j) = \begin{cases} max(i,j) & \text{if } min(i,j)=0, \\ min\begin{Bmatrix} lev_{a,b}(i-1,j)+1 \\ lev_{a,b}(i,j-1)+1 \\ lev_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{Bmatrix} & \text{otherwise.} \end{cases} \quad (1)$$

The algorithm flow, accordingly, consists in:

1. Pre-processing on the single transactions and creation of a textual corpus;
2. Search in the dictionary of the key corresponding to the clean string obtained from preprocessing;
3. Assignment of the value corresponding to the key of the dictionary, that identifies the category. If this is not found, the Levenshtein distance of the string is calculated from all the keys in the dictionary and the nearest dictionary value is assigned, with a maximum threshold of 10. This means that whether the distance is greater, neither category will be assigned (at the end of the investigation, the percentage of Na is around 15%);
4. After assigning a category to each transaction of the dataset, we created the variables to be included in the credit scoring model.

The categories have been grouped into macro-categories, the macro-categories chosen for the analysis are 5:

- Non-essential goods: including expenses for goods such as shopping, travel and living.
- Essential goods: including expenses in markets, pharmacies.
- Financial services and utilities: including expenses related to banks, payment services, telco companies, petrol stations.
- Revenue: including incomes such as transfers and dividends.
- Salaries: including wages and pensions.

For each of the previous 5 variables, we have therefore created 2 further variables: frequency and the total amount spent by the client for the specific category. We have thus obtained 10 new variables through the processing of textual data.

The second step of the analysis is the application of credit scoring model. The textual based categories created were added to the financial ones and the new dataset was used in the model. The chosen model is the lasso logistics. Lasso logistic model is a shrinkage method that allows obtaining a subset of variables that are strongly associated with the dependent variable, through regularization of the coefficients bringing them to values very close or even exactly equal to zero. Since the L1 penalty is used, the variables with a coefficient equal to zero are excluded from the model (Hastie et al., 2009). Mathematically:

$$L_{lasso}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m}\left|\hat{\beta}_j\right|. \quad (2)$$

where y_i are the n-observations for the target variable (default/no default), x_i are the n-observations for the covariates, λ is the penalization parameter chosen by cross validation and β are the coefficient of the model.

Along with Lasso we fitted Elastic net as well. Since there were no statistical significant differences, we preferred to focus on Lasso because of an easier interpretation of the results. Before applying the lasso logistic algorithm, we pre-selected the relevant variables through the Kruskal–Wallis test. We decided to apply this test due to the size of the dataset: too few observations (400) with respect to the number of available variables (52). When the size of available data is limited, as in our case, Lasso can be not efficient enough in fitting parameters. Lasso is good at dropping out not significant variables if it can use an appropriate number of observations compared to the number of variables. Thus, we pre-selected the most relevant variables (without being too restrictive) through Kruskal–Wallis paying attention to the division in training and test. Kruskal–Wallis is a non-parametric method (no assumptions on the distribution of the data) that states if there is a significant difference between the groups. The null hypothesis states that the k samples come from the same population and the alternative hypothesis states that they come from different ones (Siegel and Castellan, 1988).

The KW test (Conover, 1971) is the non-parametric version of the well-known ANOVA test and represents a multivariate generalization of the Wilcoxon test for two independent samples, that can be adapted to our problem as follows. On the basis of C independent samples (each containing the transactions of a client) of size n_1, n_C (the frequency of transactions for each client), a unique large sample L of size $N = \sum_{i=1}^{C} n_i$ is created by means of collapsing the original C samples. L can be organized as a matrix that contains a number of rows equal to N and a number of columns equal to W (the number of variables). Each entry of the matrix contains the frequency count of a specific variable along with each transaction. The KW test is then applied columns, in order to evaluate the discriminant power of each variable with respect to the client classification task. For each variable, the frequency vector corresponding to each column of the matrix L is ordered from the smallest frequency value to the largest one, and a rank is assigned to each transaction in the sample accordingly. Finally, one should calculate R_i as the mean of the ranks in each of the C original clients categories samples. The multivariate KW test can then be shown to take the following form:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{C} n_i \frac{R_i^2}{n_i} - 3(n+1) \tag{3}$$

After having selected only the significant variables, we applied the lasso logistics comparing the model keeping only the financial variables, the model with only the textual variables and the one obtained by the combination of the variables. For each analysis, the dataset was divided into 2 parts, 70% for training and the remaining 30% for the test. In addition, the model has been cross-validated using 5 folds. We applied the cross validation in the training set and then validated in the test set.

The comparison of the 3 models is based on mean misclassification error (mmce), area under the curve roc (auc), accuracy (acc), and roc curve (Krzanowski and Hand, 2009; Agresti and Kateri, 2011). Mmce is a prediction error metrics for a binary classification problem. The Roc curve is a graphical representation, along the two axes we find the sensitivity and 1-specificity, respectively represented by True Positive Rate and False Positive Rate. It is, therefore, the true positive rate as a function of the false positive rate. AUC is the area under the Roc curve, an aggregate measure of performance across all possible classification thresholds. Accuracy is the degree of correspondence of the estimated data with the real one.

3. DATA

We have undertaken this analysis starting from 2 datasets: loans and transactions relating to an Italian bank. The paper was executed in collaboration with Moneymour. Moneymour is a FinTech startup that offers a payment method to provide instant loans for online purchases. It allows client to buy immediately and pay in installments.

In the former the original variables were:

- Date of the loan request,
- Loan ID,
- Default status,
- Amount,
- Number and amount of loan payments.

In the latter, there were:

- Client,
- Accounting date,
- Value date,
- Transaction amount,
- Reason code and reason text.

This study analyzed 164931 transactions and 400 loans from 2015 to 2018.

The financial variables extracted are:

- Sum of income,
- Sum of outcome,
- Average income,
- Average outcome,
- Number of income,
- Number of outcomes,
- Total number of movements,
- Sum of salary and average salary.

All listed variables referring to the first month, three months, six months and the previous year respectively to the request for financing, and financing obtained.

Summary statistics of financial variables are reported in **Table 1**.

Through the use of text mining, the transactions carried out by each client were analyzed and the new variables were created:

- Salary,
- Total output non-essential goods,
- Total output essential goods,
- Total financial services and utilities,

TABLE 1 | Descriptive financial features.

	prev. funding	num. rev. month 1	num. rel. month 1	num. mov. month 1	num. rev. month 3	num. rel. month 3	num. mov. month 3	num. rev. month 6	num. rel. month 3	num. mov. month 6	num. rev. month 12	num. rel. month 12	num. mov. month 12
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1st Qu.	0.00	0.00	0.75	1.00	0.00	3.00	4.00	1.00	4.00	6.00	1.00	4.00	6.00
Median	0.00	1.00	6.00	7.00	3.00	16.50	21.00	6.00	34.00	39.00	11.00	55.00	71.00
Mean	0.23	1.46	13.78	15.23	4.22	41.38	45.59	8.10	81.39	89.50	14.93	150.50	165.40
3rd Qu.	0.00	2.00	23.25	26.00	6.00	74.25	83.25	12.00	149.25	165.20	22.00	248.20	271.80
Max.	0.00	9.00	90.00	93.00	43.00	253.00	265.00	59.00	466.00	485.00	97.00	873.00	911.00

- Total salaries,
- Total income,
- Frequency output non-essential goods,
- Frequency output essential goods,
- Frequency financial services and utilities,
- frequency salaries,
- Frequency income.

Summary statistics of textual variables are reported in **Table 2**.

The target variable is default or non-default of the client defined as follows: default means the non-fulfillment of loan payment installments for 3 months in a row.

4. RESULTS

In this section, we discuss the results obtained. The data set was divided into 2 portions: training and testing. Seventy percent of the data was used for training and the remaining 30% for the test. The data in both samples were distributed as follows: 37% defaulting and 63% non-defaulting. The target variable is the default status indicated with the value 0 for the non-default and with 1 the default.

We recall that the starting dataset presented 400 observations and 53 variables. The issue regarding the high number of variables with regards to the number of observations has been overcome by selecting the most significant variables through the Kruskal–Wallis test. The variables selected after applying the test are 21 and reported in the following list: previous funding, number revenue month 1, number releases month 1, number movements month 1, number revenue month 3, number releases month 3, number movements month 3, number revenue month 6, number releases month 6, number movements month 6, number revenue month 12, number releases month 12, number movements month 12, salary, total output essential goods, total financial services and utilities, frequency output non-essential goods, frequency output essential goods, frequency financial services and utilities, frequency salaries, frequency income. The model chosen for the credit scoring analysis is lasso logistics which represents an efficient choice in data analysis problems like ours when a variable selection step is needed.

For greater accuracy of the metrics obtained, we conducted the analysis using k-fold cross validation which partitions the dataset in subsets of equal size, where each subset is used as a test and the others as training. Moreover, we have conducted an out of sample analysis, training models on 2014-2015-2016 data and testing them on 2017 and 2018.

Parameters estimates of the 3 fitted logistic lasso models are reported in **Tables 3–5** referred respectively to financial variables, textual variables and the mixed one. In particular, from **Table 5** we can infer that several variables both financial and textual are significant. The textual ones, of course, are of major interest being the new ones. We observe that the largest parameter is obtained by the salary flag variable: having a negative sign means that the presence of the salary on the bank account decreases the probability of default. In particular if we calculate the odds ratio we get that the probability of non-defaulting is 12 times higher than the probability of defaulting. Thus, such

TABLE 2 | Descriptive textual features.

	Salary	Total output essential goods	Total financial services and utilities	Frequency output non essential goods	Frequency output essential goods	Frequency financial services and utilities	Frequency salaries	Frequency income
Min.	0.00	-31084	-224137.0	0.00	0.00	0.00	0.00	0.00
1st Qu.	0.00	-1862	-19456.8	3.75	0.00	0.00	0.00	1.00
Median	0.00	0.00	-699.5	50.00	0.00	5.50	0.00	11.00
Mean	0.31	-2511	-15107.2	107.97	47.01	94.86	7.082	21.29
3rd Qu.	1.00	0.00	0.00	152.25	39.00	142.00	6.000	26.00
Max.	1.00	0.00	0.00	647.00	717.00	944.00	118.00	237.00

TABLE 3 | Important variables selected by Lasso Model on financial dataset.

Variables	Parameter estimated
Previous funding	-0.4548
Number revenue month 1	-0.1553
Number releases month 1	·
Number movements month 1	-0.1759
Number revenue month 3	·
Number releases month 3	-0.0000
Number movements month 3	-0.0150
Number revenue month 6	·
Number releases month 6	-0.0000
Number movements month 6	-0.0001
Number revenue month 12	0.0028
Number releases month 12	·
Number movements month 12	·

TABLE 4 | Important variables selected by Lasso Model on textual dataset.

Variables	Parameter estimated
Salary flag	-2.5998
Freq. salaries	·
Freq. income	-0.1111
Total output essential goods	0.0015
Total output financial services and utilities	0.0006
Freq. output non essential goods	-0.0083
Freq. output essential goods	·
Freq. output financial services and utilities	-0.1294

a simple information, that can be derived by the analysis of bank transactions, can add very useful information to the credit holder. Other two textual variables are worth of mentioning: frequency of income characterized by a negative sign and the total output for financial services and utilities with a positive sign. According to the former the larger is the frequency in the income the lower is the probability of default. On the other way around, a higher number of transactions for financial services and utilities increases the probability of default. This to say that having several incomes helps in affording financial loans but the impact of expenses for services and utilities is not negligible. Regarding purely financial variables, all of them but one shows negative signs, meaning that they reduce the probability of default. The three largest parameters are shown by "previous funding" "number of movements at month 1" and "number of revenue at month 1." What affects largely the chance of repaying loans is the presence of previous loans request to the bank. This ensures a previous capability of respecting financial obligations. The same applies to the number of movements in the nearest month: having more movements can be considered a symptom of financial health as if we take into account the number of revenues.

Finally in **Table 6** we report the comparison among the 3 models: the first which considers only the financial variables, the second which considers only the textual variables and the third

TABLE 5 | Important variables selected by Lasso Model on mixed dataset.

Variables	Parameter estimated
Previous funding	-0.4984
Number revenue month 1	·
Number releases month 1	·
Number movements month 1	-0.1650
Number revenue month 3	·
Number releases month 3	-0.0020
Number movements month 3	-0.0240
Number revenue month 6	·
Number releases month 6	·
Number movements month 6	·
Number revenue month 12	·
Number releases month 12	-0.0007
Number movements month 12	-0.0012
Salary flag	-2.5011
Freq. salaries	·
Freq. income	-0.0022
Total output essential goods	·
Total output financial services and utilities	0.0003
Freq. output non essential goods	·
Freq. output essential goods	·
Freq. output financial services and utilities	·

TABLE 6 | Results from Lasso logistic regression.

	Accuracy	Auc	False positive rate	Specificity
Financial features	0.75	0.931	0.479	0.521
Text features	0.84	0.939	0.233	0.767
Mix	0.80	0.946	0.356	0.644

one which combines both. The comparison is measured through Accuracy, Auc, type 1 error and specificity. As can be seen from **Table 6**, the results obtained are pretty high for the three models and even the roc curves tend to overlap many times as shown in **Figure 1**. If we focus on the accuracy, it is interesting to note that the model restricted to the textual variables component overcomes largely the model with only financial variables (84 vs. 75%). As a consequence the mixed model is an average of the two. On the other hand, the models are perfectly comparable in terms of AUC.

Nevertheless, there are important elements, first of all, classification errors. As you can see from **Table 6**, the mixed model and the text based one have an improvement over the type 1 errors. Not all errors have the same impact, some mistakes have higher implications than others. For a bank, type I error is the most dangerous one as it represents the probability of giving a loan to those who will not pay. The costs associated with type I errors are higher than type II errors.

5. CONCLUSIONS

In this paper, we use textual data to enhance the traditional credit scoring model. We evaluated the models performance by comparing the basic model in which only financial variables were included, against one in which there are only those extracted through text mining and the last one containing the mix of the two types of variables. From the analysis, we conclude that the addition of textual variables is relevant in the model. Although accuracy and AUC do not vary much, it should be emphasized first of all the distribution of errors. We observe an improvement over type 1 error and that both the textual based model and the mixed one show better results regarding accuracy. This is a promising result that encourages to further test the methodology with a larger dataset.

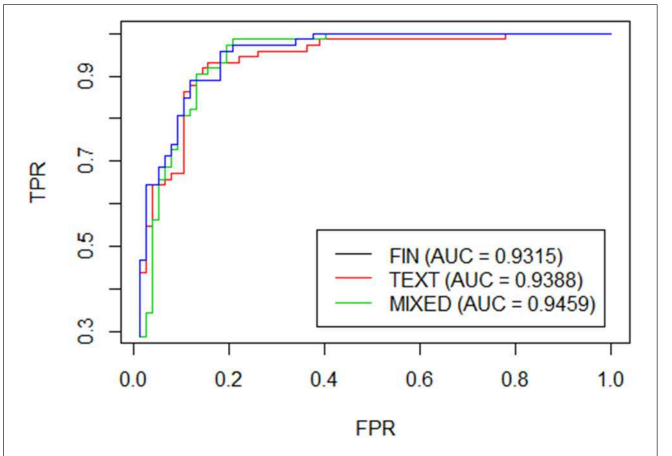

FIGURE 1 | ROC curve.

We can, therefore, state that despite the small size of the dataset, the analysis carried out shows how the textual analysis can be used in a credit scoring model to improve its accuracy.

Future research perspectives concern the application of the model to a larger dataset not only in terms of observations but also of variables based on textual information. More data can offer other types of information not available in the data at hand. Moreover the application of other text analysis technique like topic modeling (Cerchiello and Nicola, 2018) rather than the creation of manual dictionaries to make the process more automated and therefore decrease the time spent in pre-processing, can improve even more the quality of the analysis.

DATA AVAILABILITY STATEMENT

The data analyzed in this study was obtained from Moneymour and it is restricted by a non-disclosure agreement. Requests to access these datasets should be directed to RS, roberta.scaramozzino01@universitadipavia.it.

AUTHOR CONTRIBUTIONS

The paper is the product of full collaboration between the authors, however PC inspired the idea, the methodology and wrote sections Result and Conclusion. RS run the analysis and wrote sections Introduction, Methodology, and Data.

ACKNOWLEDGMENTS

This research has received funding from the European Union's Horizon 2020 research and innovation program FIN-TECH: A Financial supervision and Technology compliance training program under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA). This research was executed in collaboration with Moneymour (https://moneymour.com/).

REFERENCES

Aggarwal, C. C., and Zhai, C. (2012). (ed.). *Mining Text Data.* Springer Science & Business Media.

Agresti, A., and Kateri, M. (2011). *Categorical Data Analysis.* Berlin; Heidelberg: Springer.

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B. et al. (2017). A brief survey of text mining: classification, clustering and extraction techniques. *arXiv:1707.02919*.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *J. Operat. Res. Soc.* 54, 627–635. doi: 10.1057/palgrave.jors.2601545

Cerchiello, P., Nicola, G. (2018). Assessing news contagion in finance. *Econometrics* 6:5. doi: 10.3390/econometrics6010005

Chan, S. W. K., and Franklin, J. (2011). A text-based decision support system for financial sequence prediction. *Decis. Support Syst.* 52, 189–198. doi: 10.1016/j.dss.2011.07.003

Conover, W. J. (1971). "The Kruskal-Wallis test," in *Practical Nonparametric Statistics* (New York, NY: John Wiley and Sons Inc.), 256–264.

Cornée, S. (2019). The relevance of soft information for predicting small business credit default: evidence from a social bank. *J. Small Bus. Manag.* (2019) 57, 699–719. doi: 10.1111/jsbm.12318e

Efron, B. (1979). Bootstrap methods : another look at the jackknife. *Ann. Statist.* 1, 1–26.

Emekter, R., Tu, Y., Jirasakuldech, B., and Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Appl. Econ.* 47, 54–70. doi: 10.1080/00036846.2014.962222

Fei, W., Gu, J., Yang, Y., and Zhou, Z. (2015). Credit risk evaluation based on social media. *Proc. Comp. Sci.* 55, 725–731.

Groth, S. S., and Muntermann, J. (2011). An intraday market risk management approach based on textual analysis. *Decis. Support Syst.* 50, 680–691. doi: 10.1016/j.dss.2010.08.019

Grunert, J., Norden, L., and Weber, M. (2005). The role of non-financial factors in internal credit ratings. *J. Bank. Fin.* 29, 509–531. doi: 10.1016/j.jbankfin.2004.05.017

Gupta, V., and Lehal, G. (2009). A survey of text mining techniques and applications. *J. Emerg. Technol. Web Intell.* 1, 60–76.

Hand, D. J., and Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *J. R. Stat. Soc. Ser. A* 160, 523–541.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* London: Springer Science & Business Media.

Jayasree, V., and Balan, R. V. S. (2013). A review on data mining in banking sector. *Am. J. Appl. Sci.* 10:1160. doi: 10.3844/ajassp.2013.1160.1165

Jiang, C., Wang, Z., Wang, R., and Ding, Y. (2018). Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending. *Ann. Operat. Res.* 266, 511–529. doi: 10.1007/s10479-017-2668-z

Krzanowski, W. J., and Hand, D. J. (2009). *ROC Curves for Continuous Data.* London: Chapman and Hall/CRC.

Kumar, B. S., and Ravi, V. (2016). A survey of the applications of text mining in financial domain. *Knowl. Based Syst.* 114, 128–147. doi: 10.1016/j.knosys.2016.10.003

Levenshtein, V. I. (1966). "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet Physics Doklady,* 707–710.

Liberti, J. M., and Petersen, M. A. (2018). Information: hard and soft. *Rev. Corporate Fin. Stud.* 8, 1–41. doi: 10.3386/w2507

Siegel, S., and Castellan, N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences, 2nd Edn.* New York, NY: McGraw-Hill.

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.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 16 April 2020
doi: 10.3389/frai.2020.00018

Initial Coin Offerings: Risk or Opportunity?

Anca Mirela Toma and Paola Cerchiello*

Department of Economics and Management Science, University of Pavia, Pavia, Italy

Initial coin offerings (ICOs) are one of the several by-products in the world of the cryptocurrencies. Start-ups and existing businesses are turning to alternative sources of capital as opposed to classical channels like banks or venture capitalists. They can offer the inner value of their business by selling "tokens," i.e., units of the chosen cryptocurrency, like a regular firm would do by means of an IPO. The investors, of course, hope for an increase in the value of the token in the short term, provided a solid and valid business idea typically described by the ICO issuers in a white paper. However, fraudulent activities perpetrated by unscrupulous actors are frequent and it would be crucial to highlight in advance clear signs of illegal money raising. In this paper, we employ statistical approaches to detect what characteristics of ICOs are significantly related to fraudulent behavior. We leverage a number of different variables like: entrepreneurial skills, Telegram chats, and relative sentiment for each ICO, type of business, issuing country, team characteristics. Through logistic regression, multinomial logistic regression, and text analysis, we are able to shed light on the riskiest ICOs.

Keywords: ICOs, cryptocurrencies, fundraising, classification models, text analysis, scam

OPEN ACCESS

Edited by:
Emanuela Raffinetti,
University of Milan, Italy

Reviewed by:
Shatha Qamhieh Hashem,
An-Najah National University, Palestine
Branka Hadji Misheva,
Zurich University of Applied Sciences,
Switzerland

****Correspondence:***
Anca Mirela Toma
ancamirela.toma01@universitadipavia.it

Specialty section:
This article was submitted to Artificial Intelligence in Finance, a section of the journal Frontiers in Artificial Intelligence

Received: *19 September 2019*
Accepted: *13 March 2020*
Published: *16 April 2020*

Citation:
Toma AM and Cerchiello P (2020) Initial Coin Offerings: Risk or Opportunity? Front. Artif. Intell. 3:18. doi: 10.3389/frai.2020.00018

1. INTRODUCTION

Initial Coin Offerings (ICOs) can be considered as an innovative way of obtaining funding, promoted by entrepreneurial companies that base their business projects on a new technology known as blockchain. Up to the present date, more than 1,700 cryptocurrencies have been created but not all of them are successful or characterized by a significant impact. ICOs issue "tokens," i.e., the unit of a chosen cryptocurrency, in exchange of a flat cryptocurrency, in order to participate in the crowd-funding of the company. Tokens can be bought directly on the web platform of the company, at different stages of the ICO commonly referred as pre-sale and sale. Later, the amount of bought tokens can be sold or used in the future to obtain products or services. The portal Tokendata.io has estimated that until 2017 ICOs raised as much as $5.3 billion around the world; if we consider venture capitalist, in 2016, they invested $71.8 billion in the United States and $4.3 billion in Europe (National Venture Capital Association and Invest Europe). Start-ups and existing businesses are turning to alternative sources of capital as opposed to classical channels like banks or venture capitalists. They can offer the inner value of their business by selling "tokens," i.e., units of the chosen cryptocurrency, like a regular firm would do by means of an Initial Public Offering (IPO). When we say cryptocurrency, we refer to a digital currency, a new means of exchange, the most popular examples of which are Bitcoin and Ethereum. Blockchain (chain of blocks) is the technology at the basis of a cryptocurrency; it is a Distributed Ledger Technology defined as a distributed, shared, encrypted database that serves as an irreversible and incorruptible repository of information (Wright and De Filippi, 2015). Bitcoin is currently the largest blockchain network followed by, Ethereum, XRP, Litecoin, EOS and Bitcoin Cash (Coinmarketcap, 2018).

ICOs favor open-source project development and decentralized business, generating a built-in customer base and positive network effects. They also create a secondary market where tokens can be employed as rewards for using the app of the company or the offered services (Subramanian, 2018). This work aims at addressing the specific characteristics of ICOs using relevant variables that play a key role in determining the success of the ICO.

As it stands there is no database with the information we are looking for, thus we have been building and constantly maintaining a dataset that is currently composed of 196 ICOs that occurred between October 2017 and November 2018 (Cerchiello et al., 2019). The database comprises companies from European countries namely France, Germany, Switzerland, Estonia, Latvia, and non European countries such as Russia, United Kingdom, United States, Japan, Singapore, and Australia. The most common sectors in which ICOs operate are: high-tech services, financial services, smart contract, gambling platforms, marketplaces, and exchanges.

2. INITIAL COIN OFFERINGS

Most of the ICOs projects are related to the development of a blockchain, the issuance of new cryptocurrencies or somehow related fintech services. ICOs tokens grant contributors the right to access platform services in 68.0% of the cases, governance powers in 24.9% of the cases and profit rights in 26.1% of the cases. The secondary market for ICOs tokens is quite liquid on the first day of trading, and the initial return is large (mean value +919.9% compared to the offer price, median value +24.7%). The success of such decentralized technology lays on the fact that it works without the commitment and the control of a central authority: the blockchain is a Peer-to-Peer technology. A Peer-to-Peer (P2P) system represents a way of structuring distributed applications such that the individual nodes can act as both a client and a server. A key concept for P2P systems is to allow any two peers to communicate with each other in such a way that either ought to be able to initiate the contact (Peer-to-Peer Research Group, 2013). Then, the more a P2P network is distributed, scalable, autonomous, and secure, the more is valuable.

All of these precious features have enabled the fast growth of cryptocurrencies not just *per se* but also as a tool for crow-funding purposes, giving birth to the so-called Initial Coin Offerings. Moreover, what is further fueling the development of ICOs, according to BIS Annual Economic Report (2018) is the absence of regulation (even if many countries are currently working on it) and, at the moment, there are just a few examples of banning acts (namely China, India, South Korea). Investors buy ICO tokens in the hope of very high returns, sometimes even before the business is put in place, since the corresponding cryptocurrencies (typically Ethereum) can be immediately traded. In the first 6 months of 2018, there have been 440 ICOs, with a peak in May (125) raising more than 10 billion US, where Telegram ICO (Presale 1 and 2) is by far the most reworded one with 1.7 billion US (Coinschedule, 2018). In 2017, the total amount raised by 210 ICOs was about 4 billion US and overcame venture capital funneled toward high tech initiatives in the same period. The first token sale was held by Mastercoin in July 2013 but one of the most successful and still operative is Ethereum which raised 3,700 BTC in its first 12 h in 2014, equal to approximately 2.3 million dollars at that time.

Recently, there has been a growing literature studying the ICOs drivers aiming to predict their future outcome. A previous study offers an exploratory empiric classification of ICOs and the dynamics of voluntary disclosures. It examines to what extent the availability and quality of the information disclosed can explain the characteristics of success and failure among ICOs and the corresponding projects (Blaseg, 2018). Another important research focuses on the effectiveness of signaling ventures and ICOs projects technological capabilities to attract higher amounts of funding (Fisch, 2019). Momtaz aims at identifying the likelihood and possible timeframe of value creation for investors by combining several factors (financial return, amount of capital raised, listing, and delisting alternatives, industry events study etc.) to analyse the ICOs success drivers (Momtaz, 2018a).

Other streams of research concentrate on the impact of managers quality on the ICOs. Momtaz studies the impact of CEOs loyalty disposition and the magnitude of asymmetry of information between managers and investors on ICOs performance (Momtaz, 2018b). Moreover, to remain in the management area, an interesting spark comes from a research specifically directed on CEOs role and effects on ICOs results (Momtaz, 2018c). Finally, another area of studies focuses on the driving factors impacting the liquidity and trading volume of crypto tokens listed after the ICOs. Among those factors have been identified the quality level of disclosed documentation (source code public on Github, white paper published, an intended budget published for use of proceeds), the community engagement (measured by the number of Telegram group members), the level of preparation of the management (using as proxy the entrepreneurial professional background of the lead founder or CEO), and other outcomes of interest (i.e., the amount raised in the ICO, outright failure—delisting or disappearance, abnormal returns, and volatility) (Howell et al., 2018).

Despite the interest that has been peaked by ICOs and the constantly growing trends, it is worth mentioning that almost half of ICOs sold in 2017 failed by February 2018 (Hankin, 2018). In fact, what should drive more attention to ICOs is the consistent presence of scam activities only devoted to raising money in a fraudulent way. According to Cointelegraph, the Ethereum network (the prevalent blockchain platform for ICOs) has experienced considerable phishing, Ponzi schemes, and other scam events, accounting for about 10% of ICOs (Ethereumscamdb, 2018). On the other hand, it is interesting to assess what factors affect the probability of success of an ICO. Adhami et al. (2017), based on the analysis of 253 ICOs, showed that the following characteristics contribute: the availability of the code source, the organization of a token presale and the possibility for contributors to access to a specific service (or to share profits).

The boom of the ICOs projects and their interesting characteristic brought an important rise of interest from the general audience, many scientific studies have been conducted and published in the last years. Besides the aforementioned Adhami et al. (2017), we should mention the working paper by Zetzsche et al. (2017), that is focused on legal and financial risk aspects of ICOs, moreover a taxonomy is provided and some additional data on ICOs that the authors claim to be continuously updated. Recently, Subramanian in 2018 quoted the ICOs as an example of the decentralized blockchain-based electronic marketplace. The main source of information about blockchain, tokens, and ICOs is obviously the Web. Here we can find sites enabling to explore the various blockchains associated to the main cryptocurrencies, including Ethereum's one. We can also find websites giving extensive financial information on prices of all the main cryptocurrencies and tokens, sites specialized in listing the existing ICOs and giving information about them. Often, these sites evaluate the soundness and likeliness of success of the listed ICOs. One of the most popular among these sites is icobench.com, which evaluates all the listed ICOs and provides an API (Application Programming Interface) to automatically gather information on them. ICOs are usually characterized by the following features: a business idea, most of the time explained in a white paper, a proposed team, a target sum to be collected, a given number of tokens to be given to subscribers according to a predetermined exchange rate with one or more existing cryptocurrencies. Nowadays, a high percentage of ICOs are managed through Smart Contracts running on Ethereum blockchain, and in particular through ERC-20 Token Standard Contract (Fenu et al., 2018).

On top of all the characteristics explained so far, there is a further and not yet explored point of interest: the Telegram chats. Telegram is a cloud-based instant messaging and voice over IP service developed by Telegram Messenger founded by the Russian entrepreneur Pavel Durov. In March 2018, Telegram stated that it has 200 million monthly active users—"This is an insane number by any standards. If Telegram were a country, it would have been the sixth largest country in the world (Telegram, 2018)." Telegram is completely free and has no ads, users can send any kind of media or documents and can program messages to self-destruct after a certain period of time. Some characteristics are imposing Telegram among the first social networks, indeed it intentionally does not collect data about where its clients live and what they use the platform for. This is one of the main reason why, according to AppAnnie rankings, Telegram is particularly popular in countries like Uzbekistan, Ukraine, and Russia, where Internet access may be limited or closely monitored by the government. As of October 2017, Telegram was by far the most popular official discussion platform for current and upcoming ICOs, with 75%+ of these projects employing it. This means that retrieving Telegram discussions associated with each and every ICO would produce a huge amount of textual information potentially useful for understanding the chance of success and more interestingly possible signs of fraudulent activities.

3. METHODOLOGY

In this paper we leverage two kinds of information: structured and unstructured ones. Regarding the former, we take advantage of classical statistical classification models to distinguish the stauts of an ICO that made up of 3 classes, intended as follows:

- **Success**: the ICO collects the predefined hard cap within the time horizon of the campaign;
- **Failure**: the ICO does not collect the predefined hard cap within the time horizon of the campaign;
- **Scam**: the ICO is discovered to be a fraudulent activity with malicious intent during the campaign and described as such by all the platforms we use for data gathering (namely ICObench and Telegram). A robustness check for the scam labeling come by checking if regulatory bodies announced legal actions against the issuers (e.g., official SEC announcements of legal infringement).

Logistic regression aims at classifying the dependent variable into two groups, characterized by a different status [1 = scam vs 0 = success or 1 = success vs 0 = failure] according to the following model:

$$ln(\frac{p_i}{1-p_i}) = \alpha + \sum_j \beta_j x_{ij}, \quad (1)$$

where p_i is the probability of the event of interest, for ICO i, $x_i = (x_{i1}, \ldots, x_{ij}, \ldots, x_{iJ})$ is a vector of ICOs-specific explanatory variables, the intercept parameter α, as well as the regression coefficients β_j, for $j = 1, ..., J$, are to be estimated from the available data. It follows that the probability of success (or scam) can be obtained as:

$$p_i = \frac{1}{1 + exp-(\alpha + \sum_j \beta_j x_{ij})}, \quad (2)$$

Since the target variable is naturally categorized according to three classes, success, failure, and scam we extend the aforementioned binary logistic regression to a multinomial one. Such model assesses all the categories of interest at the same time as follows:

$$ln(\frac{p_k}{1-p_K}) = \alpha_k + \sum_j \beta_k x_{ij}, \quad (3)$$

where p_k is the probability of kth class for $k = 1, ..., K$ given the constraint that $\sum_K p_k = 1$.

Considering the textual analysis of Telegram chats, we take advantage of quantitative analysis of human languages to discover common features of written text. In particular the analysis of relatively short text messages like those appearing on micro-blogging platform presents a number of challenges. Some of these are, the informal conversation (e.g., slang words, repeated letters, emoticons) and the level of implied knowledge necessary to understand the topics of discussion. Moreover, it is important to consider the high level of noise contained in the chats, witnessed by the fact that only a fraction of them with respect to the total number available is employed in our sentiment analysis.

We have applied a Bag of Word (BoW) approach, according to which a text is represented as an unordered collection of words, considering only their counts in each comment of the chat. The word and document vectorization has been carried out by collecting all the word frequencies in a Term Document Matrix (TDM). Afterwards, such matrix has been weighted by employing the popular TF-IDF (Term Frequency Inverse Document Frequency) algorithm. Classical text cleaning procedures have been put in place like stop-words, punctuation, unnecessary symbols and space removal, specific topic words addition. For descriptive purposes we have used word-clouds for each and every Telegram chat according to the general content and to specific subcategories like sentiments and expressed moods. The most critical part of the analysis relies on the sentiment classification. In general, two different approaches can be used:

- Score dictionary based: the sentiment score is based on the number of matches between predefined list of positive and negative words and terms contained in each text source (a tweet, a sentence, a whole paragraph);
- Score classifier based: a proper statistical classifier is trained on a large enough dataset of pre-labeled examples and then used to predict the sentiment class of a new example.

However, the second option is rarely feasible because in order to fit a good classifier, a huge amount of pre-classified examples is needed and this represents a particularly complicated task when dealing with short and extremely non conventional text like micro-blogging chats (Cerchiello and Nicola, 2018). Insofar, we decided to focus on a dictionary based approach, adapting appropriate lists of positive and negative words relevant to ICOs topics in English language. We employ three vocabularies from the R package "tidytext":

- AFINN from Finn Årup Nielsen;
- BING from Bing Liu and collaborators;
- NRC from Saif Mohammad and Peter Turney.

These lexicons are based on unigrams, i.e., single words, they contain many English words and the words are labeled with scores for positive/negative sentiment and also possibly emotions like joy, anger, sadness, and so forth. The NRC lexicon categorizes words in a binary fashion ("yes"/"no") into categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The BING lexicon categorizes words into a binary manner into positive and negative categories. The AFINN lexicon assigns words with a score that runs between −5 and 5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment. By applying the above described lexicons, we produce for each and every ICO a sentiment score as well as counts for positive and negative words. All these indexes are used as additional predictors within the logistic models.

4. DATA

In this paper, we examine 196 ICOs starting from January 2017 till November 2018. For each project, we gather information from web-based sources, mainly rating platforms such as icobench.com, TokenData.io, ICO Drops.com, CoinDesk.com and project's websites. The process of building up the ICOs data set reflects the main phases that an ICO follows to be launched: from the birth of the business idea, the team building, the purpose of the tokens, the technical requirements (white paper), the promotion and the execution phase.

4.1. Collection of Structured Data

The first step in collecting data about each project is to gather information from the most used ICO related platforms as Icobench, TokenData, Coinschedule, or similar. During such phase, we look for general characteristics such as the name, the token symbol, start, and end dates of the crowdfunding, the country of origin, financial data such as the total number of issued token, the initial price of the token, the platform used, data on the team proposing the ICO, data on the advisory board, data on the availability of the website, availability of white paper and social channels.

