Chaos, Solitons and Fractals 88 (2016) 24–37

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Trading volume in financial markets: An introductory review

Sílvio M. Duarte Queirós[1,*]

Centro Brasileiro de Pesquisas Físicas, Rua Dr Xavier Sigaud 150, Urca, Rio de Janeiro - RJ 22290–180, Brazil

ARTICLE INFO

Article history:
Available online 27 January 2016

Keywords:
Trading volume
Quantitative finance
Econophysics
Complex systems
MDH
SIAH

ABSTRACT

In this article, I introduce a short review on the statistical and dynamical properties of the high-frequency trading volume and its relation to other financial quantities such as the price fluctuations and trading value. In addition, I compare these results — which were obtained within the framework of applications of Physics to quantitative financial analysis —with the mainstream financial hypotheses of mixture of distributions (MDH) and sequential arrival of information (SIAH).

1. Introduction

Although most of the work carried out for the characterisation of financial markets within the context of applications of Physics to quantitative financial analysis has been devoted to understanding fundamental features of price fluctuations (*aka* returns) or the volatility [1,2], there is another pivotal quantity in the definition of the dynamics of a given asset: the amount thereof that changes hands in a certain period of time – the *trading volume*, v. The reason for its relevance is simply understandable: If a price might not change when there is a transaction, in order to have a price change — stock splits apart — there must be a seller and a buyer agreeing to make a transaction at some price, S, that differs from the previous quote.[2] In the case of a liquid market — for which there is traditionally a large number of both bidders and askers mainly concentrated and the first levels of the order book[3] — large price fluctuations were associated with strong buying and selling pressures that would make the price to significantly hike up or slump, respectively. Mainly for this reason it loomed the perception in the meanwhile turned into an adage that goes as "it takes volume to make prices move". In other words, the emergence of an significant imbalance between bidders and askers is bound to be caused by some information flow that makes the agent make up her mind and ultimately the establishment of herding phenomena in the market, *i.e.*, trading volume could be a reasonable proxy for information. This reasoning is intimately related to one of the two currents about relation between price fluctuations, volume and information in financial theory, known as Mixture of Distributions Hypothesis (MDH). The MDH, was introduced by Clark [33] on qualitative grounds, who conjectured the dynamics of both quantities were dependent on latent events leading to a joint distribution where the volatility and the trading volume are both described by log-Normal distributions. In his formulation, Clark also introduced a distinction which is appealing to physicists interested in complex systems: the difference between physical (clock) time and proper (event) time. That distinction is fundamental in the discovery of several properties and laws in complex phenomena like earthquakes and avalanches only emerge [34]. According to Clark the proper time in a financial market would be the arrival of information.

* Corresponding author. Tel.: +55 21 21417213.
E-mail address: sdqueiro@gmail.com

[1] Associate to National Institute of Science and Technology for Complex Systems, Brazil.

[2] Besides stock splits there is also an automatic discount after the payment of a dividend.

[3] Herein, I define a market as liquid following Black's criteria [3] that it must be extremely tight, with a little deep order book and sufficiently resilient so that prices are prone to converge to the underlying value.

http://dx.doi.org/10.1016/j.chaos.2015.12.024
0960-0779/

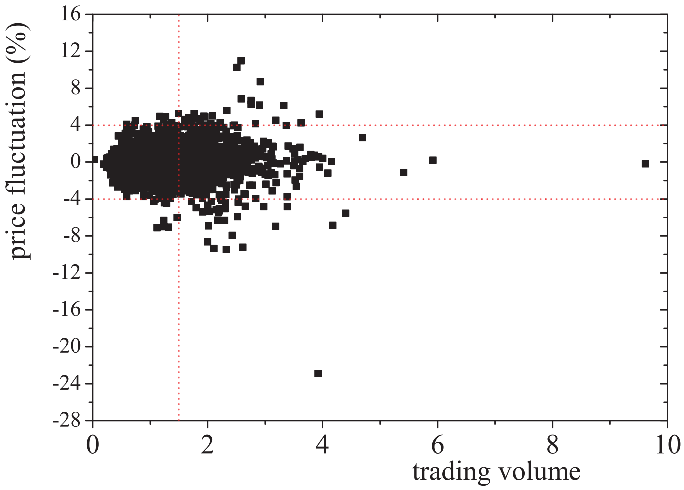

Fig. 1. Daily price fluctuations vs (detrended) trading volume in NASDAQ between 1951 and 2010. The horizontal lines define the limits of four standard deviations and the vertical line locates 3/2 of average trading volume.

This scenario was later set in a quantitative framework in Refs. [35] using a stochastic volatility approach. Other modelling proposals assumed heteroskedastic (G)ARCH regression [36], which add trading volume in the regression formula of return calculations finding evidence over the fact that volatility and trading volume would have the same underlying [61].

In opposition to the MDH, there is the scenario of Sequential Arrival of Information Hypothesis (SAIH) introduced by Copeland [37] and extended in Refs. [38]. The SAIH conjectures that information arrives to agents at different times so that the final steady state in the market is led by a sequence of local steady states. Thence, they would lead different empirical features, namely for MDH there is contemporaneous correlation between volatility and volume whereas for SIAH correlation is lagged.

While recent studies on order book dynamics suggest that the adage is closer to an urban legend than an actual empirical fact because large price fluctuations are caused by fluctuations in liquidity [6], the reality is that when we look at a scatter plot of the daily log-price fluctuations, $r_{1\mathrm{d}} \equiv \ln S_{\mathrm{d}} - \ln S_{\mathrm{d}-1}$ (d stands for day), versus the respective trading volume, v_{d}, we understand that 76% of the daily price fluctuations with magnitude larger than 4 standard deviations are related to trading volumes which are in excess of 1.5 the daily average value (see Fig. 1); this props up the likelihood of brokers' saying. Nevertheless, in the advent big data and the access to high-frequency and ultra-high-frequency datasets many properties have been disputed, namely the relevance of intra-day properties [40].

In this paper, I review some of the results obtained in recent years over statistical and dynamical features of high-frequency trading volume and its relation to other financial quantities like the price fluctuations and the volatility. Some remarks on perspectives over studies on trading volume are presented in the end of the paper.

