Noemi Nava, Tiziana Di Matteo, and Tomaso Aste

Time-dependent scaling patterns in high frequency financial data

Article (Published version)
(Refereed)

This version available at: http://eprints.lse.ac.uk/68645/

Available in LSE Research Online: December 2016

http://eprints.lse.ac.uk

Regular Article

Time-dependent scaling patterns in high frequency financial data

Noemi Nava[1,3,a], Tiziana Di Matteo[1,2], and Tomaso Aste[1,3]

[1] Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK
[2] Department of Mathematics, King's College London, The Strand, London, WC2R 2LS, UK
[3] Systemic Risk Centre, London School of Economics and Political Sciences, London, WC2A2AE, UK

Abstract. We measure the influence of different time-scales on the intraday dynamics of financial markets. This is obtained by decomposing financial time series into simple oscillations associated with distinct time-scales. We propose two new time-varying measures of complexity: 1) an amplitude scaling exponent and 2) an entropy-like measure. We apply these measures to intraday, 30-second sampled prices of various stock market indices. Our results reveal intraday trends where different time-horizons contribute with variable relative amplitudes over the course of the trading day. Our findings indicate that the time series we analysed have a non-stationary multifractal nature with predominantly persistent behaviour at the middle of the trading session and anti-persistent behaviour at the opening and at the closing of the session. We demonstrate that these patterns are statistically significant, robust, reproducible and characteristic of each stock market. We argue that any modelling, analytics or trading strategy must take into account these non-stationary intraday scaling patterns.

1 Introduction

Financial markets are complex and dynamic systems that generate non-stationary, non-linear and noisy time series. These time series do not consist of a single driving force, but of several components that are superimposed onto each other in a hierarchical form. The oscillating components are generated by many participants with diverse interests and investment horizons that interact and create effects that seem

---

[a] e-mail: n.morales.11@ucl.ac.uk

to repeat themselves cyclically with different characteristic periods [1]. These effects can be retrieved and quantified by looking at the statistical properties of financial data at different time-scales.

A classical paradigm used to model financial markets is the efficient market hypothesis (EMH) [2], which states that financial markets are efficient if they reflect all available information, thus arbitrage conditions are quickly eliminated. According to this theory, stock prices are unpredictable. The weak form of the EMH admits a rapid price adjustment process [2]. However, in practice, prices do not tend to adjust to new information so rapidly, taking a certain amount of time. During this time investors can take actions to exploit temporary profitable opportunities arising from new information [3,4].

According to the fractal market hypothesis (FMH) [5], financial participants are heterogeneous and market stability exists if there are investors with different time-horizons that create liquidity. Market participants have investment horizons that vary from seconds to years (market makers, noise traders, hedge funds). They treat the arriving information differently and affect the price dynamics in various ways depending on their trading time-scales [6]. The existence of investors with different investment horizons induces a stable market, markets become unstable when one horizon becomes dominant since liquidity dries up [7–10].

The FMH predicts that critical events are connected to dominant investment horizons and is usually associated to self-similarity, fractality or multifractality. Self-similarity in financial data was first studied by Mandelbrot [11,12], and has been found to be present across financial markets with complex properties that are significantly related to economic and financial characteristics of the markets [13–17]. Self-similarity is related to the occurrence of similar patterns at different time-scales. In this sense, probabilistic properties of self-similar processes remain invariant when the process is viewed at different time-scales [11,18]. Specifically, a stochastic process $X(t)$ is statistically self-similar, with scaling exponent $0 < H < 1$, if for any real $a > 0$ it follows the scaling law:

$$X(at) \stackrel{d}{=} a^H X(t) \quad t \in \mathbb{R}, \tag{1}$$

where the equality ($\stackrel{d}{=}$) is in probability distribution [19].

An example of self-similar process is fractional Brownian motion (FBM), a Gaussian process with stationary increments characterized by a positive scaling exponent $0 < H < 1$ [20]. When $0 < H < 1/2$, the increments of FBM show negative autocorrelation. The case $1/2 < H < 1$ corresponds to FBM with increment process exhibiting long-range dependence, i.e., the autocorrelation of the increment process decreases as a power law. When $H = \frac{1}{2}$, the FBM is reduced to Brownian motion (BM), a process with independent increments [19]. Another self-similar process with stationary and independent increments but with heavy-tailed infinite variance distribution is the α-stable Lévy motion [19]. Empirical estimates of the scaling properties are affected by both the autocorrelation between the increments, as in the case of FBM, and the high variability, as in stochastic processes whose increments are independent and heavy-tailed (Lévy processes) [21,22].

It has been observed that more than one scaling exponent is needed to account for the scaling properties of financial time series. Indeed, each moment of the distribution of financial returns varies as a power law of the time horizon with a different H exponent [18]. This phenomenon is called multiscaling and reflects the occurrence of different dynamics at different time-scales, it can be attributed to the heterogeneity of market participants. Time-dependent scaling behaviour has also been observed in financial time series [23,24], the local variations of roughness can be described by allowing the H exponent to vary with time [25].

Entropy has also been used to measure the complexity of financial time series [26–28]. A low value of entropy indicates the presence of more predictable patterns which are therefore associated with periods of financial inefficiency. Conversely, when time series exhibit more irregular and less predictable patterns, the uncertainty level is higher and such periods are described by larger entropy values.

In this paper, we propose two new time-dependent measures of complexity based on the scaling properties of financial time series. These novel measures are obtained via the Hilbert-Huang transform [29], a methodology which as a first step decomposes the analysed time series into several oscillatory modes by means of empirical mode decomposition (EMD). Secondly, the Hilbert transform is applied to these oscillations to obtain time varying attributes. The time-dependent scaling properties of financial time series are associated with the relative weights of the amplitudes at characteristic frequencies.

The first measure that we introduce in this paper is a scaling exponent which quantifies the relative hierarchical variations of the amplitudes of the components with respect to their associated time-scales. By estimating a time-dependent scaling exponent for models with known scaling laws, we demonstrate that our measure varies consistently around the expected values set in the models. When applied to intraday financial time series this measure reveals instead non-stationary scaling patterns. The second measure introduced in this paper is an entropic measure quantifying the dispersion of the amplitudes of the components. The entropic measure is more general than the scaling exponent because does not assume any a-priori relation between the amplitudes of the components. Furthermore, through this last measure we demonstrate that non-stationary changes in the complexity of financial time series create persistent and reproducible patterns during the trading day.

The remaining of this paper is organized as follows. In Sect. 2, we introduce the Hilbert-Huang transform, the methodology used to identify the oscillating components of the data. The proposed scaling measure and some applications to self-similar process are described in Sect. 3. In Sect. 4, we propose a comparative entropy-like measure. In Sect. 5, we apply the proposed measures to intraday financial data. Section 6 details the intraday scaling properties of some stock market indices. Finally, in Sect. 7, we provide conclusions and future perspectives.

2 Hilbert-Huang transform

The Hilbert-Huang transform (HHT) [29] is a technique used to analyse non-linear and non-stationary time series. It was originally designed to study water wave evolution, but it has proven to be a useful tool for other complex signals, including financial time series [30–34].

The HHT consists of two steps: namely, the empirical mode decomposition (EMD) and the Hilbert transform (HT). The EMD decomposes the time series into a set of narrow-band intrinsic mode functions (IMFs) and the Hilbert transformation of these IMFs provides local frequency and amplitude attributes.

The EMD is a fully adaptive decomposition, differently from the Fourier and the wavelet transforms, it does not require any a priori basis system [35]. Furthermore, it can be used to analyse non-stationary time series. The purpose of the method is to identify oscillating components of the signal with scales defined by the local maxima and the minima of the data itself. Hence, given a time series $x(t)$, $t = 1, 2, \ldots, T$, the EMD decomposes it into a finite number of IMFs denoted as $c_k(t)$, $k = 1, \ldots, n$ and a residue function, $r(t)$. The number of IMFs, n, is approximately of the order of $\log_2(T)$ [29]. The IMFs are components oscillating around zero and obtained through a sifting process which uses the local extrema to separate oscillations starting with

the highest frequency. At the end of the sifting process, the time series $x(t)$ can be expressed as:

$$x(t) = \sum_{k=1}^{n} c_k(t) + r(t). \tag{2}$$

The residue function, $r(t)$, is the non-oscillating drift of the data. For more details about this decomposition refer to [29]. The EMD was proposed as a way to pre-process time series before applying the Hilbert transform. It generates components of the time series whose Hilbert transform can lead to physically meaningful definitions of instantaneous amplitude and frequency. The Hilbert transformation of each function c_k is defined as:

$$\hat{c}_k(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_k(\tau)}{t - \tau} \, d\tau, \tag{3}$$

where the integral has a singular point at $\tau = t$ and it is defined as a Cauchy principal value [36]. Each function c_k and its Hilbert transformation, $\hat{c}_k$, produce a complex function z_k, defined as $z_k(t) = c_k(t) + \hat{c}_k(t)$ with amplitude

$$a_k(t) = \sqrt{c_k^2(t) + \hat{c}_k^2(t)}, \tag{4}$$

and phase

$$\theta_k(t) = \tan^{-1} \frac{\hat{c}_k(t)}{c_k(t)}. \tag{5}$$

The instantaneous frequency is defined as the derivative of the phase with respect to time

$$\omega_k(t) = \frac{d\theta_k(t)}{dt}. \tag{6}$$

The instantaneous amplitude results in a smooth function, an 'envelope', which takes the overall shape of the time series touching the maxima and the minima but never crossing the time series itself. This amplitude function represents the combined oscillations of all frequencies involved in the component but removes all the frequency information.