Some of these data, such as short and long description, and milestones are textual descriptions. Others are categorical variables, such as the country, the platform, and variables related to the team members (name, role, group). The remaining variables are numeric, with different degrees of discretization. Unfortunately, not all ICOs record all variables, so there are several missing data. The ICO web databases that we use are fully checked in order to minimize the missing values of one of the platforms, therefore we validate the information checking for the details on the website and on the white paper. As a result, the complete set of reliable information comes from the matching between the website and the white paper.

The variables set, continuous and categorical data, show us that the main area of origin of the projects is Europe with the highest percentage in Switzerland and Germany. The Switzerland peak is due to the national regulator approach—FINMA (the Swiss Financial Market Supervisory Authority)—which on 16 February 2018 issued clear guidance on the status of ICOs. FINMA does not categorize payment or utility tokens (provided they are not used for investment) as securities. All other tokens are categorized as securities and are subject to securities regulation. To legally issue an equity/asset token, authorization from FINMA should be sought, and appropriate compliance measures [know your customers (KYC) and anti-money laundering (AML)] must be taken. If a debt token can be classified as a deposit, then unless specific exceptions apply, a banking license is needed prior to the ICO. In the fragmented regulatory framework, this is one of the so-called "crypto-friendly" countries, that attract worldwide investors.

The presence of a team of experts as a figure of "advisors" that follows the stages of development are helpful in qualifying the ICO as more reliable. On the development of the dataset the research focused also on assessing the number of advisors for each ICO, checking their educational background and marking as a variable of interest the presence of a Ph.D. that attests a high degree of education.

The evolution of the classic Business Plan that we observe when we analyse the idea of a start-up, is called White Paper. The business plan is the document that illustrates the strategic intentions and the management of competitive strategies of the company, the evolution of key value drivers and the economic and financial results. The drawing up of the operational plan has the aim of achieving different subjects involved in the business. The content of the business plan should not be overlooked, it must be the most possible schematic and of intuitive interpretation. The feature that distinguishes a good plan is the clarity, the synthesis and the professional description of the project workflow. The WP (white paper) therefore fulfills these functions and, in our analysis, played a vital role in the statistical analysis in terms of presence or absence of it. The graphics quality with which it is produced is also important, the data contained within it and the description of the team's components.

4.2. Collection of Unstructured Data

Social channels are more personal than every database, rating platform or websites, so they are a way to reach a wide range of users, to update them constantly about the evolution of the project and in the end to create a trusty environment that can finalize in a successful crowdfunding activity. In order to conduct the textual analysis, we enrich our database with the social channels data, such as the presence of a channel, the numbers of users as a proxy of the community engagement and as mentioned in the introduction the textual chat, retrieved in reverse until the creation of the chat. The most used social channels are Telegram, Twitter, Facebook, Bitcointlak, Medium, while Linkedin, Reddit, and Slack are not frequently used.

In crowdfunding projects the entrepreneur and the community in which is embedded works as a strong control for the attractiveness of a business. Some studies have investigated the social network community and the entrepreneurial activity finding out that the amount of capital collected in crowdfunding is heavily dependent on the range of social networks the entrepreneurs belong to (Mollick, 2014).

With regards to the entrepreneurial dimension, we investigate the team components, pointing out that the members checked until now are almost 1,000, with a median size of 7 for project. For each team member we checked general information related to the social engagement, looking for the Linkedin channel activity (48% of them do not have an individual page), the numbers of connections, the job position in the project and the academic background. Moreover, the presence of advisors can play a crucial role in ensuring the reliability of an ICO, provided a wise choice of such advisors. The same applies to institutional investors doing due diligence on a potential venture. In collecting our data, we focused on the academic background and the current area of expertise of the declared advisors.

As it concerns the unstructured data, insightful information can be derived by the white papers in terms of quality of the technical report and specific content. A white paper is a summary report that provides detailed information about the project, its originality and the benefits it can give to investors and users, about the technological features, team behind the project, project's background and future plans. Besides all the above information, we collect Telegram chats associated to each ICO (if available) and apply all the text analytic techniques to produce a sentiment based score.

In **Table 1** we report the complete list of collected and employed variables.

TABLE 1 | Explanatory variables.

class0	f=failed, sc=scam su=success
class1	0=scam, 1=failed+success
class2	0=failed, 1= success
w_site	Website (dummy)
tm	Telegram (dummy)
w_paper	White paper (dummy)
usd	presale price in USD
tw	Twitter (dummy)
fb	Facebook (dummy)
ln	Linkedin (dummy)
yt	Youtube (dummy)
gith	Github (dummy)
slack	Slack (dummy)
reddit	Reddit (dummy)
btalk	Bitcointalk (dummy)
mm	Medium (dummy)
nr_team	Number of Team members (quantitative)
adv	Existence of advisors (dummy)
nr_adv	Number of advisors (quantitative)
project	Official name of the ICO (categorical)
nr_tm	Number of users in Telegram (quantitative)
tot_token	Number of Total Tokens (quantitative)
Pos_Bing	Standardized nr. of positive words for BL list (quantitative)
Neg_Bing	Standardized nr. of negative words for BL list (quantitative)
Sent_Bing	Standardized sentiment for BL list (quantitative)
Pos_NRC	Standardized nr. of positive words for NRC list (quantitative)
Neg_NRC	Standardized nr. of negative words for NRC list (quantitative)
Sent_NRC	Standardized sentiment for NRC list (quantitative)

5. EMPIRICAL EVIDENCE

In this section we report our main results obtained from classification analysis and textual analysis. In this regard, in **Tables 2**, **4** we report results respectively for logistic regression on Success/Failure (class 2 variable) and for multilogit regression estimated on failure (f) and scam (sc) compared to success as baseline. Regarding the first model, in **Table 2** we report the final configuration after several stepwise selection steps[1]. The reader can see that the only two relevant dummy variables are: the presence of a white paper (Paper_du) and of a Twitter account (tw). Both present positive coefficients showing their impact on increasing the probability of success of an ICO. It should be stressed that the influence of Twitter channel is much higher than the presence of a white paper, indeed if we calculate the associated odds ratio we would get, respectively 11.94 and 3.85. In other words, if the ICO has a Twitter account the probability of success is almost 12 times higher (almost 4 times higher for the white paper). Regarding the three continuous variables, number of elements of the team (Nr_team), number of advisors

[1]The full model is available in **Table A1** in the appendix and it evidently contains several not significant variables.

(Nr_adv), and scaled sentiment score based on NRC lexicon (Sent_NRC_sc), they are all highly significant and again positive suggesting that increasing people and advisors in the team has a positive impact. Regarding the sentiment, we notice a particularly high positive value, stressing the importance of the perception of possible investors which interact with the ICO proposer by means of a social media, namely Telegram.

To further evaluate such configuration, we have explored the VIF index that accounts for the level of multicollinearity brought by each and every variable. The VIF results for the two model configurations are reported in **Table 3** (logistic) and 5 (multinomial), with useful insights in defining the lack of multicollinearity[2]. Therefore, in **Table 3** we can see low values for the VIF index associated to the estimated logistic model (given in **Table 2**). The reader can easily notice that there is not any multicollinearity effect, making robust the model. Moreover, reported performance indexes, namely AIC and pseudo R^2, present good values above 50%.

In **Table 4**, we report results for fraudulent and scam ICOs compared to successful ones, on the basis of a multilogit regression. Looking at the estimated parameters, we can infer that the patterns are different. The presence of a website has a positive impact on the probability of being a successful ICO and not a scam. In other words, the absence of this characteristic is a driver of scam activity suspects. Instead the website does not differentiate successful from failures ones. With regards to the presence of advisors and of a white paper, both the variables are significant in differentiating fraudulent from successful ICO, confirming results of logistic regression. No statistical significance for fraudulent ICOs. Lastly, variable on the sentiment score is relevant and with negative sign for both the classes, in other words an increasing in the sentiment causes an increasing in the probability of success when we consider both failed and fraudulent ICOs.

In this regard, we should stress that the incidence of scam ICOs in our database is extremely low, this due to the fact that collecting information about such ICOs is particularly complex. Most of the information is completely deleted from the Web as soon as the activity is recognized as illegal and/or fraudulent. The overall model performance, assessed again in terms of AIC and pseudo R^2, is pretty good although inferior to the previous one.

In **Table 5**, we also report VIF index, so to check the absence of multicollinearity in the reported model. Please note that, multilogit model reported in **Table 4** is a final configuration obtained through stepwise selection. The full models are available in the **Appendix** (Supplementary Material)[3].

6. DISCUSSION AND CONCLUSIONS

Initial coin offerings (ICOs) are one of the several by-products of the cryptocurrencies world. IPOStart-ups and existing businesses are turning to alternative sources of capital as opposed to classical channels like banks or venture capitalists. They can offer the inner value of their business by selling "tokens," i.e., units of the chosen cryptocurrency, like a regular firm would do by means of an. The investors, of course, hope for an increase in the value of the token in the short term,

TABLE 2 | Logistic regression results on success/failure ICOs.

	Dependent variable
	Class 2
tw	2.481*
	(1.381)
Paper_du	1.351**
	(0.635)
nr_adv	0.461***
	(0.135)
nr_team	0.233***
	(0.088)
Sent_NRC_sc	2.187***
	(0.595)
Constant	−3.601**
	(1.458)
Observations	196
Akaike Inf. Crit.	89.41
McFadden pseudo R^2	0.63
McFadden Adj. pseudo R^2	0.57
Cox & Snell pseudo R^2	0.49

$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

TABLE 3 | VIF index for logistic regression model.

tw	Paper_du	nr_adv	nr_team	Sent_NRC_sc
1.229	1.033	1.067	1.053	1.228

TABLE 4 | Results from multilogit regression: failure and scam compared to success.

	Dependent variable	
	sc	f
	(1)	(2)
Oweb_dum	−1.962**	0.093
	(0.977)	(0.773)
adv_dum	−0.899	−1.707***
	(0.809)	(0.571)
Paper_du	−0.728	−2.158***
	(0.915)	(0.657)
Sent_NRC_sc	−1.390*	−2.606***
	(0.731)	(0.703)
Constant	−0.628	−0.572
	(0.997)	(0.925)
Akaike Inf. Crit.	166.339	166.339
Pseudo R square	McFadden 0.43 - McFadden Adj. 0.36- Cox & Snell 0.44	

$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

[2]In **Table A2**, we have VIF index obtained from the full model and there are high values for some variables, specifically those related to sentiment analysis.
[3]Full multilogit regression model is available in **Table A3** in the appendix and in **Table A4** the associated VIF index table.

TABLE 5 | VIF index for multilogit regression model.

Oweb_dum	adv_dum	Paper_du	Sent_NRC_sc
3.656	2.317	3.607	3.870

provided a solid and valid business idea typically described by the ICO issuers in a white paper. However, fraudulent activities perpetrated by unscrupulous actors are frequent and it would be crucial to highlight in advance clear signs of illegal money raising.

In this perspective, ICOs analysis can be considered a very particular type of fraud detection activity. However, in our opinion fraud detection presents some specificity that prevent us from entailing ICOs related problems as a proper instance of fraud detection. In particular, our data are not flowing in such huge amount from an on-line system as typically happens with credit card payments or banks transactions. Typical fraud detection approaches, as in Maheshwara Reddy et al. (2019), aim at discovering, almost in real times, fraudulent financial activities based on transactional data that ideally should be blocked as soon as possible. ICOs instead are characterized by a slow process of engagement of the prospect clients and establishment of consensus that goes through Telegram chats (if available), white paper and website. That being the case, we would suggest to label this specific stream of research as FinTech Fraud detection with all the relative specificity.

While analyzing success vs failure dynamic with a classification model is relatively easy since the incidence of the two classes is almost equal (50–50), it is much more complicated to highlight the key aspects that could witness a fraudulent activity since, in the last 3 years, only few scam events have been reported. In our sample made of 196 ICOs (data collection still active) we have 10 scam ICOs and we fit a multilogit regression model for comparing scam and failed ICOs toward successful ones. Results tell us that the presence of a website has a positive impact on the probability of not being a scam but does not have any impact on failed ones. In terms of sentiment expressed on Telegram chats, the impact appears to be negative both on the scam and failed ICOs. This suggests that monitoring in real time Telegram chats could represent a valid mean for collecting signs of possible problems within the ICOs. If instead, we compare Successful ICOs against Failed ones, we find that the presence of a White Paper and of a Twitter account show positive coefficients.

Regarding the three continuous variables, number of elements of the team, number of advisors, and sentiment score based on NRC lexicon, they are all highly significant and positive suggesting that increasing people in the team and advisors has a positive impact. Regarding the sentiment, we notice a particularly high positive value, stressing the importance of the perception of possible investors which interact with the ICO proposer by means of a social media.

The paper will be improved in the future by increasing the size of the sample and exploring alternative approaches for textual analysis with specific attention to sentiment analysis. We aim at producing a more refined and tailored sentiment score for each ICO, improving and increasing the vocabulary of words. Specifically regarding the textual analysis an alternative approach that we could use is the combination of words as in Bolasco and Pavone (2017).

As a final remark, authors are aware of the limits of the paper mainly due to the size of the sample. However, given the still limited literature in this field with no reference to the power of textual information collectable through Telegram chats, this contribution represents a step ahead in the process of understanding the ICOs phenomenon. Furthermore a different approach would be to study the trends of the ICOs by combining the available information from specialized websites on fraudulent activities (such as cyphertrace.com and deadcoin.com) and rating websites for the active projects.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

The paper is the product of full collaboration among the authors, however PC has inspired the idea, the methodology and wrote sections 1, 3 and 6, AT run the analysis and wrote sections Keywords, 2, 4 and 5.

FUNDING

This research has received funding from the European Union's Horizon 2020 research and innovation program FIN-TECH: A Financial supervision and Technology compliance training program under the grant agreement No. 825215 (Topic: ICT-35-2018, Type of action: CSA). The current version of the work is substantially based on refining a working paper (not reviewed) published by the same authors in DEM Working Paper Series ICOs success drivers: a textual and statistical analysis.

ACKNOWLEDGMENTS

We acknowledge Federico Campasso that technically helped us in collecting the data and Prof. Silvio Vismara for the useful advises on the work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2020.00018/full#supplementary-material

REFERENCES

Adhami, S., Giudici, G., and Martinazzi, S. (2017). *Why Do Businesses Go Crypto? An Empirical Analysis of Initial Coin Offerings*. Available online at: https://ssrn.com/abstract=3046209

BIS Annual Economic Report (2018). *V. Cryptocurrencies: Looking Beyond the Hype*.

Blaseg, D. (2018). *Dynamics of Voluntary Disclosure in the Unregulated, Market for Initial Coin Offerings*. Available online at: https://ssrn.com/abstract=3207641 or https://dx.doi.org/10.2139/ssrn.3207641

Bolasco, S., and Pavone, P. (2017). "Automatic dictionary-and rule-based systems for extracting information from text," in *Data Analysis and Classification (189-198)* (Berlin; Heidelberg: Springer).

Cerchiello, P., and Nicola, G. (2018). Assessing news contagion in finance. *Econometrics* 6:5. doi: 10.3390/econometrics6010005

Cerchiello, P., Tasca, P., and Toma, A. M. (2019). ICO success drivers: a textual and statistical analysis. *J. Altern. Invest.* 21, 13–25. doi: 10.3905/jai.2019.21.4.013

Coinmarketcap (2018). *All Coins*. Available online at: https://coinmarketcap.com/coins/views/all/ (accessed June 18, 2018).

Coinschedule (2018). *Cryptocurrency ICO Stats 2018*. Available online at: https://coinschedule.com/stats.html (accessed June 18, 2018).

Ethereumscamdb (2018). *Ethereum Scam Database*. Available online at: https://etherscamdb.info/ (accessed June 18, 2018).

Fenu, G., Marchesi, L., Marchesi, M., and Tonelli, R. (2018). "The ICO phenomenon and its relationships with ethereum smart contract environment," in *2018 International Workshop on Blockchain Oriented Software Engineering (IWBOSE)* (IEEE), 26–32.

Fisch, C. (2019). Initial Coin Offerings (ICOs) to finance new ventures. *J. Bus. Ventur.* 34, 1–22. doi: 10.2139/ssrn.3147521

Hankin, A. (2018). *Nearly Half of All 2017 ICOs Have Failed*. Fortune (Retrieved May 4, 2018).

Howell, S. T., Niessner, M., and Yermack, D. (2018). *Initial Coin Offerings: Financing Growth With Cryptocurrency Token Sales (No. w24774)*. National Bureau of Economic Research.

Maheshwara Reddy, C., Saiteja, M., Shashank, B., and Mrs.Dhikhi. (2019). *Financial Fraud Detection Using Machine Learning*. IJIRT.

Mollick, E. (2014). The dynamics of crowdfunding: an exploratory study. *J. Bus. Venturing* 29, 1–16. doi: 10.1016/j.jbusvent.2013.06.005

Momtaz, P. P. (2018a). *CEO Emotions and Underpricing in Initial Coin Offerings*. Working paper, UCLA Anderson.

Momtaz, P. P. (2018b). *Initial Coin Offerings, Asymmetric Information, and Loyal CEOs*. SSRN. https://dx.doi.org/10.2139/ssrn.3167061

Momtaz, P. P. (2018c). *Initial Coin Offerings*. SSRN. Available online at: https://ssrn.com/abstract=3166709

Peer-to-Peer Research Group (2013). *Peer-to-Peer Research Group*.? Available online at: https://irtf.org/concluded/p2prg (accessed 18 June 2018).

Subramanian, H. (2018). Decentralized blockchain-based electronic marketplaces. *Commun. ACM* 61, 78–84. doi: 10.1145/3158333

Wright, A., De Filippi, P. (2015). *Decentralized Blockchain Technology and the Rise of Lex Cryptographia*. Available online at: https://ssrn.com/abstract=2580664

Zetzsche, D. A., Buckley, R. P., Arner, D. W., and Für, L. (2017). *The ICO Gold Rush:It's a Scam, It's a Bubble, It's a Super Challenge for Regulators*. Univ.Luxembourg Law Working Paper No. 11/2017. Available online at: https://ssrn.com/abstract=3072298

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.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 24 April 2020
doi: 10.3389/frai.2020.00026

Explainable AI in Fintech Risk Management

Niklas Bussmann[1,2], Paolo Giudici[3]*, Dimitri Marinelli[1] and Jochen Papenbrock[1]

[1] FIRAMIS, Frankfurt, Germany, [2] Department of Economics and Management, University of Pavia, Pavia, Italy, [3] Fintech Laboratory, Department of Economics and Management, University of Pavia, Pavia, Italy

The paper proposes an explainable AI model that can be used in fintech risk management and, in particular, in measuring the risks that arise when credit is borrowed employing peer to peer lending platforms. The model employs Shapley values, so that AI predictions are interpreted according to the underlying explanatory variables. The empirical analysis of 15,000 small and medium companies asking for peer to peer lending credit reveals that both risky and not risky borrowers can be grouped according to a set of similar financial characteristics, which can be employed to explain and understand their credit score and, therefore, to predict their future behavior.

Keywords: credit risk management, explainable AI, financial technologies, peer to peer lending, logistic regression, predictive models

OPEN ACCESS

Edited by:
Shatha Qamhieh Hashem, An-Najah National University, Palestine

Reviewed by:
Paolo Barucca, University College London, United Kingdom
Bertrand Kian Hassani, University College London, United Kingdom

Correspondence:
Paolo Giudici
giudici@unipv.it

Specialty section:
This article was submitted to Artificial Intelligence in Finance, a section of the journal Frontiers in Artificial Intelligence

Received: *18 December 2019*
Accepted: *30 March 2020*
Published: *24 April 2020*

Citation:
Bussmann N, Giudici P, Marinelli D and Papenbrock J (2020) Explainable AI in Fintech Risk Management. Front. Artif. Intell. 3:26. doi: 10.3389/frai.2020.00026

1. INTRODUCTION

Black box Artificial Intelligence (AI) is not suitable in regulated financial services. To overcome this problem, Explainable AI models, which provide details or reasons to make the functioning of AI clear or easy to understand, are necessary.

To develop such models, we first need to understand what "Explainable" means. During this year, some important benchmark definitions have been provided, at the institutional level. We report some of them, in the context of the European Union.

For example, the Bank of England (Joseph, 2019) states that "Explainability means that an interested stakeholder can comprehend the main drivers of a model-driven decision." The Financial Stability Board (FSB, 2017) suggests that "lack of interpretability and auditability of AI and ML methods could become a macro-level risk." Finally, the UK Financial Conduct Authority (Croxson et al., 2019) establishes that "In some cases, the law itself may dictate a degree of explainability."

The European GDPR (EU, 2016) regulation states that "the existence of automated decision-making, should carry meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject." Under the GDPR regulation, the data subject is therefore, under certain circumstances, entitled to receive meaningful information about the logic of automated decision-making.

Finally, the European Commission High-Level Expert Group on AI presented the Ethics Guidelines for Trustworthy Artificial Intelligence in April 2019. Such guidelines put forward a set of seven key requirements that AI systems should meet in order to be deemed trustworthy. Among them three related to XAI, and are the following.

- Human agency and oversight: decisions must be informed, and there must be a human-in-the-loop oversight.
- Transparency: AI systems and their decisions should be explained in a manner adapted to the concerned stakeholder. Humans need to be aware that they are interacting with an AI system.
- Accountability: AI systems should develop mechanisms for responsibility and accountability, auditability, assessment of algorithms, data and design processes.

Following the need to explain AI models, stated by legislators and regulators of different countries, many established and startup companies have started to embrace Explainable AI (XAI) models.

From a mathematical viewpoint, it is well-known that, while "simpler" statistical learning models, such as linear and logistic regression models, provide a high interpretability but, possibly, a limited predictive accuracy, "more complex" machine learning models, such as neural networks and tree models provide a high predictive accuracy at the expense of a limited interpretability.

To solve this trade-off, we propose to boost machine learning models, that are highly accurate, with a novel methodology, that can explain their predictive output. Our proposed methodology acts in the post-processing phase of the analysis, rather than in the preprocessing part. It is agnostic (technologically neutral) as it is applied to the predictive output, regardless of which model generated it: a linear regression, a classification tree or a neural network model.

More precisely, our proposed methodology is based on Shapley values (see Lundberg and Lee, 2017 and references therein). We consider a relevant application of AI in financial technology: peer to peer lending.

We employ Shapley values to predict the credit risk of a large sample of small and medium enterprises which apply for credit to a peer to peer lending platform. The obtained empirical evidence shows that, while improving the predictive accuracy with respect to a standard logistic regression model, we maintain and, possibly, improve, the interpretability (explainability) of the results.

In other words, our results confirm the validity of this approach in discriminating between defaulted and sound institutions, and it shows the power of explainable AI in both prediction accuracy and in the interpretation of the results.

The rest of the paper is organized as follows: section 2 introduces the proposed methodology. Section 3 shows the results of the analysis in the credit risk context. Section 4 concludes.

2. METHODOLOGY

2.1. Credit Risk in Peer to Peer Lending

Credit risk models are useful tools for modeling and predicting individual firm default. Such models are usually grounded on regression techniques or machine learning approaches often employed for financial analysis and decision-making tasks.

Consider N firms having observation regarding T different variables (usually balance-sheet measures or financial ratios). For each institution n define a variable γ_n to indicate whether such institution has defaulted on its loans or not, i.e., $\gamma_n = 1$ if company defaults, $\gamma_n = 0$ otherwise. Credit risk models develop relationships between the explanatory variables embedded in T and the dependent variable γ.

The logistic regression model is one of the most widely used method for credit scoring. The model aims at classifying the dependent variable into two groups, characterized by different status (defaulted vs. active) by the following model:

$$ln(\frac{p_n}{1-p_n}) = \alpha + \sum_{t=1}^{T} \beta_t x_{nt} \tag{1}$$

where p_n is the probability of default for institution n, $\mathbf{x}_i = (x_{i,1}, ..., x_{i,T})$ is the T-dimensional vector of borrower specific explanatory variables, the parameter α is the model intercept while β_t is the t-th regression coefficient. It follows that the probability of default can be found as:

$$p_n = (1 + exp(\alpha + \sum_{t=1}^{T} \beta_t x_{nt}))^{-1} \tag{2}$$

2.2. Machine Learning of Credit Risk

Credit risk can be measured with very different Machine Learning (ML) models, able to extract non-linear relations among the financial information in the balance sheets. In a standard data science life cycle, models are chosen to optimize the predictive accuracy. In highly regulated sectors, like finance or medicine, models should be chosen balancing accuracy with explainability (Murdoch et al., 2019). We improve the choice selecting models based on their predictive accuracy, and employing *a posteriori* an explanations algorithm. This does not limit the choice of the best performing models.

To exemplify our approach we consider, without loss of generality, the XGBoost model, one of the most popular and fast algorithm (Chen and Guestrin, 2016), that implements gradient tree boosting learning models.

2.3. Learning Model Comparison

For evaluating the performance of each learning model, we employ, as a reference measure, the indicator $\gamma \in \{0, 1\}$, a binary variable which takes value one whenever the institutions has defaulted and value zero otherwise. For detecting default events represented in γ, we need a continuous measurement $p \in [0, 1]$ to be turned into a binary prediction B assuming value one if p exceeds a specified threshold $\tau \in [0, 1]$ and value zero otherwise. The correspondence between the prediction B and the ideal leading indicator γ can then be summarized in a so-called confusion matrix.

From the confusion matrix we can easy illustrate the performance capabilities of a binary classifier system. To this aim, we compute the receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC). The ROC curve plots the false positive rate (FPR) against the true positive rate (TPR), as follows:

$$FPR = \frac{FP}{FP + TN} \tag{3}$$

$$TPR = \frac{TP}{TP + FN} \tag{4}$$

The overall accuracy of each model can be computed as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

and it characterizes the proportion of true results (both true positives and true negatives) among the total number of cases.

2.4. Explaining Model Predictions

We now explain how to exploit the information contained in the explanatory variables to localize and cluster the position of each individual (company) in the sample. This information, coupled with the predicted default probabilities, allows a very insightful explanation of the determinant of each individual's creditworthiness. In our specific context, information on the explanatory variables is derived from the financial statements of borrowing companies, collected in a vector $\mathbf{x}_n$ representing the financial composition of the balance sheet of institution n.

We propose calculate the Shapley value associated with each company. In this way we provide an agnostic tool that can interpret in a technologically neutral way the output from a highly accurate machine learning model. As suggested in Joseph (2019), the Shapley values of a model can be used as a tool to transfer predictive inferences into a linear space, opening a wide possibility of using the toolbox of econometrics, hypothesis testing, and network analysis.

We develop our Shapley approach using the SHAP (Lundberg and Lee, 2017) computational framework, which allows to express each single prediction as a sum of the contributions of the different explanatory variables.

More formally, the Shapley explanation model for each prediction $\phi(\hat{f}(x_i))$ is obtained by an additive feature attribution method, which decomposes them as:

$$\phi(\hat{f}(\mathbf{x}_i)) = \phi_0 + \sum_{k=1}^{M} \phi_k(x_i). \tag{6}$$

where M is the number of available explanatory variables, $\phi \in \mathbb{R}^{\mathbb{M}}$, $\phi_k \in \mathbb{R}$. The local functions $\phi_k(x_i)$ are called Shapley values.

Indeed, Lundberg and Lee (2017) prove that the only additive feature attribution method that satisfies the properties of *local accuracy*, *missingness*, and *consistency* is obtained attributing to each feature $x_k, k = 1, \ldots, M$, a SHapley Additive exPlanation (SHAP) defined by

$$\phi_k(x_i) = \sum_{x' \subseteq \mathcal{C}(x) \setminus x_k} \frac{|x'|!(M - |x'| - 1)!}{M!} [\hat{f}(x' \cup x_k) - \hat{f}(x')], \tag{7}$$

where $\mathcal{C}(x) \setminus x_k$ is the set of all the possible models excluding variable x_k (with $m = 1, \ldots, M$), $|x'|$ denotes the number of variables included in model x', M is the number of the available variables, $\hat{f}(x' \cup x_k)$ and $\hat{f}(x')$ are the predictions associated with all the possible model configurations including variable x_k and excluding variable x_k, respectively.

The quantity $\hat{f}(x' \cup x_k) - \hat{f}(x')$ defines the contribution of variable x_k to each individual prediction.

3. APPLICATION

3.1. Data

We test our proposed model to data supplied by European External Credit Assessment Institution (ECAI) that specializes in credit scoring for P2P platforms focused on SME commercial lending. The data is described by Giudici et al. (2019a) to which we refer for further details. In summary, the analysis relies on a dataset composed of official financial information (balance-sheet variables) on 15,045 SMEs, mostly based in Southern Europe, for the year 2015. The information about the status (0 = active, 1 = defaulted) of each company 1 year later (2016) is also provided. Using this data, Giudici (2018), Ahelegbey et al. (2019), and Giudici et al. (2019a,b) have constructed logistic regression scoring models that aim at estimating the probability of default of each company, using the available financial data from the balance sheets and, in addition, network centrality measures that are obtained from similarity networks.

Here we aim to improve the predictive performance of the model and, for this purpose, we run an XGBoost tree algorithm (see e.g., Chen and Guestrin, 2016). To explain the results from the model, typically highly predictive, we employ Shapley values.

The proportion of defaulted companies within this dataset is 10.9%.

3.2. Results

We first split the data in a training set (80%) and a test set (20%).

We then estimate the XGBoost model on the training set, apply the obtained model to the test set and compare it with the optimal logistic regression model. The ROC curves of the two models are contained in **Figure 1** below.

From **Figure 1** note that the XGBoost clearly improves predictive accuracy. Indeed the calculation of the AUROC of the two curves indicate an increase from 0.81 (best logistic regression model) to 0.93 (best XGBoost model).

We then calculate the Shapley values for the companies in the test set.

To exemplify our results, **Figure 2** we provide the interpretation of the estimated credit scoring of four companies: two that default and two that do not default.

Figure 2 clearly shows the advantage of our explainable model. It can indicate which variables contribute more to the prediction. Not only in general, as is typically done by feature selection models, but differently and specifically for each company in the test set. Note how the explanations are rather different ("personalized") for each of the four considered companies.

4. CONCLUSIONS

The need to leverage the high predictive accuracy brought by sophisticated machine learning models, making them interpretable, has motivated us to introduce an agnostic, post-processing methodology, based on Shapley values. This allows to explain any single prediction in terms of the potential contribution of each explanatory variable.

Future research should include a better understanding of the predictions through clustering of the Shapley values. This can

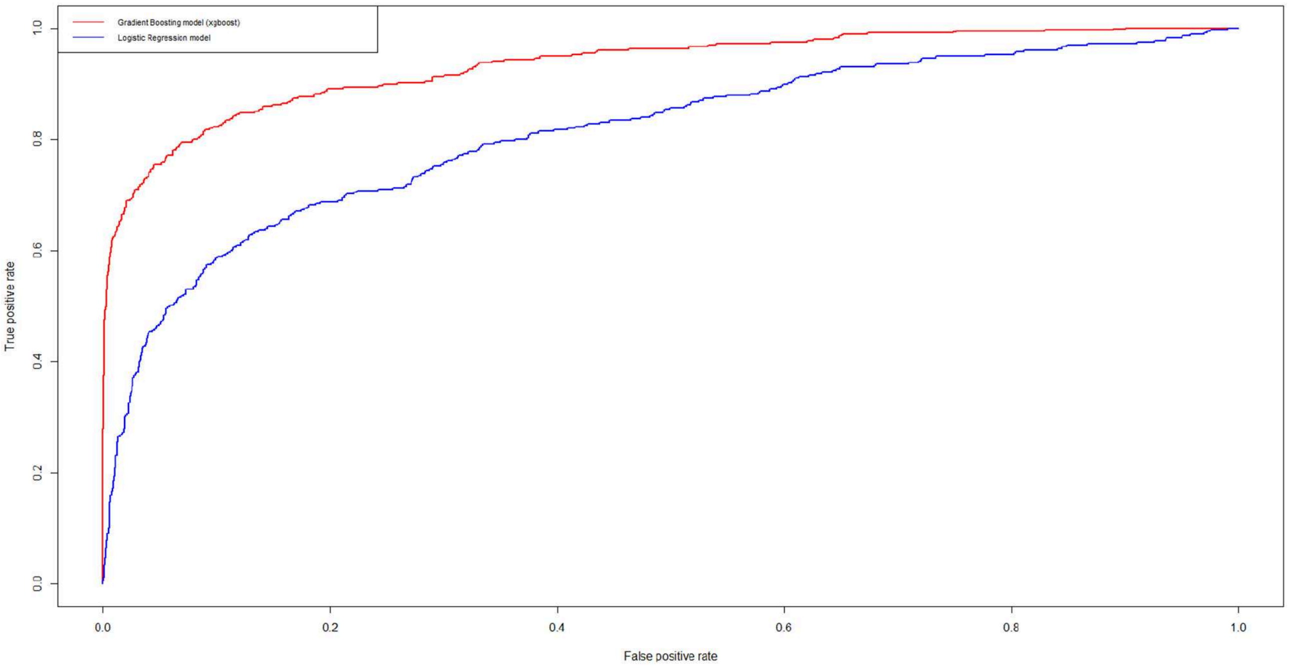

FIGURE 1 | Receiver Operating Characteristic (ROC) curves for the logistic credit risk model and for the XGBoost model. In blue, we show the results related to the logistic models while in red we show the results related to the XGBoost model.

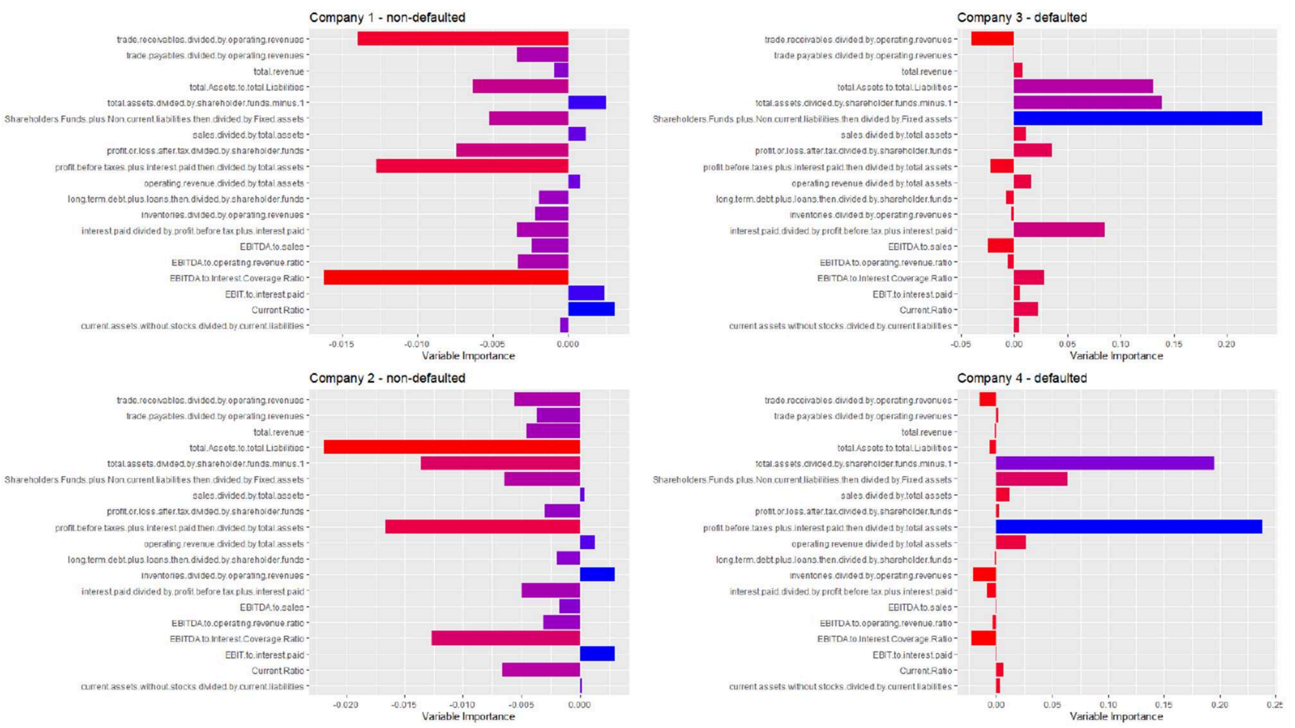

FIGURE 2 | Contribution of each explanatory variable to the Shapley's decomposition of four predicted default probabilities, for two defaulted and two non-defaulted companies. A red color indicates a low variable importance, and a blue color a high variable importance.

be achieved, for example, using correlation network models. A second direction would be to extend the approach developing model selection procedures based on Shapley values, which would require appropriate statistical testing. A last extension would be to develop a Shapley like measure that applies also to ordinal response variables.

