



OPEN ACCESS

EDITED BY

Andrea Mazzon,
University of Verona, Italy

REVIEWED BY

Dragos Bozdog,
Stevens Institute of Technology, United States
Denis Velu,
Metropolitan University of Tirana, Albania

*CORRESPONDENCE

Massimiliano Frezza
✉ massimiliano.frezza@uniroma1.it

RECEIVED 13 November 2024

ACCEPTED 29 January 2025

PUBLISHED 17 February 2025

CITATION

Bianchi S, Pianese A, Frezza M and Angelini D
(2025) A new tool to detect financial data
scaling. *Front. Appl. Math. Stat.* 11:1527750.
doi: 10.3389/fams.2025.1527750

COPYRIGHT

© 2025 Bianchi, Pianese, Frezza and Angelini.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

A new tool to detect financial data scaling

Sergio Bianchi¹, Augusto Pianese², Massimiliano Frezza^{1*} and
Daniele Angelini¹

¹Department of Methods and Models for Economics, Territory and Finance (MEMOTEF), Sapienza University of Rome, Rome, Italy, ²Quantitative Laboratory (QUANTLAB), University of Cassino and Southern Lazio, Cassino, Italy

The assumption of frictionless markets has long been debated, drawing interest from scholars and practitioners alike. Market liquidity is a central theme in this regard; it is traditionally assessed through transaction costs, volume, price-based, and market-impact measures. In contrast, the Fractal Market Hypothesis (FMH) suggests that liquidity emerges from the heterogeneity of investment time scales among participants, with liquidity shortages arise when traders converge on the same time horizons, particularly the short-term one which typically occurs during volatile periods. While current methods to assess liquidity often rely on single moments, which may provide limited insights, a novel methodology that considers the whole distributions and compares log-returns across pairs of time scales is discussed and implemented in this work. A Matlab-based algorithm is built that provides as output a dynamical estimation of the pairwise self-similarity of the scaled distributions. The lower the self-similarity parameter the higher the potential liquidity shortage.

KEYWORDS

market frictions, (i)liquidity, scaling, self-similarity, financial market

1 Introduction

The assumption of frictionless markets has been a longstanding topic of debate, attracting interest from both scholars and market practitioners. Broadly speaking, financial market frictions encompass any factors that impede natural trading and may include transaction costs, taxes, regulatory expenses, asset indivisibility, and liquidity constraints. Notably, the concept of market liquidity is a cornerstone in financial literature, with traditional liquidity measures generally categorized into four groups [1]: (a) transaction cost measures, such as the bid-ask spread and its extensions [2]; (b) volume-based measures, including transaction volume, turnover [3], and the Hui–Heubel Liquidity Ratio [4]; (c) price-based measures, for instance, the Amihud measure (ILLIQ) [5] and the Market Efficiency Coefficient (MEC) [6]; (d) market-impact measures, exemplified by the *Market-Adjusted Liquidity Model* developed by Hui and Heubel in 1984 [4].

The introduction of electronic trading systems in recent years have increased the attention toward another measure of liquidity: the order book. It captures the orders placed by traders to buy and sell stocks at different price points. In this direction, a very interesting work by Libman et al. [7] concludes that the deeper layers of the order book possess valuable information in the context of liquidity, a finding that is supported by other studies.

A radically different approach to modeling market liquidity is proposed by the Fractal Market Hypothesis (FMH) [8]. This posits that liquidity is driven by the heterogeneity of investment time horizons among market participants. Under normal conditions, investors operate on a variety of time horizons, from intraday (fractions of seconds or minutes) to several years (pension funds, institutional investors); liquidity originates from the different

expectations (and probability assessments) of market participants looking at different time-scaled distributions of financial returns. Therefore, a lack of liquidity can occur when the distributions among time scales appear similar or when trading activity ceases on some relevant time scales and investors start trading on common (typically short-term) horizons. In these cases, usually occurring during periods of high volatility where rising uncertainty results in a shift toward quick buy-and-sell operations, trading can slow or even be frozen, simply because most of traders—who are supposedly rational—base their trading decisions on similar distributions which induce similar behaviors. Typically, short-term investors rely more heavily on technical information, whereas long-term investors focus on fundamental indicators [9]. In the event of a significant market crash that challenges fundamental assumptions, long-term investors may either exit the market or shift to short-term strategies [10]. The Fractal Market Hypothesis can also be used to assess the causes of financial (in)stability, as suggested by Anderson and Noss [11]. In particular, the analysis reveals some relevant implications to a number of ongoing debates regarding the regulation of financial markets and of their major participants. A review of the Fractal Market Hypothesis together with the basic principles of fractal geometry is provided by Blackledge and Lamphiere [12]. Specifically, they consider the intrinsic scaling properties that characterize a random walk and the Efficient Market Hypotheses and exploring the Lévy index, coupled with other metrics, such as the Lyapunov exponent and the volatility, a long-term forecasts analysis is developed. In a study of two developed market indices and one emerging market index, Karp and Van Vuuren [13] find a relationship between the change in a time series' fractal dimension (before breaching a threshold) and both the magnitude and direction of the subsequent change in the time series. This relationship was found to be prevalent during times of strong price persistence, characterized by elevated Hurst exponents, suggesting potential investment strategies.