In this paper the Hilbert transform is numerical estimated by using a FFT-based method described in [37], for a different numerical estimation based on the Sinc function refer to [38].

3 Time-dependent scaling exponent

The proposed time-dependent scaling exponent, denoted as $H^*(t)$, is constructed by observing how the local amplitudes $a_k(t)$ (Eq. (4)) change with respect to the local periods $\tau_k(t) = \omega_k^{-1}(t)$ (Eq. (6)) for all $k = 1, 2, \ldots, n$.

We first applied the method to FBM and empirically observed that the amplitude function obtained through the HHT follows a power-law behaviour with respect to the instantaneous period:

$$a_k(t) \propto \tau_k^{H^*(t)}(t), \tag{7}$$

where the exponent $H^*(t)$ describes the local scaling properties of the IMF amplitudes and takes values distributed around the self-similar exponent H of FBM. The
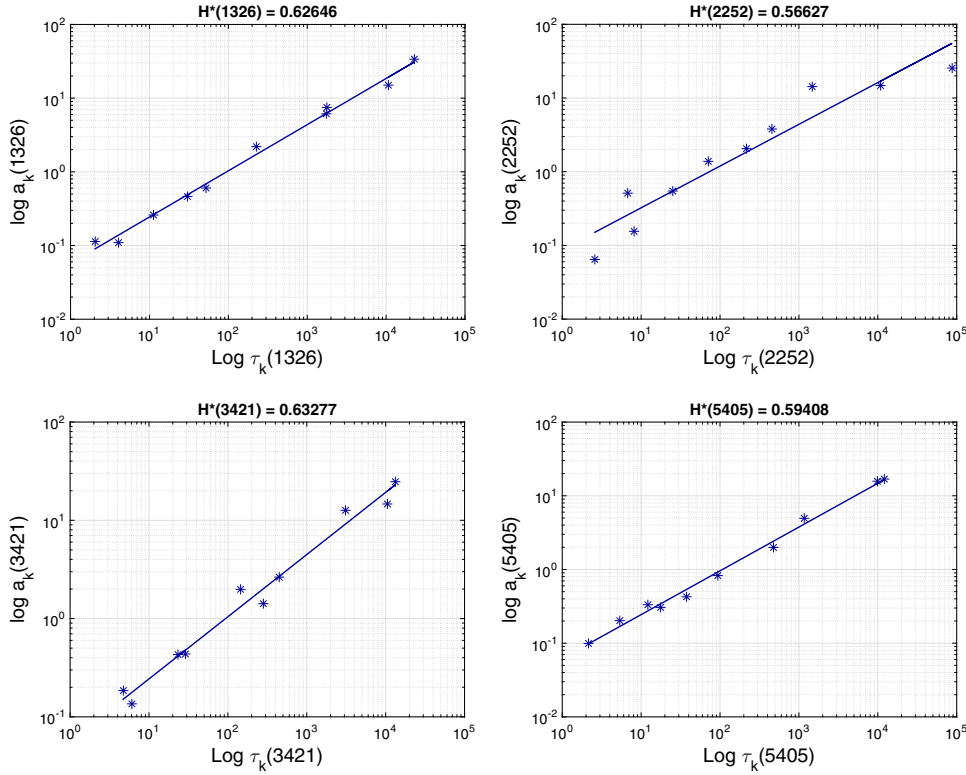
Fig. 1. Illustration that for FBM the local amplitudes $a_k(t)$ and the periods $\tau_k(t)$ follow Eq. (7): $a_k(t) \propto \tau_k^{H^*(t)}(t)$. The plots report the instantaneous amplitude as a function of period for the following four randomly chosen time instances: $t = 1\,326, 2\,252, 3\,421, 5\,405$. The simulated process is a FBM with self-similar exponent $H = 0.6$ and length $T = 10\,000$ points. The straight lines are best-fit linear regressions.

exponent $H^*(t)$ is based on the scaling properties of the absolute value of the fluctuations, therefore comparisons with other estimators such as the generalized Hurst exponent [39] must be done for the first order moments i.e., $q = 1$ [40].

In Fig. 1, we exemplify Eq. (7) showing the linear fit between $\log a_k(t)$ and $\log \tau_k(t)$ for four randomly chosen time instances of a FBM with self-similar exponent $H = 0.6$ and length $T = 10\,000$. The values of $H^*(t)$ reported in the plots are obtained from the slope of the regression line. We observe that they are all consistently close to the self-similar value $H = 0.6$. For the chosen values of t, we calculated the goodness of the linear fit by estimating the coefficient of determination R^2 [41] (values of this coefficient range from 0 to 1, with 1 indicating a perfect fit between the data and the linear model). Results for the four randomly chosen time instances $t = 1\,326, 2\,252, 3\,421, 5\,405$ are as follow: $R^2(1\,326) = 0.99$, $R^2(2\,252) = 0.90$, $R^2(3\,421) = 0.98$, $R^2(5\,405) = 0.99$, indicating therefore that for those time instances, the data are well represented by the power law of Eq. (7).

Similar linear scaling results are obtained across all times, but the scaling exponent is different at each time step, making $H^*(t)$ a time-dependent estimator. A value of $H^*(t) > 0.5$ is obtained when around time t, the amplitude of long cycles is larger than in a pure random walk. This can be interpreted as a persistent behaviour in the amplitudes of the process, meaning that in a neighbourhood of time t the process is in a cycle indistinguishable from a trend. On the contrary, values of

$H^*(t) < 0.5$ represent a rougher and more chaotic behaviour around time t. These processes are composed of oscillations with more similar amplitudes across time-scales than in Brownian motion, creating a complex and uncertain behaviour. In this case, high frequency components are more active and their contribution to the total variance is more significant than in a random walk process.

3.1 Extended simulation of self-similar and long memory processes

In order to test the power-law relation of Eq. (7), we extended the simulation set of FBM and we considered other two different self-similar processes, namely α-stable Lévy motion (SLM) [19] and autoregressive fractionally integrated moving average (ARFIMA) processes [42].

For each stochastic process, we simulated $m = 1\,000$ paths of length $T = 10\,000$ points[1]. We estimated $H^*(t)$ and calculated the time-dependent sample mean over the number of simulations, i.e., we calculated $\langle H^*(t) \rangle = \frac{1}{m} \sum_{i=1}^m H_i^*(t)$.

We also estimated the sample mean of $H^*(t)$ over time and over the number of simulations, $\langle\langle H^* \rangle\rangle = \frac{1}{T} \sum_{t=1}^T \langle H^*(t) \rangle$. We then estimated the standard deviation of $H^*(t)$ as:

$$\sigma_{H^*} = \sqrt{\sum_{t=1}^T \sum_{i=1}^m \left(H_i^*(t) - \langle\langle H^* \rangle\rangle \right)^2 / (Tm - 1)}.$$

For each process, the average over time and over the number of simulations of the coefficient of determination is also estimated and denoted as $\langle\langle R^2 \rangle\rangle$. We compared the estimated $H^*(t)$ with the generalized Hurst exponent with $q = 1$ [39] here denoted as H_G.

– **Fractional Brownian motion**. We simulated stochastic processes with different self-similar exponents, $H = 0.1, 0.2, \ldots, 0.9$. All the simulations were done using the Matlab[®] wavelet toolbox. Results for different values of H are reported in Fig. 2(a). We observe that $\langle H^*(t) \rangle$ consistently varies around the input self-similar value of H. Some disturbances can be observed at the beginning and at the end of these time series due to the boundary effects of the EMD [44].

In Table 2(b), we report $\langle\langle H^* \rangle\rangle$ and the standard deviation of $H^*(t)$. We observe a good agreement with the self-similar parameter H, but we also observe large values for σ_{H^*} that could be attributed to the local characteristics of $H^*(t)$. We notice that FBM with small scaling exponents ($H < 0.3$) have smaller $\langle\langle R^2 \rangle\rangle$ coefficients indicating significant deviations from the scaling law of Eq. (7).

The comparison of $\langle H^*(t) \rangle$ with the generalized Hurst exponent shows equivalent values for the estimated exponents. We also considered FBM paths with shorter length, $T = 1\,000$ and $T = 500$ points, noting that the longer the time series, the better the estimation of the scaling exponent H. Likewise, the standard deviation and the goodness of the fit improve with the length of the time series. However, all results are consistent with the ones reported here for length $T = 10\,000$.

---

[1] The length of the SLM is set to $T = 2^{14} - 6\,000$, following the algorithm proposed in [43].