2. Stylised facts about trading volume

Like the prices, which are assumed to grow geometrically because of the inflation and effective economic growth, trading volume has followed in the long term the same sort of evolution, but with a significantly larger rate (see left panel in Fig. 2). For instance, the trading volume of the companies composing the SP500 index increased between 1951 and 2010 at an yearly average rate of 14% whereas the value of that index soared at 3%. Leaving aside that non-stationarity, the trading volume exhibits seasonalities as well, with the most evident of them related to the dynamics of trading along the business day [7], *i.e.*, by the actions of chartists — who are not inclined to be exposed to overnight variations — there is a significantly higher level of trading in the beginning and in the end of the sessions, a feature that is known as the ∪-shape of trading in stock market dynamics first reported in for absolute price fluctuations [8] and trading volume [9] for hourly data. In Fig. 2, I show the typical intra-day profile of the volume of a US blue chip equity, namely Alcoa (AA), that I will use to illustrate several of the matters in this article. [4] In the same figure note the emergence of two outliers at the time of European markets closing. This gives a clear indication about the existence of correlations between indices and hence among equities; many companies are listed in different markets — namely London and New York — and so latter takes into consideration the final part the session of the former in order to subdue arbitrage strategies and at the same time to let investors adjust their positions (and portfolio composition) taking into

[4] The reason for using this company are the following: it is a large market capitalisation corporation (around US$ 28bi), close to the average market capitalisation of the companies listed in the larger SP500 index and more importantly its quantitative behaviour is aligned with mean results. Furthermore, being a mining business Alcoa is able to act as a proxy for the dynamics of commodities as well, providing a inkling of the so-called behavioural universality of financial markets.

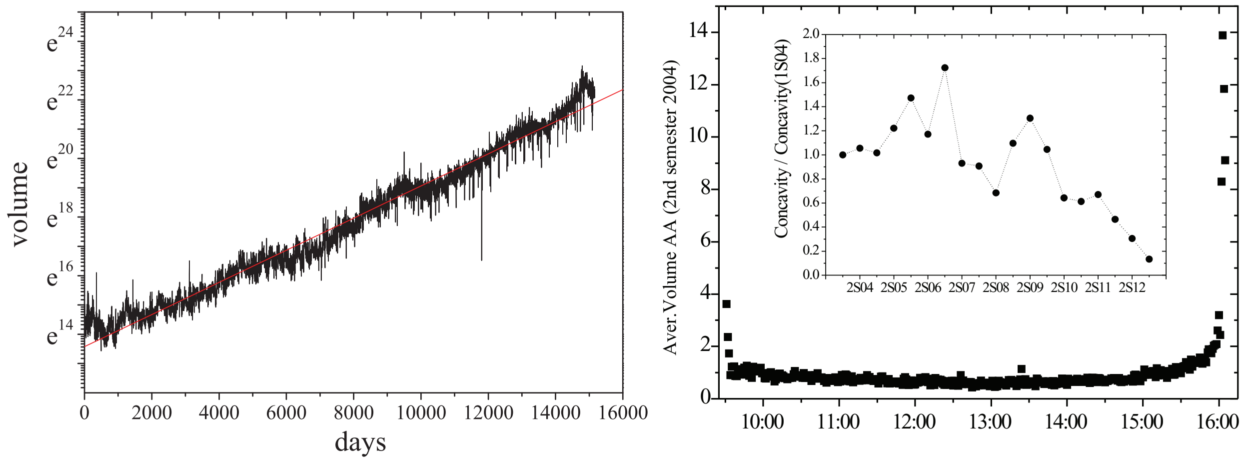

Fig. 2. Left panel: Geometric walk of the trading volume vs time (days since 3rd January 1951) The red line depicts a growth rate of 14% per year with a linear regression coefficient $R = 0.99$. Right panel: Average 1-min trading volume for AA in the 2nd semester 2004 vs time within a trading session. The inset is represents the evolution of the concavity of the ∪-shape by semester with the 1st semester of 2004 taken as default (. courtesy of M.B. Graczyk.) [10]. The last ten years present a fade out of that concavity.(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

consideration the sentiment of the world markets. It should be noted that the shape of the ∪-shape is nonstationary in itself and depends on the level of trading (or information in the market) though. In the inset in the right panel of Fig. 2, I show how the average concavity of the intraday trading profile has been changing in time. One can observe that the curves getting less concave and ever flatter; in other words, they are no longer ∪-shape to become ⊔-shape [10].

Most of the theoretical physics arsenal becomes easier to employ when one assumes stationarity in the system; thus, when treating inter/intraday trading volume data it is convenient to remove such nonstationarities, which can be added back after understanding the underlying stochastic mechanisms *a posteriori*. Two simple ways of separating them off are as follows: the inflationary/growth trend can be removed by computing the trading volume growth rate, r_v, for the dataset starting at t_0 and then:[5]

$$v(t) \to v(t)\, \exp\left[-r_v(t - t_0)\right]. \tag{1}$$

In respect of the intraday profile, it can be removed by performing an average of the volume with the same time stamp, s, over all days, d,[6]

$$v(d, s) \to \frac{v(d, s)}{\langle v(s)\rangle_d}. \tag{2}$$

2.1. Distribution

To the best of our knowledge the first analysis of the distribution of high-frequency (less than 1 h sampling rate) trading volume was presented in Ref. [15] for which they have found a consistent power-law decay of the probability density distribution (PDF), $p(v) \sim v^{-\zeta}$, with $\zeta = 2.7 \pm 0.1$, that is consistent with a Lévy regime, *i.e.*, it does not yield a finite standard deviation.[7] Similar exponents were found for other liquid markets such as the London Stock Exchange and Paris Bourse by the same group in [16]. The fat tail of the distribution was long known in finance, but it was assumed as a fingerprint of a log-Normal distribution. The reason for this assertion was based on the cascade mechanisms leading to the log-Normal: the logarithm of the trades was a random variable with finite variance so that the aggregated trading volume for a given period of time would converge to the log-Normal distribution in accordance with the central limit theorem. However, the log-Normal form was not ideal for describing the entire curve, especially typical (within the same order of magnitude of the average) and small values. A different PDF was then proposed in [17],

$$\begin{aligned} p(v) &= \frac{1}{Z}\left(\frac{v}{\theta}\right)^{\alpha}\left[1 + (q-1)\frac{v}{\theta}\right]^{\frac{1}{1-q}} \\ &= \frac{1}{Z}\left(\frac{v}{\theta}\right)^{\alpha} \exp_q\left[-\frac{v}{\theta}\right]. \end{aligned} \tag{3}$$

This corresponds to a generalised form of the Gamma distribution — which is reobtained in the limit $q \to 1$ — defined within Tsallis nonadditive entropy and matches the F-distribution when α and q are related to integer numbers. Considering the top ten liquid stocks traded in NASDAQ they have found for 1-min$\{q = 1.19, \alpha = 0.93, \theta = 0.23\}$ and for 2-min$\{q = 1.16, \alpha = 1.36, \theta = 0.2\}$, which lead to an asymptotic decay of $p(v)$ with an exponent larger than three, beyond Lévy limit. This same distribution was tested to

[5] More sophisticated approaches can be the application of moving average techniques, e.g., AR(I)MA models for cases where the simple linear regression does not present good enough linear correlation coefficients.