From a mathematical perspective, while various methods—such as DFA, wavelet power spectra, variance scaling, and R/S analysis—have been proposed to quantify liquidity in terms of scale invariance (see [14] or [15] for a comprehensive discussion), most traditional approaches primarily examine single moments rather than full distributions, which may yield incomplete insights. The aim of this work is to present the implementation of a novel methodology that examines self-similarity and (il)liquidity by comparing the entire distributions of log-returns over pairs of time scales, rather than only specific moments. In particular, an algorithm and a ready-to-use Matlab software are discussed to evaluate liquidity through the analysis of a time indexed self-similarity matrix whose entries are the pairwise self-similarity parameters relating the log-returns distributions across different investment time horizons (from 1 day up to 6 months). Roughly speaking, the estimated parameters tell how diverse the paired distributions are.

The methodology we implement provides insights into both market liquidity and the evolution of return distributions across time scales. Thus, it can provide insights to market makers and regulators particularly during periods of market turbulence. Indeed, the 2007–2009 global financial crisis, as well as the recent pandemic and geopolitical crises, have underscored the

importance of understanding liquidity and its underlying drivers in financial markets.

The remainder of this article is organized as follows: Section 2 outlines the theoretical framework and the self-similarity testing procedure employed in this study. Section 3 presents the proposed algorithm, and Section 4 concludes.

2 Theoretical background and methodology

In subSection 2.1 the definition of self-similarity is introduced along with the justification of its usage in financial context. SubSection 2.2 shortly discusses how self-similarity can be evaluated using the Kolmogorov–Smirnov distribution. Both topics are introduced to clarify how the methodology has been implemented in the software introduced in Section 3.

2.1 Self-similarity

Definition 2.1. [16] A stochastic process $\{X_t, t \geq 0\}$ in $\mathbb{R}^k$, continuous at $t = 0$, is *self-similar* with parameter $H_0 \geq 0$ (denoted H_0 -ss) if, for all scaling factors $a > 0$, the following holds:

$$\{X_{at}\} \stackrel{d}{=} \{a^{H_0} X_t\}, \quad (1)$$

where $\stackrel{d}{=}$ denotes that the finite-dimensional distributions of $\{X_{at}\}$ and $\{a^{H_0} X_t\}$ are equal.

Remark 2.1. The process is trivial ($H = 0$) if and only if $X_t = X_0$ almost surely for all $t > 0$. Additionally, if $\mathbb{E}(|X_1|) < \infty$, then $H \leq 1$ [17], meaning non-degenerate processes have $H \in (0, 1]$. A common example is Brownian motion, which is $\frac{1}{2}$ -self-similar.

Consider now a H_0 -self-similar process $\{X(t)\}$ with stationary increments, and let $Y(t, a) = X(t+a) - X(t)$ denote the increments over lag a . Due to stationarity, $\{Y(t, a)\} \stackrel{d}{=} \{Y(T, a)\}$ for any times t and T . It is easy to check that $Y(t, a)$ is also self-similar with the same parameter H_0 , i.e. that

$$\{Y(t, a)\} \stackrel{d}{=} \{a^{H_0} Y(t, 1)\}, \quad (2)$$

and X_t is said H_0 -sssi (self-similar of parameter H_0 with stationary increments).

Remark 2.2. In financial applications, let $X_t = \ln P_t$, where P_t is the price of an asset at time t . If X_t is (at least locally) self-similar and with stationary increments, the a -lagged log-price change scales as a^{H_0} times the one-lag change. This observation forms the theoretical basis for the widely-used practice of annualizing volatility in finance. Specifically, from self-similarity, we have $\mathbb{E}[X_t^2] = t^{2H_0} \mathbb{E}[X_1^2]$, and assuming $\mathbb{E}[X_t] \approx 0$, it follows that $\sigma_t^2 \approx t^{2H_0} \sigma_1^2$. For volatility, this implies $\sigma_t \approx t^{H_0} \sigma_1$. Under the assumption of Brownian motion (with $H_0 = 1/2$), this simplifies to the familiar relation $\sigma_t \approx \sqrt{t} \sigma_1$, which is commonly used by practitioners to scale volatility over different time horizons.

TABLE 1 Estimates of H and K_α and summary statistics.