H	H^*			GHE(1)	
	$\langle\langle H^*\rangle\rangle$	σ_{H^*}	$\langle\langle R^2\rangle\rangle$	$\langle H_G\rangle$	σ_G
0.1	0.07	0.08	0.20	0.15	0.01
0.2	0.18	0.08	0.39	0.22	0.01
0.3	0.28	0.08	0.59	0.31	0.01
0.4	0.38	0.09	0.73	0.40	0.01
0.5	0.50	0.09	0.81	0.50	0.01
0.6	0.60	0.09	0.86	0.60	0.01
0.7	0.70	0.10	0.89	0.70	0.01
0.8	0.80	0.10	0.92	0.79	0.01
0.9	0.90	0.11	0.93	0.87	0.01

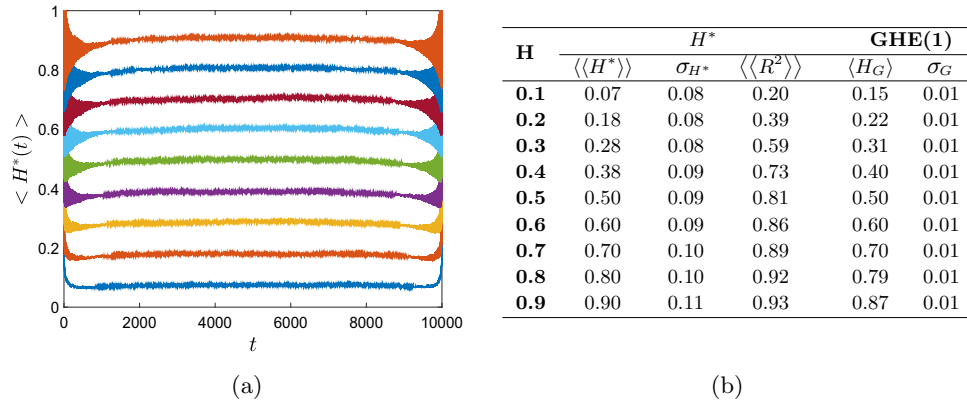(a)   (b)

Fig. 2. (a) Illustration that for FBM the scaling exponent $H^*(t)$ is on average close to the self-similar exponent H. The plot reports the sample mean of $H^*(t)$, denoted as $\langle H^*(t)\rangle$ and computed over $m = 1\,000$ simulations of FBM with self-similar exponent $H = 0.1, 0.2, \ldots, 0.9$ (bottom to top) and length $T = 10\,000$ points. (b) Sample mean of $H^*(t)$ over time and over the number of simulations, denoted as $\langle\langle H^*\rangle\rangle$, standard deviation of $H^*(t)$ and sample mean of the coefficient of determination, $\langle\langle R^2\rangle\rangle$. The values of GHE(1) denote the generalized Hurst exponent with $q = 1$.

- **α-stable Lévy motion (SLM)**. This process is a generalization of the Brownian motion to the α-stable distribution with $0 < \alpha \leq 2$. The case $\alpha = 2$ corresponds to Brownian motion. The SLM is a $1/\alpha$-self-similar process ($H = 1/\alpha$) with stationary and independent increments [19].

  We generated SLM processes using the toolbox provided by [43], sample paths are of length $T = 10\,384$, with parameters for the generation $m = 128$ and $M = 6\,000$, making $m(M+T)$ to be a power of 2, see [43] for more details. We considered the case $H = 1/\alpha$ for values of $H = 0.5, 0.55, \ldots, 0.95$.

  The time-dependent sample mean over the number of simulations is displayed in Fig. 3(a). We observe a noisier estimator than the one obtained for FBM. In Table 3(b), we report $\langle\langle H^*(t)\rangle\rangle$, noticing a fair estimate of the self-similar parameter H, with better results for SLM with smaller H. The means of the coefficient of determination suggest that the scaling relation of Eq. (7) is indeed satisfied. We compared the proposed estimator with the generalized Hurst exponent with $q = 1$, obtaining consistent results and agreement with the expected scaling law $H_G = 1/\alpha$ [40].

- **ARFIMA**. We tested the linear relationship of Eq. (7) in ARFIMA(p, d, q) processes with Gaussian innovations, $p, q \in \mathbb{N}$, autoregressive and moving average coefficients respectively [42]. This model is an extension to the ARIMA(p, d, q) model, allowing the differencing exponent d to take fractional values, $-1/2 < d < 1/2$. The correspondence between the two parameters H and d is given by $H = d + 1/2$. The interval $0 < d < 1/2$ of long-range dependence corresponds to $1/2 < H < 1$ [45]. We considered the simple case of ARFIMA(0,d,0) with fractional order $d = -0.4, -0.3, \ldots, 0.4$ and length $T = 10\,000$. We calculated $H^*(t)$ for the integrated ARFIMA time series.

  From Fig. 4(a), we observe that the estimator $\langle H^*\rangle$ is a good approximation of the exponent H. In Table 4(b), we report $\langle\langle H^*\rangle\rangle$, the standard deviation of $H^*(t)$ and the sample mean of the coefficient of determination, $\langle\langle R^2\rangle\rangle$. Similar to the FBM case, the estimation is more accurate for larger H with larger coefficient of determination.
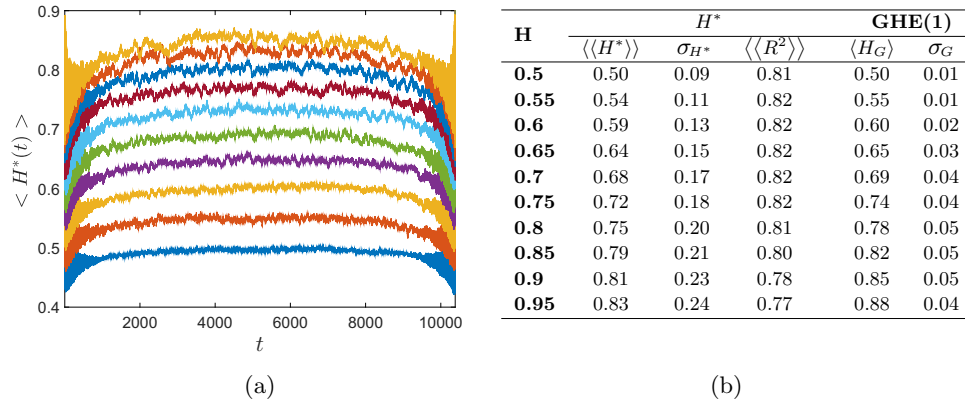
H	H^*			GHE(1)	
	$\langle\langle H^*\rangle\rangle$	σ_{H^*}	$\langle\langle R^2\rangle\rangle$	$\langle H_G\rangle$	σ_G
0.5	0.50	0.09	0.81	0.50	0.01
0.55	0.54	0.11	0.82	0.55	0.01
0.6	0.59	0.13	0.82	0.60	0.02
0.65	0.64	0.15	0.82	0.65	0.03
0.7	0.68	0.17	0.82	0.69	0.04
0.75	0.72	0.18	0.82	0.74	0.04
0.8	0.75	0.20	0.81	0.78	0.05
0.85	0.79	0.21	0.80	0.82	0.05
0.9	0.81	0.23	0.78	0.85	0.05
0.95	0.83	0.24	0.77	0.88	0.04

(a)   (b)

Fig. 3. (a) Illustration that for SLM the scaling exponent $H^*(t)$ is on average close to the value $H = 1/\alpha$. The plots report the sample mean of $H^*(t)$, denoted as $\langle H^*(t)\rangle$ and computed over $m = 1\,000$ simulations of SLM with self-similar exponent $H = 0.5, 0.55, \ldots, 0.95$ (bottom to top) and length $T = 10\,384$ points. (b) Sample mean of $H^*(t)$ over time and over the number of simulations, denoted as $\langle\langle H^*\rangle\rangle$, standard deviation of $H^*(t)$ and sample mean of the coefficient of determination, $\langle\langle R^2\rangle\rangle$.
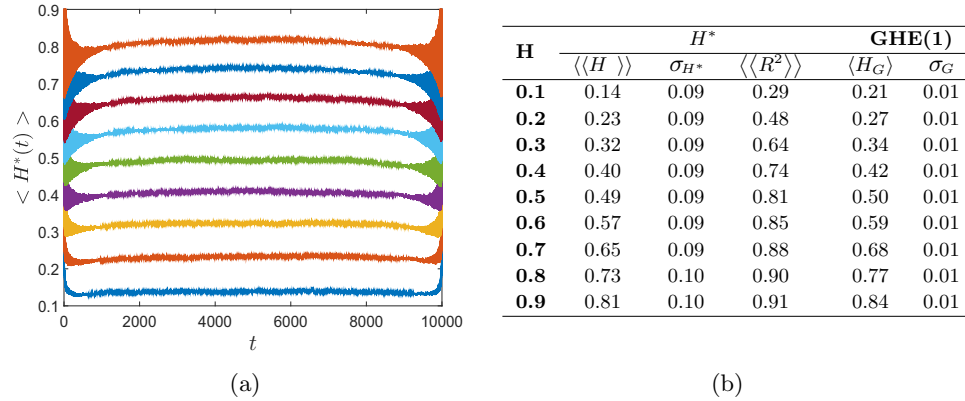

H	H^*			GHE(1)	
	$\langle\langle H^*\rangle\rangle$	σ_{H^*}	$\langle\langle R^2\rangle\rangle$	$\langle H_G\rangle$	σ_G
0.1	0.14	0.09	0.29	0.21	0.01
0.2	0.23	0.09	0.48	0.27	0.01
0.3	0.32	0.09	0.64	0.34	0.01
0.4	0.40	0.09	0.74	0.42	0.01
0.5	0.49	0.09	0.81	0.50	0.01
0.6	0.57	0.09	0.85	0.59	0.01
0.7	0.65	0.09	0.88	0.68	0.01
0.8	0.73	0.10	0.90	0.77	0.01
0.9	0.81	0.10	0.91	0.84	0.01

(a)   (b)

Fig. 4. (a) Illustration that for ARFIMA(0,1,0) the scaling exponent $H^*(t)$ is on average close to the value $H = d+0.5$. The plots report the sample mean of $H^*(t)$, denoted as $\langle H^*(t)\rangle$ and computed over $m = 1\,000$ simulations of ARFIMA(0,d,0) with self-similar exponent $H = 0.1, 0.2, \ldots, 0.9$ (bottom to top) and length $T = 10\,000$ points. (b) Sample mean of $H^*(t)$ over time and over the number of simulations, denoted as $\langle\langle H^*\rangle\rangle$, standard deviation of $H^*(t)$ and sample mean of the coefficient of determination, $\langle\langle R^2\rangle\rangle$.