[6] Another possibilities are to power the trading volume [11] within the spirit of the analysis carried out in Refs. [12,13] that aimed at weighing large and small values of v differently, to compute averages using escort distributions [14] or other alternative statistical measures.

[7] In order to cope with this fact it was assumed the existence of truncated Lévy flights [1].

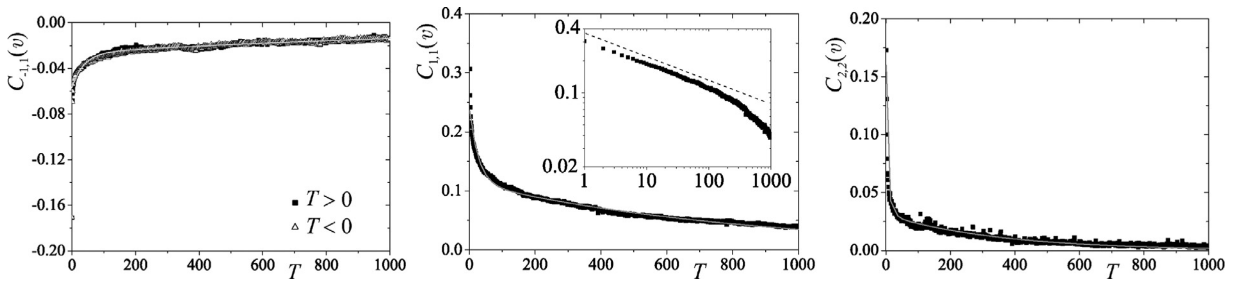

Fig. 3. $C_{\alpha,\beta}(v)$ *vs T*. On each panel black symbols are the values obtained from time series and the grey line represents the numerical fit of $C_{\alpha,\beta}(v)$ for a double exponential. In panel for $C_{-1,1}(v)$ the curves for $T > 0$ and $T < 0$ concur, which goes along the lines of time symmetry. The inset on $C_{1,1}(v)$ panel is a log–log representation of the main panel. As it is visible the correlation function does not present power-law behaviour, nay a power law decay holds for less than a decade in the ordinate. The same happens for all the other values of (α, β) studied. Reproduced from [13].

describe the trading volume of the blue chip equities of the Dow Jones Industrial Average [19], in (quite) less liquid stocks like those composing the Brazilian IBOVESPA index [20,21] as well as maturing markets such as the Korean [22] and the Chinese [23] including the size of the orders in the book [24]. In all these cases — and systematically using several different algorithms and tests — it was found an exponent larger than three, that is consistent with a finite variance without the need for imposing a cut-off in the PDF. In the case of Ref. [23], it was also found out that around the q-Gamma distribution there is a preference for certain numbers quantities namely trading volumes starting in 1 and 5, a phenomenon known as order size clustering.

2.2. Correlation and dependence

Systems with short memory, or no memory at all, are certainly simpler than other for which (some) information prevails over significant periods of time. Memory, *i.e.*, propagation of information in time, basically takes place due to linear and non-linear correlations. The former is presented in systems with simple mechanisms of memory like Markovian processes and it can be computed resorting to Pearson's correlation coefficient (or function),

$$C_{xy}(t,\tau) = \frac{\langle x(t+\tau)\, y(t)\rangle - \langle x(t+\tau)\rangle \langle y(t)\rangle}{\sigma_{x(t+\tau)}\sigma_{y(t)}}, \tag{4}$$

where $\sigma_{x(t)} \equiv \sqrt{\left\langle x(t)^2\right\rangle - \langle x(t)\rangle^2}$.

Usually a series is assumed (or made) stationary and therefore the dependence on t does not exist. Apart from the direct application of Eq. (4), there are other forms to compute the linear correlation such as the power spectrum, variograms, or analysing the "diffusive" character of the time series using Detrended Fluctuation Analysis (DFA) [18], when $y = x$, or its cross version otherwise. For DFA, the detrended fluctuation function is expected to behave as, $F(\tau) \sim \tau^{\hat{\alpha}}$, and in that case $C(\tau) \sim \tau^{2(\hat{\alpha}-1)}$. In Ref. [15], it was presented a DFA analysis of the trading volume of the 1000 largest stocks of NYSE for which it was found an average value $\hat{\alpha} = 0.83 \pm 0.02$. In another paper [19], it was used a correlation function approach to the trading volume of the stocks of the DJIA and it was observed that a power-law decay description was constrained to short regime. Therein, it was shown that $C(\tau)$ is best described by the superposition of two exponentials,

$$C_v(\tau) = a \exp\left[-\frac{\tau}{\tau_1}\right] + b \exp\left[-\frac{\tau}{\tau_2}\right], \tag{5}$$

with $b = 1 - a$. From that analysis, it was verified the existence of two clear different scales with a ratio $\tau_2/\tau_1 = 28 \pm 11$. In Ref. [20], it was verified for the IBOVESPA that the trading volume also presents fast (exponential) decay. Similar results were verified for Chinese markets as well [25].

Eq. (4) treats the quantities under analysis equally. Nevertheless, it is well known that complex systems tend to be hierarchical; hence, it is likely that in the case of financial quantities, namely the trading volume, large volumes will correlate differently with large and small volumes. With that suspicion in mind, the correlation function of trading volume was slightly modified in Ref. [13] so that it was considered $x(t) = v(t)^\alpha$ and $y(t) = v(t)^\beta$ (α and β reals). Those correlation functions were still found to be well described by Eq. (5). Recalling that the values of the trading volume are detrended and normalised, this means that for $\alpha(\beta) < 1$—especially when they are negative — one gives extra weight to small values whereas when $\alpha(\beta) > 1$ one emphasises large values. The analysis carried out in [13] pointed out that small values of trading volume are consistently anti-correlated with frequent and large values and that frequent and large values are positively correlated between them, although the scales of correlation are smaller in the case of large volumes. Moreover, the values for a and b (in absolute value) are smaller when at least one of the exponents is equal to 1. (Fig. 3)

As mentioned in this subsection, correlation splits into linear effects[8] and non-linear effects, where the latter cannot be measured by means of the correlation function. In that case the right way to carry that out is by using entropic forms, namely the application of the relative entropy[9],

$$K\left(p, p'\right) = -\sum_j p_j \ln\left(\frac{p'_j}{p_j}\right), \tag{6}$$

which provides the mean change of information related to any two probability distributions, p and p'. In the case of

[8] Or effects that can be represented in a linear case.