H	$\hat{H}$						$\hat{K}_\alpha$			
	Mean	Min	Max	StDev	Skewness	Kurtosis	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.1$
$N = 512$										
							0.1031	0.0937	0.0860	0.0775
0.1000	0.0975	0.0250	0.1750	0.0242	0.0497	2.9471	0.0589	0.0545	0.0504	0.0464
0.2000	0.1981	0.1050	0.3000	0.0339	0.1134	2.8240	0.0686	0.0628	0.0574	0.0516
0.3000	0.2943	0.1800	0.4250	0.0380	0.0165	2.9079	0.0769	0.0702	0.0637	0.0572
0.4000	0.3945	0.2350	0.5300	0.0450	−0.0498	2.8527	0.0904	0.0817	0.0737	0.0656
0.5000	0.4883	0.3500	0.6700	0.0490	−0.0936	2.8769	0.1032	0.0931	0.0830	0.0742
0.6000	0.5852	0.4350	0.7250	0.0523	−0.0841	2.8295	0.1114	0.1005	0.0897	0.0789
0.7000	0.6718	0.4650	0.8550	0.0546	−0.1532	3.3839	0.1175	0.1071	0.0973	0.0875
0.8000	0.7525	0.5200	0.9650	0.0610	−0.2348	3.2544	0.1304	0.1168	0.1036	0.0924
0.9000	0.8268	0.6350	0.9700	0.0582	−0.2180	2.9001	0.1311	0.1122	0.1063	0.0942
$N = 4,096$										
							0.0360	0.0328	0.0301	0.0271
0.1000	0.0998	0.0700	0.1300	0.0089	0.0342	2.9869	0.0223	0.0199	0.0186	0.0169
0.2000	0.1998	0.1650	0.2350	0.0116	−0.0365	2.9448	0.0249	0.0229	0.0209	0.0184
0.3000	0.2989	0.2500	0.3450	0.0141	−0.1273	3.1095	0.0280	0.0250	0.0230	0.0209
0.4000	0.3992	0.3500	0.4650	0.0158	0.0198	3.5140	0.0326	0.0290	0.0269	0.0239
0.5000	0.4992	0.4450	0.5550	0.0171	−0.0440	3.0551	0.0342	0.0323	0.0294	0.0258
0.6000	0.5978	0.5350	0.6550	0.0182	−0.1257	3.0960	0.0418	0.0378	0.0330	0.0295
0.7000	0.6921	0.6300	0.7600	0.0211	0.0093	3.2052	0.0426	0.0396	0.0344	0.0305
0.8000	0.7837	0.7100	0.8650	0.0232	0.1619	2.9807	0.0463	0.0433	0.0398	0.0353
0.9000	0.8622	0.7850	0.9450	0.0248	0.1242	3.0289	0.0532	0.0467	0.0420	0.0376
$N = 16,384$										
							0.0180	0.0164	0.0150	0.0135
0.1000	0.1001	0.0850	0.1150	0.0042	0.0031	3.0950	0.0108	0.0101	0.0094	0.0085
0.2000	0.2002	0.1850	0.2200	0.0059	0.0987	2.8413	0.0125	0.0115	0.0102	0.0094
0.3000	0.3000	0.2750	0.3200	0.0068	−0.0697	3.0795	0.0148	0.0132	0.0119	0.0106
0.4000	0.4002	0.3750	0.4350	0.0079	0.0553	3.0387	0.0167	0.0149	0.0134	0.0118
0.5000	0.4996	0.4750	0.5250	0.0087	−0.0489	2.6184	0.0188	0.0162	0.0147	0.0129
0.6000	0.5993	0.5700	0.6300	0.0097	0.0793	3.0243	0.0211	0.0188	0.0168	0.0151
0.7000	0.6971	0.6600	0.7250	0.0105	−0.0290	3.0093	0.0218	0.0204	0.0184	0.0163
0.8000	0.7916	0.7550	0.8300	0.0126	−0.0375	2.9096	0.0245	0.0217	0.0199	0.0178
0.9000	0.8743	0.8350	0.9500	0.0143	0.1586	3.7706	0.0288	0.0238	0.0220	0.0198

In this regard, the methodology is particularly well-suited to classes of financial instruments that are continuously traded in financial markets and exhibit a certain degree of risk (including commodities or cryptocurrencies), while it proves less applicable to instruments with a low level of risk, such as, for example, the fixed income.

2.2 Testing for Self-Similarity

To estimate the self-similarity parameter H_0 , one needs to test the equality in [Equations 1](#), or [2](#). To this aim, the tool which will be described in next sections follows the methodology proposed in [\[18–20\]](#) focusing on the increment process $\{Y(t, a)\}$ for simplicity.

Let $\mathcal{A}$ be a compact set of time scales, and define $\underline{a} = \min(\mathcal{A})$ and $\bar{a} = \max(\mathcal{A})$, with $\underline{a} < \bar{a}$. For each $a \in \mathcal{A}$, let $F_{Y(a)}(x)$ denote the cumulative distribution function (CDF) of $Y(t, a)$. For any $a, b \in \mathcal{A}$, the self-similarity condition from Equation 2 can be rewritten as:

$$F_{a^{-H_0} Y(a)}(x) = F_{b^{-H_0} Y(b)}(x). \quad (3)$$

Let H be a candidate parameter. From Equation 3, it follows that:

$$F_{(a/b)^{-H} Y(a)}(x) = F_{Y(b)}\left(\left(a/b\right)^{H-H_0} x\right). \quad (4)$$