From the analysis of these three different stochastic processes, we observe that our proposed method produces a fair estimate of the self-similar parameter H. The chosen models concern two different properties which contribute to the self-similarity of the processes. The first property is the long-range autocorrelation of the increments. The second property is the high variability or the heavy tails in the distribution of the increments of the α-stable Lévy motion. We have demonstrated empirically that the analytics of the relative amplitude of the oscillating components of the signal capture both of these properties. Let us remark that this scaling exponent is not intended as an alternative method to estimate H, which can instead be obtained with more reliable tools [15,46,47]. The aim of this method is instead to compute the time-dependent amplitude contribution of the prevalent fluctuations present in a time

series, distinguishing between periods when high or low frequencies are contributing more than what could be expected from Brownian motion.

4 Time-dependent complexity measure

We define a time-dependent Shannon entropy-like measure based on the square of the amplitude of the IMFs. This measure provides a time-varying quantification of complexity spectrum that offers an alternative to the scaling exponent to measure the strength of cycles present in financial time series. Making use of the functions a_k, described in Eq. (4), we define a time-scale relative distribution of amplitudes as:

$$p_k(t) = \frac{a_k^2(t)}{\sum\limits_{k=1}^{n} a_k^2(t)}, \tag{8}$$

where n is the number of IMFs excluding the residue. Similarly to Shannon entropy [48], we define the time-dependent complexity measure as:

$$C^*(t) = -\sum_{k=1}^{n} p_k(t) \ln p_k(t). \tag{9}$$

Equation (9) provides a measure of the distribution of amplitudes between the oscillating components. If the total amplitude at time t is concentrated in one oscillation mode, we observe a low complexity value, implying that around time t the process is following a prevalent trend. On the contrary, if at time t all the oscillation modes have similar amplitudes, we obtain a large complexity value which indicates a more erratic and unpredictable behaviour.

Thus, $C^*(t)$ provides a time-varying estimation of disorder and adapts closely to our visual perception of complexity. Moreover, Eq. (9) offers a more general measure of uncertainty than the variance since the latter measures the dispersion around the mean, while $C^*(t)$ measures the dispersion of energy around the different IMFs. Similar to an entropy measure, the value of the proposed complexity at time t varies between zero, if one IMFs dominates the energy of the process, and $\log(n)$ if the energy is uniformly distributed between the n IMFs.

The choice of weights equal to the square of the amplitudes in Eq. (8) is arbitrary, although it is in agreement with other measures of entropy that have been defined, for example in [49]. We tested alternative choices, such as the linear weight $a_k(t)$, obtaining analogous results to the ones reported here.

Let us note that, although not independent, the two measures convey different information. The estimator $C^*(t)$ is an information quantifier of uncertainty, it is obtained from the distribution of the amplitudes regardless of their time-scales and it only quantifies the homogeneity of the components. On the other hand, the scaling exponent, $H^*(t)$, measures the change in the amplitudes across time-scales, testing the scaling law of Eq. (7). In this way, $H^*(t)$ is a more restrictive measure which assumes a linear relationship between the logarithm of the amplitudes and the logarithm of the oscillating periods.

5 Time-dependent scaling in stock markets

We applied the proposed measures to intraday prices of four stock market indices: (1) S&P 500 (USA), (2) IPC (Mexico), (3) Nikkei 225 (Japan) and (4) XU 100
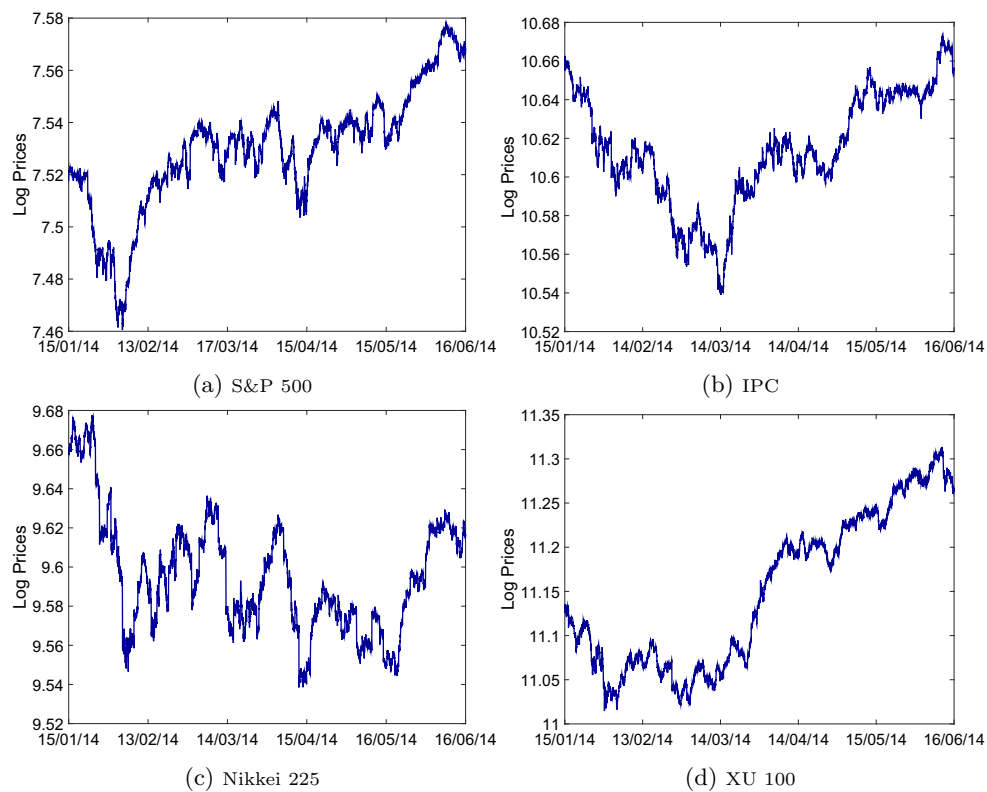
Fig. 5. 30-second sampled log-prices for different stock market indices for the period January 15th, 2014 to June 16th, 2014. (a) S&P 500, (b) IPC, (c) Nikkei 225 and (d) XU 100.

Table 1. Number of days and length of each financial time series.

Country	Index	No. of Days	Length
USA	S&P 500	105	81.900
Japan	Nikkei 225	104	62.400
Mexico	IPC	101	78.780
Turkey	XU 100	106	78.440

(Turkey). We intentionally chose two financial markets that are classified as developed markets (USA and Japan) and two emerging markets (Mexico and Turkey) with the additional feature that the Japanese and the Turkish stock exchanges have two trading sessions separated by a lunch break.