Our research has important policy implications for policy makers and regulators who are in their attempt to protect the consumers of artificial intelligence services. While artificial intelligence effectively improve the convenience and accessibility of financial services, they also trigger new risks, and among them is the lack of model interpretability. Our empirical findings suggest that explainable AI models can effectively advance our understanding and interpretation of credit risks in peer to peer lending.

Future research may involve further experimentation and the application to other case studies.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request 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 research has received funding from the European Union's Horizon 2020 research and innovation program FIN-TECH: A Financial supervision and Technology compliance training programme under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA), and from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 750961.

ACKNOWLEDGMENTS

In addition, the authors thank ModeFinance, a European ECAI, for the data; the partners of the FIN-TECH European project, for useful comments and discussions.

REFERENCES

Ahelegbey, D., Giudici, P., and Hadji-Misheva, B. (2019). Latent factor models for credit scoring in P2P systems. *Phys. A Stat. Mech. Appl.* 522, 112–121. doi: 10.1016/j.physa.2019.01.130

Chen, T., and Guestrin, C. (2016). "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: ACM), 785–794.

Croxson, K., Bracke, P., and Jung, C. (2019). *Explaining Why the Computer Says 'no'*. London, UK: FCA-Insight.

EU (2016). Regulation (EU) 2016/679–general data protection regulation (GDPR). *Off. J. Eur. Union*.

FSB (2017). *Artificial Intelligence and Machine Learning in Financial Services–Market Developments and Financial Stability Implication*. Technical report, Financial Stability Board.

Giudici, P. (2018). "Financial data science," in *Statistics and Probability Letters*, Vol. 136 (Elsevier). 160–164.

Giudici, P., Hadji Misheva, B., and Spelta, A. (2019a). Correlation network models to improve P2P credit risk management. *Artif. Intell. Finance*.

Giudici, P., Hadji-Misheva, B., and Spelta, A. (2019b). Network based credit risk models. *Qual. Eng.* 32, 199–211. doi: 10.1080/08982112.2019.1655159

Joseph, A. (2019). *Shapley Regressions: A Framework for Statistical Inference on Machine Learning Models*. Resreport 784, Bank of England.

Lundberg, S., and Lee, S.-I. (2017). "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al. (New York, NY: Curran Associates, Inc.), 4765–4774.

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. U.S.A.* 116, 22071–22080. doi: 10.1073/pnas.1900654116

Conflict of Interest: NB, DM, and JP are employed by the company FIRAMIS.

The remaining 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.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 24 April 2020
doi: 10.3389/frai.2020.00022

Network Models to Enhance Automated Cryptocurrency Portfolio Management

Paolo Giudici[1]*, *Paolo Pagnottoni*[1]* *and Gloria Polinesi*[2]

[1] Department of Economics and Management, University of Pavia, Pavia, Italy, [2] Department of Economics and Social Sciences, Universitá Politecnica delle Marche, Ancona, Italy

The usage of cryptocurrencies, together with that of financial automated consultancy, is widely spreading in the last few years. However, automated consultancy services are not yet exploiting the potentiality of this nascent market, which represents a class of innovative financial products that can be proposed by robo-advisors. For this reason, we propose a novel approach to build efficient portfolio allocation strategies involving volatile financial instruments, such as cryptocurrencies. In other words, we develop an extension of the traditional Markowitz model which combines Random Matrix Theory and network measures, in order to achieve portfolio weights enhancing portfolios' risk-return profiles. The results show that overall our model overperforms several competing alternatives, maintaining a relatively low level of risk.

Keywords: cryptocurrencies, correlation networks, network centrality, portfolio optimization, random matrix theory, minimal spanning tree

OPEN ACCESS

Edited by:
Shatha Qamhieh Hashem, An-Najah National University, Palestine

Reviewed by:
Audrius Kabasinskas, Kaunas University of Technology, Lithuania
Jirakom Sirisrisakulchai, Chiang Mai University, Thailand

****Correspondence:***
Paolo Giudici
paolo.giudici@unipv.it
Paolo Pagnottoni
paolo.pagnottoni01@universitadipavia.it

Specialty section:
This article was submitted to Artificial Intelligence in Finance, a section of the journal Frontiers in Artificial Intelligence

Received: *06 November 2019*
Accepted: *24 March 2020*
Published: *24 April 2020*

Citation:
Giudici P, Pagnottoni P and Polinesi G (2020) Network Models to Enhance Automated Cryptocurrency Portfolio Management. Front. Artif. Intell. 3:22. doi: 10.3389/frai.2020.00022

1. INTRODUCTION

FinTech innovations are rapidly expanding nowadays, with applications including payments, lending, insurance and asset management, among others. Two technical reports from the Financial Stability Board (FSB) (FSB, 2017a,b)—establish several key drivers for FinTech, i.e., the shift of consumer preferences on the demand side, the change of financial regulations on the supply side and the technology evolution.

In this context, services of automated financial consulting are widely spreading and, in particular robo-advisors[1]. They are supposed to match the investors' risk profile with specific class of financial assets and thereby build an efficient portfolio allocation for each specific client. However, the mechanisms underlying the portfolio construction are often obscure, as well as they arguably do not properly take into account for multivariate dependencies across securities which are key to achieve diversification and, therefore, mitigate financial risk. This is particularly true when dealing with peculiarly volatile markets, such as the cryptocurrency one, which could be one of the future target market of robo-advisors, given its rapidly growing influence in the financial world.

Indeed, after its introduction by Nakamoto (2008), Bitcoin was launched online in 2009 and paved the way for many other cryptocurrencies. As a matter of fact, as of 17 October 2019, the cryptocurrency market capitalization amounts to ∼220 billion USD, with a daily trading volume of roughly 52 billion USD.

Along with descriptive and qualitative studies, many researches dealt with quantitative analysis applied to the cryptocurrency market. In particular, a stream of research focuses on price discovery on Bitcoin markets, aiming to determine which are the leaders and followers of the Bitcoin

[1] An article published on "Statista" in 2019 states that assets under management in the robo-advisory segment amounts to roughly 981 billion USD, as well as that they are expected to grow at an annual growth rate (CAGR 2019–2023) of 27% (source: https://www.statista.com/outlook/337/100/robo-advisors/worldwide).

price formation process (see Brandvold et al., 2015; Pagnottoni and Dimpfl, 2018; Giudici and Abu-Hashish, 2019). Other related researches studied the interconnectedness and spillover in the cryptocurrency market (such as Corbet et al., 2018b; Giudici and Pagnottoni, 2019a,b). Another important area regards the study of Bitcoin derivatives—i.e., options and futures written on Bitcoin, with studies conducted by Corbet et al. (2018a), Baur and Dimpfl (2019), Giudici and Polinesi (2019), and Pagnottoni (2019).

From a methodological viewpoint, we base our analysis on an important stream of literature, which focuses on stock and financial networks built on correlation matrices. The seminal paper by Mantegna (1999) uses correlation matrices to infer the hierarchical structure of stock markets, deriving a distance measure based on correlation matrices and building the so called Minimal Spanning Tree (MST), a graphical representation able to connect assets which are similar in terms of returns in a pairwise manner. After that, a research by Tola et al. (2008) uses the Random Matrix Theory (RMT) together with several clustering techniques and show that this significantly lowers portfolio risks. Subsequently, other papers about portfolio construction involving the network structure of financial assets followed (see Zhan et al., 2015; León et al., 2017; Raffinot, 2017; Ren et al., 2017).

To the best of our knowledge, there are no papers yet that exploit network topologies to build portfolios composed by cryptocurrencies. We fill this gap proposing a model that exploits the network structure of cryptocurrencies to provide a portfolio asset allocation that well compares with traditional ones. Following Mantegna (1999) we use Markowitz' asset allocation as a benchmark, and we check whether our proposal is able to improve on it, in terms of risk/return profile.

Indeed, the originality of the current paper is 2-fold. From a methodological point of view, we improve the traditional (Markowitz, 1952) portfolio allocation strategy by means of RMT and MST and by taking network centralities specifically into account. Moreover, throughout this technique we are able to set a parameter of systemic risk aversion that investors can tune to better match their investment strategies with their own risk profile. From an empirical viewpoint, we apply our methodology to data coming from a nascent and highly volatile market, i.e., the cryptocurrency one. This is particularly interesting, as the cryptocurrency market is rapidly expanding and its opportunities due to the high uncertainty (and volatility) around it are quite appealing, and thus a greater number of investors will likely enter it in the short run.

Our empirical findings confirm the effectiveness of our model in achieving better cumulative portfolio performances, while keeping a relatively low level of risk. In particular, we show that our proposed model which employs RMT, MST and centrality measures rapidly adapts to market conditions, and is able to yield satisfactory performances during bull market periods. During bear market periods—instead—our Network Markowitz model employing RMT and MST realizes the best performances, protecting investors from relatively high losses which are instead generated by many other asset allocation strategies tested. Furthermore, the riskiness of our strategy is still lower than most of the competing model we analyze. These outcomes suggest that a sound combination of the proposed models should be employed in order to achieve an efficient cryptocurrency allocation strategy, which could be also used as robo-advisory toolboxes to improve automated financial consultancy.

The paper proceeds as follows. Section 2 presents our methodology and, particularly, the Random Matrix Theory, the Minimal Spanning Tree and the portfolio construction. Section 3 illustrates our empirical results. Section 4 concludes.

2. METHODOLOGY

2.1. Random Matrix Theory

Random Matrix Theory (RMT) is widely employed in several fields, such as quantum mechanics (Beenakker, 1997), condensed matter physics (Guhr et al., 1998), wireless communications (Tulino et al., 2004), as well as economics and finance (Potters et al., 2005). This technique is able to remove the noise component from the pure signal which is embedded into correlation matrices.

The algorithm tests subsequent empirical eigenvalues of the correlation matrix: $\lambda_k < \lambda_{k+1}; k = 1, \ldots, n$, against the null hypothesis that they are equal to the eigenvalues of a random Wishart matrix $\mathbf{R} = \frac{1}{T}\mathbf{A}\mathbf{A}^{\mathbf{T}}$ of the same size, being $\mathbf{A}$ a $N \times T$ matrix containing N time series of length T. The elements of $\mathbf{A}$ are *i.i.d.* random variables, with zero mean and unit variance.

Marchenko and Pastur (1967) show that as $N \to \infty$ and $T \to \infty$, and the ratio $Q = \frac{T}{N} \geq 1$ is fixed, there is convergence of the sample eigenvalues' density to:

$$f(\lambda) = \frac{T}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{1}$$

with $\lambda \in (\lambda_-, \lambda_+)$, $\lambda_\pm = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}$.

Provided that, if $\lambda_k > \lambda_+$ the null hypothesis is rejected from the k-th eigenvalue onwards. Hence, through a singular value decomposition the RM approach builds up a filtered correlation matrix (see Eom et al., 2009).

In our specific case, consider the continuous log return time series r_i of a generic cryptocurrency i at any time point t. i.e.,

$$r_i^t = logP_i^t - logP_i^{t-1}, \tag{2}$$

where P_i^t is the price of the cryptocurrency i at time t.

Considering a bunch of N cryptocurrency return time series, let $\mathbf{C}$ be the $N \times N$ correlation matrix of the cryptocurrency return time series. The random matrix approach filters the correlation matrix, thus obtaining a new matrix $\mathbf{C}^*$ as:

$$\mathbf{C}^* = \mathbf{V}\Lambda\mathbf{V}^{\mathbf{T}}, \tag{3}$$

with

$$\Lambda = \begin{cases} 0 & \lambda_i < \lambda_+ \\ \lambda_i & \lambda_i \geq \lambda_+ \end{cases}$$

and $\mathbf{V}$ being the matrix of the deviating eigenvectors linked to the eigenvalues which are larger than λ_+.

2.2. The Minimal Spanning Tree

In order to simplify the relationships given by the filtered correlation matrix $\mathbf{C}^*$ obtained from the random matrix approach, we apply the Minimal Spanning Tree representation of the cryptocurrency return time series. This is consistent with the literature on stock similarities (i.e., Mantegna and Stanley, 1999; Bonanno et al., 2003; Spelta and Araújo, 2012).

Given the filtered correlation matrix obtained in the step above, we may derive an Euclidean distance for each pairwise correlation element in the matrix, i.e.,

$$d_{ij} = \sqrt{2 - 2c^*_{ij}}, \tag{4}$$

where c^*_{ij} is a generic element (i, j) of the matrix $\mathbf{C}^*$, with $i, j = 1, ..., N$. Each pairwise distance can be inserted in the so-called distance matrix $\mathbf{D} = \{d_{ij}\}$. The MST algorithm is able to reduce the number of links between the assets from $\frac{N(N-1)}{2}$ to $N - 1$ linking each node to its closest neighbor. In particular, we initially consider N clusters associated to the N cryptocurrencies and, at each subsequent step, we merge two generic clusters l_i and l_j if:

$$d\left(l_i, l_j\right) = \min\left\{d\left(l_i, l_j\right)\right\},$$

with the distance between clusters being defined as:

$$\hat{d}\left(l_i, l_j\right) = \min\left\{d_{pq}\right\},$$

being $p \in l_i$ and $q \in l_j$. This procedure is iteratively repeated until we remain with just one cluster at hand.

Moreover, with the aim of explaining the evolution of relationships evolve over time, Spelta and Araújo (2012) proposed the so-called residuality coefficient, which compares the relative strength of the connections above and below a threshold distance value, i.e.,

$$R = \frac{\sum\limits_{d_{i,j} > L} d_{i,j}^{-1}}{\sum\limits_{d_{i,j} \leq L} d_{i,j}^{-1}} \tag{5}$$

with L being the highest threshold distance value ensuring connectivity of the MST. Intuitively, the residuality coefficient R increases when the number of links increases—meaning the network becomes more sparse, and viceversa lowers with decreasing number of links.

2.3. Network Centrality Measures

In this paper we employ of centrality measures in order to develop a portfolio allocation that takes into account the centrality of a node (cryptocurrency) in the system. Network theory includes several centrality measures, such as the degree centrality, counting how many neighbors a node has, as well centrality measures based on the spectral properties of graphs (see Perra and Fortunato, 2008). Among the spectral centrality measures we remark Katz's centrality (see Katz, 1953), PageRank (Brin and Page, 1998), hub and authority centralities (Kleinberg, 1999), and the eigenvector centrality (Bonacich, 2007).

In this paper we employ of the eigenvector centrality, as it measures the importance of a node in a network by assigning relative scores to all nodes in the network. Relative scores are based on the principle that being connected to few high scoring nodes contributes more to the score of the node in question than equal connections to low scoring nodes. In other words, considering a generic node i, the centrality score is proportional to the sum of the scores of all nodes which are connected to it, i.e.,

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} \hat{d_{i,j}} x_j \tag{6}$$

where x_j is the score of a node j, $\hat{d_{i,j}}$ is the element (i, j) of the adjacency matrix of the network, λ is a constant. The equation from above can be rewritten in a compact form as:

$$\hat{\mathbf{D}x} = \lambda \mathbf{x} \tag{7}$$

where $\hat{\mathbf{D}}$ is the adjacency matrix, λ is the eigenvalue of the matrix $\hat{\mathbf{D}}$, with associated eigenvector $\mathbf{x}$, a vector of scores of dimension N, meaning one element for each node. Note that as our networks are based on distances between returns, the higher the centrality measure associated to a node, the more the node behaves dissimilarly with respect to the other nodes in the network.

2.4. Portfolio Construction

Asset correlations are key items in investment theory and risk measurement, in particular for optimization problems as in the case of the widely known portfolio theory described by Markowitz (1952). As a consequence, correlation based graphs are useful tool to build optimal investment strategies. In this subsection we show how portfolio construction can be enhanced by means of a combination of the RMT, MST, and network centrality measures described above.

Several researches have investigated the relationship between the network structure of financial assets and portfolio strategies. The study (Onnela et al., 2003) shows how a portfolio constructed via Markowitz theory is mainly composed by assets that lie in the periphery of the asset network structure, i.e., outer node assets, and not in its core. Pozzi et al. (2013) find that peripheral assets in the network yield to better performances and lower portfolio risk with respect to central ones. Peralta and Zareei (2016) show that the centrality of assets within a network are negatively related with the optimal weights obtained through the Markowitz technique. Building on that, Výrost et al. (2018) conclude that asset allocation strategies including the network structure of financial asset are able to improve a portfolio's risk-return profile.

Another stream of literature focused on proposing alternative portfolio allocation strategies based on the network structure of financial assets. To illustrate, Plerou et al. (2002) and Conlon et al. (2007) use the random matrix theory to filter the correlation matrix to be inserted in the Markowitz minimization problem, while Tola et al. (2008) add the MST obtaining improvements with respect to the raw model.

In the present context we aim to study the differences in the risk-return profiles of our strategy, which includes topological measures in the optimization problem, with respect to the traditional Markowitz model, possibly yielding to better risk-return characteristics of the portfolios. The originality of our approach builds on the fact that we do not only use RMT and MST as alternative approaches to quantify risk diversification, but we employ an extension of the traditional Markowitz method by including these techniques in the minimization problem. Indeed, in the present case we want to solve the following problem:

$$\min_{\mathbf{w}} \mathbf{w}^{\mathbf{T}} \Sigma^{*} \mathbf{w} + \gamma \sum_{i=1}^{n} x_i w_i \tag{8}$$

subject to

$$\begin{cases} \sum_{i=1}^{n} w_i = 1 \\ \mu_P \geq \frac{\sum_{i=1}^{n} \mu_i}{n} \\ w_i \geq 0 \end{cases}$$

where $\mathbf{w}$ is the vector of portfolio weights, being w_i the weight associated to the cryptocurrency i, Σ^* is the filtered variance-covariance matrix with generic element (i, j) represented by $\sigma_i \sigma_j c^*_{i,j}$, γ is the parameter representing the risk aversion of the investor, x_i is the eigenvector centrality associated with the cryptocurrency i, μ_P indicates the return of the portfolio and μ_i the return of the generic cryptocurrency i.

Generally speaking, portfolios built upon the traditional Markowitz theory are such that the risk is minimized for a given expected return, using as input the raw variance-covariance matrix of returns. In our case, the methodological improvement is 2-fold. Firstly, we modify the input variance-covariance matrix, which is filtered by both RMT and MST. Secondly, we add a component derived from the MST structure which relates to an extra risk component the investor may want to control for. Indeed, by modulating γ the investor can set its own level of risk aversion toward systemic risk specifically, and not just to the portfolio risk as in the Markowitz framework. As a matter of fact, being centralities inversely related with distances, a small value of γ yields to portfolios composed by less systemically risky cryptocurrencies, which generally lie in the peripheral part of the network. Conversely, a large value of γ makes the algorithm select more systemically relevant cryptocurrencies, meaning those who are in the center of the network structure. For the sake of completeness, we will test different values of the systemic risk aversion parameter in the course of the current application.

Starting from the cryptocurrency return time series, the steps of the algorithm can be summarized as follows:

1. Estimation of the filtered correlation matrix $\mathbf{C}^*$ by RMT
2. Reduction of the number of links in the filtered correlation matrix $\mathbf{C}^*$ by MST
3. Computation of the filtered variance-covariance matrix Σ^* associated to the filtered correlation matrix $\mathbf{C}^*$ in step 2
4. Computation of the eigenvector centralities x_i
5. Computation of the portfolio weights by solving the minimization problem:

$$\min_{\mathbf{w}} \mathbf{w}^{\mathbf{T}} \Sigma^{*} \mathbf{w} + \gamma \sum_{i=1}^{n} x_i w_i \; s.t. \begin{cases} \sum_{i=1}^{n} w_i = 1 \\ \mu_P \geq \frac{\sum_{i=1}^{n} \mu_i}{n} \\ w_i \geq 0 \end{cases}$$

The weights calculation finally yields to the portfolio returns which we use to evaluate the performance of our allocation method.

3. EMPIRICAL FINDINGS

3.1. Data Description and Network Topology Analysis

In our empirical application we consider 10 time series of returns referred to cryptocurrencies traded over the period 14 September 2017–17 October 2019 (764 daily observations). In particular, we consider the first 10 cryptocurrencies in terms of market capitalization as of 17 October 2019[2]. To be precise, we analyze the return time series of the following cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Tether (USDT), Bitcoin Cash (BCH), Litecoin (LTC), Binance Coin (BNB), Eos (EOS), Stellar (XLM), Tron (TRX).

We provide some basic descriptive statistics of our data in **Table 1**. From **Table 1** one may notice that average daily returns are all close to zero, in line with the general economic theory regarding asset returns. However, the 10 cryptocurrencies exhibit different standard deviations, meaning that the variability in returns differs quite strongly among cryptocurrencies. To illustrate, USDT is the one showing the lowest relative variability; this is in line with the fact that this cryptocurrency is classified as stable coin, therefore its price should not deviate too much on a daily basis. On the other hand, TRX is the one showing the highest standard deviation; indeed, this particular cryptocurrency witnessed a period of high fluctuations during the considered sample period. As far as kurtosis is concerned, most of the cryptocurrencies exhibit values which reflects the non-Gaussian and heavy tailed behavior of their associated distribution. This is particularly true for XLM and XRP, whose kurtosis are relatively larger than the ones of the other time series.

To better understand the dynamics of the cryptocurrency time series, we plot the normalized price series in **Figures 1**, **2**[3]. The two figures confirm well-known features of cryptocurrencies, such as their overall high volatility (with TRX being the most volatile), the stability of the stable coin (USDT) as well as the low liquidity that some of them exhibit (such as TRX).

In order to apply the filter through RMT, we divide the dataset into consecutive overlapping windows having a width $T = 120$ (4 trading months). We set the window step length to 1 week (7 trading days), which makes up a total of 93 weekly 4-months windows.

[2] We exclude Bitcoin SV (BSV) in order to achieve a sufficiently large timespan, meaning a more than 2-years time period.

[3] We split the plot in two different figures for scale reasons.

TABLE 1 | Summary statistics.

	Mean	Std	Kurtosis	Skewness
BTC	0.0009	0.04	3.35	−0.07
ETH	−0.0007	0.05	2.90	−0.33
XRP	0.0004	0.07	15.73	1.80
USDT	0.0000	0.01	4.28	0.22
BCH	−0.0011	0.08	6.47	0.49
LTC	−0.0003	0.06	8.02	0.66
BNB	0.0033	0.07	7.74	0.78
EOS	0.0017	0.07	3.93	0.60
XLM	0.0021	0.10	26.19	2.03
TRX	0.0021	0.15	13.15	0.66

The table shows relevant summary statistics for the 10 cryptocurrencies considered related to the whole sample period, i.e., 13 September 2017–10 October 2019.

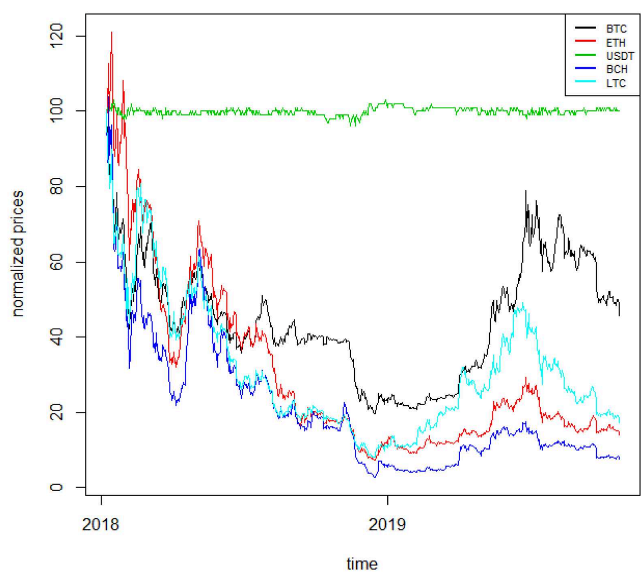

FIGURE 1 | Normalized cryptocurrency price series I. This figure shows the normalized price series for five cryptocurrencies: XRP, BNB, EOS, XLM, TRX, relative to the period 7 January 2018–17 October 2019.

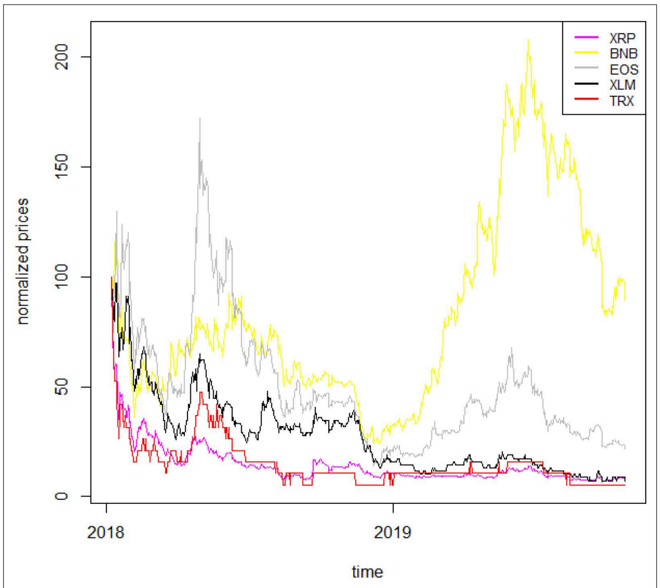

FIGURE 2 | Normalized cryptocurrency price series II. The figure shows the normalized price series for five cryptocurrencies: BTC, ETH,USDT, BCH, LTC, relative to the period 7 January 2018–17 October 2019.

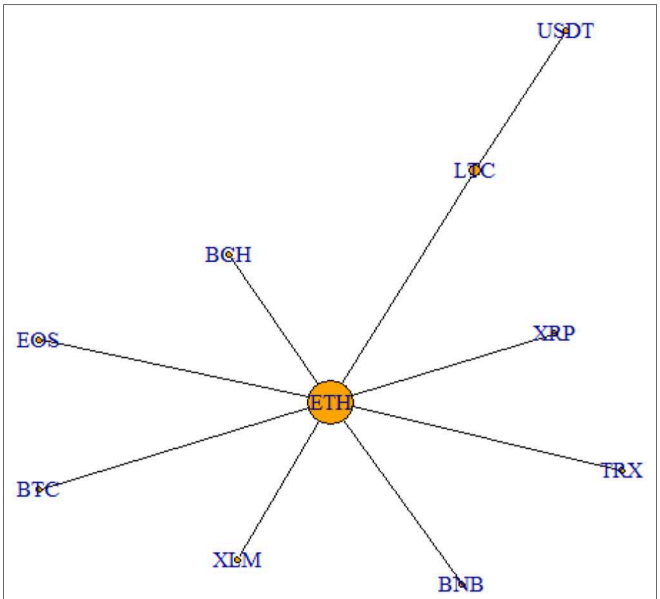

FIGURE 3 | MST September 2017–January 2018. This figure shows the MST representation relative to the period of the speculative bubble.

For each time window considered, we use 15 weeks of daily observations to estimate the model, while the last week is used for validation purposes. In other words, we compute 93 correlation matrices between the 10 cryptocurrency return time series, each one based on 15 weeks of daily returns and then filter them by means of the Random Matrix approach. Applying the Random Matrix filtering, correlation matrices are rebuilt considering only the eigenvectors corresponding to the deviating eigenvalues.

In order to have a better understanding of the links existing between cryptocurrencies, the filtered correlation matrices are then used to derive the MST representation over two main periods of interest. In particular, we plot the MST structure emerging from the period of the cryptocurrency price hype (September 2017–January 2018) in **Figure 3**, while the MST structure related to the latest trading period analyzed (June 2019–October 2019) in **Figure 4**.

As it is clear from the graph, the two networks show quite similar features. Indeed, ETH is the cryptocurrency which always lies in the center of the structure, indicating its central role in the cryptocurrency market. The only difference between the two graphical representations concerns USDT, which during the price hype is not connected directly to ETH as the other cryptocurrencies, but to LTC. This is linked to the fact that USDT

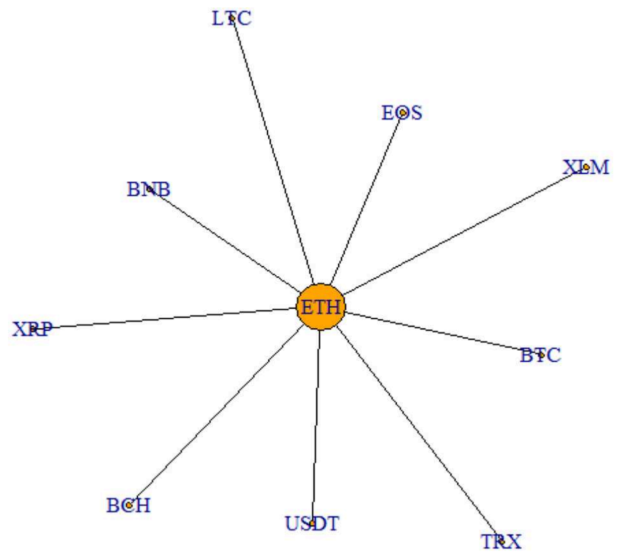

FIGURE 4 | MST June 2019–October 2019. This figure shows the MST relative to the period June 2019–October 2019.

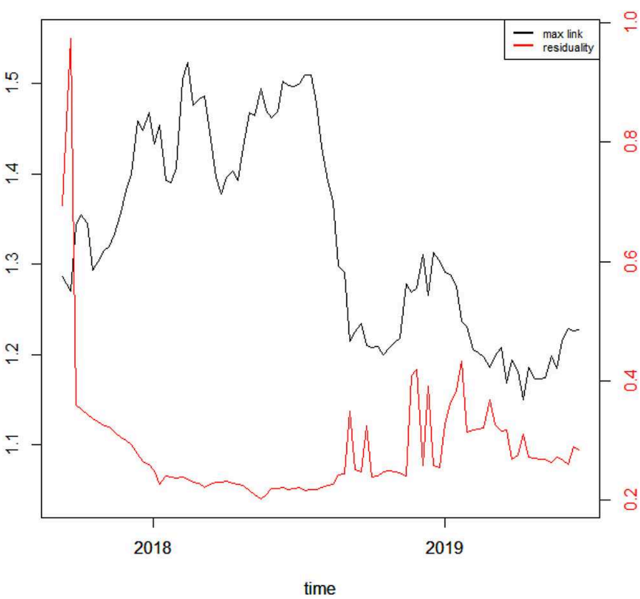

FIGURE 5 | MST thresholds and residuality coefficients. The black line shows the Max link distance, while the red line shows the residuality coefficient, whose values are reported, respectively on the left and right y axis.

is a stable coin and, therefore, behaves dissimilarly from the other cryptocurrencies considered, being it much less volatile. However, this difference in behavior levels out during the latest period, as it emerges from **Figure 4**.

To better understand the dynamics of the MST among cryptocurrencies, we investigate the evolution of the links over time. Indeed, we compute two different measures: the Max link, i.e., the value of the maximum distance between two pairs of

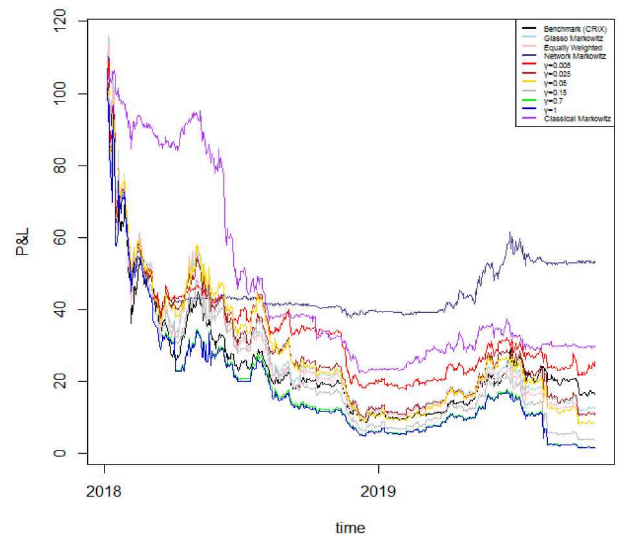

FIGURE 6 | Performances of different portfolio strategies. The plot reports the profits and losses of a portfolio with initial value of 100 USD obtained by the CRIX benchmark index [Benchmark (CRIX)], the optimization using the Markowitz approach with the variance-covariance matrix filtered by Glasso (Glasso Markowitz), the naive portfolio (Equally Weighted), our optimization using RMT and MST applied to the variance-covariance matrix (Network Markowitz), our model based on different values of γ (γ=0.005, 0.025, 0.05, 0.15, 0.7, 1), and the standard Markowitz portfolio (Classical Markowitz). The portfolio values are plotted for the period 7 January 2018–17 October 2019.

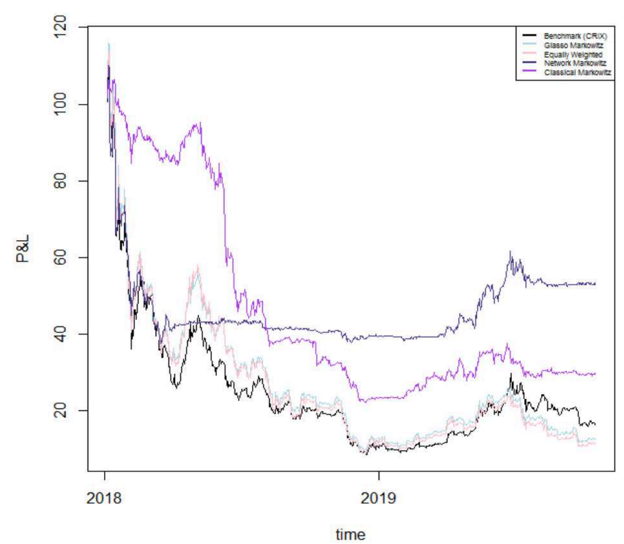

FIGURE 7 | Performances of selected portfolio strategies. The plot reports the profit and losses of a portfolio with initial value of 100 USD obtained by the CRIX benchmark index [Benchmark (CRIX)], the optimization using the Markowitz approach with the variance-covariance matrix filtered by Glasso (Glasso Markowitz), the naive portfolio (Equally Weighted), our optimization using RMT and MST applied to the variance-covariance matrix (Network Markowitz) and the standard Markowitz portfolio (Classical Markowitz). The portfolio values are plotted for the period 7 January 2018–17 October 2019.

TABLE 2 | Cumulative Profits and Losses.

Period	CRIX	GM	EW	CM	NW	$\gamma = 0.005$	$\gamma = 0.025$	$\gamma = 0.05$	$\gamma = 0.15$	$\gamma = 0.7$	$\gamma = 1$
Jan-2018	−0.14	−0.13	−0.16	0.04	−0.22	−0.21	−0.26	−0.27	−0.36	−0.43	−0.43
May-2018	−0.67	−0.62	−0.60	−0.12	−0.79	−0.78	−0.73	−0.66	−0.83	−1.08	−1.10
Sep-2018	−1.37	−1.37	−1.43	−0.88	−0.83	−1.02	−1.24	−1.23	−1.40	−1.60	−1.64
Jan-2019	−1.85	−1.78	−1.78	−1.32	−0.87	−1.50	−1.86	−1.98	−2.19	−2.29	−2.31
May-2019	−1.35	−1.25	−1.27	−1.01	−0.74	−1.22	−1.33	−1.29	−1.44	−1.55	−1.57
Sep-2019	−0.99	−1.45	−1.49	−1.02	−0.54	−1.19	−1.34	−1.44	−1.86	−2.13	−2.15

This table shows the cumulative 4-months Profits and Losses of portfolios under different strategies. Particularly, Profits & Losses are computed for the CRIX benchmark index (CRIX), the Glasso Markowitz (GM), the naive portfolio (EW), the Network Markowitz (NW), the classical Markowitz (CM), and the proposed models with different values of γ (γ = 0.005, 0.025, 0.05, 0.15, 0.7, 1). All values are expressed in percentage terms.

nodes in the tree, and the residuality coefficient, meaning the ratio between the number of links which are dropped and the number of those who are kept by the MST algorithm. The two metrics, computed over the whole sample period, are illustrated in **Figure 5**.