[9] Also known as Kullback–Leibler divergence, information divergence, information gain.

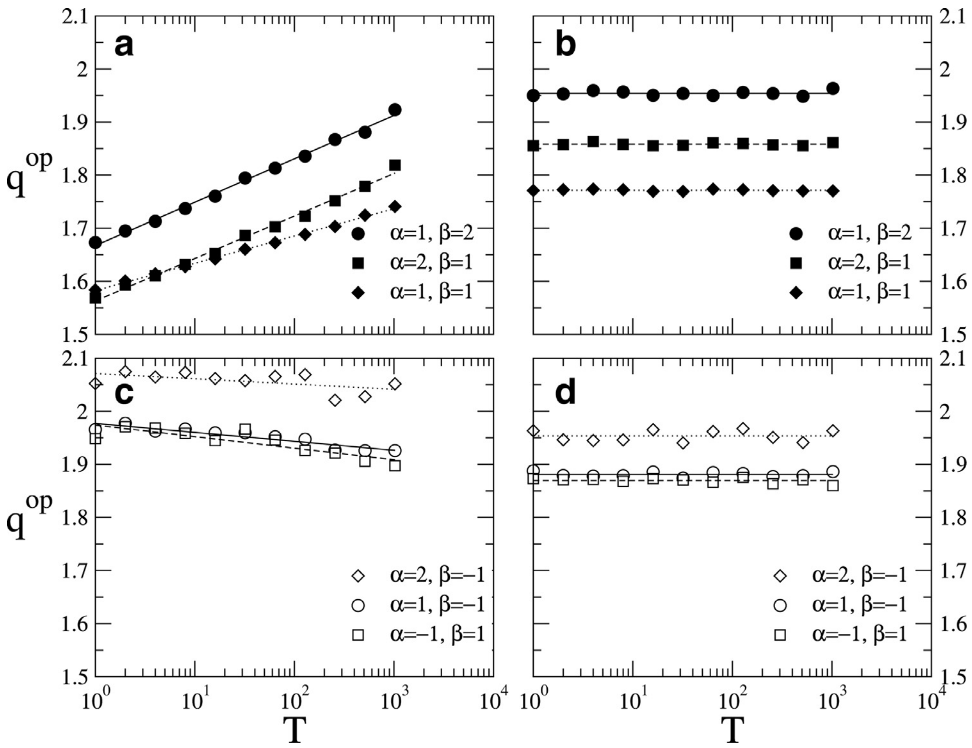

Fig. 4. Optimal index q^{op} versus lag T. Panel a: Lines correspond to fitting function $q^{op} = A + B \log T$, where (A, B) is (1.667 ± 0.003, 0.035 ± 0.001) (solid line), (1.563 ± 0.004, 0.035 ± 0.001) (dashed line) and (1.583 ± 0.001, 0.0223 ± 0.0003) (dotted line) for (α, β)= (2,1), (1, 2) and (1, 1) respectively. Panels b: Same as in a on the shuffled version of the time series. Constant values are: 1.954 ± 0.002 (solid line), 1.858 ± 0.001 (dashed line) and 1.7713 ± 0.0008 (dotted line). Panel c: Logarithmic fitting as in a: (A, B) is (1.977 ± 0.004, −0.007 ± 0.001) (solid line), (1.974 ± 0.008, −0.009 ± 0.002) (dashed line), (2.07 ± 0.01, −0.004 ± 0.002) (dotted line) for (α, β)= (2,−1), (1, −1) and (−1, 1) respectively. Panel d: Same as in c on the shuffled version of the time series. Constant values are: 1.881 ± 0.002 (solid line), 1.869 ± 0.002 (dashed line) and 1.953 ± 0.005 (dotted line). Reproduced from Ref. [13].

multidimensionality of j, particularly $j \to (x, y)$, the most interesting case is the analysis of the distance between distribution $p(x, y)$ and the product of marginal distributions representing independence $p'(x, y) = p_1(x)\, p_2(y)$, which has been applied in quantitative finance and econophysics [26]. In spite of the several results and applications, relative entropy $K(p, p')$ has some problems, namely in the classification of the degree of dependence between variables among some other issues. Putting it differently, it does present a means linking the quantitative result given by Eq. (6) with a qualitative assessment of strong/weak non-linearities. These issues are largely sorted out by considering a generalised version of the relative entropy that was introduced within the context of Tsallis entropy. In this generalisation [27], the usual logarithm must be replaced by the q-logarithm, $\ln_q(x) \equiv \left(x^{1-q} - 1\right)/(1-q)$ and Eq. (6) is recovered in the limit $q \to 1$. For $q > 0$, there exist well defined minimum and maximum values of $K_q(p, p')$ corresponding to minimum and maximum dependence degrees between random variables (x and y). This allows us to define a criterion for statistical testing [26] through the normalised quantity $R_q \equiv K_q\left(p, p'\right)/K_q^{\max}\left(p, p'\right)$, which is limited between 0 and 1 corresponding to inexisting and full dependence between x and y, respectively. Given the two quantities x and y, R_q can be simply calculated as a function of q yielding a continuous and monotonically increasing function the inflexion point of which defines the value of q for which R_q is most sensible in detecting changes in the dependence between the two variables. For this reason, the value is called optimal value, q^{op}.

In quantitative finance, this form was first applied to explain the slower than expected convergence to the Gaussian shown by uncorrelated price fluctuations [28]. Regarding the trading volume, the generalised mutual information R was first employed in [19] with x being the time series and y the same time series with a lag in time, T. In that work, it was found that the non-linearities in trading volume decay in a very slow way, *i.e.*, the index q^{op} goes logarithmically with the lag. In a subsequent paper [13] the nonlinearities were tested in accordance with the study of the generalised correlation function, *i.e.*, $x_\alpha = v(t)^\alpha$ and $y_\beta = v(t+T)^\beta$. Again a logarithmic behaviour of q^{op} as a function of T was obtained for different values of α and β, but with different parameters. The results reproduced in Fig. (4) show that, in this case, q^{op} diminishes as a function of the lag, but the rate of change is not as high as in the $\alpha > 0$, $\beta > 0$ case (see the respective caption). Note that this result occurs for the same exponents where anti-correlated behaviour is found indicating that the anti-dependence is associated with a negative slope in the logarithmic behaviour of q^{op} with the lag. In both cases, positive and negative dependent converge to independence that is indicated by the computation of q^{op} for shuffled series for which causality is totally destroyed.