We now define the set $\Psi_H := \{F_{a^{-H} Y(a)}(x)\}$ and introduce a distance function ρ based on the sup-norm $\|\cdot\|_\infty$ over the set $\mathcal{A}$. The metric space (Ψ_H, ρ) has a diameter given by:

$$\begin{aligned} \delta(\Psi_H) &= \sup_{x \in \mathbb{R}} \sup_{a, b \in \mathcal{A}} |F_{a^{-H} Y(a)}(x) - F_{b^{-H} Y(b)}(x)| \\ &= \sup_{x \in \mathbb{R}} |F_{\underline{a}^{H-H_0} Y(1)}(x) - F_{\bar{a}^{H-H_0} Y(1)}(x)|. \end{aligned} \quad (5)$$

The diameter $\delta(\Psi_H)$ is minimized when $H = H_0$, as it decreases for $H \leq H_0$ and increases for $H \geq H_0$ (see [18] for the proofs). Thus, we can estimate the self-similarity parameter as:

$$\hat{H}_0 = \arg \min_{H \in (0,1]} \hat{\delta}(\Psi_H), \quad (6)$$

where $\hat{\delta}(\Psi_H)$ is the empirical counterpart of $\delta(\Psi_H)$, computed using the empirical cumulative distribution function $\hat{F}$.

Remark 2.3. Notice that the last line of Equation 5 is the two-sample Kolmogorov–Smirnov (KS) statistic, which provides the distribution of the maximum of the absolute differences between two empirical CDF, under the null hypothesis that both are drawn from the same population distribution. In particular, the goodness-of-fit of the empirical distribution can be evaluated with respect to a theoretical distribution F (Kolmogorov, D_n) or between two empirical distributions (Smirnov, $(D_{n,m})$) of n and m observations:

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|, \quad (7)$$

where $F_{1,n}$ and $F_{2,m}$ are the empirical cumulative distribution functions of two independent samples. The statistics is non parametric and for large samples, the null hypothesis is rejected at level α if

$$D_{n,m} > K_\alpha = \sqrt{-\log(\alpha/2) \frac{1 + \frac{m}{n}}{2m}} \quad (8)$$

When applied to the estimation of the self-similarity parameter H_0 of a sample of size n , Equation 7 becomes

$$\hat{\delta}(\Psi_H) = \sup_{x \in \mathbb{R}} \left| \frac{1}{n - \underline{a}} \sum_{i=1}^{n-\underline{a}} 1_{X_{1,i} \leq \underline{a}^H x} - \frac{1}{n - \bar{a}} \sum_{i=1}^{n-\bar{a}} 1_{X_{1,i} \leq \bar{a}^H x} \right| \quad (9)$$

The distribution (Equation 8) of the KS statistic is well-known when the samples are independent, i.e. when $H_0 = 1/2$. For values of $H_0 > 1/2$, the process exhibits positive autocorrelation increasing with the distance from $1/2$ and decreasing with the

lag at such a slow rate that the autocorrelation function becomes non-summable (the so-called long-term memory). For $H < 1/2$, the process displays short-term memory characterized by negative autocorrelation. These positive and negative autocorrelations result in a Type I error for $H > 1/2$ and a Type II error for $H < 1/2$, i.e. the KS critical value tends to be too restrictive for moderate to high levels of positive autocorrelation and too lenient for moderate to high levels of negative autocorrelation. Since in these cases the distribution of the test statistic is not explicitly known and it is also difficult to be derived analytically, we have implemented Monte Carlo simulations to deduce the p-values for different values of H_0 . The results are summarized in Table 1.

Given the daily distribution of returns of the all stocks forming a stock index, the value $\hat{H}_0$ of Equation 11 can be calculated at any time t and any pair of time scales a, b in the set $\mathcal{A} = \{1, 2, \dots, N\}$. Thus, a symmetric self-similarity matrix can be built as

$$\hat{\mathbf{H}}(\mathbf{t}) = [\hat{H}(a, b, t)] \quad (10)$$

with

$$\hat{H}(a, b, t) = \arg \min_{H \in (0,1]} \left(\max_{x \in \mathbb{R}} \left| \hat{F}_{Y(t,a)}(x) - \hat{F}_{Y(t,b)}\left(\left(b/a\right)^H x\right) \right| \right). \quad (11)$$

Of course, since $\hat{H}_0(a, b, t) = \hat{H}_0(b, a, t)$ the resulting estimation will consist in a time-indexed sequence of lower (upper) triangular matrices.

3 Algorithm for testing self-similarity

The pseudocode Algorithm 1 provides a structured overview of the function SelfSimilarity, detailing the nested loop structure, the determination of rescaled sequences, and the two-sample Kolmogorov–Smirnov (KS) test implementation.