From **Figure 5** one may notice that the Max link increases during the Bitcoin price hype and fluctuates around relatively large values until roughly mid 2018, meaning that during this period correlations between cryptocurrency returns are strongly misaligned. After that, the index bounces back toward its previous values and even below, suggesting that cryptocurrency returns start to behave more similarly during the latest period. Furthermore, the residuality coefficient increases during the very beginning of the sample period, while it sharply declines during the price hype phase. After the decrease, the coefficient stays quite stable and then gently increases not without fluctuations from mid 2018 to the end of the sample period. This suggests that the number of links until mid 2018 was quite limited, and therefore, returns misaligned, whereas the same number started to increase after that phase, meaning there were more connections and thus more synchronicity across cryptocurrency returns.

3.2. Portfolio Construction

In this subsection we illustrate the results related to the proposed portfolio strategies. The optimal portfolio weights are obtained through the constrained minimization of the objective function in Equation 8. For the sake of completeness, we use different values of the systemic risk aversion parameter γ, meaning $\gamma = 0.005, 0.025, 0.05, 0.15, 0.7, 1$. These values have been chosen, without loss of generality, to be representative of different aversion profiles. While $\gamma = 0$ indicates no aversion, $\gamma = 1$ indicates a high aversion, with systemic risk being given the same importance as non-systemic one.

We use fifteen weeks, i.e., to compute the optimal portfolio weights as described in section 2. We then use the last week associated to each window to evaluate the out-of-sample performance of our technique, meaning to compute the portfolio returns and, therefore, the resulting Profit & Losses. We then compute portfolio returns for the period 7 January 2018–17 October 2019, accounting for rebalancing costs, which are supposed to amount to 10 basis points.

In **Figure 6** we plot the returns of our investment strategies for the different values of γ mentioned above as well as for $\gamma =$

TABLE 3 | VaR.

Period	CRIX	EW	NW	GM	CM
Jan-2018	0.11	0.13	0.15	0.14	0.03
May-2018	0.04	0.05	0.02	0.05	0.03
Sep-2018	0.11	0.11	0.10	0.12	0.02
Jan-2019	0.07	0.10	0.05	0.07	0.01
May-2019	0.04	0.02	0.03	0.02	0.04
Sep-2019	0.05	0.05	0.02	0.05	0.01

This table shows the 4-months Value at Risk of portfolios under different strategies for a confidence interval of 95%. In particular, the VaR is computed for the CRIX benchmark index (CRIX), the naive portfolio (EW), the Network Markowitz (NW), the Glasso Markowitz (GM), and the classical Markowitz (CM). All values are expressed in absolute terms multiplied by a scale factor of 100.

0 (Network Markowitz), meaning the results of the Markowitz portfolio strategy using the variance-covariance matrix filtered by RMT and MST. In doing so, we plot portfolio performances under the hypothesis of investing 100 USD at the beginning of the period, and examining how much is lost along time. The results of our strategies are compared with the performance of several strategies and indicators: the benchmark portfolio (CRIX[4]), the Markowitz portfolio with variance-covariance matrix filtered by the Glasso[5] technique (Glasso Markowitz), the naive portfolio (Equally Weighted) and the traditional Markowitz portfolio (Classical Markowitz). To better highlight the results of our best proposed model, we plot the results only for a selection of portfolio strategies in **Figure 7**. To complement this information, we report the 4-months cumulative Profits and Losses of each of the considered strategy in **Table 2**.

Overall, we are considering a period in which the cryptocurrency market witnesses a down period—except for the first part of our analyzed timespan and several short periods consequently occurring. Therefore, as the market is not profitable during the studied period, we aim to achieve through

[4]The CRIX is a cryptocurrency market index following the Laspeyres methodology for the construction of indexes. More information about CRIX can be found at https://thecrix.de/

[5]The sparsity parameter ρ has been set to 0.01, as in the reference paper by Friedman et al. (2008).

TABLE 4 | Sharpe ratio.

Period	GM	EW	CM	NW	$\gamma = 0.005$	$\gamma = 0.025$	$\gamma = 0.05$	$\gamma = 0.15$	$\gamma = 0.7$	$\gamma = 1$
Jan-2018	−0.05	−0.05	−0.03	−0.13	−0.12	−0.08	−0.06	−0.08	−0.09	−0.10
May-2018	−0.14	−0.14	−0.19	−0.03	−0.04	−0.08	−0.09	−0.08	−0.07	−0.07
Sep-2018	−0.10	−0.09	−0.17	−0.04	−0.17	−0.17	−0.20	−0.20	−0.18	−0.17
Jan-2019	0.10	0.09	0.11	0.09	0.06	0.08	0.11	0.12	0.12	0.12
May-2019	−0.02	−0.02	0.01	0.08	0.02	0.02	−0.00	−0.03	−0.04	−0.04
Sep-2019	−0.06	−0.06	−0.03	0.03	0.07	−0.11	−0.14	−0.14	−0.14	−0.14

This table shows the 4-months values of Sharpe ratio of portfolios under different strategies. In particular, the SR is computed for the Glasso Markowitz (GM), the naive portfolio (EW), the classical Markowitz (CM), the Network Markowitz (NW), and for all the value of γ.

TABLE 5 | Rachev ratio.

Period	GM	EW	CM	NW	$\gamma = 0.005$	$\gamma = 0.025$	$\gamma = 0.05$	$\gamma = 0.15$	$\gamma = 0.7$	$\gamma = 1$
Jan-2018	0.74	0.75	0.63	0.64	0.69	0.77	0.79	0.78	0.77	0.99
May-2018	0.73	0.75	0.95	0.83	0.74	0.77	0.83	0.87	0.87	0.55
Sep-2018	0.81	0.84	0.87	0.61	0.80	0.75	0.76	0.80	0.80	0.48
Jan-2019	1.16	1.11	1.47	1.24	1.34	1.36	1.39	1.40	1.40	1.26
May-2019	0.80	0.80	1.05	0.97	0.93	0.84	0.75	0.72	0.72	0.98
Sep-2019	0.75	0.78	1	1.14	0.43	0.38	0.38	0.38	0.37	0.78

This table shows the 4-months values of Rachev Ratio (RR) of portfolios under different strategies. In particular, the RR is computed for the Glasso Markowitz (GM), the naive portfolio (EW), the classical Markowitz (CM), the Network Markowitz (NW), and for all the value of γ.

our allocation strategies losses which are lower than those yielded by other competing methodologies.

On the one hand, during a first phase which lasts roughly until mid 2018, the traditional Markowitz portfolio seems to overperform the other portfolio allocation strategies. Indeed, the allocation by Markowitz' technique yields to positive (cumulative) returns until January 2018 and just slightly negative ones until May 2018, however still lower than the losses provided by the other strategies in absolute terms.

On the other hand, from September 2018 onwards all portfolios start providing strong negative returns. Indeed, the returns yielded by the portfolio constructed via Markowitz start to decline dramatically, together with those of the model including the systemic risk aversion parameter. This is because the latter model takes into account the centrality of the cryptocurrencies in the network and is therefore more adaptive to market conditions, regardless of whether they are favorable or not. Indeed it can be noticed that—overall—during bull market periods our model taking into account for risk aversion reacts very fast to upward movements and yields to good cumulative performances; conversely, during down market periods, the same model yields to worse relative performances due to declining market conditions.

However, during the second half of our sample period our proposed model with the systemic risk aversion parameter γ set to 0 (Network Markowitz) clearly overwhelms the other portfolio allocation strategies. To illustrate, if we look at the cumulative performance of the above mentioned method, we can see that it more than halves losses with respect to the equally weighted portfolio, to the Glasso Markowitz portfolio and to all portfolios including a risk aversion parameter $\gamma > 0$. Moreover, it almost halves the losses with respect to the benchmark index (CRIX) and to the traditional Markowitz methodology. This suggests that this model is capable to provide a stronger coverage for losses in case of down market periods with respect to all other considered asset allocation strategies[6].

In **Table 3** we compute the 4-months Value at Risk (VaR) with a confidence level of 0.05% for the benchmark index (CRIX), the equally weighted portfolio, our Network Markowitz portfolio, the Glasso Markowitz and the traditional Markowitz portfolios. This is done in order to compare, together with cumulative returns, the potential riskiness of our strategy with respect to the alternative portfolio allocation methods considered.

Table 3 shows that, except for the price hype period, our proposed Network Markowitz approach generally yields to lower values at risk with respect to the benchmark index (CRIX), the naive portfolio and the Glasso Markowitz. The aforementioned model is instead more risky than the traditional Markowitz model, although the latter, overall, yields too far way larger negative returns. In general, the riskiness of our strategy seems to be quite satisfactory with respect to the alternative allocation strategies analyzed.

To further support our conclusions, **Table 4** presents the Sharpe ratio under the different strategies.

Table 4 gives further evidence to support our conclusions: the proposed Network Markowitz approach yields better Sharpe Ratios.

To strengthen the robustness of our conclusions, **Table 5** presents the Rachev ratio, with a confidence level of 10%, under the different strategies. The Rachev ratio is a useful supplement of

[6]A sensitivity analysis reported in the **Appendix** confirms that results are robust with respect to different choices of the starting points and rolling estimation windows.

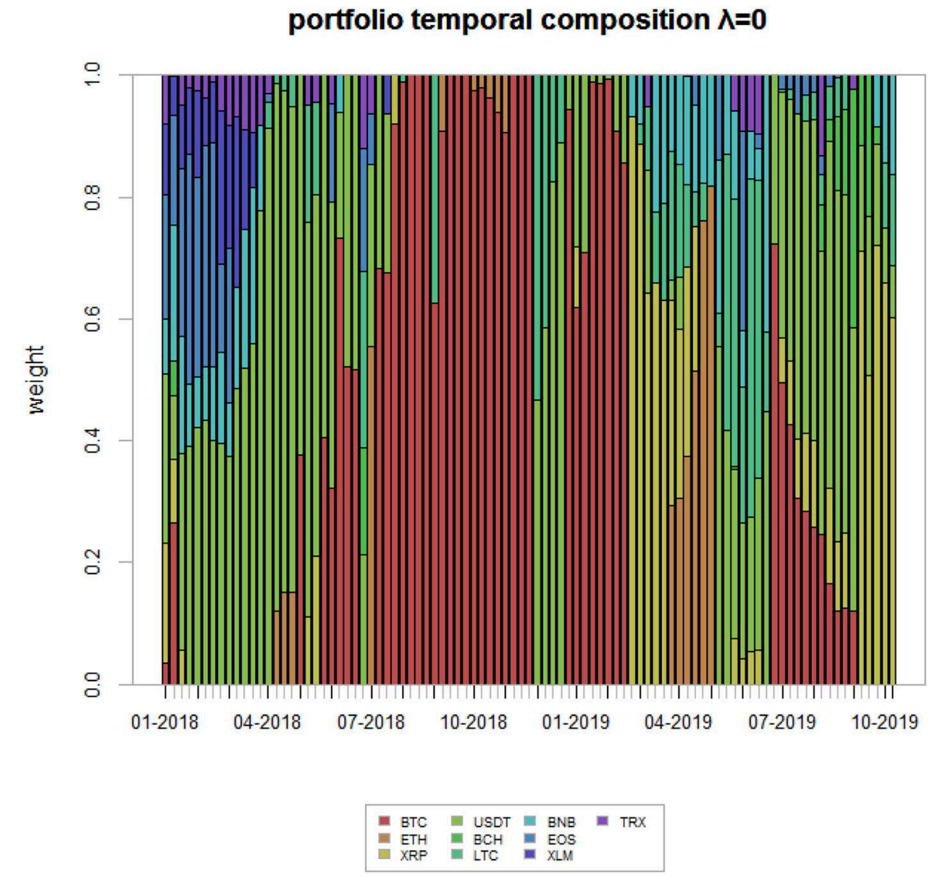

FIGURE 8 | Winning strategy portfolio weights. The figure shows the portfolio weights associated to the winning strategy—i.e., the Network Markowitz (NW)—for the analyzed time period.

the Sharpe ratio, when data is non-symmetric, as in our context. It is calculated as the ratio between an extreme gain and an extreme loss.

Table 5 shows that the Network Markowitz approach yields the best performances in the initial and final periods, and the Classic Markowitz in all other periods. The other strategies generally show worse performances. This is consistent with our previous findings, and with the fact that the Rachev ratio takes higher values during periods characterized by decreasing returns, such as the quarter preceding January 2019.

Overall, we cannot say that the proposed model overperforms traditional approach (such as Glasso Markowitz and Classical Markowitz). It does so in certain periods and according to certain risk aversion parameterizations.

For the sake of completeness, we plot the portfolio weights of the winning strategy over the evaluation time horizon in **Figure 8**. As one can clearly see, the composition of the portfolio varies quite much over time. Indeed, during the first period of the sample, approximately until February 2018, the portfolio is composed by various assets, with USDT gaining a high share over time, being it the most stable across all. After that, BTC is the cryptocurrency which is mostly selected by our algorithm, roughly until October 2018 (with some exceptions), as it is considered a proxy of the whole market. Finally, the algorithm selects different cryptocurrency compositions until the end of the sample, being the latter a highly uncertain period for the market.

Last, we present, for comparison purposes, the portfolio weights associated with $\gamma = 1$.

While **Figure 8** gives the weights relative to the situation of no systemic risk aversion, **Figure 9** gives the weight corresponding to a very high systemic risk aversion, in which it has the same importance as non-systemic risk.

4. CONCLUSIONS

In this paper we have proposed a methodology that aims to build an allocation strategy which is suitable for highly volatile markets, such as cryptocurrency ones. In particular, we have applied our models to a set of 10 cryptocurrency return time series, selected in terms of market capitalization. We have shown that the use of network models can enhance portfolios' risk-return profiles and mitigate losses during down market periods.

We have demonstrated how the use of centrality measures, together with tuning an investor's systemic risk aversion, is a suitable methodology to make profits during bull market periods, as this method is rapidly adaptive to market conditions. We have also shown that, to protect investors from losses during bear market periods, the combination of Random Matrix Theory

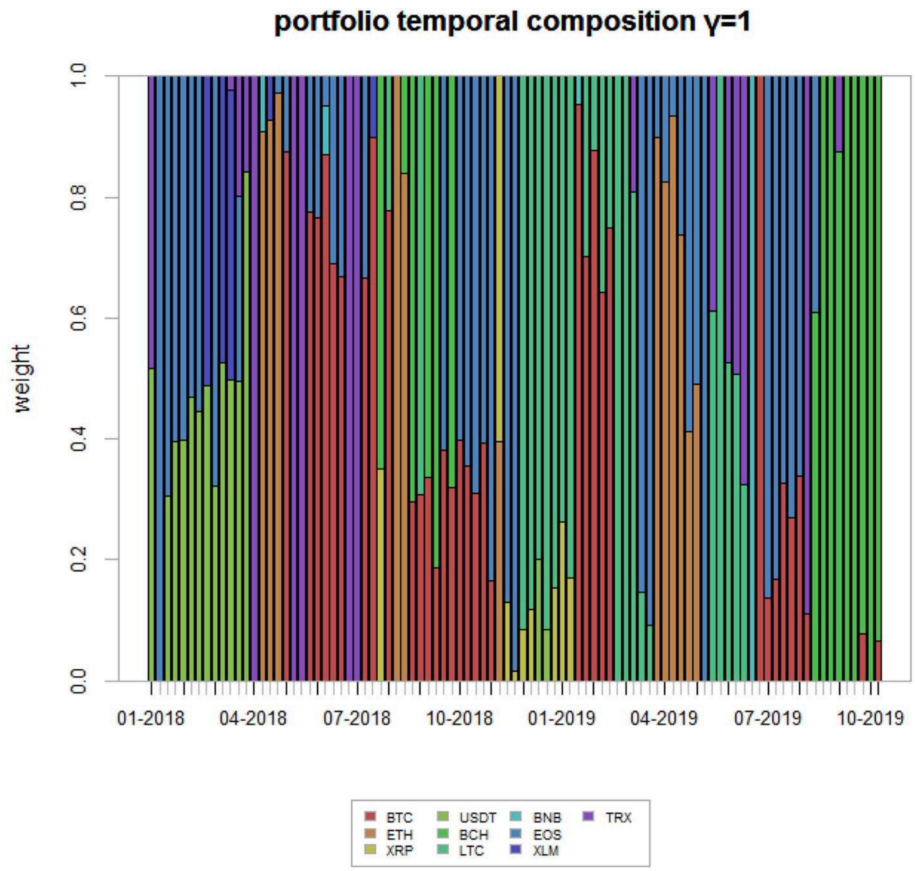

FIGURE 9 | Highly risk adverse strategy portfolio weights. This figure shows the portfolio weights associated to a highly systemic risk adverse strategy—$\gamma = 1$—for the analyzed time period.

and Minimal spanning trees can yield to acceptable risk-return profiles and/or mitigate losses.

Our empirical findings show that, overall, the proposed method is acceptable, even during downturn periods. However, we cannot claim that this proposed model should always be used in automated consultancy. It should always be compared with competing alternatives, according to different market conditions and different risk aversions.

We strongly believe that the proposed model should be further tested in different contexts. For this purpose, we provide at https://www.fintech-ho2020.eu a link to the used data and software, so the proposed methods can be fully reproduced. The software is written in the R language, and allows the methods to be extended to other data and contexts.

Further research should involve, besides the application to other contexts, the consideration of different base portfolio allocation models. We have used Markowitz' as is the most employed by robot advisory platforms.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study can be found in Coinmarketcap (https://coinmarketcap.com/). The software used in this article is not publicly available because based on a proprietary source. Requests to access the datasets should be directed to Gloria Polinesi (glopol@hotmail.it).

AUTHOR CONTRIBUTIONS

The paper was written in close collaboration between the authors. However, sections 2 and 4 have been written by GP, section 1 and 3 by PP, while PG has supervised and coordinated the work.

FUNDING

This research has received funding from the European Union's Horizon 2020 research and innovation program FIN-TECH: A Financial supervision and Technology compliance training programme under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2020.00022/full#supplementary-material

REFERENCES

Baur, D. G. and Dimpfl, T. (2019). Price discovery in bitcoin spot or futures?. *J. Fut. Markets*. 39.7, 803–817. doi: 10.2139/ssrn.3171464

Beenakker, C. W. (1997). Random-matrix theory of quantum transport. *Rev. Mod. Phys*. 69:731. doi: 10.1103/RevModPhys.69.731

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Soc. Netw*. 29, 555–564. doi: 10.1016/j.socnet.2007.04.002

Bonanno, G., Caldarelli, G., Lillo, F., and Mantegna, R. N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Phys. Rev. E* 68:046130. doi: 10.1103/PhysRevE.68.046130

Brandvold, M., Molnár, P., Vagstad, K., and Valstad, O. C. A. (2015). Price discovery on bitcoin exchanges. *J. Int. Financ. Markets Inst. Money* 36, 18–35. doi: 10.1016/j.intfin.2015.02.010

Brin, S., and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst*. 30, 107–117. doi: 10.1016/S0169-7552(98)00110-X

Conlon, T., Ruskin, H. J., and Crane, M. (2007). Random matrix theory and fund of funds portfolio optimisation. *Phys. A Stat. Mech. Appl*. 382, 565–576. doi: 10.1016/j.physa.2007.04.039

Corbet, S., Lucey, B., Peat, M., and Vigne, S. (2018a). Bitcoin futures-what use are they? *Econ. Lett*. 172, 23–27. doi: 10.1016/j.econlet.2018.07.031

Corbet, S., Meegan, A., Larkin, C., Lucey, B., and Yarovaya, L. (2018b). Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Econ. Lett*. 165, 28–34. doi: 10.1016/j.econlet.2018.01.004

Eom, C., Oh, G., Jung, W.-S., Jeong, H., and Kim, S. (2009). Topological properties of stock networks based on minimal spanning tree and random matrix theory in financial time series. *Phys. A Stat. Mech. Appl*. 388, 900–906. doi: 10.1016/j.physa.2008.12.006

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 432–441. doi: 10.1093/biostatistics/kxm045

FSB (2017a). *Financial Stability Implications From Fintech: Supervisory and Regulatory Issues That Merit Authorities' Attention*. Basel: FSB.

FSB (2017b). *Fintech Credit*. Financial Stability Board Report (27 June, 2017).

Giudici, P., and Abu-Hashish, I. (2019). What determines bitcoin exchange prices? A network var approach. *Finance Res. Lett*. 28, 309–318. doi: 10.1016/j.frl.2018.05.013

Giudici, P., and Pagnottoni, P. (2019a). High frequency price change spillovers in bitcoin markets. *Risks* 7:111. doi: 10.3390/risks7040111

Giudici, P., and Pagnottoni, P. (2019b). Vector error correction models to measure connectedness of bitcoin exchange markets. *Appl. Stochast. Models Bus. Ind*. 36, 95–109. doi: 10.1002/asmb.2478

Giudici, P. and Polinesi, G. (2019). Crypto price discovery through correlation networks. *Ann. Oper. Res*. 2. doi: 10.1007/s10479-019-03282-3

Guhr, T., Müller-Groeling, A., and Weidenmüller, H. A. (1998). Random-matrix theories in quantum physics: common concepts. *Phys. Rep*. 299, 189–425. doi: 10.1016/S0370-1573(97)00088-4

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika* 18, 39–43. doi: 10.1007/BF02289026

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *J. ACM* 46, 604–632. doi: 10.1145/324133.324140

León, D., Aragón, A., Sandoval, J., Hernández, G., Arévalo, A., and Nino, J. (2017). Clustering algorithms for risk-adjusted portfolio construction. *Proc. Comput. Sci*. 108, 1334–1343. doi: 10.1016/j.procs.2017.05.185

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *Eur. Phys. J. B Condens. Matter Complex Syst*. 11, 193–197. doi: 10.1007/s1005100 50929

Mantegna, R. N., and Stanley, H. E. (1999). *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge: Cambridge University Press.

Marchenko, V. A., and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematich. Sbornik* 114, 507–536. doi: 10.2307/2975974

Markowitz, H. (1952). Portfolio selection. *J. Finance* 7, 77–91. doi: 10.1111/j.1540-6261.1952.tb01525.x

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Available online at: www.bitcoin.org

Onnela, J.-P., Chakraborti, A., Kaski, K., Kertesz, J., and Kanto, A. (2003). Dynamics of market correlations: taxonomy and portfolio analysis. *Phys. Rev. E* 68:056110. doi: 10.1103/PhysRevE.68.056110

Pagnottoni, P. (2019). Neural network models for bitcoin option pricing. *Front. Artif. Intell*. 2:5. doi: 10.3389/frai.2019.00005

Pagnottoni, P., and Dimpfl, T. (2018). Price discovery on bitcoin markets. *Digit. Finance* 1, 1–23. doi: 10.1007/s42521-019-00006-x

Peralta, G., and Zareei, A. (2016). A network approach to portfolio selection. *J. Empir. Finance* 38, 157–180. doi: 10.1016/j.jempfin.2016.06.003

Perra, N., and Fortunato, S. (2008). Spectral centrality measures in complex networks. *Phys. Rev. E* 78:036107. doi: 10.1103/PhysRevE.78.036107

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., and Stanley, H. E. (2002). Random matrix approach to cross correlations in financial data. *Phys. Rev. E* 65:066126. doi: 10.1103/PhysRevE.65.066126

Potters, M., Bouchaud, J.-P., and Laloux, L. (2005). Financial applications of random matrix theory: old laces and new pieces. *arXiv* 0507111.

Pozzi, F., Di Matteo, T., and Aste, T. (2013). Spread of risk across financial markets: better to invest in the peripheries. *Sci. Rep*. 3:1665. doi: 10.1038/srep01665

Raffinot, T. (2017). Hierarchical clustering-based asset allocation. *J. Portfolio Manag*. 44, 89–99. doi: 10.3905/jpm.2018.44.2.089

Ren, F., Lu, Y.-N., Li, S.-P., Jiang, X.-F., Zhong, L.-X., and Qiu, T. (2017). Dynamic portfolio strategy using clustering approach. *PLoS ONE* 12:e0169299. doi: 10.1371/journal.pone.0169299

Spelta, A., and Araújo, T. (2012). The topology of cross-border exposures: beyond the minimal spanning tree approach. *Phys. A Stat. Mech. Appl*. 391, 5572–5583. doi: 10.1016/j.physa.2012.05.071

Tola, V., Lillo, F., Gallegati, M., and Mantegna, R. N. (2008). Cluster analysis for portfolio optimization. *J. Econ. Dyn. Control* 32, 235–258. doi: 10.1016/j.jedc.2007.01.034

Tulino, A. M., and Verdú, S. (2004). Random matrix theory and wireless communications. *Found. Trends Commun. Inform. Theory* 1, 1–182. doi: 10.1561/0100000001

Výrost, T., Lyócsa, Š., and Baumöhl, E. (2018). Network-based asset allocation strategies. *N. Am. J. Econ. Finance* 47, 516–536. doi: 10.1016/j.najef.2018.06.008

Zhan, H. C. J., Rea, W., and Rea, A. (2015). An application of correlation clustering to portfolio diversification. *arXiv* 1511.07945.

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.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 30 April 2020
doi: 10.3389/frai.2020.00021

The Bitcoin as a Virtual Commodity: Empirical Evidence and Implications

Cinzia Baldan and Francesco Zen*

Department of Economics and Management, University of Padova, Padua, Italy

The present work investigates the impact on financial intermediation of distributed ledger technology (DLT), which is usually associated with the blockchain technology and is at the base of the cryptocurrencies' development. "Bitcoin" is the expression of its main application since it was the first new currency that gained popularity some years after its release date and it is still the major cryptocurrency in the market. For this reason, the present analysis is focused on studying its price determination, which seems to be still almost unpredictable. We carry out an empirical analysis based on a cost of production model, trying to detect whether the Bitcoin price could be justified by and connected to the profits and costs associated with the mining effort. We construct a sample model, composed of the hardware devices employed in the mining process. After collecting the technical information required and computing a cost and a profit function for each period, an implied price for the Bitcoin value is derived. The interconnection between this price and the historical one is analyzed, adopting a Vector Autoregression (VAR) model. Our main results put on evidence that there aren't ultimate drivers for Bitcoin price; probably many factors should be expressed and studied at the same time, taking into account their variability and different relevance over time. It seems that the historical price fluctuated around the model (or implied) price until 2017, when the Bitcoin price significantly increased. During the last months of 2018, the prices seem to converge again, following a common path. In detail, we focus on the time window in which Bitcoin experienced its higher price volatility; the results suggest that it is disconnected from the one predicted by the model. These findings may depend on the particular features of the new cryptocurrencies, which have not been completely understood yet. In our opinion, there is not enough knowledge on cryptocurrencies to assert that Bitcoin price is (or is not) based on the profit and cost derived by the mining process, but these intrinsic characteristics must be considered, including other possible Bitcoin price drivers.

Keywords: Bitcoin, FinTech, Vector Autoregression model, distributed ledger technology, cryptocurrencies price determination
JEL Codes: G12, C52, D40

OPEN ACCESS

Edited by:
Alessandra Tanda,
University of Pavia, Italy

Reviewed by:
Jürgen Hakala,
Leonteq Securities AG, Switzerland
Marika Vezzoli,
University of Brescia, Italy

****Correspondence:***
Cinzia Baldan
cinzia.baldan@unipd.it

Specialty section:
This article was submitted to Artificial Intelligence in Finance, a section of the journal Frontiers in Artificial Intelligence

Received: *19 November 2019*
Accepted: *20 March 2020*
Published: *30 April 2020*

Citation:
Baldan C and Zen F (2020) The Bitcoin as a Virtual Commodity: Empirical Evidence and Implications. Front. Artif. Intell. 3:21. doi: 10.3389/frai.2020.00021

INTRODUCTION

A strict definition of FinTech seems to be missing since it embraces different companies and technologies, but a wider one could assert that FinTech includes those companies that are developing new business models, applications, products, or process based on digital technologies applied in finance.

Financial Stability Board (FSB) (2017) defines FinTech as "technology-enabled innovation in financial services that could result in new business models, applications, processes, or products with an associated material effect on the provision of financial services."

OECD (2018) analyzes instead various definitions from different sources, concluding that none of them is complete since "FinTech involves not only the application of new digital technologies to financial services but also the development of business models and products which rely on these technologies and more generally on digital platform and processes."

The services offered by these companies are indeed various: some are providing financial intermediation services (FinTech companies), while others offer ancillary services relating to the financial intermediation activity (TechFin companies). Technology is, for FinTech firms, an instrument, a productive factor, an input, while for TechFin firms, it is the final product, the output. The latter are already familiar with different technologies and innovation; hence, they could easily diversify their production by adding some digital and financial services to the products they already offer. They enjoy a situation of privileged competition because they are already known in the market due to their previous non-financial services and thus could take advantage of their customers' information to enlarge their supply of financial services. TechFin firms are the main competitors for FinTech companies (Schena et al., 2018). Indeed FinTech, or financial technology, is changing the way in which financial operations are carried out by introducing new ways to save, borrow, and invest, without dealing with traditional banks.

FinTech platforms, firms, and startups rose after the global financial crisis in 2008 as a consequence of the loss of trust in the traditional financial sector. In addition, digital natives (or millennials, born between 1980 and 2000) seemed interested in this new approach proposed by FinTech entrepreneurs. Millennials were old enough to be potential customers, who feel much more related to these new, fresh mobile services offered through mobile platforms and apps, rather than bankers. The strength of these new technologies lies in their transparent and easy-to-use interfaces that was seen as an answer to the trust crisis toward banks (Chishti and Barberis, 2016).

After the first Bitcoin (Nakamoto, 2008) has been sent in January 2009, hundreds of new cryptocurrencies started being traded in the market, whose common element is to rely on a public ledger (or blockchain technology; Hileman and Rauchs, 2017). In fact, in addition to Bitcoin, other cryptocurrencies gained popularity, such as: Ethereum (ETH), Dash, Monero (XMR), Ripple (XRP), and Litecoin (LTC). Ethereum (ETH) was officially launched in 2015 and is a decentralized computing platform characterized by its own programming language. Dash was introduced in 2014 but its market value was rising in 2016. The peculiarity of this digital coin is that, in contrast with other cryptocurrencies, block rewards are equally shared among community participants and a revenue percentage (equal to 10%) is stored in the "treasury" to fund further improvements, marketing, and network operations. Monero (XMR), launched in 2014, is a system that guarantees anonymous digital cash by hiding the features of the transacted coins. Its market value raised in 2016. Ripple (XRP) has the unique feature to be based on a "global consensus ledger" rather than on blockchain technology. Its protocol is adopted by large institutions like banks and money service businesses. Litecoin (LTC) appeared for the first time in 2011 and is characterized by a large supply of 84 million LTC. Its functioning is based on that of Bitcoin, but some parameters were altered (the mining algorithm is based on Scrypt rather than Bitcoin's SHA-265).

Despite the creation of these new cryptocurrencies, Bitcoin remains the main coin in terms of turnover. The main advantage of this new digital currency seems to be the low cost of transaction (even if this is actually a myth, since BTC transactions topped out at 50 USD per transaction in 2017–2018, while private banks charge less these days) and, contrary on what many people think, anonymity was not one of its main features when this network was designed. An individual could attempt to make his identity less obvious but the evidences available by now do not support the claim that it could be hidden easily; it may be probably impossible. To this purpose, fiat physical currencies remain the best option.

Hayes (2015, 2017, 2019) analyzes the Bitcoin price formation. In particular, he assumes the cryptocurrency as a virtual commodity, starting from the different ways by which an individual could obtain it. A person could buy Bitcoins directly in an online marketplace by giving in exchange fiat currencies or other types of cryptocurrencies. Alternatively, he can accept them as payment and finally an individual can decide to "mine" Bitcoins, which consists in producing new units, by using computer hardware designed for this purpose. This latter case involves an electrical consumption and a rational agent would not be involved in the mining process if the marginal costs of this operation exceed its marginal profits. The relation between these values determines price based on the cost of production that is the theoretical value underlying the market price, around which it is supposed to gravitate. Abbatemarco et al. (2018) resume Hayes' studies introducing further elements missed in the previous formulation. The final result confirms Hayes' findings: the marginal cost model provides a good proxy for Bitcoin market price, but the development of a speculative bubble is not ruled out.

We study the evolution of Bitcoin price by considering a cost of production model introduced by Hayes (2015, 2017, 2019). Adding to his analysis some adjustment proposed by Abbatemarco et al. (2018), we recover a series for the hypothetical underlying price; then, we study the relationship between this price and the historical one using a Vector Autoregression (VAR) model.

The remainder of the paper proceeds as follows: in section Literature Review, we expose a literature overview, presenting those papers that investigate other drivers for Bitcoin price formation, developing alternative approaches. In section Materials and Methods, we exploit the research question, describing the methodology behind the implemented cost of production model, the sources accessed to collect data, the hardware sample composition, and the formula derivations. In section Main Outcomes, we analyze and comment on the main findings of the analysis; section Conclusions concludes the work with our comments on main findings and their implications.

LITERATURE REVIEW

Researchers detect a number of economic determinants for Bitcoin price; it seems that given the new and particular features of this cryptocurrency, price drivers will change over time. For this reason, several authors analyze various potential factors, which encompass technical aspects (such as the hashrate and output volume), user-based growth, Internet components (as Google Trends, Wikipedia queries, and Tweets), market supply and demand, financial indexes (like S&P500, Dow Jones, FTSE100, Nikkei225), gold and oil prices, monetary velocity, and exchange rate of Bitcoin expressed in US dollar, euro, and yen. Among others, Kristoufek (2015) focuses on different sources of price movements by examining their interconnection during time. He considers different categories: economic drivers, as potential fundamental influences, followed by transaction and technical drivers, as influences on the interest in the Bitcoin. The results show how Bitcoin's fundamental factors, such as usage, money supply and price level, drive its price over the long term. With regard to the technical drivers, a rising price encourages individuals to become miners but this effect eclipses over time, since always more specialized mining hardware have increased the difficulty. Evidences show that price is even driven by investors' interest. According to previous studies (Kristoufek, 2013; Garcia et al., 2014), the relationship appears as most evident in the long run, but during episodes of explosive prices, this interest drives prices further up, while during rapid declines, it pushes them further down. He then concludes that Bitcoin is a unique asset with properties of both a speculative-financial asset, and a standard one and because of his dynamic nature and volatility, it is obvious to expect that its price drivers will change over time. The interest element seems to be particularly relevant when analyzing the behavior of Bitcoin price, leading many researchers to study its interconnection with Internet components, such as Google Trends, Wikipedia queries, and Tweets.

Even Matta et al. (2015) investigate whether information searches and social media activities could predict Bitcoin price comparing its historical price to Google Trends data and volume of tweets. They used a dataset based only on 60 days, but, in addition to the other papers regarding this topic, they implement an automated sentiment analysis technique that allows one to automatically identify users' opinions, evaluations, sentiments, and attitudes on a particular topic. They use a tool called "SentiStrength," which is based on a dictionary only made by sentiment words, where each of them is linked to a weight representing a sentiment strength. Its aim is to evaluate the strength of sentiments in short messages that are analyzed separately, and the result is summed up in a single value: a positive, negative, or neutral sentiment. The study reveals a significant relationship between Bitcoin price and volumes of both tweets and Google queries.

Garcia et al. (2014) study the evolution of Bitcoin price based on the interplay between different elements: historical price, volume of word-of-mouth communication in on-line social media (information sharing, measured by tweets, and posts on Facebook), volume of information search (Google searches and Wikipedia queries), and user base growth. The results identify an interdependence between Bitcoin price and two signals that could form potential price bubbles: the first concerns the word-of-mouth effect, while the other is based on the number of adopters. The first feedback loop is a reinforcement cycle: Bitcoin interest increases, leading to a higher search volume and social media activity. This new popularity encourages users to purchase the cryptocurrency driving the price further up. Again, this effect would raise the search volume. The second loop is the user adoption cycle: after acquiring information, new users join the network, growing the user base. Demand rises but since supply cannot adjust immediately but changes linearly with time, Bitcoin price would increase.