2.3. Multiscaling

I now introduce the last of our three (quantitative) cornerstones of a complexity assessment. As early as the introduction of the concept of fractal by B. Mandelbrot, it was verified the existence of scaling properties, particularly multiscaling, in the form of self-affinity, for financial quantities such as the price fluctuations and the volatility [1,2,29]. This multiscaling represents a composition of several sub-sets, each one with a certain local exponent, h, and all supported onto a main structure, which is self-affine as well. Although other methods can be used to assess multiscaling (or multifractality), the most applied of them is the multi-fractal detrended fluctuation analysis [30] and its variants. In MF-DFA, the fluctuation function is generalised to take into account different q moments scale with differently. That leads to a functional dependence $F_q(\tau) \sim \tau^{h(q)}$ which can be related to the so-called singularity spectrum, $f(\alpha)$ by means of the relations,

$$f(\alpha) = q[\alpha - h(q)] + 1, \quad \alpha = h(q) + q\frac{dh(q)}{dq}. \tag{7}$$

Along these lines, existence of multiscaling is characterised by the determination of a set of values $f(\alpha) \neq 0$ between $\alpha_{\min}$ and $\alpha_{\max}$, or equivalently by the difference $\Delta h \equiv h(q_{\min}) - h(q_{\max})$. The multifractality that is measured in a time series is then assumed as the superposition of different independent features: linear (lin) and non-linear (nlin) correlations, (asymptotic) scale-invariant behaviour of the PDF [10], $\Delta h = \Delta h_{\mathrm{lin}} + \Delta h_{\mathrm{nlin}} + \Delta h_{\mathrm{PDF}}$. The results for all the cases are qualitatively similar independently of the market analysed.

For trading volume the first studies on high-frequency data were reported in [22,31,32]. Considering the results in Ref. [31], it was verified the existence of a relevant level of multifractality with $\alpha_{\min} = 0.32 \pm 0.04$ and $\alpha_{\max} = 1.09 \pm 0.04$. Regarding its components, it was verified that the largest stake of trading volume multifractality stems from the power-law behaviour of its PDF (66%), indicating a strong bifractality of trading volume [30] and a residual contribution of linear dependencies (4%). In addition, as visible in Fig. 5, there is a strong asymmetry of the curve $f(\alpha)$ indicating that large and small fluctuations appear due to different dynamical mechanisms.

3. Dynamical approaches

3.1. Stochastic differential approach

The description of the properties I have presented, namely the correlations, based on dynamical approaches has been done since the 1970s [4,15,33,35,39]. After the introduction of the all-inclusive distribution Eq. (3) a dynamical scenario was naturally welcome. Although the story actually runs backwards we verify that the correlation function, which is described by Eq. (5), suggests the existence of different dynamical regimes composing the rules governing the evolution of trading volume. Furthermore,

[10] There can be also spurious multifractality due to the finiteness of the series.

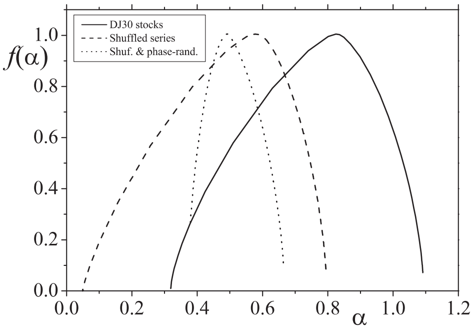

Fig. 5. Multi-fractal spectra $f(\alpha)$ *vs.*α. The "original" and shuffled time series (linear correlations destroyed) present a strong multi-fractal character whereas the shuffled plus phase randomised (multifractality from the distribution and nonlinearities destroyed) time series presents a narrow width in α. Reproduced from Ref. [31].

the multiscaling analysis showed that the trading volume is mainly a bi-fractal due to is asymptotic power-law behaviour. Last but not least, trading volume is formally non-stationary, particularly due to variations in the number of active agents in the market [41]. That being so, let us follow the reasoning introduced in Ref. [42] which is inspired by the concept of superstatistics that emergence within the framework of non-equilibrium statistical mechanics [43]. Locally, it was assumed that the trading volume follows a dynamics given by the Feller process,

$$dv = -\gamma(v - \omega)\, dt + \phi\sqrt{v} dW_t, \tag{8}$$

where W_t is a standard Wiener process, and $\phi = \sqrt{\frac{2}{\varphi}\omega\gamma}$ which is associated to a Gamma distribution,

$$p(v;\omega) = \frac{\varphi^{\varphi}}{\omega\,\Gamma[\varphi]}\left(\frac{v}{\omega}\right)^{\varphi-1} \exp\left[-\frac{\varphi}{\omega}v\right], \tag{9}$$

with average value $\langle v\rangle = \omega$. Let now us take into account the changes in the number of active agents with time which implies a change in ω that happens in a long scale (in comparison with the local scale of relaxation). In other words, ω stops being a parameter and turns into a variable, but since it evolves in a slowly way it can be considered locally constant. This concurs with empirical observations and can be connected to the two significantly different scales of the correlation function (see page 4): the first, local, of the order of $\gamma^{-1} \simeq \tau_1$, which corresponds to the accommodation of v to a new local mean value ω, and the second scale, equal to $T \simeq \tau_2$, representing the evolution in the number of active agents. Moreover, if one assumes that in the long term the local average trading volume follows an inverse gamma distribution,

$$f'(\omega) = \left(\frac{\varphi}{\lambda}\right)^{\delta} \frac{\omega^{-\delta-1}}{\Gamma[\delta]} \exp\left[-\frac{\varphi}{\omega\,\lambda}\right], \tag{10}$$

then the long term distribution of v is given by $p(v) = \int p(v;\omega)\, f(\omega)\, d\omega$ that is equal to Eq. (3) with,

$$\lambda = \frac{q-1}{\theta}, \quad \varphi = (q-1)^{-1} - \delta, \quad \alpha = \varphi - 1.$$