The data set, obtained from Bloomberg, consists of prices recorded at 30-second intervals. This sampling frequency provides a reasonable balance between the effects of market micro-structure frictions at the very highest sampling frequencies on the one hand, and the analytics on a wide frequency range on the other. The data set covers a period of 5 months, from January 15th, 2014 to June 16th, 2014. The number of days and the number of data points for every trading day depend on the opening hours of each stock exchange. The logarithm of prices for each stock market index are plotted in Fig. 5. Table 1 shows the number of days and the length of each time series.
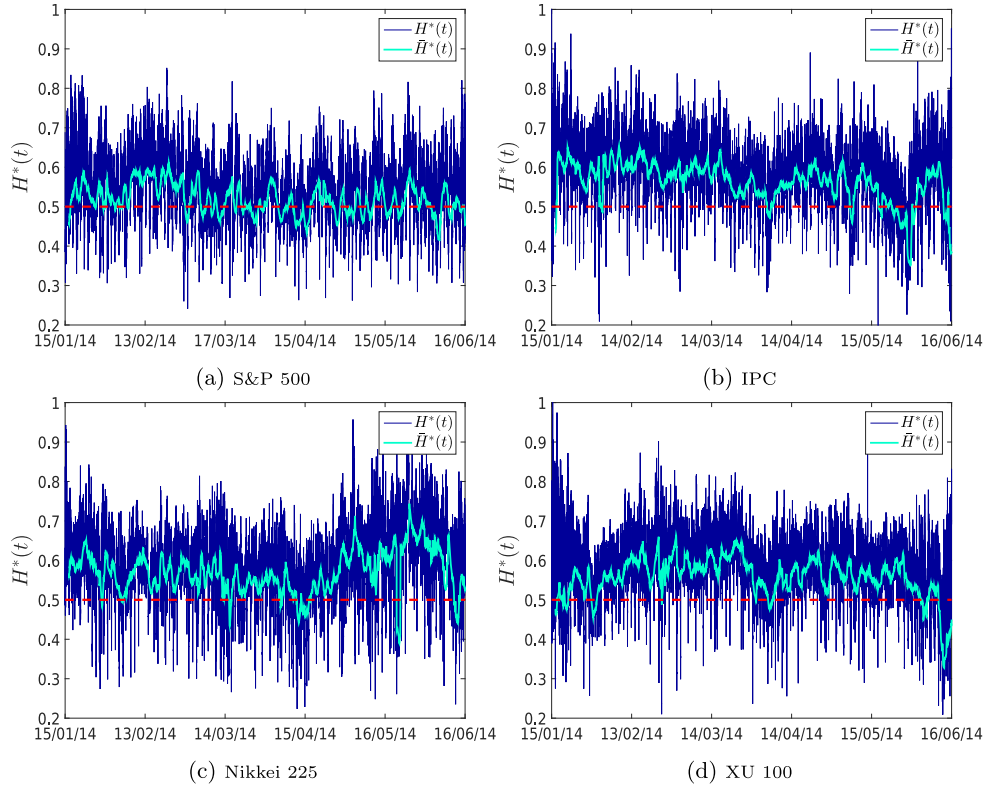
Fig. 6. Time-dependent scaling exponent for different stock market indices for the period January 15$^{\text{th}}$, 2014 to June 16$^{\text{th}}$, 2014. The scaling exponent $H^*(t)$ is depicted by a dark-blue line. The light-blue line represents $\bar{H}^*(t)$, a rolling window average over the length of a trading day. The dashed red line indicates the value $H = 0.5$. (a) S&P 500, (b) IPC, (c) Nikkei 225 and (d) XU 100.

The time evolution of $H^*(t)$ over the 5-month period for each financial index is shown in Fig. 6 (dark-blue line). We note that $H^*(t)$ has large intraday variations which obscure any possible trend over longer periods. For this reason, we report a moving average version of $H^*(t)$ denoted as $\bar{H}^*(t)$ (light-blue line in the same Figure). More specifically, $\bar{H}^*(t)$ is calculated from the relation, $\bar{a}_k(t) \propto \bar{\tau}_k^{\bar{H}^*(t)}(t)$, where $\bar{a}_k(t)$ and $\bar{\tau}_k(t)$ are the averages over a rolling window of the size of a trading day. The dashed red line in this Figure indicates the value $H = 0.5$.

By comparing the values of $H^*(t)$ and $\bar{H}^*(t)$ for the four different stock market indices, we observe that the S&P 500 index is the one closest to the value $H = 0.5$ as expected for Brownian motion, with values of $\bar{H}^*(t)$ fluctuating around 0.5 exposing only some brief departures from it (see Fig. 6(a)). For instance, we can detect a period around February 2014 where the scaling parameter results in significantly larger values. In this period of time, the S&P 500 index was indeed in a rising rally, see Fig. 5(a). Therefore, this suggests that the identified persistent behaviour could be attributed to the recognition of a long time-scale cycle with larger amplitudes than in the case of pure random walk.

In Fig. 6(c), we report the scaling dynamics for the Nikkei 225 index. We observe that $\bar{H}^*(t)$ has values constantly above 0.5, specially at the end of the analysed period.

Table 2. Average goodness-of-fit coefficient (R^2) for the amplitude versus period log-linear model for different financial indices. First, the average is calculated for all the time instances t, then it is calculated separately for those instances where $H^*(t) < 0.45$, $0.45 < H^*(t) < 0.55$ and $H^*(t) > 0.55$.

Index	$\langle R^2 \rangle_{All}$	$\langle R^2 \rangle_{H^*<0.45}$	$\langle R^2 \rangle_{0.45<H^*<0.55}$	$\langle R^2 \rangle_{H^*>0.55}$
S&P 500	0.8753	0.825	0.8716	0.8916
IPC	0.8812	0.7971	0.8703	0.8915
NIKKEI 225	0.8072	0.7345	0.7829	0.8198
XU100	0.9196	0.7987	0.8737	0.9209

It should be noted that this market has lunch breaks which affect the intraday values of $\bar{H}^*(t)$.

For the IPC index the values of $H^*(t)$ and $\bar{H}^*(t)$ are consistently closer to $H = 0.6$, see Fig. 6(b). This suggests that the IPC returns show intervals where the amplitudes display a persistent behaviour. Similarly, the Turkish scaling exponents take values larger than $H = 0.5$, see Fig. 6(d).

We tested the validity of Eq. (7) when applied to financial data by computing for every time instance t the coefficient of determination $R^2(t)$. The mean over the whole period is reported in the second column of Table 2. We also considered the three cases: $H^*(t) < 0.45$, a window around $0.45 < H^*(t) < 0.55$ and $H^*(t) > 0.55$. We observe that the goodness-of-fit is generally better for $H^*(t) > 0.5$, see Table 2.

For a comparative analysis, we calculated the complexity measure $C^*(t)$ described by Eq. (9). The obtained values for each stock market index are illustrated in Fig. 7. We observe that the complexity values for the S&P 500 index, Fig. 7(a), are overall the largest among the four indices.

The IPC index shows an increasing evolution of $C^*(t)$, suggesting a more uniform distribution of amplitudes at the beginning of 2014, see Fig. 7(b). On the contrary, the Nikkei 225 index presents a decreasing measure of complexity, indicating a period of higher complexity at the beginning of the sample period, Fig. 7(c). Finally, the XU 100 index presents alternate intervals of high and low complexity, displaying regularly large values in the last two months of the analysed period. This higher randomness is also visible from the scaling exponent that displays relative lower values of $\bar{H}^*(t)$, Fig. 7(d).

Overall, the functions $H^*(t)$ and $C^*(t)$ vary in opposite directions. For each market, the correlation between these two measures is negative with values $\rho_{S\&P} = -0.21$, $\rho_{IPC} = -0.23$, $\rho_{Nikkei} = -0.28$, $\rho_{XU} = -0.15$. We note that these correlations values are relatively small, indicating a weak linear dependence between these variables. This is to be expected as the underlying measures are associated with rather different properties.

6 Intraday analysis of scaling patterns

We investigated the intraday patterns by separating the paths of $H^*(t)$ and $C^*(t)$ into daily windows. Taking for example the time series $H^*(t)$ for the S&P 500 index displayed in Fig. 6(a), we separated this time series into the $n = 105$ days included in the data set, see Table 1. In Fig. 8(a), we display these daily time series (one day on top of the other). The colour bar represents the value of $H^*(t)$. This graphical representation allows us to compare trading sessions and identify patterns at specific time of the day.

We estimated the statistical mean of $H^*(t)$ across the days, resulting in an average value for each time instance of the trading session. This average, denoted
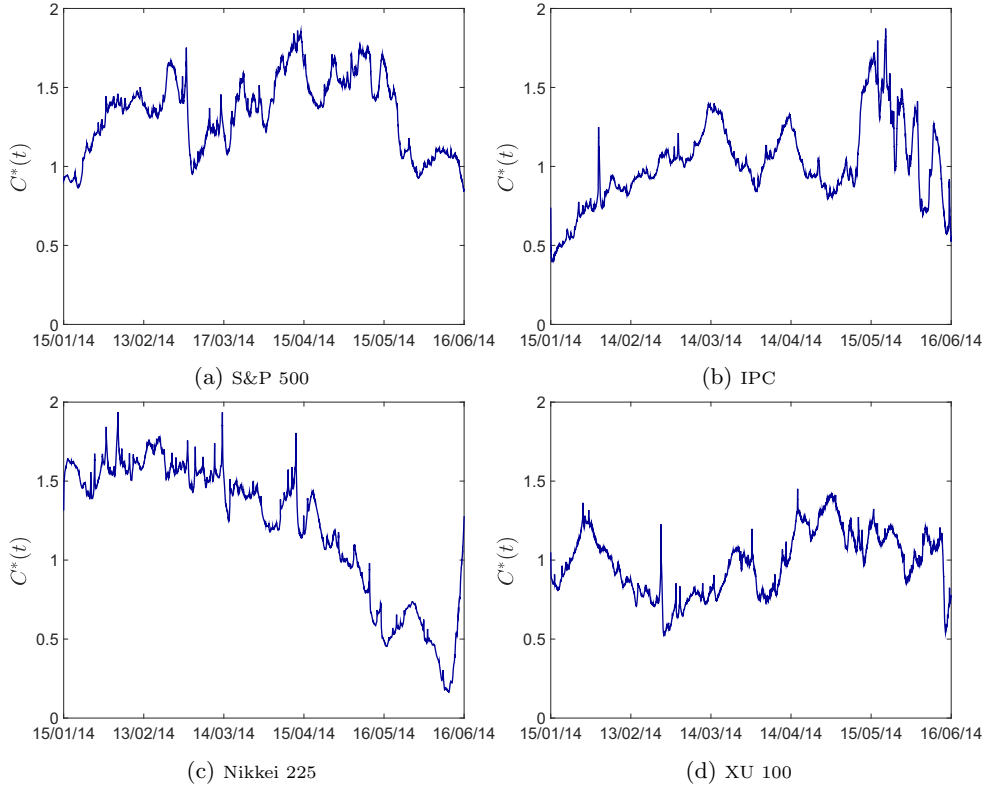
Fig. 7. Time-dependent complexity measure, $C^*(t)$, for different stock market indices for the period January 15$^{\text{th}}$, 2014 to June 16$^{\text{th}}$, 2014. (a) S&P 500, (b) IPC, (c) Nikkei 225 and (d) XU 100.

as $\langle H^*(t)\rangle_{\text{days}}$, describes the regular behaviour of $H^*(t)$ on a trading session, see Fig. 8(b).