In detail, the algorithm iteratively estimates self-similarity on different time scales and intervals, aiming to find the parameter H that minimizes the KS statistic, thereby quantifying the self-similarity index between two paired lagged sequences. The function receives four input values: the time sequence $X(t, m)$ (log-price for financial applications), which stores m paths of length t ; the maximum number of time scales N and the indices of the startpoint and endpoint of the analysis. Within nested loops that iterate over time indices, time scales and a range of potential H values, the algorithm generates lagged sequences for each pair of scales and performs the KS test to assess distributional similarity. The KS statistic and its associated p -value are stored, enabling the identification of the optimal H for each pair of scales. Finally, the function outputs matrices containing the estimated self-similarity parameter, the KS statistic, and the p -value across all analyzed time scales and indices, providing a robust measure of self-similarity in the time series data.

Regarding the two-sample Kolmogorov–Smirnov test used by the algorithm, it should be noted that this test is readily available in the standard function libraries of most major programming languages. Implementations in languages such as Python, MATLAB, and R provide accessible and optimized functions that allow users to efficiently perform the KS test to

```

function SelfSimilarity(X, N, t_start, t_end)
    ▷ SelfSimilarity estimates the self-similarity parameter using the two-sample Kolmogorov–Smirnov test. <
    ▷ <
    ▷ Arguments: <
    ▷ X: Input time series (matrix) <
    ▷ N: Maximum number of time scales (cardinality of set A) <
    ▷ t_start: Start index for analysis, must be greater than N <
    ▷ t_end: End index for analysis, must be greater than t_start <
    ▷ <
    ▷ Returns: <
    ▷ H_est: Matrix containing estimated self-similarity parameter values <
    ▷ KS_stat: Matrix containing the Kolmogorov–Smirnov test statistic <
    ▷ KS_prob: Matrix containing the asymptotic p-value from the Kolmogorov–Smirnov test <

    Initialize i to 0
    for all time index t from t_start to t_end do
        Increment i by 1

        for all time scale a from 1 to N do
            ▷ Calculate a-lagged sequence and mean normalization <
            Set  $X_1$  to  $(X(t, :) - X(t - a, :)) - \text{mean}(X(t, :) - X(t - a, :))$ 

            for all time scale b from a+1 to N do
                ▷ Calculate b-lagged sequence and mean normalization <
                Set  $X_2$  to  $(X(t, :) - X(t - b, :)) - \text{mean}(X(t, :) - X(t - b, :))$ 

                Initialize j to 0
                for all H ranging from 0 to 1 with step size 0.01 do
                    Increment j by 1
                    ▷ Rescale sequences based on current H value <
                    Set  $Y_1$  to  $X_1 \cdot a^{-H}$ 
                    Set  $Y_2$  to  $X_2 \cdot b^{-H}$ 
                    Perform Kolmogorov–Smirnov test on  $Y_1$  and  $Y_2$ , obtaining:
                    - P as the p-value
                    - KSSTAT as the test statistic
                    Store H, KSSTAT, and P in matrix d at row j

                ▷ Find the minimum KSSTAT value in d and the corresponding H <
                Retrieve index j of minimum KSSTAT in d
                Set  $H\_est(a, b, i)$  to H at row j in d
                Set  $KS\_stat(a, b, i)$  to KSSTAT at row j in d
                Set  $KS\_prob(a, b, i)$  to P at row j in d

```

Algorithm 1. Estimation of the self-similarity parameter.