The probability distribution $f(\omega)$ using average daily values was verified for the IBOVESPA index [20]. In addition, if one considers a linear relation between ω and the number of active agents within a certain time window, N, then we find that its PDF goes as, $p(N) \sim N^{-\delta-1}$, with $\delta = 3.33 \pm 0.15$ that is consistent with empirical values [44]. Despite the fact that the present account showed effective for IBOVESPA trading volume in closer scrutiny and in line with the exponent of the distribution of active traders, an analysis of the Kramers–Moyal moments of DJIA trading volume evinced a parabolic profile for the second order moment that is different from the linear form implicit in Eq. (8). To accommodate this property the scenario was changed in a subsequent work [19]. Explicitly, it was assumed a local dynamics given by,

$$dv = -\gamma\left(v - \frac{\omega}{\delta}\right) dt + \sqrt{2\frac{\gamma}{\delta} v dW_t}, \tag{11}$$

and,

$$f(\omega) = \frac{\varphi^{\varphi}}{\lambda\Gamma[\varphi]}\left(\frac{\omega}{\lambda}\right)^{\varphi-1} \exp\left[-\frac{\omega}{\lambda}\right], \tag{12}$$

which leads to an inverse $q-$Gamma distribution,

$$P(v) = \frac{1}{Z}\left(\frac{v}{\theta}\right)^{-\delta-2}\left[1 + (q-1)\frac{\theta}{v}\right]^{\frac{1}{1-q}}, \tag{13}$$

where $\lambda = \theta(q-1)$, $\delta = \frac{1}{q-1} - \varphi - 1$. It should be noted that redefining the parameters there is a correspondence between Eqs. (3) and (13), so that an inverse q-Gamma distribution can be recast into a q-Gamma distribution and vice-versa. A first analysis on the local average value of traded volume [45] (slide 23) suggested a Gamma distribution.

The superstatistical scenario can be further appraised taking into consideration time dependent quantities, such as the return intervals a quantity that was also analysed for the volatility and volume fluctuations [46]. In that case, it was found that the time taken by such quantities to get back to some value $v*$ follows a stretched exponential. Despite a fit to the return intervals distribution of trading volume was not made I show in Fig. 7 a comparison between empirical results and the results obtained from the stochastic differential equation approach where is visible a long term agreement. This is noteworthy taking into account a simplicity of the model.

3.1.1. Statistical approach

As I have mentioned in the previous subsection, the hypothesis of a two-scale dynamics is able to reproduce a series of empirical properties high-frequency trading volume. However, there was still a point urging further specification: the dynamics of changes of the local regime as it is unlikely that the variations of the local average volume occur at a fixed scale; one expects that the larger scale presents fluctuations.

That assumption was probed in Ref. [47] using the Kolmogorov–Smirnov algorithm of segmentation of nonstationary time series into local steady states intervals introduced by the same authors in [48]. It was then verified that the dynamics of trading volume can actually be considered as a composition of contiguous segments of unequal duration, ℓ, with different average values. The distribution of the duration of the stationary patches is well described by an exponential distribution with a typical scale of $\lambda = 116 \pm 12$ min, *i.e.*, around two hours. With that segmentation in hand, the ansatz related to the local distribution of trading volume and the distribution of local mean were also surveyed. Applying statistical significance testing to several standard distributions, it was verified that in 81% of the segments the Gamma distribution was proven statistical significant whereas for the inverse Gamma distribution that figure was only 41%. That result underpins the validity of the proposal conveyed in [42] and suggests that the Kramers–Moyal analysis requires supplemental investigation in agreement with sample rating results [49]. Moreover, looking at the distribution of the parameter $\delta = \omega/\varphi$ in Eq. (10), we verify that using a Kolmogorov–Smrinov (times square degree of freedom) test — although not the best result — is very close to the best result, which is given by the Gamma distribution. Specifically, it was obtained 1.39 ± 0.47 for the former and 1.05 ± 0.58 for the latter. This closeness between the results of the goodness of fitting partially support the q-inverse Gamma approach [19], especially the distribution $p(\omega)$ in Fig. 6 (lower panel).

The analysis in [48] shed further light on the nonstationary properties of trading volume:

- the longer the patch, the smaller the local average value, μ_ℓ. This can be understood as a proxy for the agitation in the market. When the market is in a higher state of nervousness, it is clear that it is more nonstationary and therefore the segments of local stationary should be shorter. Alternatively, periods considered quiet in the market show small average values and little fluctuation in the number of active traders (and hence in the activity). Coherently, the stationarity will be longer and have smaller average values. These segmentation result and the respective analysis agrees with the empirical finding that the correlation of returns is higher on low-volume days than on high-volume days [54];
- shorter periods of local stationarity are concentrated in the last part of the trading session. That can be related to the natural agitation set by chartists that close their positions around the session in order to finish the day with a 100% cash portfolio;
- the correlation function of the local average volume as a function of the number of lags exhibits slow decay as it takes 18 segments (5.3 business days);
- there is a two power-law relation between the local average trading volume and the local variance. The threshold in terms of local trading volume is 1.23 ± 0.88.

The quality of the segmentation approach in the description of the long-term distribution of trading volume is presented in Fig. 7 of Ref. [48].

4. Relation with other quantities

4.1. Price fluctuations

As I mentioned in the Introduction, the interest in studying trading volume lies in the fact that they are considered as a quantity that contains some form of information about price movements that sustains the previously

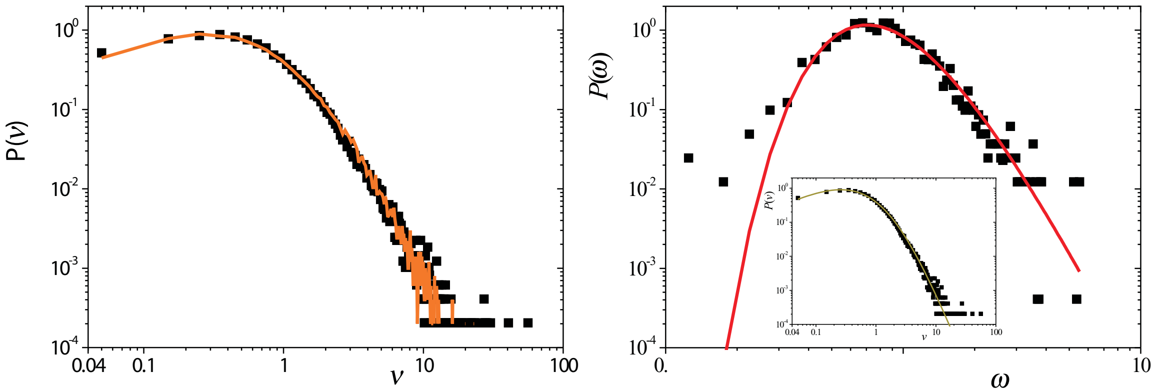

Fig. 6. Left panel: The points represent the empirical distribution function of the trading volume of Alcoa in the 2nd semester of 2004 and the line the differential stochastic dynamics given by the approach introduced in Ref. [42]. Right panel: Numerical adjustment of the local trading volume for a fixed time patch $T = 180$ min described by a Gamma distribution with $\varphi = 5.02$ and $\lambda = 1.18$. The inset shows that by using these parameters in the proposal Ref. [19] a good description of the long term distribution of trading volume is obtained as well. First presented in [45].