In order to validate that the observed dynamics of $\langle H^*(t)\rangle_{\text{days}}$ are statistically significant, we compared these dynamics with the scaling exponents, $H^*_{\text{BM}}(t)$, obtained from several realizations of Brownian motions of length equal to the analysed financial time series, see Table 1. The time series of $H^*_{\text{BM}}(t)$ were fragmented into n windows of equal length. The mean over these n windows is denoted as $\langle H^*_{\text{BM}}(t)\rangle$. The pink band reported in Fig. 8(b) corresponds to the 5$^{\text{th}}$ and the 95$^{\text{th}}$ percentiles of the empirical distribution of $\langle H^*_{\text{BM}}(t)\rangle$ computed from 100 simulations.

We compared the values of $H^*(t)$ obtained for each stock market index with the $\langle H^*_{\text{BM}}(t)\rangle$ band. At each time instance t of the trading session, we estimated the relative fraction of $H^*(t)$ values that falls outside the pink band. In Fig. 8(c), we report these results as a ratio of the number of days outside the band divided by the total number of days. This ratio is labelled as likelihood. The colour bar of this figure represents the value of the average scaling exponent, i.e., $\langle H^*(t)\rangle_{\text{days}}$ (value plotted in Fig. 8(b)). From this Figure, we observe that across the day there are periods of time with very high empirical probability of observing values of the scaling exponents significantly different from the corresponding values extracted from pure Brownian motion.
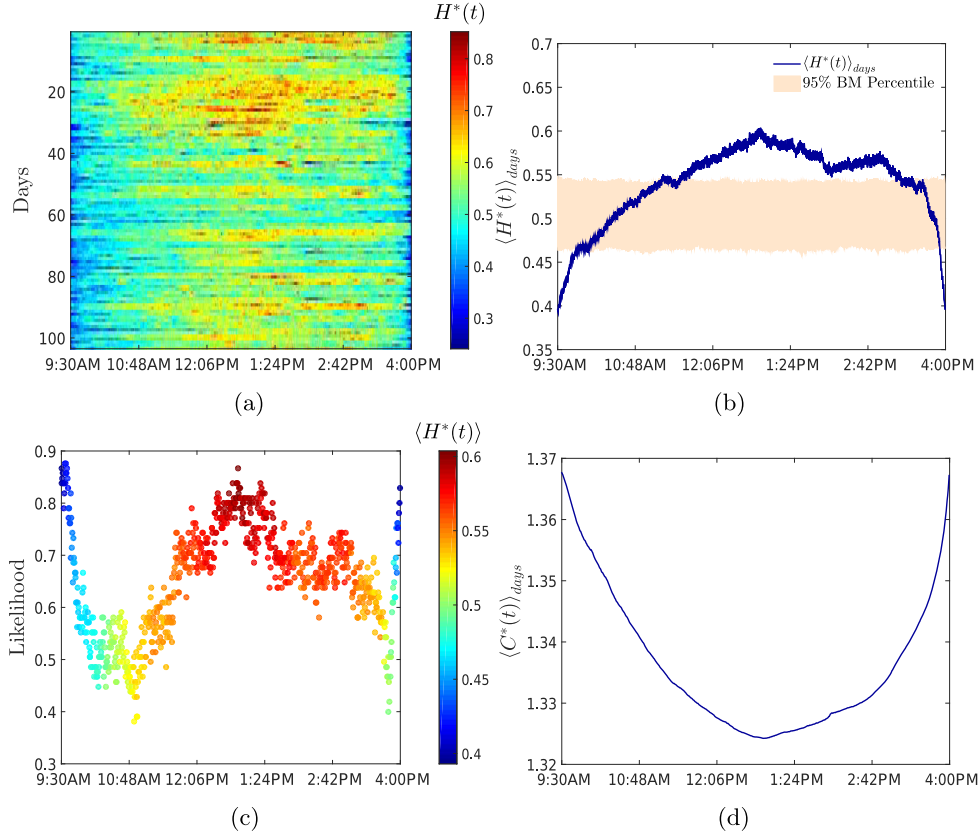
(a)

(b)

(c)

(d)

Fig. 8. Intraday analysis for the S&P 500 Index. (a) Intraday dynamics of the scaling exponent, $H^*(t)$, as a function of day and time. The colour bar indicates the value of the scaling exponent $H^*(t)$. (b) Mean of $H^*(t)$ over the 105 days, denoted as $\langle H^*(t)\rangle_{\text{days}}$. The pink band corresponds to the 5th and 95th percentile of the distribution of $\langle H^*_{\text{BM}}(t)\rangle$ computed from 100 simulations. (c) Likelihood of $H^*(t)$ to fall outside the 5th and 95th percentile band for Brownian motion (pink band of Figure (b)). The colour bar indicates $\langle H^*(t)\rangle_{\text{days}}$, the value shown in Figure (b). (d) Mean of the windowed complexity measure, denoted as $\langle C^*(t)\rangle_{\text{days}}$.

The mean of the complexity measure, $C^*(t)$, at each time t of the trading session is shown in Fig. 8(d). Equally as with the $H^*(t)$ exponent, the mean of $C^*(t)$ is computed across all n days and it is denoted as $\langle C^*(t)\rangle_{\text{days}}$. The same daily analysis for the remaining three stock market indices is reported in Figs. 9, 10 and 11, respectively. Overall, from the intraday scaling and complexity measures we observed the following patterns:

- For each stock market index, the daily average $\langle H^*(t)\rangle_{\text{days}}$ displays an inverted U-shaped form that reflects a more chaotic behaviour at the beginning and at the end of the trading session. The opposite behaviour is observed for the complexity measure, $\langle C^*(t)\rangle_{\text{days}}$, which reveals a U-shaped form.
- The S&P 500 index displays the largest values of $H^*(t)$ (a stronger amplitude persistent behaviour) during the middle of the trading session. From Fig. 8(a), we observe that most of the days present large values of $H^*(t)$ around midday. At this time, the average exponent $\langle H^*(t)\rangle_{\text{days}}$ reaches a value of 0.6, see
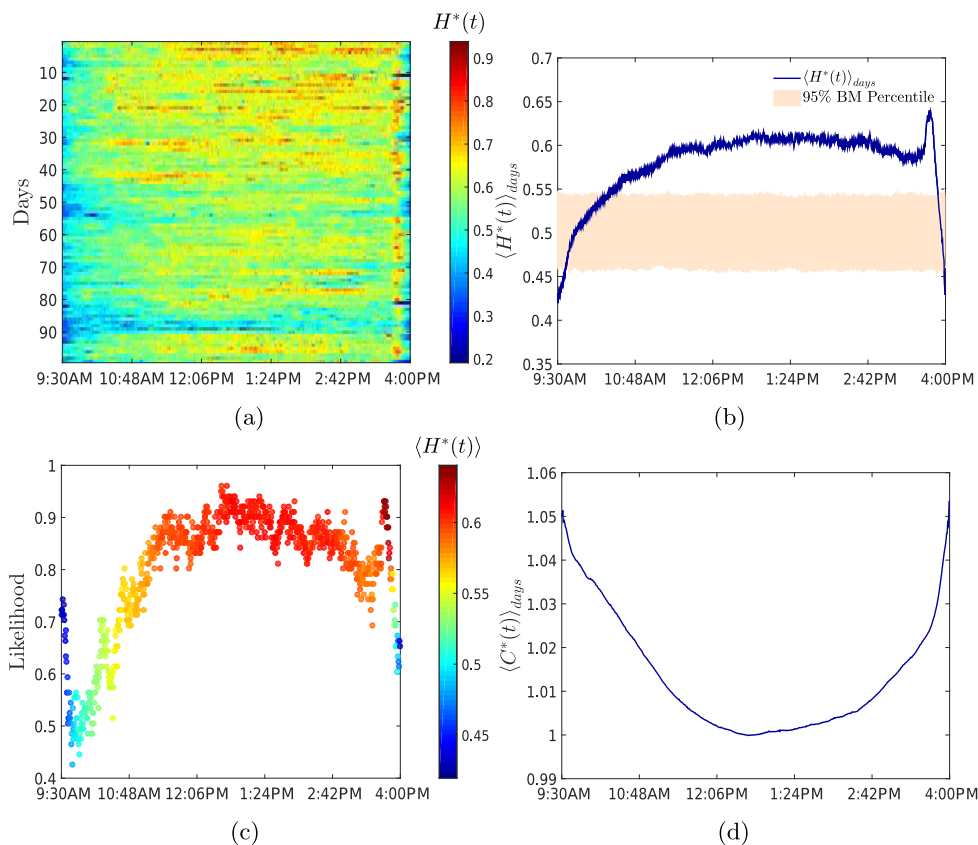
Fig. 9. Intraday analysis for the IPC index. Caption for sub-figures (a), (b), (c) and (d) are the same as Fig. 8.