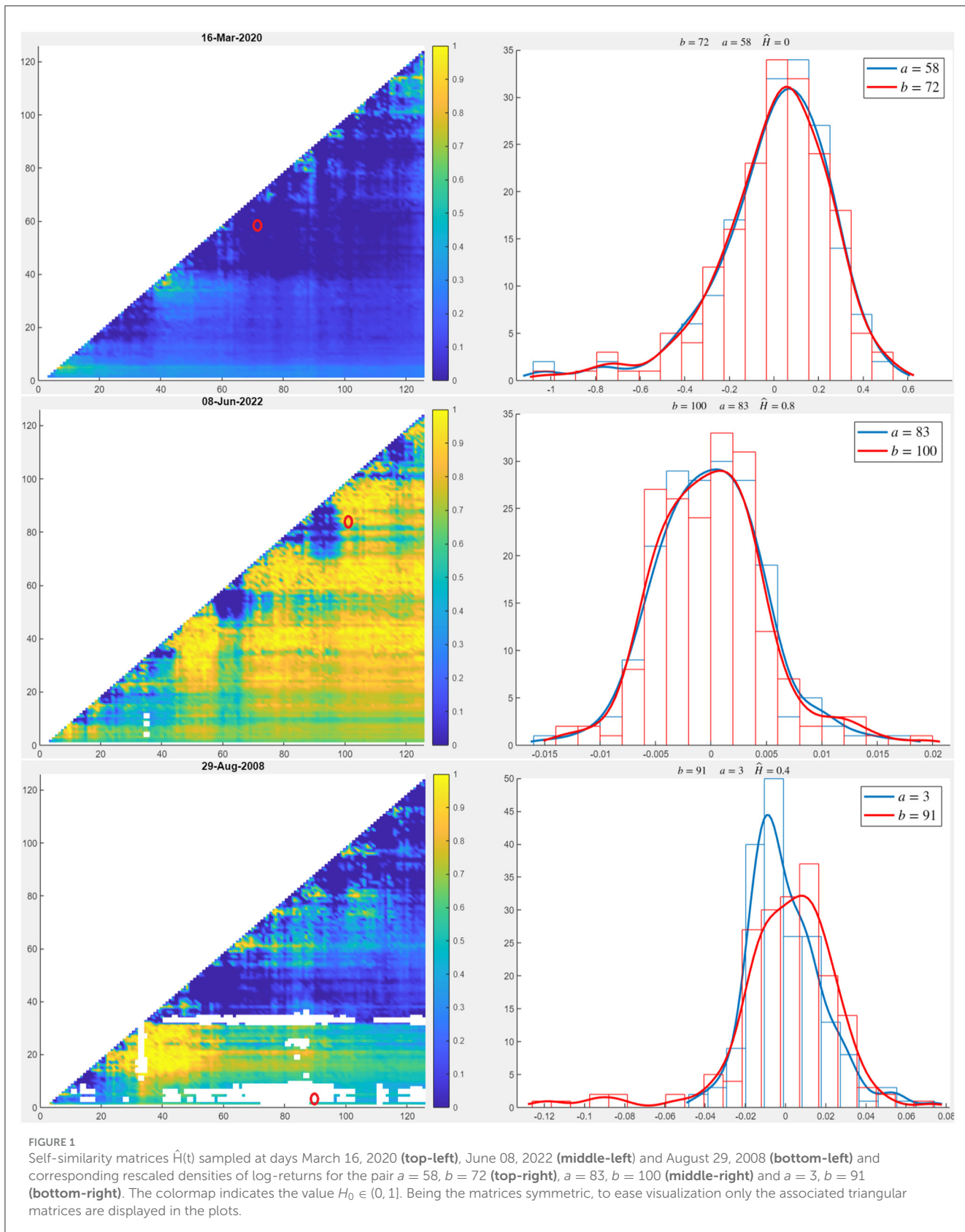
assess whether two samples come from the same distribution. These library functions typically offer flexibility in terms of input parameters, enabling control over test options such as significance levels and alternative hypotheses.

To demonstrate the effectiveness of the proposed algorithm with real data, it was implemented in the MATLAB R2024a environment and tested on a computer with a 64-bit Windows operating system, an Intel Core i7-10510U CPU (1.80 GHz–2.30 GHz), and 16 GB of RAM. The input matrix X was constructed using the logarithmic prices of 183 assets continuously listed in the S&P500 index (Standard and Poor's, USA) from January 2000 to December 2023, totaling 6,037 observations. Even if 183 out of

500 individual stocks were considered as representative of the index, constructing the distribution from the all stocks that make up the market index makes it possible to eliminate the inertial effect that would be observed if—instead of the individual stocks—only the time-marginal distribution of the index were considered.

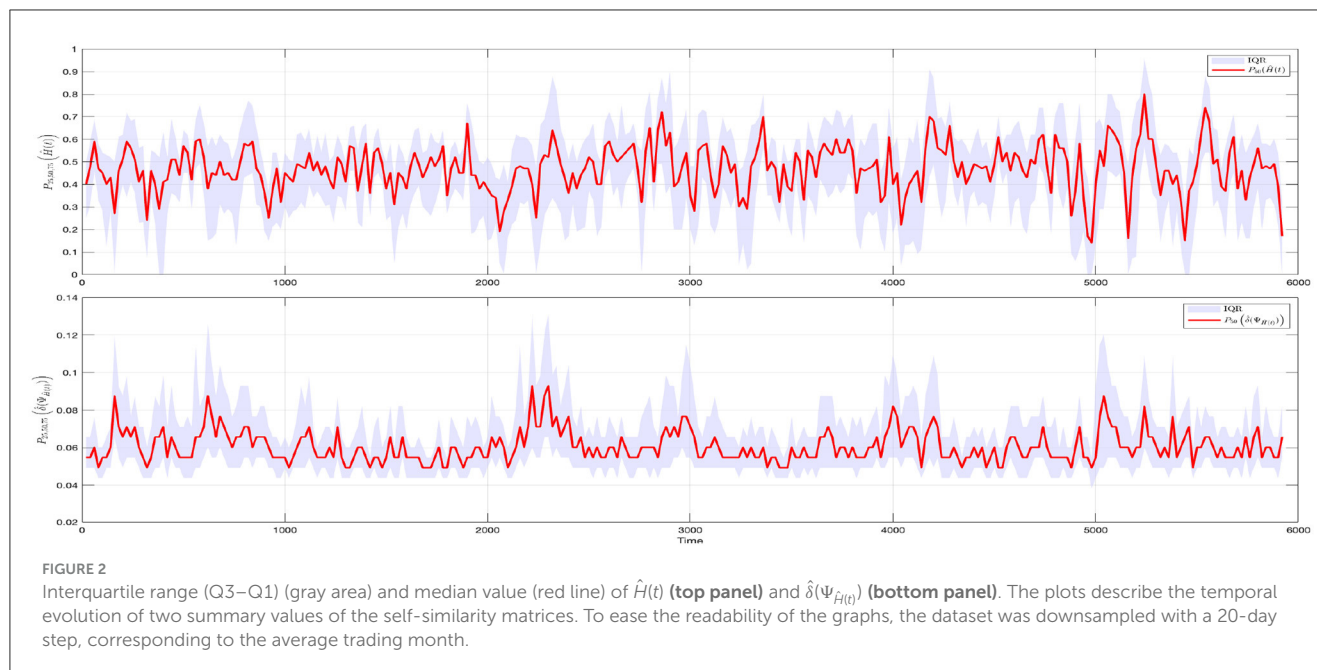
The average execution time for each iteration was ~ 30 s.