Ciaian et al. (2016) adopt a different approach to identify the factors behind the Bitcoin price formation by studying both the digital and traditional ones. The authors point out the relevance of analyzing these factors simultaneously; otherwise, the econometric outputs could be biased. To do so, they specify three categories of determinants: market forces of supply and demand; attractiveness indicators (views on Wikipedia and number of new members and posts on a dedicated blog), and global macro-financial development. The results show that the relevant impact on price is driven by the first category and it tends to increase over time. About the second category, they assert that the short-run changes on price following the first period after Bitcoin introduction are imputable to investors' interest, which is measured by online information search. Its impact eases off during time, having no impact in the long run and may be due to an increased trust among users who become more willing to adopt the digital currency. On the other hand, the results suggest that investor speculations can also affect Bitcoin price, leading to a higher volatility that may cause price bubbles. To conclude, the study does not detect any correspondences between Bitcoin price and macroeconomics and financial factors.

Kjærland et al. (2018) try to identify the factors that have an impact on Bitcoin price formation. They argue that the hashrate, CBOE volatility index (VIX), oil, gold, and Bitcoin transaction volume do not affect Bitcoin price. The study shows that price depends on the returns on the S&P500, past price performance, optimism, and Google searches.

Bouoiyour and Selmi (2015) examine the links between Bitcoin price and its potential drivers by considering investors' attractiveness (measured by Google search queries); exchange–trade ratio; monetary velocity; estimated output volume; hashrate; gold price; and Shanghai market index. The latter value is due to the fact that the Shanghai market is seen as the biggest player in Bitcoin economy, which could also drive its volatility. The evaluation period is the one from 5th December 2010 to 14th July 2014 and it is investigated through the adoption of an ARDL Bounds Testing method and a VEC Grander causality test. The results highlight the speculative nature of this cryptocurrency stating that there are poor chances that it becomes internationally recognized.

Giudici and Abu-Hashish (2019) propose a model to explain the dynamics of bitcoin prices, based on a correlation network VAR process that models the interconnections between different crypto and classic asset price. In particular, they try to assess whether bitcoin prices in different exchange markets are correlated with each other, thus exhibiting "endogenous" price variations. They select eight exchange markets, representative of different geographic locations, which represent about 60% of the total daily volume trades. For each exchange market, they collect daily data for the time period May 18th, 2016 to April 30th, 2018. The authors also try to understand whether bitcoin price variations can also be explained by exogenous classical market prices. Hence, they use daily data (market closing price) on some of the most important asset prices: gold, oil, and SP500, as well as on the exchange rates USD/Yuan and USD/Eur. Their main empirical findings show that bitcoin prices from different exchanges are highly interrelated, as in an efficiently integrated market, with prices from larger and/or more connected trading exchanges driving the others. The results also seem to confirm that bitcoin prices are typically unrelated with classical market prices, thus bringing further support to the "diversification benefit" property of crypto assets.

Katsiampa (2017) uses an Autoregressive model for the conditional mean and a first-order GARCH-type model for the conditional variance in order to analyze the Bitcoin price volatility. The study collects daily closing prices for the Bitcoin Coindesk Index from 18th July 2010 to 1st October 2016 (2,267 observations); the returns are then calculated by taking the natural logarithm of the ratio of two consecutive prices. The main findings put on evidence that the optimal model in terms of goodness of fit to the data is the AR-CGARCH, a result that suggests the importance of having both a short-run and a long-run component of conditional variance.

Chevallier et al. (2019) investigate the Bitcoin price fluctuations by combining Markov-switching models with Lévy jump-diffusion to match the empirical characteristics of financial and commodity markets. In detail, they try to capture the different sub-periods of crises over the business cycle, which are captured by jumps, whereas the trend is simply modeled under a Gaussian process. They introduce a Markov chain with the existence of a Lévy jump in order to disentangle potentially normal economic regimes (e.g., with a Gaussian distribution) vs. agitated economic regimes (e.g., crises periods with stochastic jumps). By combining these two features, they offer a model that captures the various crashes and rallies over the business cycle, which are captured by jumps, whereas the trend is simply modeled under a Gaussian framework. The regime-switching Lévy model allows identifying the presence of discontinuities for each market regime, and this feature constitutes the objective of the proposed model.

MATERIALS AND METHODS

We study the evolution of Bitcoin price by considering a cost of production model introduced by Hayes (2015, 2017). Adding to his analysis some adjustment proposed by Abbatemarco et al. (2018), we recover a series for the hypothetical underlying price, and we study the relationship between this price and the historical one using a VAR model. In detail, Hayes back-tests the pricing model against the historical market price to consolidate the validity of his theory. The findings show how Bitcoin price is significantly described by the cryptocurrency's marginal cost of production and suggest that it does not depend on other exogenous factors. The conclusion is that during periods in which price bubbles happen, there will be a convergence between the market price and the model price to shrink the discrepancy. Abbatemarco et al. (2018) resume Hayes' studies introducing further elements missed in the previous formulation. The final result confirms Hayes' findings: the marginal cost model provides a good proxy for Bitcoin market price, but the development of a speculative bubble is not ruled out. Since these studies were published before Bitcoin price raise reached its peak on 19th December 2017 (the value was $19,270), the aim of our work is to extend the analysis considering a larger time frame and verify if, even in this case, the results are unchanged. In particular, we consider the period from 9th April 2014 to 31st December 2018. We start with some unit root tests to verify if the series are stationary in level or need to be integrated and then we identify the proper number of lags to be included in the model. We then check for the presence of a cointegrating relationship to verify whether we should adopt a Vector Error Correction Model (VECM) or a VAR model; the results suggest that a VAR model is the best suited for our data[1]. We thus collect the final results of the analysis and we improve them by correcting the heteroscedasticity in the regressions.

The marginal cost function, which estimates the electrical costs of the devices used in the mining process, is presented as Equation (1):

$$COST_{\frac{\$}{day}} = H_{\frac{hash}{s}} * Eff_{\frac{J}{hash}} * CE_{\frac{\$}{kWh}} * 24_{\frac{h}{day}} \tag{1}$$

Where:

$H_{hash/s}$ is the hashrate (measured by hash/second);

$EFF_{J/hash}$ is the energy efficiency of the devices involved in the process and it is measured by Joule/hash;

$CE_{\$/kWh}$ is the electricity cost expressed in US dollar per kilowatt/hour;

24 is the number of hours in a day;

A marginal profit function, which estimates the reward of the mining activity, is instead depicted as Equation (2):

$$PROFIT_{\frac{BTC}{day}} = BR_{BTC} * \left[\frac{3,600_{\frac{s}{h}} * 24_{\frac{h}{day}}}{BTs} \right] \tag{2}$$

Where:

[1]According to Abbatemarco et al. (2018), the nature of the variables considered suggests that they probably are mutually interdependent. Lütkepohl and Krätzig (2004) state that the analysis of interdependencies between time series is subject to the endogenous problem; part of the literature proposes to specify a Vector Auto Regressive model (VAR) that analyzes the causality between the two series estimated by the model. Engle and Granger (1987), instead, demonstrated that the estimate of such a model in the presence of non-stationary variables (i.e., with mean and variance non-constant over time) can lead to erroneous model specification and hence to unconditional regressions (spurious regressions). Scholars' intuition suggests that the price trend of a cryptocurrency and that of its estimated equilibrium prices are non-stationary time series, as there is a constant increase in their values over time.

BR_{BTC} is the block reward that refers to new Bitcoins distributed to miners who successfully solved a block (hence it is measured by BTC) and it is given by a geometric progression (Equation 3):

$$BR_{BTC} = BR_1 * \frac{1}{2}^{n-1} \tag{3}$$

n increases by 1 every 210,000 blocks. At the beginning, it was $BR_1 = 50$, but during the course of time, it halved twice: on 29th November 2012 and on 10th July 2016.

3,600 is the number of seconds in an hour;

24 is again the number of hours in a day;

BT_s is the block time, which is expressed as the seconds needed to generate a block (around 600 s = 10 min), and it is computed as Equation (4):

$$BT_s = \frac{D * 2^{32}}{H} \tag{4}$$

Where H = hashrate and D = difficulty. The latter variable specifies how hard it is to generate a new block in terms of computational power given a specific hashrate. This is the value that changes frequently to ensure a BT_s close to 10 min[2].

In addition to the variables already considered, we introduce some adjustments proposed by Abbatemarco et al. (2018), who thought there were two elements missing in Hayes' formulations.

They add, on the cost side, the one required to maintain and update miners' hardware (MAN, expressed in US dollar), and on the profit side, the fees (FEES) received by miners who place transactions in a block[3].

Maintenance costs are computed as a ratio between the weighted devices' price and their weighted lifespan (5), while fees, expressed in BTC, are measured as a ratio between the daily total transaction fees and the number of daily transactions[4] (6).

$$MAN_{\$} = \frac{Weighted\ Devices\ Price_{\$}}{Weighted\ Lifespan} \tag{5}$$

$$FEES_{BTC} = \frac{Total\ Transaction\ Fees\ (BTC)}{Daily\ Transaction\ Fees} \tag{6}$$

The new equations become:

$$COST_{\frac{\$}{day}} = H_{\frac{hash}{s}} * Eff_{\frac{J}{hash}} * CE_{\frac{\$}{kWh}} * 24_{\frac{h}{day}} + MAN_{\$} \tag{7}$$

$$PROFIT_{\frac{BTC}{day}} = BR_{BTC} * \left[\frac{3,600_{\frac{s}{h}} * 24_{\frac{h}{day}}}{BTs}\right] + FEES_{BTC} \tag{8}$$

Moreover, due to the equality 1 joule = 1 watt*second, Equation (7) could be expressed as follows:

$$COST_{\$/day} = H_{hash/s} * Eff_{\frac{W*s}{hash}} * CE_{\frac{\$}{kWh}} * 24_{h/day} + MAN_{\$} \tag{9}$$

[2]Results are shown in Table A.1 (**Supplementary Material**). In order to simplify the presentation, we display only the values for the last day of each month.

[3]Bitcoin could be obtained through both the mining process and the registration of transactions but, since Bitcoin supply is limited to 21 million, once it is reached, fees become the only remuneration source in the future.

[4]Fees computation results are displayed in Table A.1 (**Supplementary Material**).

TABLE 1 | Sources.

Variables		Sources
$P_{hist\$}$	Historical price in US dollar	https://Bitcoinvisuals.com
$H_{hash/s}$	Hashrate	
BR_{BTC}	Block reward	
D	Difficulty	
BT_s	Block time	Computed using D and $H_{hash/s}$
$FEES_{BTC}$	Transaction fees	https://charts.Bitcoin.com/bch/
$CE_{\$/kWh}$	Cost of energy	Computed using data from: en.Bitcoin.it/wiki/Mining_hardware_comparison https://archive.org/web/
$MAN_{\$}$	Hardware maintaining cost	
$EFF_{J/hash}$	Hardware energy efficiency	

Source: Authors' elaboration.

By converting watt in kilowatt/hour, it can be written as:

$$COST_{\$/day} = H_{hash/s} * \frac{Eff_{\frac{W*s}{hash}}}{1000} * CE_{\frac{\$}{kWh}} * 24_{h/day} + MAN_{\$} \tag{10}$$

$$COST_{\$/day} = H_{hash/s} * Eff_{\frac{kWh*s}{hash}} * CE_{\frac{\$}{kWh}} * 24_{h/day} + MAN_{\$} \tag{11}$$

According to the competitive market economic theories, the ratio between the cost and profit functions must lead to the price under equilibrium condition (Equation 12):

$$P_{\$/BTC} = \frac{COST_{\frac{\$}{day}}}{PROFIT_{\frac{BTC}{day}}} \tag{12}$$

A historical price below the one predicted by the model would force a miner out of the market, since he is operating in loss, but at the same time, the removal of its devices from the network increases others' marginal profits (competition decreases), and at the end, the system would return to equilibrium. On the other hand, a historical price higher than what predicted by the model attracts more miners, thus increasing the number of devices operating in the network and decreasing others' marginal profits (competition increases). Again, the system would return in balance (Hayes, 2015).

We must remark that the assumption of an energy price per hemisphere is not very realistic. In fact, for large consumers, energy price is contractually set differently for peak times and less busy times. There is a lot of variation in the energy price of mines in different countries and circumstances (see, for example, Iceland with its geothermal cheap energy as a cheap energy example; Soltani et al., 2019). Taking more variation around energy prices into account would probably add a wider range of BTC prices (de Vries, 2016); due to the difficulties on collecting comparable data, we adopted a simplified proxy of the cost of energy.

Table 1 presents the sources used to collect and compute the required information.

We start the analysis by constructing a hardware sample that evolves during a chosen time window (2010–2018), which is divided in semesters associated with the introduction of a specific device (**Table 2**).

Since the first Bitcoin was traded, there has been an evolution of the devices used by miners. The first ones adopted were GPU (Graphical Processing Unit) and later FPGA (Field-Programmable Gate Array), but these days, only ASIC (Application-Specific Integrated Circuit) is suitable for mining purposes.

For each device model, we collect the efficiency, expressed in Mhash/J, and the dollar price at the release day.

Technical data were collected from the Wikipedia pages https://en.Bitcoin.it/wiki/Mining__hardware__comparison and https://en.Bitcoin.it/wiki/Non-specialized__hardware__comparison by using in addition the online archive https://archive.org/web/, which allows the recovery of different webpages at the date in which they were modified, enabling the comparison before and after reviews[5].

Since only ASIC devices were created with specifications to mining purpose, there is homogeneity among FPGA and especially among GPU hardware. Due to this fact and considering the difficulty to recover the release prices, we make some simplified assumptions about them based on the information available online. This means that given the same computational power, we assume price homogeneity among devices when they were not available for specific models[6].

Given the hardware sample, we construct a weights distribution matrix (Table A.3 in **Supplementary Material**) that represents the evolution of the devices used during each semester of the time window selected, which are replaced following a substitution rate that increases over time. In fact, until 2012, before FPGA took roots, it is equal to 0.05; until 2016, we set it equal to 0.1, and in the last 2 years of the analysis, it is equal to 0.15[7].

All computations are based on this matrix; indeed, we multiplied it by a specific column of the hardware sample table to obtain the biannual Efficiency (Table A.4 in **Supplementary Material**) (J/Hash), Weighted Devices' Prices ($) (Table A.5 in **Supplementary Material**), and Weighted Lifespans (Table A.6 in **Supplementary Material**). Regarding this latter matrix, we made further assumptions on the device lifespans by implementing Abbatemarco et al. (2018) assumptions. Hence, we set a lifespan equal to 2,880 days for GPU, 1,010 days for FPGA, and 540 days for ASIC, but after 2017, due to a supposed market growth phase, we halved these numbers (**Table 2**).

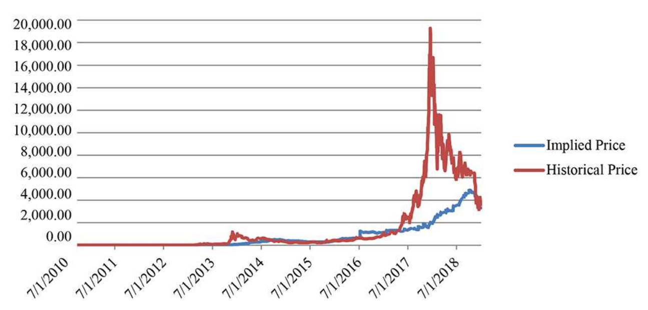

FIGURE 1 | Historical market price vs. implied model price (July 2010–December 2018). Source: Authors' elaboration.

To evaluate the cost of energy, we follow the assumptions suggested by the cited researchers and we divide the world into two parts relative to Europe: East and West, each one with a fix electricity price equal to 0.04 and 0.175 $/kWh, respectively. The weights' evolution of the mining pool is set up in 2010 equal to 0.7 for the West part and 0.3 for the East part and it changes progressively until reaching in 2018 a 0.2 for the West and 0.8 for the East. We obtained a biannual cost of energy evolution measured by $/kWh by multiplying the biannual weights to the electricity costs and summing up the value for the West and the East (Table A.7 in **Supplementary Material**).

At this point, to smooth the values across the time window, we take the differences between $biannualMAN_{\$}$, $biannualEFF_{J/Hash}$, and $biannualCE_{\$/kWh}$ at time t and $t-1$ and we divide these values by the number of days in each semester, obtaining *DeltaMAN*, *DeltaEFF*, and *DeltaCE* (Table A.8 in **Supplementary Material**). Starting the first day of the analysis with the first value of the biannual matrixes, we compute the final variables as follows:

$$MAN_{\$}(t) = MAN_{\$}(t-1) + DeltaMAN \tag{13}$$

$$EFF_{\frac{J}{hash}}(t) = EFF_{J/hash}(t-1) + DeltaEFF \tag{14}$$

$$CE_{\frac{\$}{kWh}}(t) = CE_{\frac{\$}{kWh}}(t-1) + DeltaCE \tag{15}$$

MAIN OUTCOMES

By applying Equations (8), (11), and (12), we obtain the model price[8] and compare its evolution to the historical one (**Figure 1**).

The evolution of the model (or implied) price shows a spike during the second semester of 2016, probably because on 10th July 2016, the Block Reward halved from 25 to 12.5, leading to a reduction on the profit side and a consequent price increase.

Despite this episode, the historical price seems to fluctuate around the implied one until the beginning of 2017, the period

[5]When possible, we double check Wikipedia prices with those on the websites of the companies producing mining hardware, and if they are not identical, we choose the latter.

[6]In detail, we approximate the prices of ATI FirePro M5800, Sapphire Radeon 5750 Vapor-X, GTX460, FireProV5800, Avnet Spartan-6 LX150T, and AMD Radeon 7900.

[7]Despite that ASIC devices have been released for the first time in 2013, they became the main devices used in the mining process only in 2015–2016. In the last 2 years of the analysis, we increase the substitution rate up to 0.15 because the competition among miners has been driven up as more sophisticated hardware was developed with a larger frequency.

[8]Table A.2 (**Supplementary Material**) displays all the variables required to compute the model price and compares it with the historical price. Since our time window involves 3,107 observation days, for the sake of simplicity, we present only the results for the last day of each month.

TABLE 2 | Hardware sample.

TYPE	MODEL	TIME	EFF. (Mhash/J)	PRICE (USD)	LIFESPAN	
					Before '17	After '17
GPU	ATI FirePro M5800	2 s. 2010	1.45	175	2,880	1,440
GPU	Sapphire Radeon 5750 Vapor-X	2 s. 2010	1.35	160	2,880	1,440
GPU	GTX460	2 s. 2010	1.73	200	2,880	1,440
GPU	FirePro V5800	1 s. 2011	2.08	469	2880	1,440
FPGA	Avnet Spartan-6 LX150T	2 s. 2011	6.25	995	1,010	505
FPGA	AMD Radeon 7900	1 s. 2012	10.40	680	1,010	505
FPGA	Bitcoin Dominator X5000	2 s. 2012	14.70	750	1,010	505
FPGA	X6500	1 s. 2013	23.25	989	1,010	505
ASIC	Avalon 1	2 s. 2013	107.00	1,299	540	270
ASIC	Bitmain AntMiner S1	1 s. 2014	500.00	1,685	540	270
ASIC	Bitmain AntMiner S2	2 s. 2014	900.00	2,259	540	270
ASIC	Bitmain AntMiner S3	1 s. 2015	1,300.00	1,350	540	270
ASIC	Bitmain AntMiner S4	2 s. 2015	1,429.00	1,400	540	270
ASIC	Bitmain AntMiner S5	1 s. 2016	1,957.00	1,350	540	270
ASIC	Bitmain AntMiner S5+	2 s. 2016	2,257.00	2,307	540	270
ASIC	Bitmain AntMiner S7	1 s. 2017	4,000.00	1,832	540	270
ASIC	Bitmain AntMiner S9	2 s. 2017	10,182.00	2,400	540	270
ASIC	Ebit E9++	1 s. 2018	10,500.00	3,880	540	270
ASIC	Ebit E10	2 s. 2018	11,100.00	5,230	540	270

Source: Authors' elaboration.

in which Bitcoin price started raising exponentially, reaching its peak with a value equal to \$19,270 on 19th December 2017. It declined during 2018, converging again to the model price.

Another divergence was detected at the end of 2013, but it was of a lower amount and resolved quickly.

Given the historical and implied price series, we make a further step than what Hayes (2019) and Abbatemarco et al. (2018) did, by including in the analysis a time frame even in the divergence phase. Therefore, we consider the period from 9th April 2014 to 31st December 2018. We select this time window also to base the analysis on solid data. Because of the difficulty to obtain reliable information on the hardware used in the mining process, we make some simplified assumptions on their features. By choosing this time window, we include the hardware sample whose data are more precise.

Unit Root Tests

We first try to determine with different unit root tests whether the time series is stationary or not. The presence of a unit root indicates that a process is characterized by time-dependent variance and violates the weak stationarity condition[9]. We test the presence of a unit root with three procedures: the augmented Dickey–Fuller test (Dickey and Fuller, 1979), the Phillips–Perron test (Phillips and Perron, 1988), and the Zivot–Andrews test (Zivot and Andrews, 1992).

Given a time series $\{y_t\}$, both the augmented Dickey–Fuller test (Dickey and Fuller, 1979) and the Phillips–Perron test are based on the general regression (Equation 16):

$$\Delta y_t = \alpha + \beta t + \theta y_{t-1} + \delta_1 \Delta y_{t-1} + \ldots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \tag{16}$$

Where Δy_t indicates changes in time series, α is the constant, t is the time trend, p is the order of the autoregressive process, and ε is the error term (Boffelli and Urga, 2016).

For both tests, the null hypothesis is that the time series contains a unit root; thus, it is not stationary ($H_0: \theta = 0$), while the alternative hypothesis asserts stationarity ($H_0: \theta < 0$).

Considering only the augmented Dickey–Fuller test, its basic idea is that if a series $\{y_t\}$ is stationary, then $\{\Delta y_t\}$ can be explained only by the information included in its lagged values ($\Delta y_{t-1} \ldots \Delta y_{t-p+1}$) and not from those in y_{t-1}.

For each variable, we conduct this test firstly with a constant term and later by including also a trend[10].

Table 3 presents the main findings of the test.

The Phillips–Perron test points out that the process generating y_t might have a higher order of autocorrelation than the one admitted in the test equation. This test corrects the issue, and it is

[9]The condition of weak stationarity asserts that $Var(r_t) = \gamma_o$, which means that the variance of the process is time invariant and equal to a finite constant.

[10]In order to select the proper number of lags to include in this test, we used, only for this part of the analysis, the open-source software Gretl. Its advantage is to apply clearly the Schwert criterion for the maximum lag ($p_{\max}$) estimation, which is given by: $p_{\max} = \textit{integer part of} \left[12 * \left(\frac{T}{100}\right)^{1/4}\right]$, where T is the number of observations. The test is conducted firstly with the suggested value of $p_{\max}$, but if the absolute value of the t statistic for testing the significance of the last lagged value is below the threshold 1.6, $p_{\max}$ is reduced by 1 and the analysis is recomputed. The process stops at the first maximum lag that returns a value >1.6. When this value is found, the augmented Dickey–Fuller test is estimated.

robust in case of unspecified autocorrelation or heteroscedasticity in the disturbance term of the equation. **Table 4** displays the test results.

The main difference between these tests is that the latter applies Newey–West standard errors to consider serial correlation, while the augmented Dickey–Fuller test introduces additional lags of the first difference.

Since the previous tests do not allow for the possibility of a structural break in the series, Zivot and Andrews (1992) propose to examine the presence of a unit root including the chance of an unknown date of a break-point in the series. They elaborate three models to test for the presence of a unit root considering a one-time structural break:

a) permits a one-time change in the intercept of the series:

$$\begin{aligned}\Delta y_t &= \alpha + \beta t + \theta y_{t-1} + \gamma DU_t + \delta_1 \Delta y_{t-1} + \ldots \\ &+ \delta_{p-1} y_{t-p+1} + \varepsilon_t \end{aligned} \tag{17}$$

b) permits a one-time change in the slope of the trend function:

$$\begin{aligned}\Delta y_t &= \alpha + \beta t + \theta y_{t-1} + \vartheta DT_t + \delta_1 \Delta y_{t-1} + \ldots \\ &+ \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \end{aligned} \tag{18}$$

c) combines the previous models:

$$\begin{aligned}\Delta y_t &= \alpha + \beta t + \theta y_{t-1} + \gamma DU_t + \vartheta DT_t + \delta_1 \Delta y_{t-1} + \ldots \\ &+ \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \end{aligned} \tag{19}$$

Where DU_t is a dummy variable that relates to a mean shift at a given break-date, while DT_t is a trend shift variable.

The null hypothesis, which is the same for all three models, states that the series contains a unit root ($H_0: \theta = 0$), while the alternative hypothesis asserts that the series is a stationary process with a one-time break occurring at an unknown point in time ($H_0: \theta < 0$) (Waheed et al., 2006).

The results in **Table 5** confirm what the other tests predict: both series are integrated of order 1. Since this last test identifies for Δ*lnPrice* the presence of a structural break on 18th December 2017 and after this date the Bitcoin price reaches its higher value to start declining later, we add to the analysis a dummy variable related to this observation, in order to take into account a broken linear trend in a series.

Identifying the Number of Lags

The preferred lag length is the one that generates the lowest value of the information statistic considered. We follow Lütkepohl's intuition that "*the SBIC and HQIC provide consistent estimates of the true lag order, while the FPE and AIC overestimate the lag order with positive probability*" (Becketti, 2013). Therefore, for our analysis, we select 1 lag (**Table 6**)[11].

Identifying the Number of Cointegrating Relationships

A cointegrating relationship is a relationship that describes the long-term link among the levels of a number of the non-stationary variables. Given K non-stationary variables, they can have at most $K - 1$ cointegrating relationships. Since we have only two non-stationary variables (*lnPrice* and *lnModelPrice*), we could obtain, at most, only one cointegrating relationship.

If series show cointegration, a VAR model is no more the best suited one for the analysis, but it is better to implement a Vector Error-Correction Model (VECM), which can be written as (20):

$$\Delta y_t = \mu + \delta t + \alpha \beta' u_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \tag{20}$$

Where the deterministic components $\mu + \delta t$ are, respectively, the linear and the quadratic trend in y_t that can be separated into the proper trends in y_t and those of the cointegrating relationship. This depends on the fact that in a first-difference equation: a constant term is a linear trend in the level of the variables ($y_t = \kappa + \lambda t \rightarrow \Delta y_t = \lambda$), while a linear trend derives from the quadratic one in the regression in levels ($y_t = \kappa + \lambda t + \omega t^2 \rightarrow \Delta y_t = \lambda + 2\omega t - \omega$). Therefore, $\mu \equiv \alpha \nu + \gamma$, and $\delta t = \alpha \rho t + \tau t$.

By substituting in the previous expression, the VECM can be expressed as Equation (21):

$$\Delta y_t = \alpha \left(\beta' y_{t-1} + \nu + \rho t\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \gamma + \tau t + \varepsilon_t \tag{21}$$

Where the first part $\alpha \left(\beta' u_{t-1} + \nu + \rho t\right)$ represents the cointegrating equations, while the second $\sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \gamma + \tau t + \varepsilon_t$ refers to the variables in levels.

This representation allows specifying five cases that Stata tests:

1) Unrestricted trend: allows for quadratic trend in the level of $y_t (\tau t$ appears in the equation) and states that the cointegrating equations are trend stationary, which means they are stationary around time trends.
2) Restricted trend ($\tau = 0$): excludes quadratic trends but includes linear trends (ρt). As in the previous case, it allows the cointegrating equations to be trend stationary.
3) Unrestricted constant ($\tau = 0$, $\rho = 0$): lets linear trends in y_t to present a linear trend (γ) but the cointegrating equations are stationary around a constant means (ν).
4) Restricted constant ($\tau = 0$, $\rho = 0$, $\gamma = 0$): rules out any trends in the levels of the data but the cointegrating relationships are stationary around a constant mean (ν).

[11] To identify the proper lag length to be included in the VAR model, we use the "*varsoc*" command in Stata that displays a table of test statistics, which reports for each lag length, the log of the likelihood functions (LL), a likelihood-ratio test statistic with the related degrees of freedom and p value (LR, *df*, and p), and also four information criteria: Akaike's final prediction error (FPE); Akaike's information criterion (AIC), Hannan and Quinn's information criterion (HQIC), and Schwarz's Bayesian information criterion (SBIC). Every information criteria provide a trade-off between the complexity (e.g., the number of parameters) and the goodness of fit (based on the likelihood function) of a model. Since the output is sensitive to the maximum lag considered, we try different options by changing the one included in the command computation. We tried with 4, 8, 12, 16, 20, and 24 lags. After selecting a maximum lag length equal to 16, the optimal number of lags suggested changes: while the previous results agree recommending 1 lag with each information criteria, now the FPE and AIC diverge and propose 13 lags.

TABLE 3 | Augmented Dickey–Fuller test.

	Augmented Dickey–Fuller test				
	Constant		Constant + trend		Result
	t stat	*p*-value	*t*-stat	*p*-value	
lnPrice	−0.606	0.8696	−1.839	0.6856	NO stationary
lnModelPrice	−0.467	0.8982	−1.669	0.7644	NO stationary
ΔlnPrice	−7.694	0.0000	−7.697	0.0000	Stationary
ΔlnModelPrice	−8.041	0.0000	−8.038	0.0000	Stationary

Critical values					
Constant			Constant + trend		
1%	5%	10%	1%	5%	10%
−3.430	−2.860	−2.570	−3.960	−3.410	−3.120

Source: Authors' elaboration.

TABLE 4 | Phillips–Perron test.

	Phillips–Perron test				
	Constant		Constant + trend		Result
	t stat	*p*-value	*t* stat	*p*-value	
lnPrice	−0.437	0.9037	−1.546	0.8130	NO stationary
lnModelPrice	−0.637	0.8624	−1.805	0.7021	NO stationary
ΔlnPrice	−34.394	0.0000	−34.385	0.0000	Stationary
ΔlnModelPrice	−42.972	0.0000	−42.959	0.0000	Stationary

Critical values					
Constant			Constant + trend		
1%	5%	10%	1%	5%	10%
−3.430	−2.860	−2.570	−3.960	−3.410	−3.120

Source: Authors' elaboration.

5) No trend ($\tau = 0, \rho = 0, \gamma = 0, \nu = 0$): considers no non-zero means or trends.

Starting from these different specifications, the Johansen test can detect the presence of a cointegrating relationship in the analysis. The null hypothesis states, again, that there are no cointegrating relationships against the alternative that the null is not true. H_0 is rejected if the trace statistic is higher than the 5% critical value.

We run the test with each case specification and the results agree to detect zero cointegrating equations (a maximum rank of zero). Only the unrestricted trend does not display any conclusion from the test but, since the other results matched, we consider $rank = 0$ the right solution. This implies that the two time series could be fitted into a VAR model.

VAR Model

The VAR model allows investigating the interaction of several endogenous time series that mutually influence each other. We do not only want to detect if Bitcoin price could be determined by the one suggested by the cost of production model; we also want to check if the price has an influence on the model price. This latter relation can occur if, for example, a price increase leads to a higher cost for the mining hardware. In fact, a raise in the price represents also a higher reward if the mining process is successfully conducted, with the risk to push hardware price atop, which in turn could boost the model price up.

To explain how a VAR model is constructed, we present a simple univariate AR(p) model, disregarding any possible exogenous variables, which can be written as (22):

$$y_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t \qquad (22)$$

Or, in a concise form (23):

$$\phi(L) y_t = \mu + \varepsilon_t \qquad (23)$$

where y_t depends on its p prior values, a constant (μ) and a random disturbance (ε_t).

TABLE 5 | Zivot–Andrews test.

	Zivot–Andrews test									
	Intercept			**Trend**			**Intercept + trend**			**Result**
	***t* stat**	**Break**	**Date**	***t* stat**	**Break**	**Date**	***t* stat**	**Break**	**Date**	
lnPrice	−2.964	1,083	26/03/2017	−2.049	261	25/12/2014	−2.562	1,196	17/07/2017	NON-stationary
lnModelPrice	−3.221	281	14/01/2015	−3.357	408	21/05/2015	−3.914	620	19/12/2015	NON-stationary
ΔlnPrice	−34.905	1,350	18/12/2017	−34.626	1,285	14/10/2017	−34.895	1,350	18/12/2017	Stationary
ΔlnModelPrice	−42.848	582	11/11/2015	−42.781	1,469	16/04/2018	−42.858	582	11/11/2015	Stationary

Critical values								
Intercept			**Trend**			**Intercept + trend**		
1%	5%	10%	1%	5%	10%	1%	5%	10%
−5.34	−4.8	−4.58	−4.93	−4.42	−4.11	−5.57	−5.08	−4.82

Source: Authors' elaboration.

TABLE 6 | Proper number of lags.

Lag	LL	LR	*df*	*P*	FPE	AIC	HQIC	SBIC
0	7160.95				8.0e−07	−8.36581	−8.3611	−8.35308
1	7190.57	59.237	4	0.000	7.7e−07	−8.39575	−8.38633*	−8.37029*
2	7192.42	3.7134	4	0.446	7.8e−07	−8.39325	−8.37911	−8.35506
3	7194.48	4.1059	4	0.392	7.8e−07	−8.39097	−8.37231	−8.34005
4	7195.74	2.5346	4	0.638	7.8e−07	−8.38778	−8.36422	−8.32413
5	7197.81	4.1319	4	0.388	7.8e−07	−8.38552	−8.35725	−8.30914
6	7199.73	3.8486	4	0.427	7.8e−07	−8.38309	−8.35011	−8.29399
7	7201.63	3.8014	4	0.434	7.9e−07	−8.38064	−8.34295	−8.2788
8	7204.56	5.8468	4	0.211	7.9e−07	−8.37938	−8.33698	−8.26482
9	7208.36	7.6003	4	0.107	7.9e−07	−8.37914	−8.33204	−8.25185
10	7212.23	7.7429	4	0.101	7.9e−07	−8.37899	−8.32717	−8.23897
11	7213.48	2.5086	4	0.643	7.9e−07	−8.37578	−8.31925	−8.22304
12	7225.63	24.303	4	0.000	7.8e−07	−8.38531	−8.32407	−8.21983
13	7243.57	35.872*	4	0.000	7.7e−07*	−8.4016*	−8.33565	−8.2234
14	7244.29	1.4495	4	0.836	7.7e−07	−8.39777	−8.32711	−8.20684
15	7246.50	4.4025	4	0.354	7.7e−07	−8.39567	−8.3203	−8.19201
16	7248.86	4.7357	4	0.316	7.8e−07	−8.39376	−8.31368	−8.17737

Source: Authors' elaboration.

A vector of n jointly endogenous variables is express as (24):

$$y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{n,t} \end{bmatrix} \tag{24}$$

This n-element vector can be rearranged as a function (Equation 25) of n constants, p prior values of Y_t, and a vector of n random disturbances, ϵ_t:

$$y_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t \tag{25}$$

Where μ is a vector (Equation 26) of the n-constants:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \tag{26}$$

the matrix of coefficients Φ_i is Equation (27):

$$\Phi_1 = \begin{bmatrix} \phi_{i,11} & \phi_{i,12} & \cdots & \phi_{i,1n} \\ \phi_{i,21} & \phi_{i,22} & \cdots & \phi_{i,2n} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{i,n1} & \phi_{i,n2} & \ldots & \phi_{i,nn} \end{bmatrix} \tag{27}$$

TABLE 7 | Regressions of the Vector Autoregression model.