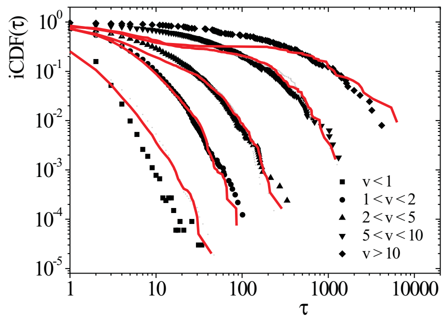

Fig. 7. The points represent the (inverse) cumulative distribution function of the return interval of the trading volume of Alcoa in the 2nd semester of 2004 for the values presented in the legend and the red lines are the results obtained using the differential stochastic model using the parameters of the previous plots. First presented in [45].(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

mentioned stock brokers' rule of thumb over relation between the two quantities.

How can we assess that empirical knowledge and put it in quantitative context for high-frequency data? In the last 15 years a lot of work has been dedicated to trying to understand the microstructure of order books, including the properties of transaction size which corresponds to the trading volume at the most elemental natural tick scale. Herein, I report results for the 1-min coarse grained scale — high-frequency still — and address the reader to Ref. [50] for a review about microstructure features in financial markets.

Assuming that a direct relation between trading volume and the returns is true to life, one should be able to express it quantitatively, *i.e.*, defining a trading volume impact as shown in Fig. 8. It is verifiable that there are different relations between the trading volume for (average) negative and positive returns, $\mathcal{I}_{(\pm)}(v) \equiv \langle r|v\rangle$, and that both are well described to good extent either by power laws (full lines) or logarithmic laws (dashed lines) of v. These two forms were proposed to describe the market impact of trades for Paris and London stock markets, respectively, and follow a prior first analysis using data from NYSE where positive and negative returns were considered together [52].[11] The exponents obtained for the DJIA data are significantly small than 1/2 incompatible with the theory introduced in Ref. [44] (I will be back to this point in the end of this article). In addition, the impact is higher for negative returns which reflects natural risk aversion factors. However, these laws clear break for values greater than trading volumes larger than ten times the average. A possible explanation for this effect can lie in the set of relations between all types of orders that actually represent all the factors that influence the market dynamics [53]. For instance, a large trading volume can be associated with a small price fluctuation as the result of a strong and balanced clash between agents (*e.g.*, fundamentalists and chartists) over the asset.

Following the claim that prices are moved by the volume one could sketch a relation between the probability density function of price fluctuations, $p_{\rm ret}(r)$, and the probability of the trading volume, $p_{\rm vol}(v)$. Using simple Bayesian statistics we have,

$$p_{\rm ret}(r) = \int p(r|v)\, p_{\rm vol}(v)\, dv. \tag{14}$$

So, what can we say about $p(r|v)$? I have already made reference to the fact that although price changes are associated with trading, the reciprocal is not necessarily true: it is possible to have a series of trades whose net result in a price variation equal to zero; that is to say, for a given volume v there is a probability, $h'(v) = 1 - h(v)$, that such volume yields $r = 0$. At the tick size, $h(v)$ is a power-law with an exponent equal to 0.25 whereas for high-frequency returns $h(v)$ evolves more slowly than that, as depicted in Fig. 9. Subsequently, in Ref. [47], $h(v)$ was split into trading volume yielding positive (+) and negative (−) price

[11] The logarithmic formula can be related to the initial formulation by Kyle [51] that the *price differences* are linear with the trading volume. In spite of this fact the power law description is favoured because it is more easier to handle in analytical models.

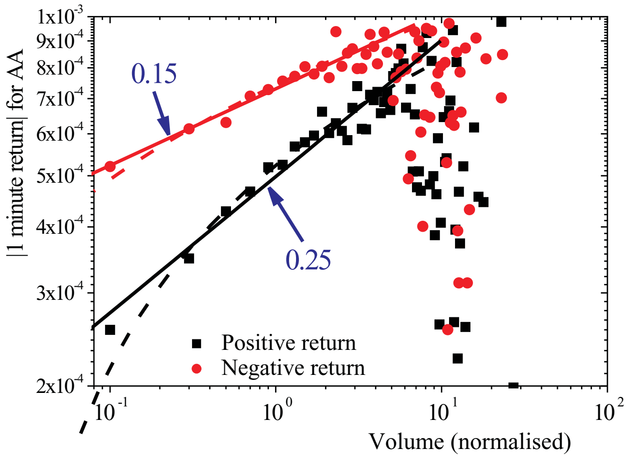

Fig. 8. Smoothed price fluctuation vs trading volume for Alcoa transactions in NYSE 2nd semester 2004 in log–log scale. The full lines are power-law fits and the dashed lines are logarithmic fits. First presented in [45].

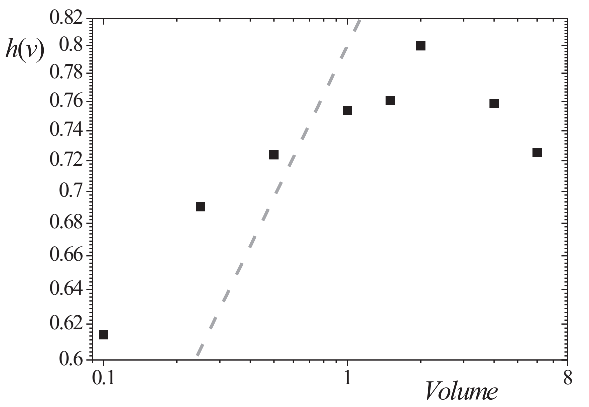

Fig. 9. Probability that a volume v of Alcoa (in average trading volume units) traded in 1-min yields a non-zero price fluctuation vs trading volume in log–log scale. The dashed grey line corresponds to the power-law with exponent equal to 0.25 that was measured for order size (trading volume at one tick scale). First presented in [45].

fluctuations which are well described by,

$$h^{(\pm)}(v) = H_{(\pm)} \tanh\left(\varpi_{(\pm)}\, v^{\zeta_{(\pm)}}\right), \tag{15}$$

where the exponents are $\zeta_{(+)} = 0.25 \pm 0.05$ and $\zeta_{(-)} = 0.3 \pm 0.1$ for positive and negative returns, respectively.

For non-zero returns, the conditioned probability $p(r|v)$ is equal to the conditioned probability distribution $f(r|v)$ time the probability of having a volume yielding a non-zero result. For single trades, it was found that $f(r|v)$ was a function of r alone [6]-a). Since $f(r|v) = f(r)$ is power-law decaying, the stylised fact of asymptotic scale-free distributions of $p(r)$ is assured.