Figure 1 illustrates three contrasting market conditions in terms of liquidity: the top row represents a period of near-total liquidity shortage, revealed by the self-similarity surface (left-hand plot) and the rescaled log-return densities, reported for the pair $a = 58$ and $b = 72$ (right-hand plot) on March 16, 2020. The self-similarity surface, dominated by a uniform blue



color across all investment pairs, indicates very low values of the self-similarity parameter, signifying an extremely low liquidity level in the market due to the fact that the distributions almost

overlap as they are, the scaling exponent being close to zero. This outcome aligns with expectations, as this date was chosen during the U.S. quarantine or lockdown period, where markets



were affected by heightened concerns regarding the COVID-19 pandemic's progression. Additional evidence appears in the right panel, where both rescaled densities exhibit a pronounced left-skewed tail.

In contrast, the plots in the mid row illustrate market liquidity conditions on June 8, 2022, when the S&P index rose by 1% or 39.25 points, with positive performance across most sectors. The self-similarity surface (left-hand plot) displays high values (yellow shading) for nearly all the considered investment pairs, indicating a high-liquidity market. In the corresponding right-hand plot, the rescaled log-return densities for the pair $a = 83$ and $b = 100$ are shown, featuring right-skewed tails. Finally, the plots in the bottom row illustrate a scenario in which self-similarity is rejected at a 95% significance level on August 29, 2008. In this case, “holes” appear in the similarity surface (blank areas), as the two densities (right-hand plot) are significantly different. At $H = 0.4$, the distance between the distributions is minimized but remains statistically significant; therefore, self-similarity is rejected (i.e., no value of the parameter H satisfies the distributional equality for the pair $a = 3$ and $b = 91$).

Figure 2 illustrates two summary measures for the daily distributions of $\hat{H}(t)$ and the estimated diameters $\hat{\delta}(\Psi_{\hat{H}(t)})$. To improve the readability of the graphs, the entire dataset was downsampled using a 20-day step, corresponding to an average trading month. The median value and the InterQuartile Range ($IQR = P_{75}(\cdot) - P_{25}(\cdot)$) were then computed for the distributions of both $\hat{H}(t)$ and $\hat{\delta}(\Psi_{\hat{H}(t)})$. The data are particularly interesting because the dynamics of $P_{50}\hat{H}(t)$ are fully consistent with the known estimates of the Hurst parameter reported in the literature (see, e.g. [21–24]). Furthermore, the fact that the values oscillate around 1/2 is of particular financial significance, as this is the only value consistent with the absence of arbitrage opportunities in the market—a sort of benchmark to which the market always returns after deviations

that can be more or less prolonged depending on the shocks it experiences.

4 Conclusions

In this paper, we have implemented a new methodology for assessing market liquidity. Starting from the Fractal Market Hypothesis's (FMH) assumption that liquidity emerges from the heterogeneity of investment time scales, the scale invariance or self-similarity between pairs of investment horizons is examined with regard to the whole distribution rather than, as usual, its individual moments. Specifically, the self-similarity exponent is obtained as the unknown argument that minimizes the diameter of the space of rescaled distributions, which is observed to reduce to the Kolmogorov–Smirnov statistic. A numerical procedure is built to determine this value with respect to a number of time scales flowing through time and an algorithm provides as output a dynamical estimation of the pairwise self-similarity parameter of the considered scaled distributions. Implemented in MatLab, the algorithm uses nested cycles that, for each trading day, iterates over both time scales and a range of potential values of H , and performs the KS test to assess distributional similarity. The KS statistic and its associated value are stored, allowing the optimal H for each pair of scales to be identified. Low values of the self-similarity parameter, typically occurring during volatile periods where long-term investors may either exit the market or shift to short-term scales, indicate potential liquidity shortage. In contrast, high values of the self-similarity parameter describe liquid market phases in which each time horizon is populated by a sufficient share of market participants to match trades. Empirical evidence that proves the effectiveness of the proposed algorithm with real data is offered by analyzing 183 assets continuously listed in the in the

S&P500 index (Standard and Poor's, USA) from January 2000 to December 2023.

Data availability statement

The datasets presented in this article are not readily available. Requests to access the datasets should be directed to Massimiliano Frezza, massimiliano.frezza@uniroma1.it.