Variables	(1)	(2)
	dlnPrice	**dlnModelPrice**
L.dlnPrice	0.18330223***	0.00799770
	(0.02359822)	(0.02055802)
L.dlnModelPrice	−0.00655017	−0.02899205
	(0.02762476)	(0.02406582)
Dummy	−0.00588960***	0.00027999
	(0.00185465)	(0.00161571)
Constant	0.00236755***	0.00149779**
	(0.00086910)	(0.00075713)
Observations	1,726	1,726
R^2	0.04178812	0.00092579

Standard errors in parentheses.
****p < 0.01, **p < 0.05, and *p < 0.1.*
Source: Authors' elaboration.

TABLE 8 | Regressions with robust standard errors.

Variables	(1)	(2)
	dlnPrice	**dlnModelPrice**
L.dlnPrice	0.18330223***	0.00799770
	(0.04306718)	(0.01592745)
L.dlnModelPrice	−0.00655017	−0.02899205***
	(0.02681078)	(0.00979148)
Dummy	−0.00588960***	0.00027999
	(0.00225058)	(0.00142356)
Constant	0.00236755***	0.00149779*
	(0.00078480)	(0.00078942)
Observations	1,726	1,726
R^2	0.04178812	0.00092579

Robust standard errors in parentheses.
****p < 0.01, **p < 0.05, and *p < 0.1.*
Source: Authors' elaboration.

and ϵ_t consists in Equation (28):

$$\epsilon_t = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} \tag{28}$$

With $E\epsilon_t = 0$ and $E\epsilon_t\epsilon'_s = \begin{cases} \Sigma, & t = s \\ 0, & t \neq s \end{cases}$

the elements of ϵ_t can be contemporaneously correlated.

Given these specifications, a pth-order VAR can be presented as Equation (29):

$$\Phi\,(L)\,u_t = \mu + \epsilon_t \tag{29}$$

To clarify this expression, the ith endogenous time series can be extracted from these basic VAR and be represented as (30):

$$\begin{aligned} y_{i,t} &= \mu_i + \phi_{1,i1}y_{1,t-1} + \ldots + \phi_{1,in}y_{n,t-1} \\ &+ \phi_{2,i1}y_{1,t-2} + \ldots + \phi_{2,in}y_{n,t-2} + \ldots \\ &+ \phi_{p,i1}y_{1,t-p} + \ldots + +\phi_{p,in}y_{n,t-p} + \varepsilon_{i,t} \end{aligned} \tag{30}$$

The result of the VAR model considering the dummy variable is presented in **Table 7**:

As expected, the dummy is significant in the *dlnPrice* function but not in *dlnModelPrice*.

Looking at the significance of the parameters, we can see how *dlnPrice* depends on its lagged value, on the dummy and on the constant term, but it seems not to be linked with the lagged value of *dlnModelPrice*. The regression of *dlnModelPrice* appears not to be explained by any variable considered in the model. We then check the stationarity of the model. The results confirm that the model is stable and there is no residual autocorrelation (Table A.9 in **Supplementary Material**).

Heteroscedasticity Correction

Given the series' path and the daily frequency of the data, the variables included in the model are probably heteroskedastic. This feature does not compromise the unbiasedness or the consistency of the OLS coefficients but invalidates the usual standard errors. In time series analysis, heteroscedasticity is usually neglected, as the autocorrelation of the error terms is seen as the main problem due to its ability to invalidate the analysis.

Since it is not possible to check and correct heteroscedasticity while performing the VAR model, we run each VAR regression separately and check the presence of heteroscedasticity by running the Breusch-Pagan test, whose null hypothesis states that the error variance are all equal (homoscedasticity) against the alternative hypothesis that the error variances change over time (heteroscedasticity).

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma^2 \tag{31}$$

The null hypothesis is rejected if the probability value of the chi-square statistic (Prob < chi2) is <0.05. The results of the test for both regressions show that the null hypothesis is always rejected, implying the presence of heteroscedasticity in the residuals (Table A.10 in **Supplementary Material**).

We try to correct the issue using heteroscedasticity-robust standard errors. The results are displayed in **Table 8**.

These new robust standard errors are different from the standard errors estimated with the VAR model, while the coefficients are unchanged. The first difference of *lnPrice* depends even in this case on its lag, but, contrary from the VAR, now the first difference of *lnModelPrice* is not independent from its previous values. This new specification confirms the previous finding that each variable does not depend on the lagged value of the other one. Therefore, it seems that during the time window considered, the Bitcoin historical price is not connected with the price derived by Hayes' formulation, and vice versa.

Recalling **Figure 1**, it seems that the historical price fluctuated around the model (or implied) price until 2017, the year in which Bitcoin price significantly increased. During the last months of 2018, the prices seem to converge again, following a common path. In our analysis, we focus on the time window in which Bitcoin experienced its higher price volatility (Figure

A.1 in **Supplementary Material**) and the results suggest that it is disconnected from the one predicted by the model. These findings may depend on the features of the new cryptocurrencies, which have not been completely understood yet.

The previous analyses, conducted on different time periods, by Hayes (2019) and Abbatemarco et al. (2018) assert that Bitcoin price could be justified by the costs and revenues of its blockchain network, leading to an opposite result from ours. We suggest that the difference could be based on the time window analyzed since we make a further step evaluating also the months in which Bitcoin price was pushed atop and did not follow a stable path. We think that there is not enough knowledge on cryptocurrencies to assert that Bitcoin price is (or is not) based on the profit and cost derived by the mining process, but these intrinsic characteristics must be considered and checked also in further analysis that include other possible Bitcoin price drivers suggested by the literature.

CONCLUSIONS

The main findings of the analysis presented show how, in the considered time frame, the Bitcoin historical prices are not connected with the price derived from the model, and *vice versa*.

This result is different from the one obtained by Hayes (2019) and Abbatemarco et al. (2018), who conclude that the Bitcoin price could be explained by the cost of production model.

The reason behind these opposite outcomes could be the considered time window. In fact, our analysis includes also those months where Bitcoin price surges up, reaching a peak of $19,270 on 19th December 2017, without following a seasonal path (Figure A.1 in **Supplementary Material**). This has a relevant impact on the results even if the historical price started declining in 2018, converging again to the model one. Looking at the overall time frame, it seems that the increasing value of the historical price from the beginning of 2017 to the end of 2018 is a unique episode that required some months to get back to more standard behavior (Caporale et al., 2019).

It seems now possible to assert that Bitcoin could not be seen as a virtual commodity, or better not only. According to Abbatemarco et al. (2018), the implemented approach does not rule out the possibility of a bubble development and, given the actual time frame, this is the reason why it would be more precise to explain Bitcoin price not only with the one implied by the model, but also with other explanatory variables that the literature seems to identify as meaningful. Therefore, to avoid misleading results, Bitcoin intrinsic characteristics must be considered and checked by adding to the profit and cost functions also these suggested parameters that range from technical aspects and Internet components to financial indexes, commodity prices, and exchange rate. This could open new horizons for research, which, despite the traditional drivers, should consider also new factors such as Google Trends, Wikipedia queries, and Tweets. These elements are related to the Internet component and appear to be particularly relevant given the social and digital Bitcoin's nature.

Kristoufek's (2013) intuition, which considers Bitcoin as a unique asset that presents properties of both a speculative financial asset and a standard one, whose price drivers will change over time considering its dynamic nature and volatility, seems to be confirmed.

The explanatory power of the VAR specification we implemented to inspect fundamental vs. market price dynamics could be quite low, which is to ascribe to missing factors and volatility. Further researches could include more tests on the VAR specification also including other controls/factors to check whether, for example, the VIX is another and important explanatory factor. More involved analyses should also explore for latent factors and/or time-varying relationships with stochastic and jump components.

Although there are highlighted elements of uncertainty, Bitcoin has undoubtedly introduced to the market a new way to think about money transfers and exchanges. The distributed ledger technology could be a disruptive innovation for the financial sector, since it can ease communication without the need of a central authority. Moreover, the spread of private cryptocurrencies, which enter into competition with the public forms of money, could affect the monetary policy and the financial stability pursued by official institutions. For these reasons, central banks all over the world are seeking to understand if it is possible to adopt this technology in their daily operations, with the aim of including it in the financial system and controlling its implementations, enhancing its benefits, and reducing its risks (Gouveia et al., 2017; Bank for International Settlements, 2018).

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

FZ: Introduction, Literature Review, and Conclusions. CB: Materials and Methods, Main Outcomes, and Conclusions.

ACKNOWLEDGMENTS

We acknowledge useful comments and suggestions from two anonym ous referees that have helped to substantially improve the paper. We are also grateful to Alessia Rossi, who has helped us in collecting and processing data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2020.00021/full#supplementary-material

REFERENCES

Abbatemarco, N., De Rossi, L., and Salviotti, G. (2018). *An econometric model to estimate the value of a cryptocurrency network. The Bitcoin case.* Association for Information Systems. AIS Electronic Library (AISeL) ECIS. Available online at: https://aisel.aisnet.org/ecis2018_rp/164?utm_source=aisel.aisnet.org%2Fecis2018_rp%2F164&utm_medium=PDF&utm_campaign=PDFCoverPages@@uline@ (accessed February 20, 2020).

Bank for International Settlements (2018). Central bank digital currencies. *Committee on Payments and market Infrastructures*. Markets Committee.

Becketti, S. (2013). *Introduction to Time Series Using Stata*. College Station, TX: Stata Press.

Boffelli, S., and Urga, G. (2016). *Financial Econometrics Using Stata*. College Station, TX: Stata Press, 17–30.

Bouoiyour, J., and Selmi, R. (2015). What does bitcoin look like? *Ann. Econ. Financ.* 16, 449–492. Available online at: http://aeconf.com/Articles/Nov2015/aef160211.pdf

Caporale, G. M., Plastun, A., and Oliinyk, V. (2019). Bitcoin fluctuations and the frequency of price overreactions, *Financ. Mark. Portf. Manag.* 33, 109–131. doi: 10.1007/s11408-019-00332-5

Chevallier, J., Goutte, S., Guesmi, K., and Saadi, S. (2019). *Study of the Dynamic of Bitcoin's Price, HAL*. Available online at: https://halshs.archives-ouvertes.fr/halshs-02175669 (accessed February 20, 2020).

Chishti, S., and Barberis, J. (2016). *The FinTech Book: The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries*. Chichester: Wiley.

Ciaian, P., Rajcaniova, M., and Kancs, D. A. (2016). The economics of Bitcoin price formation, *Appl. Econ.* 48, 1799–1815. doi: 10.1080/00036846.2015.1109038

de Vries, A. (2016). Bitcoin's growing energy problem, *Joule* 2, 801–809. doi: 10.1016/j.joule.2018.04.016

Dickey, D. A., and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *J. Am. Stat. Assoc. 74,* 427–431.

Engle, R. F., and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing, *Econometrica* 55, 251–276.

Financial Stability Board (FSB) (2017). *Financial Stability Implications from FinTech. Supervisory and Regulatory Issues that Merit Authorities' Attention*. Available online at: http://www.fsb.org/wp-content/uploads/R270617.pdf (accessed February 20, 2020).

Garcia, D., Tassone, C. J., Mavrodiev, P., and Perony, N. (2014). The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. *J. R. Soc. Interface* 11, 1–9. doi: 10.1098/rsif.2014.0623

Giudici, P., and Abu-Hashish, I. (2019). What determines bitcoin exchange prices? A network VAR approach. *Financ. Res. Lett.* 28, 309–318. doi: 10.1016/j.frl.2018.05.013

Gouveia, O. C., Dos Santos, E., de Lis, S. F., Neut, A., and Sebastián, J. (2017). *Central Bank Digital Currencies: Assessing Implementation Possibilities and Impacts*. BBVA Working Paper, No. 17/04.

Hayes, A. (2015). *A Cost of Production Model for Bitcoin*. Department of Economics. The New School for Social Research, Working Paper No. 5.

Hayes, A. (2017). Cryptocurrency value formation: an empirical analysis leading to a cost of production model for valuing bitcoin, *Telemat. Inform.* 34, 1308–1321. doi: 10.1016/j.tele.2016.05.005

Hayes, A. (2019). Bitcoin price and its marginal cost of production: support for a fundamental value, *Appl. Econ. Lett.* 26, 554–560. doi: 10.1080/13504851.2018.1488040

Hileman, G., and Rauchs, M. (2017). *Global Cryptocurrency Benchmarking Study*. Cambridge Centre for Alternative Finance, University of Cambridge, Judge Business School.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: a comparison of GARCH models, *Econ. Lett.* 158, 3–6. doi: 10.1016/j.econlet.2017.06.023

Kjærland, F., Khazal, A., Krogstad, E. A., Nordstrøm, F. G. B., and Oust, A. (2018). An analysis of Bitcoin's price dynamics. *J. Risk Financ Manag.* 11:63. doi: 10.3390/jrfm11040063

Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: quantifying the relationship between phenomena of the Internet era. *Sci. Rep.* 3:3415. doi: 10.1038/srep03415

Kristoufek, L. (2015). What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis. *PLoS ONE* 10:e0123923. doi: 10.1371/journal.pone.0123923

Lütkepohl, H., and Krätzig, M. (2004). (Eds.). *Applied Time Series Econometrics*. Cambridge University Press.

Matta, M., Marchesi, M., and Lunesu, M. I. (2015). "*Bitcoin* spread prediction using social and web search media," in *Proceedings of Conference on Workshop Deep Content Analytics Techniques for Personalized & Intelligent Services.* (Dublin).

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Available online at: https://Bitcoin.org/Bitcoin.pdf (accessed February 20, 2020).

OECD (2018). *Financial Markets, Insurance and Pensions.* Digitalisation and Finance.

Phillips, P. C. B., and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* 75, 335–346. doi: 10.1093/biomet/75.2.335

Schena, C., Tanda, A., Arlotta, C., and Potenza, G. (2018). *Lo sviluppo del FinTech.* CONSOB. Quaderni Fintech.

Soltani, M., Kashkooli, F. M., Dehghani-Sanij, A. R., Kazemia, A. R., Bordbar, N., Farshchi, M. J., et al. (2019). A comprehensive study of geothermal heating and cooling systems. *Sustain. Cities Soc.* 44, 793–818. doi: 10.1016/j.scs.2018.09.036

Waheed, M., Alam, T., and Ghauri, S. P. (2006). *Structural Breaks and Unit Root: Evidence from Pakistani Macroeconomic Time Series*. Available Online at: https://ssrn.com/abstract=963958 (accessed February 20, 2020).

Zivot, E., and Andrews, D. (1992). Further evidence of great crash, the oil price shock and unit root hypothesis. *J. Bus. Econ. Stat.* 10, 251–270. doi: 10.1080/07350015.1992.10509904

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.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 18 September 2020
doi: 10.3389/frai.2020.00060

How Risk Profiles of Investors Affect Robo-Advised Portfolios

Dmitri Boreiko[1]* *and Francesca Massarotti*[2]

[1] Faculty of Economics and Management, Free University of Bozen-Bolzano, Bolzano, Italy, [2] Frankfurt School of Finance and Management, Frankfurt, Germany

Automated financial advising (robo-advising) has become an established practice in wealth management, yet very few studies have looked at the cross-section of the robo-advisors and the factors explaining the persistent variability in their portfolio allocation recommendations. Using a sample of 53 advising platforms from the US and Germany, we show that the underlying algorithms manage to identify different risk profiles, although substantial variability is evident even within the same investor types' groups. The robo-advisor expertise in a particular asset class seems to play a significant role, as does the geographical location, while the breadth of the offered investment choice (number of portfolios) across the robo-advisors under study does not seem to have an effect.

Keywords: FinTech, robo-advisors, investment advise, portfolio management, portfolio optimisation

OPEN ACCESS

Edited by:
Alessandra Tanda,
University of Pavia, Italy

Reviewed by:
Flavio Bazzana,
University of Trento, Italy
Daniele Grechi,
University of Insubria, Italy

Correspondence:
Dmitri Boreiko
dmitri.boreiko@unibz.it

Specialty section:
This article was submitted to Artificial Intelligence in Finance, a section of the journal Frontiers in Artificial Intelligence

Received: *31 December 2019*
Accepted: *09 July 2020*
Published: *18 September 2020*

Citation:
Boreiko D and Massarotti F (2020) How Risk Profiles of Investors Affect Robo-Advised Portfolios. Front. Artif. Intell. 3:60. doi: 10.3389/frai.2020.00060

INTRODUCTION

Over the last decade, the financial industry experienced some radical changes. Following the financial crisis of 2007, increased regulatory burdens on incumbents and wide adaptations of new technologies led to the emergence of a new structure, where some of the areas are dominated by smaller, more efficient start-ups that use internet, blockchain, and social media to create new products for consumers. The financial technologies (FinTech) adaptation is a key strategic advantage on its way to being successful (Jung et al., 2018)[1]. Technological transformation is particularly evident when it comes to wealth management, retail banking, payments, and lending (Metha et al., 2019). The wealth management industry has not only undergone a transformation driven by technology, but there has been a change also in terms of demand that caused the overall increase of assets under management and the emergence of new players. As Blackrock's (2015) highlighted, the demand for financial advice has increased along with the household's level of cash, people's increased longevity, income gaps caused by retirement, and a general lack of financial literacy.

It is argued that Artificial Intelligence (AI) is one of the most promising technologies that would advance the transformation of the finance industry (Park et al., 2016). One of the most disruptive AI applications in finance has so far been the introduction of automated investment managers or digital advisors, more commonly known as robo-advisors (RAs). Based on each investor's characteristics, RAs deliver and execute portfolio allocation advice through automated algorithms on digital platforms. Purportedly, such a service is free from individual human adviser's biases but come at the cost of a one-size-fits-all problem and limitations introduced by the robo-advising algorithm (D'Accunto et al., 2019).

[1] For a brief history of how technological innovations impacted financial industry in the last 75 years, see Ashta and Biot-Paquerot (2018).

As of now, consumer adaptation to robo-advising services has been rather slow. Several factors are responsible for such dynamics. First is the lower familiarity with AI and robotic technologies of the investing clients who might dislike entrusting their funds to non-human control (Reuba, 2017; Belanche et al., 2019). Several other aspects behind consumer trust, such as service security, information quality, and general proficiency in Internet usage, also play a role in slower adaptation to RAs (Lee et al., 2018).

Second is the problem of supply, induced by the apprehension of investment-advice providers that the robo-advising services would cannibalize higher-margin human investment advice offered by the same firm[2]. In addition, such an online-based service is subject to the problems of consumer loyalty (Luo and Ye, 2019) that might lead to underinvestment from incumbents given the low switching costs. This problem is aggravated by industry rigidity in adopting advanced marketing methods such as, [for e.g., word-of-mouth marketing, social media, and the internet in general (Casaló et al., 2008)]. The final reason is the legal uncertainty that still surround the RAs business, starting from fundamental issues such as whether RAs are subject to "investment advice" regulations (MiFiD2 in Europe) or possible liability risks, currently of great interest as the markets are being hit hard by the COVID-19-induced crisis (Maume, 2019).

The objective of this research is to empirically examine portfolio recommendations from a diversified set of RAs. So far, academic research has mainly focused on traditional human financial advising, their advice variation, biases, and conflicts of interest. Considering the rapid growth in assets under management (AUM) over recent years[3], we believe it is necessary to empirically investigate portfolio recommendations provided by automated investment managers.

Previous research has highlighted the variability of the recommended asset allocation from different RAs, however, very few (to our knowledge) addressed the questions as to why this is the case and what, in general, affects RAs' portfolio recommendations. In this study we attempt to identify the factors behind the proposed split between asset classes. Following the research conducted by Mankowitz and Skilje (2018) and focusing on RAs offering advice to retail customers in the German and North American markets, we have investigated whether investors' risk profiles, the number of model portfolios offered by each RA, economies of scale, and RAs' target market have any influence on the final proposed asset allocation.

Based on the sample of active RAs operating in the United States and Germany in 2019 and three constructed generic investors' profiles, we obtained the proposed portfolio splits between equity and debt instruments for each combination of investor-type and RAs under study. Further econometric analysis identified significant variations in recommended equity exposure, thus confirming findings by Cerulli Associates (2015). Our results indicate that cross-firm variations are notably evident for the moderate and especially for the conservative investor types; aggressive investor profiles, on the contrary, seem to receive more uniform asset allocation proposals.

Our multivariate analysis used several plausible explanatory factors for the RAs recommendations. It emerged that the most significant factor impacting portfolio recommendation is the risk-profile of the investors, implying that the RAs included in the sample are able to identify their investors' preferences based on the data entered by the client. In addition, economies of scale have proven to be statistically significant, with equity-specialized RAs favoring equity-biased allocations and vice versa, confirming the findings of Baker and Dellaert (2018). Furthermore, in line with previous research conducted by Rieger et al. (2010), the country of origin seems to have a strong effect. Indeed, U.S.-based RAs tend to skew their recommendations toward equity, thus, probably, addressing the more risk-taking investment mentality of U.S. investors. Lastly, the number of model portfolios surprisingly does not influence portfolio recommendations and is not statistically significant under all specifications.

Recently, the robo-advising industry seemed to have lost its momentum. An absence of trust, legal uncertainty, and low profitability impacted on the rates of growth and a more wide-spread adoption of the technology. For example, ABN AMRO shuttered down its RA Prospery because of low profitability compared to the traditional private banking division[4]. However, the recent Covid-19 crisis has led to a higher participation rate and trading activities of the retail investors who have been gambling on the stock market since March 2020 (Economist, 2020; Financial Times, 2020). The imminent cost-cutting programs would increase the interest on the supply side in wider use of RAs in investment advice. Our results show that, although RAs seem to take into account the risk-profiles of investors, there is still large variability in the investment recommendations even for the same risk-type model investor or models produced by RAs in different jurisdictions. We call for the faster development of industry standards to instill more trust in consumers. Whether these would be adopted as a code of good practice within the financial industry or imposed by the legislators remains an open question.

The rest of the paper is organized as follows. In Robo-advising process and literature review, we briefly describe the robo-advising process and review accumulated academic literature. In Hypotheses, we formulate our research hypotheses. Methods presents the data set and the econometric methods used. Results discusses the main results. The last section addresses the limitations of the study, outlines potential future research agenda, and concludes the paper.

ROBO-ADVISING PROCESS AND LITERATURE REVIEW

It is argued that robotics, artificial intelligence, and blockchain are currently contributing to the transformation of many aspects of the financial industry (Bayon, 2018). Cocca (2016) identifies two streams of innovation in wealth management: virtualization

[2]This concern was voiced in several interviews with investment management firms we conducted while writing this paper.
[3]S&P (2016).
[4]https://www.finextra.com/newsarticle/33623/abn-amro-shuts-down-german-digital-wealth-manager-prospery (accessed November 08, 2019).

of the interactions with the substitution of the traditional face-to-face meetings with digital channels, and virtualization of the advisory content. The latter process is exactly what automated financial advisors or RAs offer. By leveraging the mistrust in traditional wealth management companies caused by the financial crisis, RAs are offering alternative ways to invest, purportedly free from the deficiencies of a more traditional approach.

Researchers and regulators have still not given an official definition of robo-advising. As argued by Deloitte (2016), the term "robo" indicates the reduced presence or complete absence of human interaction, with automated mathematical algorithms used to produce customized asset allocations. The term "advising" is used to talk about somebody (in this setting, something) giving advice on a matter such as wealth management. Put together, these two terms refer to online-based portfolio management solutions, tailored mainly to retail investors, and attempting to automate all advisory process stages.

The pioneers of robo-advisory platforms were the US-based firms Betterment and Wealthfront, that began offering investment advice to retail investors in 2010. The American market continues to be the largest and most profitable one. Statista (2019) estimates the assets under management (AuM) in Northern America to amount to circa $740 billion, whereas Central and Western Europe AuM is only $26 billion.

Several factors have facilitated the international proliferation of RAs, such as increased investors' protection regulation, higher usage of smartphones and internet access, and increasing awareness and sophistication of retail investors (Haffenden and Melone, 2016). Yet, despite RAs' robust industry growth rates, many still question the viability of the model. As Morningstar (2018) reports, it costs circa £300 to get a new advised client for a robo-advising business, which then generates only £70 in annual revenue.

When the first RAs emerged in the US in 2010, they represented rather basic online interfaces used by financial managers to control their clients' assets. Further evolution underwent four stages, as described by Deloitte (2016). The first stage envisaged the RAs issuing recommendations based on the results of an online questionnaire filled in by investors. Trades were conducted by investors on a different platform, without banks or brokers supporting the robo-advising process and executing orders. In addition, RAs also issued recommendations on individual stocks and bonds.

RAs 2.0 executed investors' trades in addition to providing them with portfolio recommendations. However, it is still not possible to talk about automated investment managers as there is still a human component; indeed, an investment manager is responsible for the supervision of the investment algorithm and oversees setting the investment rules. RAs 3.0 are currently a mainstream in the market with 80% of active players executing investment decisions and portfolio rebalancing automatically via the algorithms. Fund managers only oversee the whole process with limited human intervention. RAs 4.0 employ self-learning artificial intelligence tools for investment algorithms, with automatic rebalancing between asset classes in reaction to market movements and conditions, always complying with investors' preferences expressed via the questionnaire.

Given the different attitudes of investors toward digitalization, robo-advising can be segmented into two main sectors. The first one is pure robo-advising, which is completely free from human intervention in the advisory process. This results in considerably lower fees compared to traditional advisory services, attracting lower-income clientele. As reported by Ringe and Ruof (2018), pure RAs charged fees ranging between 0.4% (US market) and 0.8% (European markets), compared to human financial advising costing circa 1–2%. Pure RAs have become quite popular due to their propensity to avoid conflict of interests due to automation. Fisch et al. (2017) highlight that RAs are less exposed to conflict of interests due to their higher independence, smaller bias to recommend actively managed funds that generate commissions as a potential additional expense, more transparent cost structures, lower minimum investment requirements, and 24/7 availability.

Once the risk profile has been identified, the RAs usually employ modern portfolio theory to construct an optimal mean-variance allocation (Markowitz, 1952). Several optimization algorithms were tried that would work better for an automated advice design (Chen et al., 2019). The investment assets chosen are usually exchange-traded funds, that allow for passive cheap and liquid indexing strategies when investing in different asset classes. Moreover, continuous rebalancing, monitoring, and 24/7 accessibility can also be automatized (Sironi, 2016; Jung et al., 2019).

However, despite the continuous improvement of RAs and the substantial growth in AuM, the value of assets switching to automated investment managers from human financial advisors remains relatively low (Fisch et al., 2017). As was argued by Faloon and Scherer (2017), the modern RAs' questionnaires fail to uncover individual risk aversion and thus are not suited to model the clients' investment problems. Indeed, there is a tendency to retreat from robo-advising (Murray-West, 2018). A survey conducted by IW Capital (2018) has reported that 38% of investors would not count on digital advisers for managing their assets; many investors have discarded automated solutions because of increased market volatility following some economic events, such as Brexit. Several solutions aimed to alleviate these problems were offered that focused on the types of interactions between the algorithms and consumers (Glaser et al., 2019) or suggestions to demonstrate a higher perceived level of automation (Ruhr et al., 2019). However, the industry response was a move backward to a standard investment model, where investment managers utilize digital services for portfolio-rebalancing or asset allocation to optimize their quality of advisory services within a shorter time. Such a model was termed "hybrid robo-advising."

Surprisingly, the RAs phenomenon has received more attention in psychological and information-technologies scientific literature than in finance and economics research. D'Accunto et al. (2019) is a noticeable exception. In their study, the authors show that RAs help investors to diversify their portfolios and help to mitigate a set of well-known and frequent behavioral biases.

HYPOTHESES

Little academic research has addressed the question of the factors that cause advice variability across Ras, which has been well-documented in industry reports. Cerulli Associates (2015) analyzed the proposed asset allocations made by seven different RAs for a 27-years-old investor, whose investment goal was saving for retirement; the proportion of recommended equity obtained by them displayed substantial variation across the RAs under study, with recommendations ranging between 51 and 90%. The recommended exposure to fixed income also seemed to be non-uniform, fluctuating from 10 to 40%.

Foerster et al. (2015) have conducted a study on traditional financial advisors: using regression analysis, they have been able to demonstrate that some advisors, employed at traditional financial firms, fail at tailoring portfolio recommendations to their clients' individual needs and financial situations. Indeed, they have demonstrated that personal characteristics, such as the risk profile, only explain 12.2% of cross-firm variations in recommended equity exposure. According to Lam (2016), this issue can be explained by the fact that portfolio recommendation of traditional financial advisors is driven by their own beliefs, thus implying that they might impose their own opinions on clients' preferences. In contrast to human advisers, RAs tend to provide their recommendations systematically and respect the inputs of the clients; this implies that the proposed asset allocation should be highly influenced by the investors' observable characteristics, in particular by their risk profile. Following the results found by Mankowitz and Skilje (2018), the key factor explaining different weights in asset allocation was found to be investors' risk profiles. According to the authors, digital advisors were able to categorize investors based on their risk-tolerance and they were likely to give them different portfolio recommendations. This might stem from the requirement of financial regulators to provide investors with an asset allocation suitable to them[5]. However, the hypothesis sustained by Lam (2016) and Mankowitz and Skilje (2018) is in contrast with other research and opinions on Ras; automated investment managers have in fact been criticized for their methodology, often considered simplistic and inefficient (Tertilt and Scholz, 2017). Many believe RAs' questionnaires are not as detailed as the ones filled in by human financial advisors. In order to test whether RAs fail to assess their clients' risk profiles, we have formulated our first hypothesis.

H1: Variations in portfolio recommendation across RAs are explained by their ability to successfully identify investors' different risk profiles.

A plausible explanation for cross-variation in portfolio recommendation could be the variability in questionnaires' structure and format. One of the more general and frequent differences is the number of questions in the questionnaire. However, previous research (Tertilt and Scholz, 2017) found that it could not explain the investment advice variability. We instead decide to focus our analysis on the RAs' predominance to allocate investors into certain risk categories depending on their answers to the questionnaires. These risk-categories are associated with a certain number of model portfolios, which will be recommended once questionnaires have been answered. Thus, hypothesis two is formulated as follows:

H2: The ability of the RAs to fully reflect the investor's risk profiles in portfolio recommendation depends on the number of model portfolios offered. Differences in the number of model portfolios offered lead to higher variations in portfolio recommendations.

As has been mentioned before, RAs are very price-competitive–this is due to economies of scale as the variable costs per additional client are relatively unimportant. Bayon (2018) stresses that RAs also exploit economies of scale as clients' assets are managed based on a limited number of financial products. Baker and Dellaert (2018) put forward a hypothesis that RAs are not more transparent and honest than human financial advisors, and that digital advisors might be programmed to recommend products with the highest margins to the sponsor institution. As a result, it seems that RAs' developers could produce recommendations that would be skewed toward asset classes in which they have higher expertise. Thus, our third hypothesis is as follows:

H3: Cross-firm variations in portfolio recommendation could be explained by the RAs sponsors' expertise in different asset classes.

Rieger et al. (2010) have demonstrated that risk behavior varies across countries and cultural regions. Results showed that American investors tolerate more risk than European investors do. We have selected our sample of RAs advising US and German residents, hoping to see some considerable differences across the two regions. Our last hypothesis therefore is as follows:

H4: German RAs tend to recommend more conservative allocations than their United-States-based competitors, as manifested by the suggested proportion of investment in fixed income products.

METHODS

Sample

The focus of this research is on testing what causes variations in portfolio recommendations between RAs based in the United States and in Germany. Our choice of countries was motivated by interest to compare US-based RAs against non-US-based ones, and Germany featured the highest number of active players[6]. The first key and challenging step was in identifying the relevant market players. In the absence of any coherent database of operational RAs, we have relied on market reports (CBInsights, 2017; Fintechnews Switzerland, 2018) and on various reviews and comparisons of RAs found on dedicated blogs (Robo-Advisor Comparison for the United States and ExtraETF for Germany)[7]. From these sources we have constructed an initial sample of 84 active digital advisors, based either in Germany or

[5]See European Securities Markets Authority (ESMA) (2018). Guidelines on Certain Aspects of The MiFID II Suitability Requirements.

[6]Business Insider Intelligence, "The US still has the robo-advisor lead," Business Insider. (2017). Available online at: www.businessinsider.de/the-us-still-has-the-robo-advisor-lead-2017-4 (accessed August 20, 2019).

[7]https://www.roboadvisorpros.com/category/comparisons/ and https://de.extraetf.com/robo-advisor

TABLE 1 | Average management fees charged.

	United States $N = 25$ %	Germany $N = 28$ %	Diff. in means (Germany vs. US) %
Pure robo-advisors ($N = 34$)	0.34	0.90	+0.56***
Hybrid robo advisors ($N = 19$)	0.36	1.06	+0.80***
All ($N = 53$)	0.35	0.95	+0.60***

*The Table reports the level of management fees by country and by RA types. Pure/Hybrid RAs stand for services offered without / with possibility of human adviser interaction. *** denote the significance of the respective coefficients at 1% level.*

in the United States. We excluded RAs that were based in other countries but offered the services to US and German citizens, B2B advisors, or the ones that serviced a restricted group of investors[8]. The final sample consists of 62 B2C-oriented RAs. We also had to exclude some services that required a social security number as an input, leaving 53 RAs in the final sample. Of the RAs, 28 are based in Germany, while 25 of them are in the United States.

The pure RA model is prevalent in our sample–only 19 RAs offer the possibility to talk to a human financial advisor at some stage. This is in line with the increasing tendency to switch to hybrid robo-advisory, as more established financial institutions are launching their own robo-advising platforms, such as Charles Schwab in the United States and Castell'sche Bank in Germany.

We controlled the fees charged by RAs in our sample across the industry reported figures. As it can be seen in **Table 1**, U.S.-based pure-robo advisors included in the dataset show an average of 0.34% management fee, while German-based players tend to have higher fees of circa 0.9%. Hybrid RAs in our sample tend to be more expensive than pure RAs, regardless of the country of origin. Both findings are in line with industry-reported numbers.

To help investors to make conscious and profitable investment decisions, it is essential for an advisor to successfully identify their risk tolerance; the failure to do so might lead to the selection of sub-optimal asset allocation. For the sake of comparing portfolio recommendations, we have created the following three general investor profiles with varying risk attitudes that we called "conservative," "moderate," and "aggressive" investor's types. The assembled investors' profiles were fed to online questionnaires from 53 RAs, resulting in a collection of 159 portfolio recommendations[9]. The detailed information about the risk profiles' construction is given in **Appendix 1**. **Table 2** provides an overview of the general investors' profiles.

The recommended ratio of the investment in equity class was taken as the dependent variable. It was calculated in the

[8] For example, Ellevest, a US-based RA providing advice only for women.

[9] The exact structure of the questionnaires is available from the authors upon request.

TABLE 2 | Investors' profile descriptions.

	Conservative	Moderate	Aggressive
Age	48 years old	48 years old	48 years old
Gender	Male	Male	Male
Education	High school diploma	High school diploma	High school diploma
Marital status	Married	Married	Single
Dependents	Yes	No	No
Field of work	Logistic	Logistic	Logistic
Annual income	\$63,875[a] (€57,000)	\$63,875 (€57,000)	\$63,875 (€57,000)
Aim of investment	Saving for retirement	Saving for retirement	Saving for retirement
Investment philosophy	Minimize losses	Minimize losses and maximize returns	Maximize returns
Investment horizon	3–5 years	3–5 years	3–5 years
Risk tolerance	Low	Medium	High
Amount invested	\$6,386	\$6,386	\$6,386
per year	€5,700	€5,700	€5,700

[a] *EUR/USD exchange rate as per 14.05.2019.*

following way:

$$y_{ij} = \frac{RE_{ij}}{RE_{ij} + RFI_{ij} + ROtA_{ij}} \tag{1}$$

$$i \epsilon \{1,\ 2 \ldots .53\}$$
$$j \epsilon \left\{Conservative,\ Moderate,\ Agrressive\right\}$$

In Equation 1, y_{ij} is the ratio of the recommended equity (RE_{ij}) in the recommended portfolio composed of equity, fixed income, and other assets ($RE_{ij} + RFI_{ij} + \ ROtA_{ij}$).