Fig. 8(b). These values are significantly different from what would be expected for Brownian motion with more than 80% of the observations outside the 5th and the 95th percentiles, see Fig. 8(c). Consistently, the complexity measure reaches its minimum at the same time of the trading session, see Fig. 8(d).

– The Mexican stock exchange is characterized by some large scaling values at the middle of the day. However the most noticeable pattern is the peak formed just before the end of the trading session, see Fig. 9(a). The mean of the windowed values reaches a maximum of 0.65, creating an upswing shape at this time, see Fig. 9(b). This could be associated with an increase of trades in the last few minutes of the trading session which creates a more drastic change in the amplitudes. This pattern is only present in the Mexican stock exchange.

From Fig. 9(c), we observe that some minutes before the closing of the market, more than 90% of the local scaling indices fall outside the 5th and the 95th percentile band for Brownian motion. The complexity measure also reflects a steep increase of disorder at the end of the trading session, see Fig. 9(d).

– The Japanese and the Turkish stock exchanges display two regions of large values for the scaling exponent. These regions are separated by the lunch break, see Figs. 10(a) and 11(a) respectively. The mean of the scaling exponent reflects a quasi-double inverted U-shaped form that is associated with the opening and closing of the morning and afternoon sessions, see Figs. 10(b) and 11(b). It is worth noting that the two trading sessions do not display exactly the same
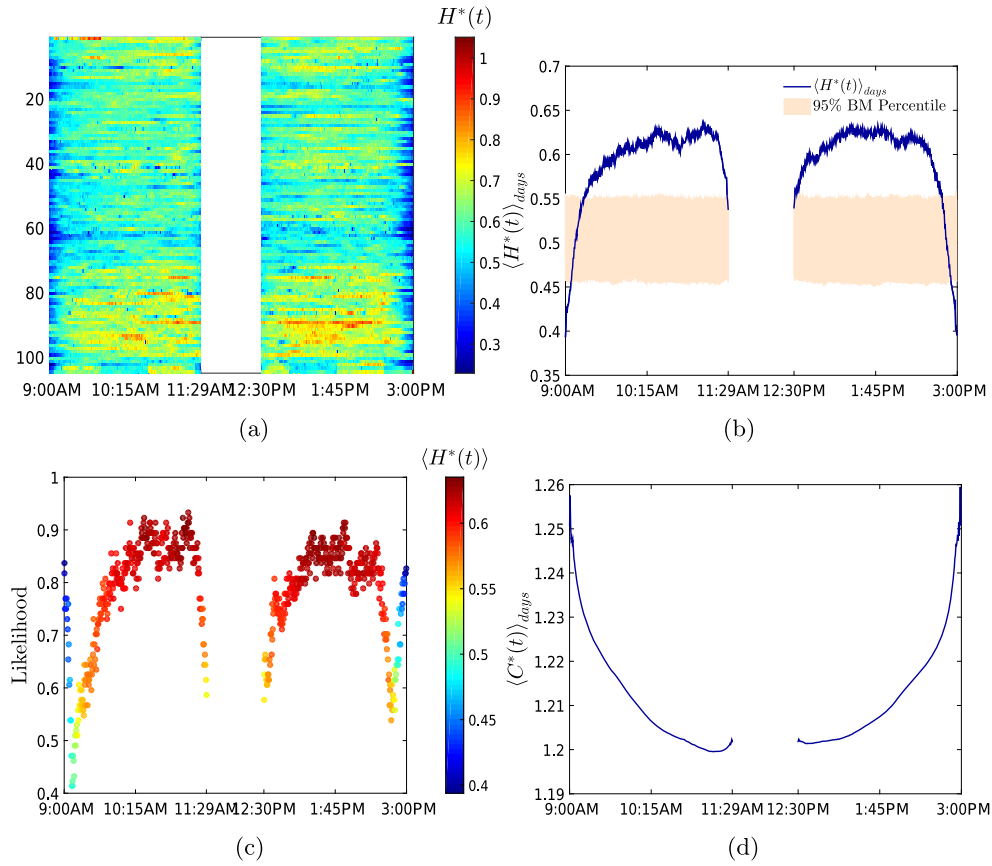
Fig. 10. Intraday analysis for the Nikkei 225 index. Caption for sub-figures (a), (b), (c) and (d) are the same as Fig. 8. The white vertical band in each sub-figure corresponds to the lunch break in this stock exchange.

profile. For the Japanese financial market, the inverted U-shaped form of the first trading session is slightly skewed to the right, in comparison with the more symmetric shape of the second trading session, see Fig. 10(b).

– For the Turkish stock exchange, we observe that the first part of the trading session shows larger values of $H^*(t)$. More than 90% of the analysed days present local scaling exponents that surpass the value of 0.6, see Fig. 11(c). The dominance of one IMF amplitude in the first trading session is also reflected in the lower values of the complexity measure, which reaches the lowest value when compared to the other stock market indices, see Fig. 11(d).

Overall the intraday patterns of $H^*(t)$ and $C^*(t)$ confirm the well known fact that activity on financial markets is not constant throughout the day. The uncovered patterns corroborate the hectic buy and sell activity affecting different markets at the opening and closing of trading sessions [50,51]. Smaller values of $H^*(t)$ (large values of $C^*(t)$) imply a non-persistent and rougher behaviour which is reflected in higher volatility. The exposed daily patterns of $\langle H^*(t)\rangle_{\text{days}}$ and $\langle C^*(t)\rangle_{\text{days}}$ are in agreement with the results which document the existence of a distinct U-shaped pattern in market activity and volatility over a trading day, i.e., volatility is higher
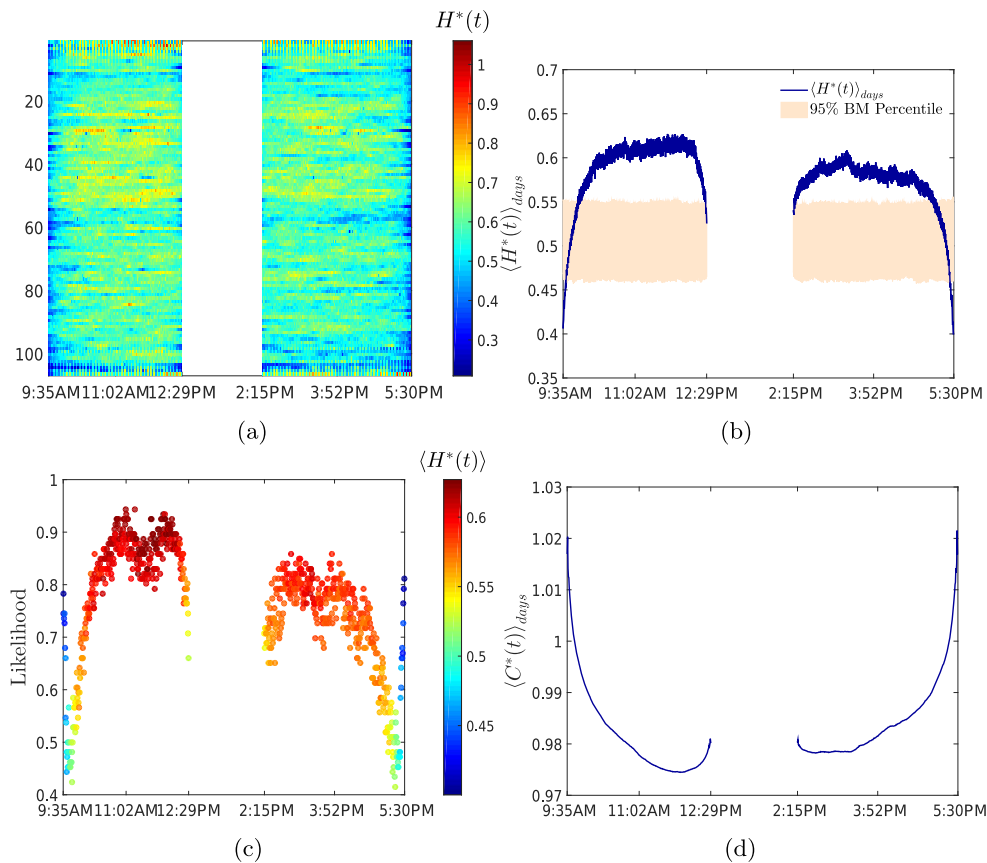
Fig. 11. Intraday analysis for the XU 100 index. Caption for sub-figures (a), (b), (c) and (d) are the same as Fig. 8. The white vertical band in each sub-figure corresponds to the lunch break in this stock exchange.

at the opening and at the closing of the trading session and low in the middle of the day, see for example [52–54].