Author contributions

SB: Conceptualization, Formal analysis, Supervision, Writing – review & editing. AP: Data curation, Methodology, Software, Writing – original draft. MF: Data curation, Formal analysis, Methodology, Writing – original draft. DA: Investigation, Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

References

- Bianchi S, Frezza M. Liquidity, Efficiency and the 2007 – 2008 global financial crisis. *Ann Econ Finance*. (2018) 19:375–404.
- Amihud Y, Mendelson H. Asset pricing and the bid-ask spread. *J Financ Econ*. (1986) 17:223–49. doi: 10.1016/0304-405X(86)90065-6
- Dennis P, Strickland D. The effect of stock splits on liquidity and excess returns: evidence from shareholder ownership composition. *J Financ Res*. (2003) 26:355–70. doi: 10.1111/1475-6803.00063
- Hui B, Heubel B. *Comparative Liquidity Advantages among Major U.S. Stock Markets*. Working Paper, DRI Financial Information Group Study Series, No. 84081. (1984).
- Amihud Y. Illiquidity and stock returns: cross-section and time-series effects. *J Financ Mark*. (2002) 5:31–56. doi: 10.1016/S1386-4181(01)00024-6
- Hasbrouck J, Schwartz RA. Liquidity and execution costs in equity markets. *J Portfolio Manag*. (1988) 14:10–6. doi: 10.3905/jpm.1988.409160
- Libman D, Haber S, Schaps M. Forecasting quoted depth with the limit order book. *Front Artif Intell*. (2021) 4:667780. doi: 10.3389/frai.2021.667780
- Peters E. *Fractal Market Analysis — Applying Chaos Theory to Investment and Analysis*. Hoboken, NJ: Wiley (1994).
- Kristoufek L. Fractal markets hypothesis and the global financial crisis: scaling, investment horizons and liquidity. *Adv Complex Syst*. (2012) 15:1250065. doi: 10.1142/S0219525912500658
- Li DY, Nishimura Y, Men M. Fractal markets: liquidity and investors on different time horizons. *Phys A: Stat Mech Appl*. (2014) 407:144–51. doi: 10.1016/j.physa.2014.03.073
- Anderson N, Noss J. The Fractal Market Hypothesis and its implications for the stability of financial markets. *Bank of England, Financial Stability Paper n23*. (2013).
- Blackledge J, Lamphiere M. A review of the fractal market hypothesis for trading and market price prediction. *Mathematics*. (2022) 10:117. doi: 10.3390/math10010117
- Karp A, Van Vuuren G. Investment implications of the fractal market. Hypothesis. *Ann Financ Econ*. (2019) 14:1950001. doi: 10.1142/S2010495219500015
- Mandelbrot BB. *Fractals and scaling in finance. Discontinuity, Concentration, Risk Selecta Volume E*. Cham: Springer (1997). doi: 10.1007/978-1-4757-2763-0
- Cont R, Potters M, Bouchaud JP. Scaling in stock market data: stable laws and beyond. In: Dubrulle B, Graner F, Sornette D, editors. *Scale Invariance and Beyond*. Berlin, Heidelberg: Springer Berlin Heidelberg (1997), p. 75–85. doi: 10.1007/978-3-662-09799-1_5
- Lamperti J. Semi-stable stochastic process. *Trans Am Math Soc*. (1962) 104:62–78. doi: 10.1090/S0002-9947-1962-0138128-7
- Embrechts P, Maejima M. *Selfsimilar Processes*. Princeton, NJ: Princeton University Press (2002).
- Bianchi S. A new distribution-based test of self-similarity. *Fractals*. (2004) 12:331–46. doi: 10.1142/S0218348X04002586
- Bianchi S, Pianese A, Frezza M. A distribution-based method to gauge market liquidity through scale invariance between investment horizons. *Appl Stochastic Models Bus Ind*. (2020) 36:809–24. doi: 10.1002/asmb.2531
- Bianchi S, Bruni V, Frezza M, Marconi S, Pianese A, Vantaggi B, et al. An information theory approach to stock market liquidity. *Chaos*. (2024) 34:061102. doi: 10.1063/5.0213429
- Cajueiro DO, Tabak BM. The Hurst exponent over time: testing the assertion that emerging markets are becoming more efficient. *Phys A Stat Mech Appl*. (2004) 336:521–37. doi: 10.1016/j.physa.2003.12.031
- Bianchi S, Pantanella A. Pointwise regularity exponents and well-behaved residuals in stock markets. *Int J Trade Econ Finance*. (2011) 2:52–60. doi: 10.7763/IJTEF.2011.V2.78
- Frezza M. Modeling the time-changing dependence in stock markets. *Chaos Solitons Fractals*. (2012) 45:1510–20. doi: 10.1016/j.chaos.2012.08.009
- Bianchi S, Pantanella A, Pianese A. Modeling stock prices by multifractional Brownian motion: an improved estimation of the pointwise regularity. *Quant Finance*. (2013) 13:1317–30. doi: 10.1080/14697688.2011.594080

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

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