Estimation Methods

We used the ordinary least square (OLS) regression model with corrections for autocorrelation and heteroscedasticity based on the Newey–West method. Considering the full sample and the formulated hypothesis, the following model has been tested:

$$\begin{aligned} y_{ij} = {} & \alpha_i + \beta_1 Country_i + \beta_2 Moderate_i + \ \beta_3 Conservative_i \\ & + \ \beta_4 Portfolios\ Offered_i + \beta_5 Equity\ Expertise_i \\ & + \beta_6 Fixed\ Income\ Expsrtise_i \\ & + \beta_7 Other\ Asset\ Expertise_i + \varepsilon_i \end{aligned} \tag{2}$$

Appendix 2 defines all the variables and their data sources. In order to test for the hypotheses formulated earlier, the independent variables for assessing the number of portfolios offered, the equity, and the fixed income expertise have been introduced. The numbers of portfolio offered, the key variable for testing for H2, is determined by manually counting the model portfolios offered by each RA. This information can be generally found on the website of the digital advisors; in some cases, however, it has been necessary to directly contact the provider for

TABLE 3 | Expertise in asset classes vs. geographical location.

Country	Variable	Mean	Std. Dev	Min	Max
United States	Expertise in equity	0.51**	0.24	0.01	1.00
	Expertise in FI	0.28	0.18	0.00	0.71
	Expertise in other assets	0.15	0.14	0.00	0.46
Germany	Expertise in equity	0.44**	0.21	0.14	1.00
	Expertise in FI	0.36**	0.15	0.00	0.60
	Expertise in other assets	0.20	0.14	0.00	0.50

*The Table reports the statistics of expertise in various asset classes by geographical location. Pure/Hybrid RAs stand for services offered without/with possibility of human adviser interaction. ** denote the significance of the respective coefficients at 5% level.*

more detailed information. The number was in all cases double-confirmed during the portfolio recommendation phase. The RAs in our sample offer on average seven model portfolios, with German RAs offering 7.7 portfolios, and US ones only 6.2.

A proxy capturing the expertise in investing in equity, fixed income, and other assets of each RAs was created in order to test the third hypothesis. RA's expertise has been proxied with the weights, $w_{ij,}$ of each investment class in the RAs investment universe. These have been calculated as the following:

$$w_{ij} = \frac{n_{ij}}{m_j} \tag{3}$$

Where, in Equation 3, n_{ij} indicates the number of the assets within asset class i and per RA j, while m_j represents the total number of assets in each RA investment universe. Securities were categorized following the industry convention by dividing them into equity, fixed income, and other assets groups. The categorization has been done manually based on the information provided by the digital advisors. In most cases, the list of investment vehicles is publicly available; when this was not the case, RAs were directly contacted.

Table 3 shows that US-based RAs tend to have more exposure to equity than their Germen peers. It is known that the capital markets' participants in the United States tend to be more likely to be risk-takers than their counterparts in Germany, who generally have a more conservative approach. In addition, it could be said that, in both countries, lower importance is given to the other assets; this could be explained by the fact that most of the digital advisers taken into consideration for this study do not include "other assets" in their investment universe.

TABLE 4 | Descriptive statistics for portfolio recommendations.

Investor style	Advice	Mean	Std. Dev	Min	Max
Aggressive	Equity	0.73***	0.23	0.18	1.00
	FI	0.21	0.21	0.00	0.75
	Others	0.06	0.07	0.00	0.33
Moderate	Equity	0.56**	0.23	0.14	1.00
	FI	0.38*	0.21	0.00	0.75
	Others	0.06	0.08	0.00	0.41
Conservative	Equity	0.35	0.28	0.00	1.00
	FI	0.59**	0.27	0.00	1.00
	Others	0.06	0.09	0.00	0.51

*The Table reports average values of recommended allocations to various asset classes by the investor risk profile. FI stands for Fixed Income investment class. ***, **, * denote the significance of the respective coefficients at 1,5, and 10% levels.*

TABLE 5 | Recommended allocation to equity.

	(1)	(2)	(3)	(4)
Moderate	−0.174*** (0.028)	−0.174*** (0.027)	−0.174*** (0.028)	−0.167*** (0.027)
Conservative	−0.376*** (0.042)	−0.376*** (0.041)	−0.376*** (0.041)	−0.373*** (0.039)
Location: USA		0.175** (0.053)	0.179** (0.053)	0.084* (0.043)
N of Portfolios Offered			0.002 (0.005)	0.000 (0.002)
Expertise in Equity				0.367*** (0.074)
Expertise in FI				−0.351*** (0.104)
Expertise in Other Assets				−0.154 (0.193)
Intercept	0.073*** (0.035)	0.647*** (0.044)	0.627*** (0.053)	0.649*** (0.084)
Adj. R^2	0.28	0.37	0.37	0.58
N obs.	159	159	159	159

The Table reports the results of the OLS regressions of recommended portfolio allocation to equity (in % to total investment) against the risk profile of the investors and other explanatory variables. All variables are defined in ***Appendix 2****. ***, **, * denote the significance of the respective coefficients at 1,5, and 10% levels.*

RESULTS

Univariate Analysis

In line with results found by Cerulli Associates (2015), even for the same risk profile, significant variations in asset allocation across RAs have been found. This is particularly true for equities and fixed income, characterized by higher standard deviations than the other assets, with the results reported in **Table 4**. The table displays the descriptive statistics for portfolio recommendation subdivided by asset classes and by above-described general investors' profile. Interestingly, results display that the conservative investors experience higher variances in recommended allocation across all asset classes.

Regression Results

Table 5 reports the results of the OLS regressions. The first regression excludes control variables. As predicted, more risk-averse profiles result in smaller share equity allocation. As seen in regression 2, US-based RAs generally recommend a higher investment in equity, even after controlling for all other factors. The breadth of the portfolio choice does not play a significant role, still RA's equity expertise positively affect allocation to equity asset class.

As can be seen in **Table 5**, the risk profiles are strongly statistically significant and the intercept, e.g., the *Aggressive$_i$* risk profile variable, is positively correlated with the equity investment recommendations. In the case of risk-taking investors, the share of equity in the recommended portfolio is likely to increase by 64.9%. On the contrary, if the investor has a moderate risk-tolerance or is risk-averse, RAs tend to recommend the final portfolio that feature a smaller equity stake (16.7 and 37.3%, respectively). This said, we fail to reject the first hypothesis that variations in portfolio recommendation across RAs are explained by the digital advisers' ability to successfully identify investors' different risk profiles.

We also demonstrate that the choice of portfolios offered does not influence the portfolio recommendation, hence they do not cause cross-firm variations. The variable *N. of Portfolios offered* is insignificant in all the regressions and this provides evidence against the second hypothesis. We find that differences in the number of model portfolios offered across various RAs do not lead to variations in portfolio recommendations.

In line with the economies of scale hypothesis, it has been found that RAs with more expertise in equities tend to base their recommendations more on this particular asset class. The regressions show that for these RAs the recommended share of equity is higher (36.7%). On the contrary, if the RA proves to have more expertise in fixed income, the weight toward equity for the recommended asset allocation is lower (−35.1%). As can be seen from **Table 5**, both *Expertise in Equity* and *in FI* variables are found to be strongly statistically significant across all regressions. Therefore, we find support for the third hypothesis for equity and fixed income expertise, but not for the *other assets* class.

Lastly, it also emerges that the USA domicile dummy is statistically significant. The US-based RAs generally recommend higher equity allocations (by 8.4% on average). Thus, in line with previous expectations and the related literature, geographical location does play a role in recommended portfolios. Therefore, we also confirm our fourth hypothesis of US-based RAs advice to be skewed to the equity assets.

We also run similar regressions for conservative, moderate, and aggressive investor groups separately. As can be seen in **Table 6**, the only variable that is statistically significant across all the three profiles is the equity expertise. This implies that the main factor affecting the increase of the equity portion in portfolio recommendation is RAs' experience in investing in equity products, thus confirming the economies-of-scale hypothesis (H3). Moreover, the expertise in fixed income has proven to be strongly statistically significant for both the conservatives and the moderate investors. Some expertise in FI investment reduces the recommended equity exposure by more than 40% for the conservative and moderate investors.

TABLE 6 | OLS Regression Results for each investor style.

	Conservative	Moderate	Aggressive
Intercept	−0.011	0.700***	0.716***
	(0.083)	(0.065)	(0.089)
Country	0.037	0.074	0.142*
	(0.064)	(0.052)	(0.064)
Portfolios offered	0.004	0.000	−0.004
	(0.004)	(0.003)	(0.004)
Equity expertise	0.800***	0.142*	0.156*
	(0.08)	(0.061)	(0.076)
Fixed income expertise	−0.414***	−0.548***	−0.090
	(0.112)	(0.102)	(0.11)
Other assets expertise	0.349	−0.404	−0.409
	(0.244)	(0.223)	(0.240)
R^2	0.61	0.52	0.33
N obs.	53	53	53

*The Table reports the results of the OLS regressions of recommended portfolio allocation to equity (in % to total investment) for each risk profile subsample against the selected explanatory variable. All variables are defined in **Appendix 2**. *** and * denote the significance of the respective coefficients at 1% and 10% levels.*

DISCUSSION AND CONCLUSIONS

In this study we have identified some of the factors influencing portfolio recommendations provided by RAs and thus causing cross-firm variations. Using a sample of cross-sectional data containing the asset allocation recommendations provided by 53 different digital advisors based either in the United States or in Germany, we analyzed whether RAs comply with financial regulations and recommend investors with different risk-preferences different portfolios. We find that in our sample, RAs successfully recognize investors' style and provide them with different portfolio recommendations, thus complying with financial regulations. Still, there is large variability in the investment recommendations even for the same risk-type model investor or produced by RAs in different jurisdictions. We call for faster development of the industry standards to instill more trust from consumers. Whether these would be adopted as a code of good practice within the financial industry or imposed by the legislators remains an open question.

We also confirm that the number of portfolios offered is not statistically significant in explaining the recommended equity weight in a portfolio, in line with the results of Mankowitz and Skilje (2018).

Furthermore, the equity ratio is found to have a positive and negative association with RA's equity and fixed income expertise, respectively, providing evidence for the direct effect of the economies of scales. In addition, the study demonstrates the existence of large inconsistencies in portfolio recommendations, especially for moderate and conservative investors. It could be concluded that economies of scale are considered a key factor affecting portfolio recommendation and, it being a firm-specific capability, it can also be considered the main factor causing cross-firm variations.

Lastly, results demonstrate that RAs based in the United States recommend higher equity exposure than German digital advisers. The fact that US-based RAs recommend 8.4% more equity that their German counterparts can be interpreted as a proof of existence of some essential beliefs or investment preferences in the United States that are not shared in Germany.

Certainly, our research design is not without limitations. First of all, our sample might suffer from selection bias since we were not able to include all active RAs in the USA and Germany. However, we believe that we have identified and included in the sample virtually all that are being advertised and that have open access to the general public.

As the study is focused only on the two countries, given the significance of country effect on portfolio recommendation and considering that investors' mentality is highly influenced by their country of origin, other conclusions could be drawn when changing the geographical location or when including other countries in the sample. If, however, an official database with information given by financial regulators would be available to academia, the sample could be enriched, and the generalizability of the study would improve.

Further research on robo-advising industry should focus on more geographically dispersed samples, preferably adding the time-series dimension by looking at the variation of recommendations for each RA across time. Comparing the rates of return earned by investing in recommended portfolios against the human-adviser averages or market indices also merits further attention. It is highly desirable to enrich the set of controls and add more dimensions to the typical risk profiles of investors. The impact of social media certainly deserves special attention. If more information about and from RAs were available, it would be interesting to control for the impact on performance, client attraction, and client retention of new publicity provided by articles published on a dedicated blog such as Extra ETF. However, given the secrecy and confidentiality of these start-ups, it is not possible to capture this effect.

Other potential suggestions for future research would be to focus on the development of the RAs market with the entrance of the established financial service sector actors. However, despite further improvements of the regression model via the enlargement of the sample or a wider choice of explanatory variables, the most challenging aspect of the research on RAs is still the lack of transparency and its effect on trust; considering that RAs are primarily private companies, finding relevant information about their operations, profitability, and business models is quite challenging. To sum up, little is still known about the future of the automated-advice industry and AI applications in finance. The COVID-19 shock to the financial industry is still to be gauged. Whether FinTech will be a victim or a savior remains to be seen.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

DB general structure, literature review, and discussion. FM data collection, empirical analysis, and hypotheses set-up. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Open Access Publishing Fund of the Free University of Bozen-Bolzano.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2020.00060/full#supplementary-material

REFERENCES

Ashta, A., and Biot-Paquerot, G. (2018). FinTech evolution: strategic value management issues in a fast-changing industry. *Strateg. Change* 27, 301–311. doi: 10.1002/jsc.2203

Baker, T., and Dellaert, B. G. C. (2018). Regulating robo-advice across the financial services industry. *Iowa Law Rev.* 103, 713–750. doi: 10.2139/ssrn.2932189

Bayon, P. S. (2018). "A Legal Framework for Robo-Advisors," in *Datenschutz/LegalTech: Tagungband des 21 Internationalen Rechtsinformatik Symposions*, eds E. Schweighofer, F. Kummer, A. Saarenpää, and B. Schafer (Bern: Weblaw), 311-318.

Belanche, D., Casalo, L., and Flavian, C. (2019). Artificial intelligence in fintech: understanding robo-advisors adoption among customers. *Ind. Manage. Data Syst.* 119, 1411–1430. doi: 10.15444/GFMC2019.04.02.03

Blackrock (2015). *Global Investor Pulse (Investor Pulse)*. Available online at: www.blackrockinvestorpulse.com (accessed June 11, 2020).

Casaló, L. V., Flavián, C., and Guinalíu, M. (2008). The role of satisfaction and website usability in developing customer loyalty and positive word-of-mouth in the e-banking services. *Int J Bank Market.* 26, 399–417. doi: 10.1108/02652320810902433

CBInsights (2017). *A Wealth Tech World: Mapping Robo-Advisors Around the Globe*. Available online at: www.cbinsights.com/research/robo-advisor-global-fintechmap/ (accessed June 11, 2020).

Cerulli Associates (2015). *Retail Direct Firms and Digital Advice Providers. Addressing Millennials, The Mass Market, and Robo Advice*. Boston: Cerulli Associates.

Chen, L., Liu, K., Wang, Y., and Zhang, H. (2019). "A portfolio selection model *for* robo-advisor," in *Proceedings of 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems, CCIS* (Hoboken: IEEE), 693–698.

Cocca, T. D. (2016). Potential and limitation of virtual advice in wealth management. *Capco Inst J Financ Transform.* 44, 45-57.

D'Accunto, F., Prabhala, N., and Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *Rev Financ Stud.* 32, 1983–2020. doi: 10.1093/rfs/hhz014

Deloitte (2016). *Investors 'Attitudes Towards Robo-Advisors*. London: Evidence from the US and the UK. Deloitte

Economist (2020). *Most Investors and Some Firms are Upbeat About the World Economy*. Economist.

European Securities and Markets Authority (ESMA) (2018). *Guidelines on Certain Aspects of The MiFID II Suitability Requirements*. Paris: European Securities and Markets Authority.

Faloon, M., and Scherer, B. (2017). Individualization of robo-advice. *J. Wealth Manage.* 20, 30–36. doi: 10.3905/jwm.2017.20.1.030

Financial Times (2020). *Coronavirus: Retail Investors Attempt to Navigate Crisis*. Financial Times.

Fintechnews Switzerland (2018). *Robo-Advisory Market in Germany*. Available online at: http://fintechnews.ch/roboadvisor_onlinewealth/robo-advisory-market-in-germany/16815/ (accessed June 11, 2020).

Fisch, J. E., Laboure, M., and Turner J. A. (2017). *The Economics of Complex Decision Making: The Emergence of the Robo Adviser*. Available online at: https://www.geog.ox.ac.uk/events/170911/Robo-vs-Human-Advisers-Aug-28.pdf (accessed 11 June 2020).

Foerster, S., Linnainmaa, J. T., Melzer, B. T., Previtero, A. (2015). *Retail Financial Advice: Does One Size Fit All? Chicago Booth Paper No.* 14-38. Chicago. doi: 10.3386/w20712

Glaser, F., Iliewa, Z., Jung, D., and Weber, M. (2019). "Towards designing robo-advisors for unexperienced investors with experience sampling of time-series data," in *Lecture Notes in Information Systems and Organisation* (Cham: Springer).

Haffenden, E., and Melone, C. (2016). *Investors' Attitudes towards Robo-Advisors*. Kreuzlingen: Evidence from the US and the UK. My Private Banking Research.

IW Capital (2018). *Investment in 2018: The Human Factor*. Available online at: https://iwcapital.co.uk/investments-2018-human-factor/ (accessed June 11, 2020).

Jung, D., Dorner, V., Weinhardt, C., and Pusmaz, H. (2018). Designing a robo-advisor for risk-averse, low-budget consumers. *Electronic Markets* 28, 367–380. doi: 10.1007/s12525-017-0279-9

Jung, D., Glaser, F., and Kopplin, W. (2019). "Robo-advisory: opportunities and risks for the future of financial advisory," in *Advances in Consulting Research: Recent Findings and Practical Cases*, ed. *V*. Nissen (Cham: Springer International Publishing), 405–427.

Lam, J. (2016). *Robo-Advisors: A Portfolio Management Perspective*. https://www.coursehero.com/file/46506470/Robo-Advisors-A-Portfolio-Management-Perspectivepdf/ (accessed 11 June 2020).

Lee, S., Choi, J., Ngo-Ye, T. L., and Cummings, M. (2018). "Modelling trust in the adoption decision process of robo-advisors: An agent-based simulation approach," *in Proceedings of the Tenth Midwest Association for Information Systems Conference (Saint Louis, MO),* 17-18.

Luo, Y., and Ye, Q. (2019). Unders0tanding consumers' loyalty to an online outshopping platform: the role of social capital and perceived value. *Sustainability* 11:5371. doi: 10.3390/su11195371

Mankowitz, J., and Skilje, J. (2018). *Cross-Firm Variations in Portfolio Recommendations for Robo-Advisors*. Stockholm: Stockholm School of Economics Working Paper.

Markowitz, H. (1952). Portfolio selection. *J Finance* 1, 77–91. doi: 10.1111/j.1540-6261.1952.tb01525.x

Maume, P. (2019). Reducing legal uncertainty and regulatory arbitrage for *roboadvice*. *Eur Company Financ Law Rev*. 16, 622–651. doi: 10.2139/ssrn.3420011

Metha, D., Striapunina, K., and Hamke A. (2019). In-Depth: FinTech 2019. *Statista Digital Market Outlook – Trend Report*. https://www.statista.com/study/45600/statista-report-fintech/ (accessed June 11, 2020).

Morningstar (2018). *Financial Services Observer: Robo-Advisor Upgrade! Installing a Program for Profitability: Digital Advice Raises Profits for Investment Services Industry*. https://files.pitchbook.com/website/files/pdf/Morningstar_Presentation_20180608_RoboAdvisor_Upgrade.pdf (accessed June 11, 2020).

Murray-West, R. (2018). *Robo Advice suffers Trust Crisis. FT Adviser*. https://ftadviser.com/investments/2018/06/19/robo-advice-suffers-trust-crisis/ (accessed June 11, 2020).

Park, J. Y., Ryu, J. P., and Shin, H. J. (2016), Robo advisors for portfolio management. *Adv. Sci. Technol. Lett.* 141, 104–108. doi: 10.14257/astl.2016.141.21

Reuba, K. (2017). *Robo-Advisors: Early Disruptors in Private Wealth Management. Allianz Investment Themes and Strategy*. Available online at: https://sg.allianzgi.com/en/insights/artificial-intelligence/robo-advisors-early-disruptors-in-private-wealth-management (accessed June 11, 2020).

Rieger, M. O., Wang, M., and Hens, T. (2010). *Prospect Theory Around the World. NCCR-Finrisk Working Paper Number*, 731.

Ringe, W. G., and Ruof, C. (2018). *A Regulatory Sandbox for Robo Advice. EBI Working Paper Series*, Frankfurt am Main. 26.

Ruhr, A., Berger, B., and Hess, T. (2019). "Can I control my robo-advisor? trade-offs in automation and user control in (digital) investment management," in *25th Americas Conference on Information Systems, AMCIS 2019* (Atlanta).

S&P (2016). *SandP Global Market Intelligence Report*. Available online at: https://spglobal.com/our-insights/US-Digital-advisor-Forecast-AUM-To-Surpass-450B-By-2021.html (accessed June 11, 2020).

Sironi, P. (2016). *FinTech Innovation: From Robo-Advisors to Goal Based Investing and Gamification. New York, NY: John Wiley and Sons Ltd.*

Statista (2019). *Digital Market Outlook: Robo-Advisors*. Available online at: https://statista.com/outlook/337/100/robo-advisors/worldwide (accessed June 11, 2020).

Tertilt, M., and Scholz, P. (2017). To advise, or not to advise—how robo-advisors evaluate the risk preferences of private investors. *J. Wealth Manage.* 21, 70–84. doi: 10.3905/jwm.2018.21.2.070

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.

frontiers in Artificial Intelligence

PERSPECTIVE
published: 15 October 2020
doi: 10.3389/frai.2020.00063

FinTech: A New Hedge for a Financial Re-intermediation. Strategy and Risk Perspectives

*Anna Omarini**

Department of Finance, Bocconi University, Milan, Italy

The emergence of new technologies and players, along with a favorable regulatory framework (PSD2 Directive), is changing the banking industry. FinTechs and TechFins have allowed the introduction of new services and changed the way customers interact to satisfy their financial needs. The FinTech landscape is constantly evolving in the market. Different business value propositions are entering the financial services industry, moving from increasing the user's experience to developing a time to market framework for banks to innovate products, processes, and channels, increasing the cost efficiency and looking for a "partnering on order" to lighten the regulatory burdens for banks. The many businesses of banks are changing their value chains, and banks' business models should do the same accordingly. Strategists could no longer take their value chains as a given; choices have to be made on what needs to be protected and maintained, what abandoned and the new on coming to make banks evolve and become more resilient in doing their job. Banking is shifting significantly from a pipeline, vertical paradigm, to open banking business models where open innovation, modularity, and ecosystem-based bank's business model may become the ongoing mainstream and paradigm to follow and develop. Opportunities and threats for banks are many and new ones to re-gaining their role in the market throughout a re-intermediation process.

Keywords: FinTech, open banking, platform, ecosystem, APIs, digitalization, re-intermediation, bank business model

Edited by:
Peter Schwendner, Zurich University of Applied Sciences, Switzerland

Reviewed by:
Bertrand Kian Hassani, University College London, United Kingdom
Jon Williams, Bangor University, United Kingdom

****Correspondence:***
Anna Omarini
anna.omarini@unibocconi.it

Specialty section:
This article was submitted to Artificial Intelligence in Finance, a section of the journal Frontiers in Artificial Intelligence

Received: *02 April 2020*
Accepted: *17 July 2020*
Published: *15 October 2020*

Citation:
Omarini A (2020) FinTech: A New Hedge for a Financial Re-intermediation. Strategy and Risk Perspectives. Front. Artif. Intell. 3:63. doi: 10.3389/frai.2020.00063

INTRODUCTION

The rise of the ever-increasing relation between technology and financial services is bringing significant change to the banking industry. Shifting market conditions, customer needs, the entrance of new players, and digital technologies, along with new regulations—such as the Payment Service Directive 2 (PSD2) in Europe that aims to increase innovation, competition, and transparency—are all reshaping the banking industry and the financial intermediation model as well (Brueggemann, 2017).

There are many definitions of FinTech (Omarini, 2019, p. 198); however, it can be summarized that one main feature regards any technology that may reduce or eliminate the costs of financial intermediation especially in three broad areas of finance (Das, 2019, p. 981): (a) raising capital, (b) allocating capital, and (c) transferring capital.

FinTechs seem to be disrupting all the banks' primary functions of maturity transformation (through competition in lending), allocation (through robo-advisor and crowd investing platforms), payment services (through the introduction of new payment platforms and interfaces), and information processing (through the use of big data, machine learning, and artificial intelligence—AI), as well as liquidity provision and risk pooling.

The Financial Stability Board (Financial Stability Board – FSB, 2017) describes well that FinTechs enable financial innovation so that it could result "in new business models, applications, processes, or products with an associated material effect on financial markets and institutions and the provision of financial services." Since FinTechs started entering the market, they have worked on two important selling points:

- Better use of data and
- Frictionless customer experience (speed of sign-up, no-fee transparency, peace of mind through activity notification, rich in choice, etc.) to deepen relationships

They have brought to the traditional banking industry a wave of competition and broken pipeline value chains, unbundling them into different modules of products or services, which may be combined among themselves. These companies on the one hand and the BigTechs (Google, Facebook, Apple, Samsung, Alibaba, etc.) on the other have been forcing the industry to change, transform, and evolve in a set of new financial intermediation directions.

Use of data and customer experience are both FinTechs' major assets and threats as well. On the one hand, they please the customers as individuals and introduce the paradigm of contextual banking. On the other, the two selling points are threatening both the incumbent players and regulators in different ways. For banks, it is even more urgent to react actively because their "no fee zone" is expanding, due to new regulations from the Consumer Financial Protection Bureaus (CFPB) and similar entities in different countries.

Advances in digitization "are increasing opportunities to create new products and services and transform businesses." The competitive landscape is experiencing major discontinuities, such as "ubiquitous connectivity, industry deregulation, technology convergence. All this is blurring industry boundaries and product definitions. These discontinuities are releasing worldwide flows of information, capital, product, and ideas, allowing non-traditional competitors to upend the status quo. At the same time, competition is intensifying, and profit margins are shrinking (KPMG, 2016). Managers can no longer focus solely on costs, product and process quality, speed and efficiency. For profitable growth, managers must also strive for new sources of innovation and creativity. Thus, the paradox of the twenty-first century economy: consumers have more choices that yield less satisfaction. Top management has more strategic options that yield less value" (Prahalad and Ramaswamy, 2004, p. 1–2).

The industry is also experiencing new risks (data privacy, cyber risks, data protection, etc.) and a changing framework of some old ones (operational—because of cloud compliance features—reputational risks, etc.). Different geographies are developing specific regulatory frameworks, and this is going to impact the way and the degree to which the industry is becoming most adaptable to change. In particular, "the strength and nature of the competitive advantages created by advances in AI could also harm the operations of efficient and competitive markets if consumers' ability to make informed decisions is constrained by high concentrations amongst market providers. Some analysts caution that the path of AI-based financial services technology may be similar to the path of other technology-based platforms (De Reuver et al., 2017) that have trended toward high-levels of market concentration (e.g., in Internet search and messaging). An AI/machine learning performance model improves through an abundance of data. Models that have a large market presence, therefore, have a built-in self-reinforcing advantage as their gains in market share improve the performance model, which could in turn further their gain in market share" [Department of the Treasury (The U.S.), 2018, p. 57].

The article is developed as follows: paragraph 2 and related 2.1 look at understanding how FinTechs are impacting value chains in the financial industry; paragraph 3 outlines open banking, platforms, and ecosystems as the main paradigms for banking and banks. Paragraph 4 describes the oncoming framework of risks. Paragraph 5 develops a brief conclusion and develops a discussion on the true next challenges for each actor of the market. Under these circumstances, having vision and build strategies, business models, and organizations is fundamental to standing the test of time.

FinTechs AND THE VALUE CHAINS IN THE FINANCIAL INDUSTRY

It is beneficial to remember how things worked before and after FinTechs and TechFins or big techs in the financial industry.

Banking models are shifting significantly from a pipeline, vertical, paradigm, to modular solutions that pave the way to new banking paradigms that entail higher levels of openness toward third parties and a growing number of modular services bundled together.

Value is created in platforms through economies of scope in production and innovation (Gawer, 2014). In order for platforms to work, adoption and network effects are essential. Models can go to mere compliance with the prescriptions of openness of PSD2, to the inclusion of new services, the opening of the banking core and data, and the aggregation of those within a platform experience. In particular, we assist both to the evolution of a Bank-as-a-Platform model and a tech-platform-driven model supporting banking and financial intermediation, which both constitute a new interesting field of analysis.

Since the wave of digital transformation started entering the financial industry, banking-as-a-business has started moving from a product/service perspective to more contextual solutions where providers are customer needs-driven. This is because customer-driven companies outperform the shareholder-driven ones, and this requires an outside-in approach.

Having said that, it is beneficial to remember that digital transformation implies four main categories of innovation (product, process, organizational and business model) (Omarini, 2019, p. 340); all of them require rediscovering that a new strategy paradigm exists. This regards the concept of co-creation, and because of this no single firm can unilaterally carry out a process of continuous experimentation, risk reduction, time compression, and minimizing investment while maximizing market impact. Co-creation requires access to resources from extended networks (suppliers, partners, and consumer communities).

Under these new market conditions, FinTechs have become an important piece of a bigger puzzle, each one in its own area of business (payment, lending, etc.), while at the beginning most of them started as mono-business companies. Only a few of them may become leaders in the market. On the one hand, there are those that make their strategy become international, and on the other, there are FinTechs which enlarge their services-scopes. However, the majority of them will become part of ecosystems where the direction could swing from banks to tech companies or to FinTechs as well, able to manage the network by developing kinds of conglomerate-as-a-service.

Another interesting point to outline regards this recent period where all of us have experienced lockdowns around the world, and some effects have also impacted FinTechs as well. The valuations of most unicorns have crashed overnight, while on the FinTechs side there are different situations. Some of them have experienced a dramatic reduction in their evaluation, others were quite lucky and suffered less.

There are many and different feelings on the way FinTechs will exit this situation, which as far as we understand has overall accelerated some strategic choices.

First of all, there are many and different FinTechs in the market. What is critical is to look at the fundamentals of the business. All of them are about answering what society is going to look like in the future (attitudes, behaviors, habits, etc.), so that if we no longer need to go to retail stores anymore, why do we need some services based on this situation? This, again, underlines that banking is a people business (Omarini, 2015) and this requires a business to be resilient to become adaptive to consumer changes or moves into a different market where you can still apply the service because the society is not yet ready to shift somewhere else, which means the same business in different markets. Just think of the ongoing situation where the recent wave of people is rethinking and restructuring their finances, so that they have decided to switch rates to digital banks. In this scenario, the winners are those that have enough liquidity—or better still cash-rich—to buy good technology and invest in new directions, also taking the opportunity to use the pandemic to its advantage. This is especially true for payments that are going to be increasingly contactless. However, some more lessons can be learnt from difficult times especially due to external factors such as the following:

- People costs and per-customer contribution margin are key factors, and valuable indicators. They are valuable for incumbents too. When staff costs rise, then this becomes a burden if growth is not going to move on. Then, if we move on the per-customer contribution margin (revenue, minus variable costs including credit losses), then this makes a FinTech earn more money per bank account than the cost of running those bank accounts.
- One more point has to do with the way a FinTech makes its revenues per customer, and net income is the figure to look out for here. This means that the more sources of revenues a company holds, the better it is for it. If we think of some of the best-known FinTechs, they gather their net income from interchange fees, ATM withdrawals, which can diminish during the pandemic, but gathering revenues from other sources such as lending, investing, or again from referring customers to third-party services, and earning commissions from these referrals.

Under this oncoming market structure configuration, a focus on control and ownership of resources is giving way to the importance of accessing and leveraging resources through unique ways of collaboration. "The co-creation process also challenges the assumption that only the firm's aspirations matter. (…) Every participant in the experience network collaborates in value creation and competes in value extraction. This result in constant tension in the strategy development process, especially when the various units and individuals in the network must collectively execute that strategy. The key issue is this: balancing act between collaborating and competing is delicate and crucial" (Prahalad and Ramaswamy, 2004, p. 197).

If co-creation is fundamental to the industry, this needs to leverage on a wider customer perspective that requires introducing the idea of developing ecosystems where the customer is truly free to move and choose the best deal in more competitive markets able to let consumers' ability to make informed decisions against any possible market concentrations among market providers.

A business ecosystem (Moore, 1996) reflects the new paradigm of competition in a better way. Traditional management models aimed at gaining competitive advantage, such as vertical or horizontal integration, economies of scale and scope, are not effective anymore. The value of today's companies is determined by the size of its ecosystem (Tewari, 2014). Business ecosystems consist in crossovers of a variety of industries, of which companies cooperate and embrace open innovation to satisfy new customers' needs and develop new products and services, to improve the customers' experience (Moore, 1996).

Finally, it is worth outlining that in order to increase efficiency and costs optimization, there has been an increase in the use of the cloud that has also been fundamental for FinTechs to take off. Cloud technology—a part of the new construct of software-as-a-service, SaaS—is enabling organizations across the economy to more rapidly innovate (Chesbrough, 2003, 2006, 2011; Chesbrough et al., 2014) by reducing barriers to entry and acquire high-quality computing resources. On the one hand, cloud computing enables more convenient, on-demand access to computing resources (e.g., networks, servers, storage, applications, and services) [National Institute of Standards and Technology (NIST), 2011]. On the other, it makes banks and other financial service providers rely increasingly on third-party providers by increasing some related concerns and risks.

The Nascent Business Ecosystem: Concept, Rationale, and Approaches of Analysis

Ecosystems are cross-industry entities (Moore, 1993) where there is a loose of the "networks of suppliers, distributors, outsourcing firms, makers of related products or services, technology providers, and a host of other organizations that affect, and are affected by, the creation and delivery of a company's own offerings" (Iansiti and Levien, 2004b).

They are sets of data and features that combine to create value. The ecosystem economy is therefore linked to the marketability of the information that can be produced thanks to the integrated management of its data.

Ecosystems are characterized by both symbiotic and antagonistic relationships, without which each single player would lose its own individual meaning, so that the value relies on the interdependencies among actors (Adner and Kapoor, 2010; Gawer, 2014; Gawer and Cusumano, 2014). While the boundaries of an ecosystem may be blurred, companies should try to identify the players on which their success depends. In doing this, a new intermediation model is emerging, where different players can take several roles such as the "keystone" (Iansiti and Levien, 2004a and what Moore, 1993—initially defined as "leader" or "focal firm" according to Adner and Kapoor, 2010). This is a firm that furnishes a set of common resources on which other players can leverage.

Trying to make a parallelism, the ecosystems in which banks could find themselves working in require banks to look for a new re-intermediation model. This is because ecosystems are a technology stack structure supporting different value propositions which are mediated by the presence of other participants that increase "system value through direct and indirect network externalities" (Parker et al., 2016). In the meantime, this also "increases the likelihood of serendipitous interactions between partners, which may unlock new interactions and combinations" (Parker et al., 2016). In the ecosystem, partners have to focus on reaching a threshold level of coordination and create the (endogenous) boundary of the relevant ecosystem. The coordination is the key issue of a business ecosystem that under digital transformation is increasing its dependence on digital premises (Pagani, 2013). These partners may be, among others, the FinTechs that from an initial wave of fragmentation of the financial industry are now becoming the pillars of it by offering and increasing modularity and distributed banking throughout the re-bundling of their and others' value propositions. It is also worth outlining that in this scenario, financial services behave as a strong catalyst for the nascent ecosystems. This in fact allowed a major integration among interdependent, yet distinct modules belonging to the three areas of finance (raising, allocating, and transferring capital).

This takes us back, *mutatis mutandis*, to the main reasons why banks exist in the market (transaction costs and the problem of imperfect information, market signaling) (Benston et al., 1976; Leland and Pile, 1977; Campbell and Kracaw, 1980; Fama, 1980; Diamond and Dybvig, 1983; Diamond, 1984). In a nutshell, this is because they are information specialists and liquidity providers and are also able to transform and accept risks.

While the core objectives of financial intermediation have remained the same, the methods and functionalities relating to those objectives have been changed by new technology and market developments. At present, data analytics is frequently the preferred method of choice, and automated online computer programs are the favored functionalities of choice. Automated, algorithmic computer programs are now at the forefront of financial innovation. Just think of some of the human-led efforts in finance that have been replaced by artificial intelligent programs.