By comparing the intraday complexity values across the four markets, we observe that S&P 500 index displays the largest values across all the trading session. The Nikkei 225 index is the second most complex, followed by the IPC index and lastly the XU 100 index which has the smaller complexity values. This is in agreement with the results reported for developed and emerging markets [15,28,55].

7 Conclusion

We studied the relative weight of the oscillating components present in intraday financial time series coupled with their characteristic time-scale. These components are extracted via the Hilbert-Huang transform. We have shown that the combination of the EMD and its associated Hilbert spectral analysis offers a powerful tool to uncover the time-dependent scaling patterns of intraday financial data.

We proposed two new time-dependent measures: 1) an amplitude scaling exponent and 2) an entropy-like complexity measure. Our measures are non-parametric and they do not assume any a priori stochastic process. The scaling

exponent only assumes the existence of a power-law relation between the instantaneous amplitudes and the instantaneous periods which was empirically shown to be present. By applying our methodology to different models with known scaling laws, we demonstrated that the time-dependent values of our scaling exponent vary consistently around the known exponent of the models. When applied to real financial data, our measures uncover significant variations of the scaling properties of the market during the trading day and provide evidences of non-stationary patterns. We verified that the scaling exponent and the complexity measure are related, with larger scaling exponents associated with lower complexity values. However, the proposed measures convey different information about the properties of the oscillating components of the signal. Specifically, we applied the scaling and the entropy-like measures to the study of four financial markets, two developed (US and Japan) and two emerging markets (Mexico and Turkey). We contrasted and compared the decomposition of their financial indices. With the use of intraday data, we recognized some patterns and identified periods of low and high complexity.

Compared to the other analysed stock market indices, the S&P 500 index results the most complex. The intraday analysis reveals a distinctive anti-persistent behaviour at the opening and at the closing of the trading session, contrasting with the persistent behaviour at the middle of the session. Similar intraday results are obtained for the other stock market indices. The variations observed in the scaling measure are well outside the 5$^{\text{th}}$ and the 95$^{\text{th}}$ percentiles expected for Brownian motion, suggesting strong deviations from this model that could be attributed to the presence of long-range dependence or/and heavy tails.

With the proposed measures, we are able to describe the dynamics of financial time series whose regularity changes over time. Our results suggest that financial time series have dynamic scaling properties that change during the trading day following identifiable patterns that are characteristic of each market. The origin of the scaling laws could be attributed to the autocorrelation of the process, the presence of heavy tails and the non-stationarity of the time series. It is beyond the purpose of this paper to investigate this aspect (which is discussed in [22] by using a different approach). Our aim here was to uncover non-stationary scaling patterns which we showed to be significant, reproducible and characteristic of specific stock markets. These non-stationary scaling patterns must be considered when modelling financial time series and building trading strategies.

References

1. M.M. Dacorogna, R. Gençay, U. Müller, R.B. Olsen, O.V. Pictet, *An Introduction to High Frequency Finance* (Academic Press, San Diego, 2001)
2. E.F. Fama, J. Finance **25**, 383 (1970)
3. T. Chordia, R. Roll, A. Subrahmanyam, J. Financial Econ. **76**, 271 (2005)
4. N. Visaltanachoti, T. Yang, J. Banking Finance **34**, 594 (2010)
5. E.E. Peters, *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics* (Wiley, New York, 1994)
6. A. Weron, R. Weron, Chaos Solitons Fractals **11**, 289 (2000)
7. L. Kristoufek, Adv. Complex Syst. **15**, 1250065 (2012)
8. L. Kristoufek, Sci. Rep. **3**, 2857 (2013)
9. S. Rachev, A. Weron, R. Weron, Math. Comput. Modell. **29**, 23 (1999)

10. D.-Y. Li, Y. Nishimura, M. Men, Physica A: Stat. Mech. Appl. **407**, 144 (2014)
11. B. Mandelbrot, J. Business **36**, 394 (1963)
12. B. Mandelbrot, H.M. Taylor, Oper. Res. **15**, 1057 (1967)
13. M. Bartolozzi, C. Mellen, T. DiMatteo, T. Aste, Eur. Phys. J. B **58**, 207 (2007)
14. T. Di Matteo, T. Aste, M. Dacorogna, Physica A: Stat. Mech. Appl. **324**, 183 (2003)
15. T. Di Matteo, T. Aste, M. Dacorogna, J. Banking Finance **29**, 827 (2005)
16. R.N. Mantegna, H.E. Stanley, Nature **376**, 46 (1995)
17. S. Drozdz, F. Ruf, J. Speth, M.Wjcik, Eur. Phys. J. B **10**, 589 (1999)
18. L.E. Calvet, A.J. Fisher, Rev. Econ. Stat. **84**, 381 (2002)
19. G. Samoradnitsky, M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance, Stochastic Modeling Series* (Taylor & Francis, New York, 1994)
20. B.B. Mandelbrot, J.W. van Ness, SIAM Rev. **10**, 422 (1968)
21. R. Cont, in *Fractals in Engineering* (Springer, 2005), p. 159
22. R. Buonocore, T. Aste, T.D. Matteo, Chaos Solitons Fractals **88**, 38 (2016)
23. S. Bianchi, A. Pantanella, A. Pianese, Quant. Finance **13**, 1317 (2013)
24. S. Bianchi, A. Pantanella, A. Pianese, Risk Decision Anal. **5**, 1 (2014)
25. D.O. Cajueiro, B.M. Tabak, Physica A: Stat. Mech. Appl. **336**, 521 (2004)
26. W.A. Risso, Appl. Econ. Lett. **16**, 485 (2009)
27. L. Zunino, M. Zanin, B.M. Tabak, D.G. Prez, O.A. Rosso, Physica A: Stat. Mech. Appl. **389**, 1891 (2010)
28. L. Kristoufek, M. Vosvrda, Eur. Phys. J. B **87**, 1 (2014)
29. N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, Proc. R. Soc. London, A **454**, 903 (1998)
30. N.E. Huang, M.-L. Wu, W.Qu, S.R. Long, S.S.P. Shen, Appl. Stochastic Models Bus. Ind. **19**, 245 (2003)
31. B. Zhu, P. Wang, J. Chevallier, Y. Wei, Comput. Econ. 1 (2013)
32. B. Premanode, C. Toumazou, Expert Syst. Appl. **40**, 377 (2013)
33. C.-H. Cheng, L.-Y. Wei, Econo. Modell. **36**, 136 (2014)
34. N. Nava, T.D. Matteo, T. Aste, Physica A: Stat. Mech. Appl. **447**, 434 (2016)
35. Z. Peng, P.W. Tse, F. Chu, Mech. Syst. Signal Process. **19**, 974 (2005)
36. F. King, *Encyclopedia of Mathematics and its Applications* (Cambridge University Press, Cambridge, 2009)
37. S.L. Marple Jr., Signal Process. IEEE Trans. **47**, 2600 (1999)
38. G. Fusai, G. Germano, D. Marazzina, Eur. J. Oper. Res. **251**, 124 (2016)
39. T. Di Matteo, Quant. Finance **7**, 21 (2007)
40. J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, H. Stanley, Physica A: Stat. Mech. Appl. **316**, 87 (2002)
41. C.R. Rao, *Linear Statistical Inference and its Applications* (Wiley, New York, 1973)
42. C.W.J. Granger, R. Joyeux, J. Time Ser. Anal. **1**, 15 (1980)
43. S. Stoev, M.S. Taqqu, Fractals **12**, 95 (2004)
44. N.E. Huang, S.S.P. Shen, *Hilbert-Huang Transform and its Applications*, Vol. 16 (World Scientific, Singapore, 2005)
45. M.S. Taqqu, V. Teverovsky, *A Practical Guide to Heavy Tails* (Birkhauser Boston Inc., Cambridge, MA, USA, 1998), p. 177
46. M. Fernández-Martínez, M. Sánchez-Granero, J.T. Segovia, Physica A: Stat. Mech. Appl. **392**, 5330 (2013)
47. S. Stoev, V. Pipiras, M.S. Taqqu, Signal Process. **82**, 1873 (2002)
48. C. Shannon, Bell Syst. Techn. J. **27**, 379, 623 (1948)
49. O.A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, E. Başar, J. Neurosci. Methods **105**, 65 (2001)
50. T.G. Andersen, T. Bollerslev, J. Empirical Finance **4**, 115 (1997)
51. A. Chakraborti, I.M. Toke, M. Patriarca, F. Abergel, Quant. Finance **11**, 991 (2011)
52. R. Allez, J.-P. Bouchaud, New J. Phys. **13**, 025010 (2011)
53. K. Boudt, C. Croux, S. Laurent, J. Empirical Finance **18**, 353 (2011)

54. R.A. Wood, T.H. McInish, K. Ord, J. Finance **40**, 723 (1985)
55. L. Zunino, B.M. Tabak, D.G. Pérez, M. Garavaglia, O.A. Rosso, Eur. Phys. J. B **60**, 111 (2007)