All this focuses attention on two points of analysis worth outlining. The first one is that like its traditional counterpart, new financial intermediation looks at developing the core purposes of financial intermediation, albeit by introducing new methods and functionalities (Lin et al., 2015). The second point is that (Brainard, 2017, p. 3): "More often than not, there is a banking organization somewhere in the FinTech stack. Just as third-party app developers rely on smartphone sensors, processors, and interfaces, FinTech developers need banks somewhere in the stack for such things as: (a) access to consumer deposits or related account data, (b) access to payment systems, (c) credit origination, or (d) compliance management. For instance, account comparison services rely on access to data from consumers' bank accounts. Savings and investment apps analyze transactions data from bank accounts to understand how to optimize performance and manage the funds consumers hold in those accounts." All this is due to the new ways (such as websites and apps) for intermediaries to interact with their clients.

Under these circumstances, we have to remember that financial services are fiduciary based, so that the more the ecosystem and its network are expanding, the more critical limitation of direct transactions may emerge in the market. "Taking this into consideration, there is the natural mutual distrust that derives from not knowing each other well. (...) All this requires the agents of the network to trust the network itself (so that) there is a need to reduce the trust gaps to benefit from new technologies in the presence of large trust gaps." On the one hand, banks, again, "will produce and process the information needed to enable millions of anonymous individuals to interact and trade on the web, while their reputational capital and expertise will be necessary to validate the quality of the information exchanged." On the other, "as networks bring in more participants and business opportunities, such knowledge will be useful for the intermediaries themselves to provide risk aggregation and diversification services that cannot be performed by individual agents or that may be too costly for individuals to perform" (Omarini, 2019, p. 18–19). All this seems a "win–win situation" for both incumbent banks and FinTechs.

Any further steps into the era of e-finance will make the circuit process look increasingly sophisticated, and in the meantime, it reaffirms the virtuality of bank money—based on the promised issued by specialized entities—and will always call on banks to give money a real content and preserve it.

These are the roots of the open banking paradigm, where money, production, and investment have to be considered in an integrated way, where banking and finance interrelate differently over the economic development, but performing complementary functions essential to the economy, leading to different efficiency/stability configurations, which are the next challenges for regulators and authorities to foresee and discern.

THE SHIFTING PARADIGMS OF BANKING AND BANKS: OPEN BANKING AS THE GAME CHANGER OF THE RENEWED FINANCIAL INDUSTRY

Ensuring a proper working of competitive market forces can be considered one of the main reasons for open banking (*alias* PSD2) in Europe and other countries, where the goal of promoting competition in financial services is an explicit component of the regulator's mandate (Deloitte, 2017). Its adoption is at varying stages in 35 markets relating to products that account for approximately 90 percent of revenue pools in those markets (McKinsey, 2019, p. 11, 12).

PSD2 can be described as "A legislative framework to facilitate the entry of (such) new players and ensure they provide secure and efficient payment services. (...) making it easier to shop online and enabling new services to enter the market to manage (their) bank accounts, for example to keep track of (their spending) on different accounts" (European Commission, 2015). With this, many competitive boundaries have started to loosen because of deregulation and the reduction of borders among industries so that banks have found themselves facing massive competition in many of their business areas (card payments, current accounts, consumer loans, some insurance products, financial planning, and family cash management). It is worth outlining that payment services are the entry gate for every other financial need (Omarini, 2019); transferring money is the most important pillar for any service extension. In fact, the "competition between banks and big techs is already fully visible in the area of payments where the market share of non-bank electronic payment providers, which offer alternatives to traditional credit and debit cards, is growing. Nearly 60% of retail banking transactions worldwide are now estimated to go through mobile and online providers, which offer alternatives to traditional credit and debit cards are growing" (Swiss Finance Council, 2020, p. 84).

Also, from the Basle-based BIS's annual Red Book report on payments and financial infrastructures, it is outlined that there are increasing incursions by non-bank competitors into both retail and wholesale payments, so that "The traditional bank-based ecosystem is being disrupted from below by FinTechs and from above by well-established big techs," states the report. This means that a new framework of financial intermediation system may emerge from the combination of incumbents, FinTechs, and big techs.

Big techs provide banking-like and other financial services together with their feature of being intrinsically linked to the rise of big data and data analytics and their related opportunities. All this is becoming an important driver for changing the automated decision-making process based on technologies, like artificial intelligence, and therefore make some impacts on the financial intermediation model.

However, it is worth outlining that "there are jurisdictional differences: the penetration of big techs in payments is more prominent in countries where the use of other cashless means of payments (e.g., credit cards) is low. For instance, big tech mobile payment services account for 16% of GDP in China" (Swiss Finance Council, 2020, p. 87).

If payments act as the entry level for them into the financial services industry, some big tech firms are also active in lending and asset management. Again, "there are geographical differences. For instance, the provision of credit by big techs has expanded more strongly than other FinTech credit in those jurisdictions with lighter financial regulation and higher banking sector concentration. These lending services have mainly been developed to sustain big techs' e-commerce platforms, and the data derived from e-commerce transactions have become a powerful tool for big techs in providing loans to consumers" (Swiss Finance Council, 2020, p. 87). On the expansion of big techs into asset management, this is mainly driven by their payment platforms and often regards a set of short-term investments, such as money market funds from customers' accounts' balances.

Until now, the emergence of big techs has not led to the disintermediation of the banking system. They have often acted as distribution channels relying on existing infrastructures like bank accounts or correspondent banking for cross-border transactions. Another point to make regards the fact that big techs still depend on big banks to access customers' accounts and big banks can benefit from big techs' network effect to expand their customer base, this seems to reach a win–win game, so that partnerships between them might increase. Just think of the partnering of Apple with Goldman Sachs for credit card provision to name but one. They have also become useful partners to banks by providing big banks with technological infrastructures such as cloud computing for data storage and processing. Another link between the two players is that of funding. This occurs because big techs fund themselves from financial markets and financial institutions like banks.

As a matter of fact, we can see that banks and big techs are developing different frameworks of collaboration, which are having their momentum at present. However, competition between the two players may rise, and this comes from a future question which will regard to what extent big techs will eat into big banks' revenue share and profit margins. This may be possible because these firms have low-cost structures that can easily be scaled up—they were born to be platforms—and become able to provide basic financial services, especially delivering these set of services to the underbanked and unbanked segments of population.

Their competitive edge also comes from the fact that for regulatory and reputational reasons, banks have thus far not been as effective as big techs in harnessing data, and network externalities, and if things remain like today, big techs would not have to face high capital requirements, massive and complex regulations and stringent compliance (AML/KYC), and security (data, cyber) obligations. In the long-term horizon, big techs "by partnering with licensed banks can offer financial services to their customers without having to accept deposits and become subject to strict banking regulation. The best-known example of such a collaborative platform is to be found in payments with the widespread adoption of APIs. But other forms of partnerships between global banking and big techs are emerging in, for instance, bank loans to technology firms' customers such as small and medium-sized enterprises (SMEs)" (Swiss Finance Council, 2020, p. 91). All this is creating new scenarios in the financial service landscape where barriers are diminishing, and stability and customer trust are once again becoming important issues.

Some Further Reflections on Open Banking

All the above is putting important roots for market regulators and market forces to boost the open banking (OB) agenda. As noted, UK regulators are taking a very active approach to open banking so that the Competitive Market Authority (CMA) has implemented its own reforms sometime beyond the PSD2. Further, the CMA has decided to set up the Open Banking Implementation Entity (OBIE) to support industry transformation. It is also interesting to outline the regulators' approach in China, where they developed an opposite framework by taking a more organic hands-off approach. That dichotomy shows that there is no single regulatory path or approach to open banking; local customs, standards, and expectations will dictate what is best. However, that is the direction, at present.

Incumbent banks understand that OB can assist with customer onboarding, retention, and satisfaction. As more FinTechs make their mark, there is an appetite for greater collaboration across the board. Banks are looking to improve customer value by adding some pieces of FinTech services to their existing financial expertise. This can be good, but it might not be enough to compete. This is because the greater focus on good customer outcomes means that services like categorization and aggregation will be table stakes. The winners will be the ones that place users at the heart of their approach and focus on delivering tangible customer value. Banking's holy grail is a combination of personability and relevance, and this is because this paradigm of banking will increase the number of conglomerate-as-a-platforms which are profitable and resilient only if they are able to develop themselves on a consistent and coherent customer experience evolutionary model. This evolution requires being rooted within a common framework of customer value and a strong innovative cultural organization that the entire conglomerate should outline in new rules for being an ecosystem where each part requires reliance on others' well-being.

From a technological point of view, open banking relies overall upon open Application Programming Interfaces (APIs) that are a set of codes and protocols that decide how different software components should interact. APIs are essentially allowing different applications to communicate with one another (Deutsche Bank Global Transaction Banking, 2018). APIs represent the interface through which third parties can develop and provide their services (*alias*, open innovation) defining the scope and level of access to the platforms (Microsoft Avanade and Accenture, 2017).

Through open banking, APIs are nowadays being used to issue commands to third-party providers. Before, they were used to connect developers to payment networks and display some details such as that of billings on a bank's website. They also allow for a close-to-seamless melding of services. In addition, transactional data unlocks a huge potential for greater transparency, and this also increases responsibility in the credit decision-making process.

Banks that are looking to out-pace their competitors are embedding new services into apps and websites, choosing to partner over doing it themselves. An increasing number of companies are realizing the impact a solid API strategy can have on their business, and banks are among them. This is because if 2018 signaled the huge potential presented by open banking, 2019 was the year OB started becoming realized on a more massive scale—for banks, businesses, and consumers alike. Under these circumstances, the challenge, at present, is that of balancing both endogenous and exogenous evolution.

APIs are also useful to develop Banking-as-a-Service (BaaS) to function properly; this is a key component of open banking (Zachariadis and Ozcan, 2017; Omarini, 2019).

BaaS is an end-to-end process that connects FinTechs and other third parties to banks' systems directly through the use of APIs. It helps to build up banks' offerings on top of financial providers' regulated infrastructure. However, a further step is Banking-as-a-Platform (BaaP), which is the next logical step that goes far beyond compliance with PSD2.

Banking-as-a-Platform represents just a subset of open banking, in which the choices of value and openness that banks make create several ties and roles with peculiar economics. BaaP builds on the advantages of open innovation, in putting together diverse know-how and resources (Zachariadis and Ozcan, 2017). Platforms are constructs that have the fundamental role of mediating relationships among different sides of users by reducing transaction costs and generating network effects.

They are organized around a core of elements that can constitute the basis for building innovative solutions and aggregating them toward a wider proposition. This emerging strategy acknowledges the modularization of banking services, but it tries to take advantage of the new opportunities that it spurs. Banking is indeed susceptible to migrating toward a platform model to pursue new revenue streams, as competition from FinTechs and TechFins might be unbeatable for a given set of services. The result would be an innovative proposition, supported by new business model frameworks, in which players share the costs of innovation and modules are aggregated to provide added-value services or bundles of services, and in which banks might forgo certain modules to concentrate in the orchestration of the network.

Banks, therefore, can take an active role in matching groups of users (e.g., FinTechs; developers; vendors; consumers, etc.), being the mediator through which all the groups get in contact with each other as well as become an orchestrator of the infrastructure. In doing so, they may regain their centrality in the economy and overall in their customers' everyday life. For many reasons (below zero interest rates, low profitability, increased new competition, value chains deconstructions, etc.), banks have to become an active player in a new re-intermediation open model where value is created in and through platforms and driven by nascent ecosystems business models.

In fact, open banking is an umbrella name to develop many new business frameworks over the next few years.

As a matter of fact, all the new constructs will be the frameworks for infinite interconnected financial intermediation ecosystems where banking is becoming an "enabler," and under these circumstances banks may still retain some significant strengths in entire segments as well as resources—e.g., regulation expertise, licenses etc. (Deutsche Bank, 2014; Omarini, 2017). At this point, rather than merely providing a product, FinTechs can further act as an agile consultative partner in the implementation process. In these ways, they can help incumbents to tackle their technological and organizational transformation, while keeping up with—or getting ahead of—competition.

Finally, trying to summarize all the above, Open Banking acts differently according to the player party. On the one hand, if we look at FinTech this may be a kind of detonator to scale up and become profitable. On the other, if we look at the incumbents' side, then this is an opportunity to disrupt and make them evolve from both the inside and the outside.

However, not all the incumbents are looking at Open Banking as an opportunity to change. This is because there are some pros but also some cons. The latter has to do with the different culture, organization, and skills available in FinTechs compared to those belonging to incumbents.

Among the incumbents that are making the most from the new environment, we can outline BBVA, HSBC, and Goldman Sachs, just to mention a few of the more interesting examples. All of them are undertaking different strategies and related actions to overcome the new environment.

These examples show how important it is to look at OB as a way to improve and boost their core market but also look beyond it in order to increase their resilience and develop a strong strategy. A final point, Open Banking, is not only for big firms; it can be developed under a strong commitment such as the case of Banca Sella, a medium-size Italian banking group.

RISKS FRAMEWORK IN THE ONCOMING SCENARIOS

We are currently in the early stages of transforming the banking sector and the implementation of new technologies, where both regulators and supervisors have to face the additional challenge of the digital transformation, which requires achieving the right balance between promoting new digital value propositions—and protecting against the risks inherent to the digitalization of financial services (Gonzalez-Paramo, 2017).

In the above scenarios, there are old risks as well as new ones. The latter come from the increasing use of big data, robo-advisor platforms, AI, and machine learning and other seamless tools for tutoring customers, all of them aimed at increasing customer personalization and user experience to deepen relationships.

All this is increasing attention on both consumer protection and product governance regulation, because more innovation is contextualized in other customers' needs and it is fundamental to protect a true well-informed customer choice.

If we consider financial innovation in the context of consumer protection, it can be said that innovations "do not necessarily create new problems, but they have a tendency to aggravate the existing challenges of asymmetric information, market power imbalances and other imperfections that typically characterized markets for retail financial products" (Lumpkin, 2010, p. 39).

Another interesting point is that FinTechs are promoting a massive use of open APIs through mobile devices. On the one hand, this is rising the IT interdependencies between market players and infrastructures, and by this way IT risks increase IT risks events, which could escalate into a full-blown systemic crisis (Waupsh, 2017); new forms of moral hazard and shadow banking may come into the industry. On the other, smart, connected products tech-stack driven provide a gateway for data exchange between the product and the user and integrate data from different key points, such as business systems, external sources, and other related products. All this is increasing the customization and personalization of financial services because of the changing way of customers' interactions, where those relationships are becoming continuous and open-ended (Porter and Heppelmann, 2014). However, this may raise the lock-in effect for customers and possible sub-optimization in their decision processes and selected choices.

From a managerial point of view, this is also challenging functions and related processes requiring a far more intense coordination among old and new functions and skills able to manage new forms of cross-functional collaborations. Finally, this also forces companies to redefine their industries and rethink almost everything they do, by starting with their visions and strategies.

All this has a great potential to transform the banking industry significantly, and as a consequence, most regulators and supervisors around the world have taken a closer look at this situation also monitoring both opportunities and risks that technology may bring to the industry. This is because, on the one hand, the market is experiencing new ways of using data, new types of market players, and business models. On the other, there are also new cyber threats among the top issues for regulatory bodies to focus on (see **Table 1**). The regulatory response has happened at different speeds globally and in the next few years will shape the future of financial technology and the industry as well.

On this issue, it is worth outlining a recent choice made by the Australian authorities that has delayed the introduction of open banking rules overall because of testing and security of the new provisions for account data sharing. Under the new deadline (1 July 2020), consumers will be able to ask major banks to share their credit and debit card information, as well as deposit account and transaction account data with accredited service providers. A further step regards consumers' mortgage and personal loan data that will be shared after 1 November 2020.

In the new framework, consumer data right needs to have a robust privacy protection and information securities, and this requires establishing appropriate regulatory settings and IT infrastructure around the world.

As mentioned above, for regulators, there is also the issue of protecting customers from misconduct and reassuring them

TABLE 1 | FinTech: regulators' focus.

Area of regulatory focus	Regulatori objectives	Regulatory response
Data usage	- Protect individual privacy - Ensure data is not misuses or manipulated - Prevent data leakage - Prevent unethical use of data	- Data protection and data privacy requirements - Advice on ethical aspects of using data
New market players and business models	- Support competition and innovation - Set level playing field for FinTech firms and banks - Secure the safety of the financial system as a whole	- Opening client data to FinTech firms in a secure manner - Licensing and authorization of FinTech firms - "Same services, same rules" approach - Encouraging responsible innovation - Technology-neutral rules
New cyber threats	- Ensure cyber security and client protection	- Customer awareness - Secure communication - Strong customer authentication - Technical preventive measures - Fraud monitoring and detection

Source: (Deutsche Bank Global Transaction Banking, 2018), p. 7.

about making the right choice. In fact, there is a vast body of literature showing that consumers tend to make poor financial choices, such as not buying the "best value" products on offer and so taking on too much debt, misunderstanding investment risk, and choosing financial products that do not meet their real needs.

A big issue for regulators is also to keep up both stability and competition, which might become weaker, the more consumers' freedom decreases in the market. The main reason for this depends on some developed constrains to customer mobility as well as the trap of being so well-known that for the customer it is difficult to quit the situation. This moves the focus for regulators on the systemic risk from being "too-big-to-fail," which refers to a few large financial intermediaries, to the systemic threat of "too linked to fail" (Lin et al., 2015), which includes instruments and intermediaries that are small in value and headcount but could destabilize the system because of the role they play in the networked marketplace.

It is worth outlining that most of the authorities and regulators have the approach of looking at FinTech as a single entity, on a case-by-case basis given the wide range of underlying applications (DTCC, 2017); other work is done on the definition of risks (BIS, 2017).

This can only be considered an initial step, because FinTechs are not going to remain that way for longer. They are becoming more and more part of the banking industry, and most of them are already partnering with banks or developing different frameworks of collaboration.

This wave of change finds its root in the way technological disruptions, along with regulation, could move an industry from a vertically integrated model to a multisided platform model. Therefore, moving attention from the single FinTech company to the platform framework might benefit the stability of the market as a whole, and the single company itself.

For a true competition in the ecosystems, regulators should consider the way the number and intensity of participants in the ecosystem are made possible. This fact shifts attention toward the level of openness of a platform, which is strictly linked to the need of alignment, coordination, and robustness of the platform itself, and the exact selection of the openness choices that may have to do with its stability and soundness. This factor may generate a trade-off between value capture (hindered by a too-open approach), and value proposition and platform adoption (hindered by a too-closed approach), which is a conflict between profiting from the platform and the network effects and reduction in costs it generates for all participants (West, 2003).

All this above increases systemic value, and it may produce some direct and indirect network externalities (Parker et al., 2016). This means that the resiliency of such business ecosystems requires having a threshold level of coordination to align each member in the overall value blueprint.

When it comes to the bank-specific managerial implications of such a platform choice, a bank might have to choose an approach in which it should hold restrictive terms and conditions and a burdensome due diligence process to ensure compliance with the law, of which the bank is ultimately responsible. This is because by widening the scope of the platform and augmenting the number of modules to sustain economies of scope and the related network effects, this could result in bottlenecks from ancillary activities or even lock in effects for customers. This requires a platform to compete effectively on each side to attract a fair number of members of each group not to create the incentive to subsidize those categories, which would generate most network effects for other parties, putting competitive pressure on prices (Armstrong, 2006), or/and decreasing transparency from a customer protection perspective.

In addition, the open approach toward banking may raise potential instability and risk factor, where the multiplication of the actors heightens the complexity of the system and creates potential breaches. Yet, it remains to be seen whether the technical standards and the due diligence conducted by banks will be sufficient in mitigating the risk of breaches and misuse of data and, more broadly, operational risks (large-scale theft, data corruption, etc.).

From the Basel agreements onward, the regulatory framework has changed the focus from what and how a bank can do to what a bank can do according to its capital adequacy by, first, mapping, and then managing the many risks it can undertake in doing its activity. Regulators have reinforced the prudential regulation compared to the structural regulation to reduce the risk of bank failure by prohibiting banks from getting involved in activities, which are judged by policymakers to be "too risky."

In this approach, there are two possible weaknesses. The first one is that the set of prudential measures is affected by a strong endogeneity, which is the property of being influenced within a system. The second weakness is that regulation is trying to overcome a situation through tools based on a set of linear relationships, but overall, the oncoming financial intermediation system is neither linear nor simple. On this same issue, it is interesting to report the US Treasury's recommendations, which are the following (The US Department of the Treasury, 2018 p. 15):

- Adapting regulatory approaches to changes in the aggregation, sharing, and use of consumer financial data, and to support the development of key competitive technologies;
- Aligning the regulatory framework to combat unnecessary regulatory fragmentation, and account for new business models enabled by financial technologies;
- Updating activity-specific regulations across a range of products and services offered by non-bank financial institutions, many of which have become outdated in light of technological advances; and
- Advocating an approach to regulation that enables responsible experimentation in the financial sector, improves regulatory agility (…)."

DISCUSSION ON THE MAIN CONCLUSIONS

There are some important trends that have arisen in the market over the last decade such as the following:

- The nonbank sector has become powerful in the market so that regulatory challenges placed on traditional financial institutions have increased, such as those including the launch of numerous startup platforms;
- Most of these platforms have grown fast and beyond their startup phase. They have also implemented technology-driven approaches to onboard customers as well as process consumers' requests;
- Innovative new platforms in the nonbank financial sector are, in some cases, standalone providers. However, there are also others focusing to provide support for or interconnectivity with traditional financial institutions through partnerships, joint ventures, or other means;
- Big tech-driven companies holding a huge amount of consumer data have simultaneously entered the financial services industry, primarily in payments and credit provision;
- Over time technology-enabled competitors have scaled up, and the corresponding threat of disruption have raised the crossbar for the existing firms to boost their innovation processes in a faster manner and also look for dynamic and adaptive strategies. As a result, mature firms have launched platforms aimed at reclaiming market share through alternative delivery systems and at lower costs than they were previously able to provide.

This requires new strategic thinking which is moving on regarding the future of money that has become a more complex subject.

Money is the tool through which savings, investments, and capitals are held in the economy, no matter the form (digital or otherwise). Therefore, money changes follow changes in society. However, the minting of any representation of value is backed by the full faith and credit of the issuer whoever that may be.

At present, value is in the financial needs (spending, savings, lending, etc.) and in their related and different performances (monetary and not-monetary benefits).

The new current outlook reveals nascent ecosystems made of independent actors, where the traditional supply-centered oligopoly is coupled with FinTechs, TechFins, retailers, etc. Within this lies the disruptive aspect of PSD2 in Europe and similar trends in other markets. This is a key milestone itself in the unbundling and modularization—and more recently the re-bundling—of many and different banking and non-banking services which is challenging the financial services landscape (Omarini, 2019, p. 369).

The difference with the past on the relationship between technology and banking is the stronger interdependencies from a double perspective: technological and, even more important, strategic interdependence.

As mentioned above, the challenge for regulators is to move from a single-entity perspective (FinTech or bank focus) to a broader perspective, based on the banking conglomerate-as-a-platform. In particular, because platforms develop interactions among different, new and old, stakeholders, innovating fitting might require the development of new rules. In this, there is a critical point to control, which is the balancing of the power among the different actors. This is particularly true when banking and financial intermediation is increasing reliant on technology, and on the other way around, technology is driving the banking and financial businesses, pushing them sometimes outside the traditional boundaries. This may open the door to the next wave of shadow banking the more financial services are hidden in everyday life of customers and diluted in their habits. In this new changing game, consumer trust still remains a central component for each player in working toward open banking. From the financial institution through to third-party provider relationship and potential suppliers in between, there is a necessity to build and maintain consumer trust that will act as a catalyst for building competition. This trust requires both regulators' and companies' attention to third-party risks and relationships that have augmented for many different reasons, including those related to consumer's concerns, information security concerns, and other operational risks.

There are two main reasons why banks should react to this changing environment by actively managing their business lines. The first one regards the need for them to regain their centrality in their customers' everyday lives. The second reason is that banks are expected to react because of the low interest rate situation affecting bank intermediation margins.

At present, the response depends on whether or not the situation is perceived to be long-lasting by each bank in the market.

According to BIS (2019a, p. 3), "low interest rates encourage banks to rebalance their activities from interest-generating to fee-generating and trading business lines. The impact is economically significant. According to our estimates, the long-term elasticity of fees and commissions with respect to the policy rate is 0.93, which means that for each 1% decline in the policy rate, income from fees and commissions increases by 0.93%. And the longer that low interest rates persist, the more this rebalancing effect is reinforced." This means that banks in order to move toward fee-based businesses may develop different frameworks of bank and FinTech collaboration to speed up innovation and time to market responsiveness. This is increasingly important, if we consider that as persistent low interest rates tend to reduce bank profits mainly by depressing interest margins, "banks adjust their activities in an effort to offset that reduction, at least partially. (...) And they reveal that funding tends to shift from short-term market funding toward deposits" (BIS, 2019b, p. 12).

On this outline, we underline that the construct of open banking—throughout the PSD2 directive and similar regulations around the world—is paving the way for third parties to work more and more on banks' deposits with true chances to develop mechanisms of further new disintermediation unless banks react actively in becoming good and better at innovating and offering new propositions to their customers.

This situation underlines the urgent need for banks to counteract fiercely.

While digitally native firms often have an edge on data skills, banks may retain an advantage in handling soft, context-dependent information that cannot be reliably tracked from quantitative metrics. Even if the importance of this factor varies considerably across bank business segments, it exists in many of them—including, for example, small business lending and advisory services. This is another interesting asset for banks to consider and leverage on by improving, overall, through partnering with FinTechs. In fact, 79 percent of leading banks have partnered with FinTechs to foster innovation (McKinsey, 2019, p. 34).

Now the challenge is to make these partnerships working at their best. All this goes toward the core mechanism of value co-creation, which has been mentioned above, that is, the integration of resources of several actors. If we take a service perspective, resources may regard people, systems, infrastructures, and information (e.g., Grönroos, 2006), and also knowledge and skills are becoming central resources for a company (Vargo and Lusch, 2004, 2008; Lusch and Nambisan, 2015; Omarini, 2015). All of them are important ingredients for a platform to develop its service innovation (Lusch and Nambisan, 2015) as this is going to be the new financial intermediation paradigm.

Moreover, banks also benefit from a fairly sticky customer base that is from switching deposit institutions and is likely to work with banks with which it has an existing relationship. Therefore, retaining and developing a loyal customer base will be increasingly important in the future (Omarini, 2013). While the relationship-based dimensions of banking may be on a long-term trend of erosion, due to changes in lending technology and banking regulation, they are unlikely to disappear altogether.

DATA AVAILABILITY STATEMENT

The original contributions generated for the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AO conceived the presented idea. She developed the study conception and design as well as the drafting of the manuscript and its critical revision.

REFERENCES

Adner, R., and Kapoor, R. (2010). Value creation in innovation ecosystems: how the structure of technological interdependence affects firm performance in new technology generations. *Strat. Manage. J.* 31, 306–333. doi: 10.1002/smj.821

Armstrong, M. (2006). Competition in two-sided markets. *Rand. J. Econ.* 37, 668–691. doi: 10.1111/j.1756-2171.2006.tb00037.x

Benston, G., Smith, J., and Clifford, W. Jr. (1976). A transactions cost approach to the theory of financial intermediation. *J. Finance Am. Finance Assoc.* 31, 215–231.

BIS (2017). *Implications of Fintech Developments for Banks and Bank Supervisors.* BIS.

BIS (2019a). *Bank Intermediation Activity in a Low Interest Rate Environment.* Working Papers. BIS.

BIS (2019b). *Annual Red Book Report on Payments and Financial Infrastructures.* BIS.

Brainard, L. (2017). Where Do Consumers Fit in the Fintech Stack? Governor Lael Brainard At "FinTech Risks and Opportunities: An Interdisciplinary Approach." A Conference Sponsored by the University of Michigan, Ann Arbor, MI, November 16, 2017. Retrieved from: https://www.federalreserve.gov/newsevents/speech/brainard20171116a.htm (accessed May 29, 2020).

Brueggemann, R. (2017). *Payment Services Directive (PSD2).* Swissbanking. Available online at: http://www.swissbanking.org/en/topics/current-issues/payment-services-directive-psd2 (accessed January 29, 2018).

Campbell, T. S., and Kracaw, W. A. (1980). Information production, market signalling, and the theory of financial intermediation. *J. Finance* 35, 863–882.

Chesbrough, H. W. (2003). *Open Innovation: The New Imperative for Creating and Benefiting From Technology*. Boston, MA: Harvard Business School Press.

Chesbrough, H. W. (2006). *Open Business Models: How to Thrive in the New Innovation Landscape*. Boston, MA: Harvard Business School Press.

Chesbrough, H. W. (2011). *Open Services Innovation: Rethinking your Business to Grow and Compete in a New Era*. San Francisco, CA: Wiley.

Chesbrough, H. W., Vanhaverbeke, W., and West, J. (2014). *New Frontiers in Open Innovation*. Oxford: Oxford University Press.

Das, S. R. (2019). The future of fintech. *Finan. Manag.* 48, 981–1007. doi: 10.1111/fima.12297

De Reuver, M., and Ondrus, J. (2017). When technological superiority is not enough: the struggle to impose the SIM card as the NFC secure element for mobile payment platforms. *Telecommun. Policy*. doi: 10.1016/j.telpol.2017.01.004

Deloitte (2017). *Open Banking: What Does the Future Hold*? Deloitte. https://www2.deloitte.com/content/dam/Deloitte/sk/Documents/Random/Addison-Open%20Banking%20-%20What%20The%20Future%20Holds.pdf (accessed January 3, 2018).

Department of the Treasury (The U.S.) (2018). *A Financial System That Creates Economic Opportunities*. Nonbank Financials, Fintech, and Innovation.

Deutsche Bank (2014). *Fintech – The Digital (r) Evolution in the Financial Sector*. Deutsche Bank.

Deutsche Bank Global Transaction Banking (2018). *Regulation Driving Banking Transformation#Positive Impact Insights Into The Key Regulatory Developments Shaping APIs, Artificial Intelligence, Blockchain and Cloud*. Available online at: http://cib.db.com/insights-and-initiatives/flow/regulation-driving-banking-transformation.htm (accessed May 29, 2020).

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *Rev. Econ. Stud.* 51, 393–414.

Diamond, D. W., and Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *J. Politic. Econ.* 91, 401–419.

DTCC (2017). *Fintech and Financial Stability. Exploring How Technological Innovations Could Impact the Safety and Security of Global Markets*. White Paper. Available online at: http://www.dtcc.com/~/media/Files/PDFs/Fintech%20and%20Financial%20Stability.pdf (accessed April 20, 2020).

European Commission (2015). *Competition: Commissioner Vestager Announces Proposal for E-Commerce Ector Inquiry*. Available online at: https://europa.eu/rapid/press-reease_IP-15-4701_en.htm (accessed April 20, 2020).

Fama, E. F. (1980). Banking in the theory of finance. *J. Monetary Econ.* 6, 39–57.

Financial Stability Board – FSB (2017). *Financial Stability Implications from FinTech, Supervisory and Regulatoty Issues That Merit Authorities' Attention*. Available online at: https://fsb.org/wp-content/uploads/R270617.pdf (accessed May 10, 2020).

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Res. Policy* 43, 1239–1249. doi: 10.5465/ambpp.2014.278

Gawer, A., and Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. *J. Product Innovat. Manage.* 31, 417–433. doi: 10.1111/jpim.12105

Gonzalez-Paramo, J. M. (2017). *French Payments Committee Conference on Technological Innovation in Payment Systems*. Paris: Bank of France.

Grönroos, C. (2006). Adopting a service logic for marketing. *Market. Theory*. 6, 317–333. doi: 10.1177/1470593106066794

Iansiti, M., and Levien, R. (2004a). Strategy as ecology: a synthesis and research agenda. *Harvard Bus. Rev.* 82, 68–78. doi: 10.1108/IJOA-05-2015-0865

Iansiti, M., and Levien, R. (2004b). *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability*. Boston, MA: Harvard Business School Press.

KPMG (2016). *The Profitability of EU Banks*. Available online at: https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2016/10/the-profitability-of-eu-banks.pdf (accessed May 07, 2020).

Leland, H. E., and Pile, D. H. (1977). Informational asymmentries, financial structure, and financial intermediation. *J. Finance* 32, 371–387.

Lin, Z., Whinston, A., and Fan, S. (2015). Harnessing Internet finance with innovative cyber credit management. *Finan. Innov.* 1:7. doi: 10.1186/s40854-015-0004-7

Lumpkin, S. (2010). Consumer protection and financial innovation: a few basic propositions. *OECD J.* 2010, 117–139. doi: 10.1787/fmt-2010-5km7k9tp2jxv

Lusch, R. F., and Nambisan, S. (2015). Service innovation: a service-dominant logic perspective. *Mis Q.* 39, 155–175. doi: 10.25300/MISQ/2015/39.1.07

McKinsey (2019). *The Last Pit Stop? Time for Bold Late-Cycle Moves*. McKinsey Global Banking Annual Review.

Microsoft Avanade and Accenture (2017). *PSD2 and Open Banking: Using Regulation to Kick-Start the Transformation of Banking*. Microsoft Avanade and Accenture.

Moore, J. F. (1993). Predators and prey: a new ecology of competition. *Harvard Bus. Rev.* 71, 75–86.

Moore, J. F. (1996). *The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems*. New York, NY: Harper Business.

National Institute of Standards and Technology (NIST) (2011). *The NIST Definition of Cloud Computing, Special*. Available online at: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

Omarini, A. (2013). Multichannel distribution in banking: customers perspectives and theoretical frameworks to increase user acceptance of a multiplatform banking business. *J. Banks Bank Syst*. 8, 80–98.

Omarini, A. (2015). *Retail Banking. Business Transformation and Competitive Strategies for the Future*. London: Palgrave MacMillan Publishers.

Omarini, A. (2017). The digital transformation in banking and the role of FinTechs in the new financial intermediation scenario. *Int. J. Finance Econ. Trade* 1, 1–6.

Omarini, A. (2019). *Banks and Banking: Digital Transformation and the Hype of FinTech*. Milan: McGraw-Hill Education

Pagani, M. (2013). Digital business strategy and value creation: framing the dynamic cycle of control points. *MIS Q.* 37, 617–632. doi: 10.25300/MISQ/2013/37.2.13

Parker, G. G., Van Alstyne, M. V., and Choudary, S. P. (2016). *Platform Revolution: How Networked Markets Are Transforming the Economy - and How to Make Them Work for You*.

Porter, M., and Heppelmann, J. E. (2014). How smart, connected products are transforming companies. The operations and organizational structure of firms are being radically reshaped by products' evolution into intelligent, connected devices. *Harvard Business Review, 1st* Edn (New York, NY: W.W. Norton & Co), 3–23.

Prahalad, C. K., and Ramaswamy, V. (2004). *The Future of Competition. Cocreating Unique Value With Customers*. Boston, MA: Harvard Business School Press.

Swiss Finance Council (2020). *Getting Ready for the '20s – Technology and the Future of Global Banking, Report*. Swiss Finance Council.

Tewari, A. (2014). "Platform business models and mobile ecosystems," in *Conference: University of Koblenz*. doi: 10.13140/2.1.1371.9367

The US Department of the Treasury (2018). *A Financial System That Creates Economic Opportunities Nonbank Financials, Fintech, and Innovation U.S. Report to President Donald J. Trump*. Executive Order 13772 on Core Principles for Regulating the United States Financial System. Retrieved from: https://home.treasury.gov/sites/default/files/2018-08/A-Financial-System-that-Creates-Economic-Opportunities---Nonbank-Financials-Fintech-and-Innovation_0.pdf (accessed May 11, 2020).

Vargo, S. L., and Lusch, R. F. (2004). Evolving to a new dominant logic for marketing. *J. Market*. 68, 1–17. doi: 10.1509/jmkg.68.1.1.24036

Vargo, S. L., and Lusch, R. F. (2008). Service-dominant logic: Continuing the evolution. *J. Acad. Market. Sci.* 36, 1–10.

Waupsh, J. (2017). *Bankruption - How Community Banking Can Survive Fintech*. Hoboken, NJ: Wiley.

West, J. (2003). How open is open enough? Melding proprietary and open source platform strategies. *Res. Policy* 32, 1259–1285. doi: 10.1016/S0048-7333(03)00052-0

Zachariadis, M., and Ozcan, P. (2017). *The API Economy and Digital Transformation in Financial Services: the Case of Open Banking*. SWIFT Institute.

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.

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