For 1-min data (see Fig. 10) there is a clear dependence on the trading volume. In addition, $f(r|v)$ is closer to an exponential than to an asymptotic power-law. That being so, the two properties are actually related; it is possible to obtain a power-law distribution from a set of exponentials with different characteristic scales. Looking at Fig. 10 we understand that $f(r|v) \sim \exp[-\Omega(v)\, r]$, where $\Omega(v)$ is a decreasing function of v.

We can carry on and try to understand how the relation between average price fluctuations evolve as we increase the time scale of aggregation. On the one hand, the distribution of trading volume approaches an exponential decay [17]. On the other hand, the relation between returns and trading volume, the exponent tends to become smaller, around 0.15 (see Fig. 11). It is worth recalling that exponents close to zero are often a suggestion of logarithmic relation and ultimately agree with Kyle's approach.

In order to test the relation between volume and returns we can go back to the nonstationarty behaviour of trading volume and using the results of this subsection determine the distribution $p(r)$. Explicitly, assuming that the returns are given by,

$$r = \mathcal{I}(v) + \xi_{(\pm)} \tag{16}$$

where $\xi_{(\pm)}$ is a Gaussian error, $\mathcal{G}(0, \sigma_{\xi_{(\pm)}})$, for positive (p) and negative returns (n). Testing $\mathcal{I}(v)$ in the segments of local stationary it was not found significant dependence between the mean values of the parameters of the $\mathcal{I}(v)$ and the length of the stationary patch; however, it was observed that the residuals were higher for small interval though. Plugging all the relations into a single equation it yields for $r \neq 0$,

$$p_{(\pm)}(r) = \int \mathcal{G}\left(\mathcal{I}_{(\pm)}(v), \sigma_{\xi_{(\pm)}}\right) p(v)\, dv. \tag{17}$$

The results of the Eq. (17) showed that this approach is able to reproduce the central part of the distribution but it fails by a factor of ten in trying to explain large price fluctuations. In the approach introduced in [48] the fluctuations in the impact parameters were ignored; these can be understood as a representation of volatility. Along these lines, the results point that prices are to large extent moved by the trading volume but in order to have large price fluctuations one needs volatility.

Up to here I have discussed the relation between returns and trading volume at the same time stamp. Bearing in mind that trading volume is associated with

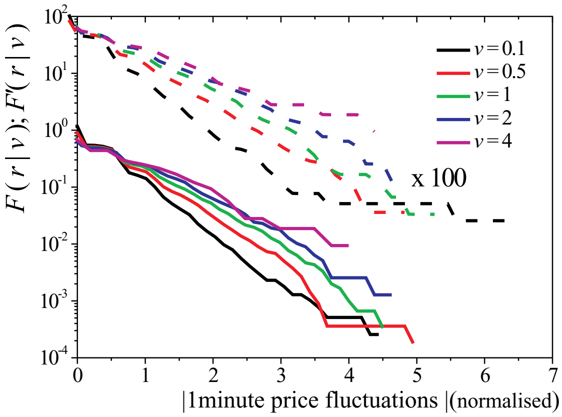

Fig. 10. Inverse cumulative conditioned probability of having a nonzero price fluctuation r given a trading volume v that vs price fluctuation r in a log-linear scale for Alcoa transactions in NYSE (2nd semester 2004). The full(dashed) lines are for positive(negative) returns. The negative curves $F'(r\,|\,v)$ are multiplied by a factor of 100 for sake of clarity. First presented in [45].

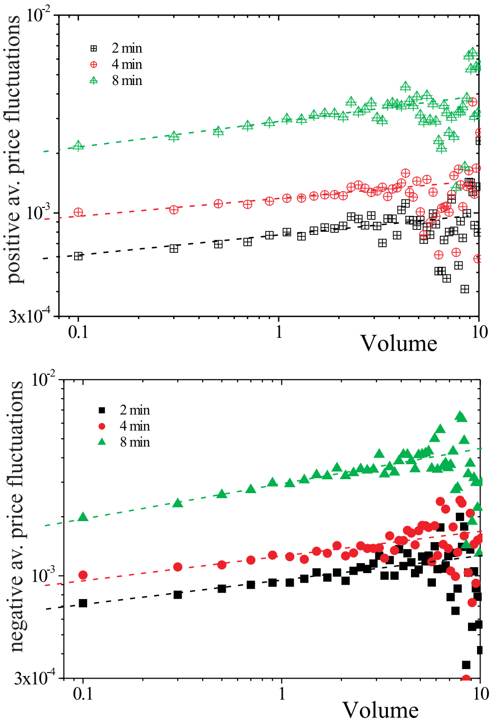

Fig. 11. Average price fluctuation vs trading volume for different time horizons as indicated in the legends. In both panels we verify a convergence to a small exponent of the order of 0.15 which is consistent to a logarithmic dependence as well. First presented in [45].

information flow it is natural to try to understand how price fluctuations and trading volumes at different times relate one another. As a matter of fact price fluctuations enhance speculation/risk and therefore influence future trading volume. For instance, from an analysis of a trading platform it was found that past returns — both market and portfolio — stoke the trading volume of individual agents [55]. Recently [56], it was verified that this correlation is nonstationary; it shown that for the SP500, the correlation has been decreasing and turned into an anti-correlation from 2000 onwards (2010) with a similar profile to the so-called leverage effect [57]; the correlation between trading volume and future returns is still at noise level though. It is important to note that the last decade of the data therein analysed (2000–2010), that index has shrunk 0.44% per year [58]. Therefore, it is very likely that such switching can be influenced by two factors [54], the economic cycles and undetrending of the data.

4.2. Volatility

In the set of the stylised facts of financial quantities, the long lasting autocorrelation of the volatility are among the most relevant because it allows the definition of profitable financial instruments and risk management. From our previous assertion Eq. (17) and subsequent analysis in page 9 we have noted that the emergence of tails in the distribution of the returns could be linked to the distribution of the volume, which in turn is related to evolution in the number of traders (certainly affected by the information flow in the market). If we recall that superstatistical scenario that a local scale-dependent distribution can yield a long-term fat tailed distribution for convenient distributions of the local variance and that the latter can be assumed as the (realised) volatility of the asset, the correlation between trading volume and volatility is an expected result (see Fig. 12). Assuming a simplistic approach we can think of the volatility and a measure of the state of anxiety in the market that is naturally ruled by the flow of information; *i.e.*, the same as the trading volume. This reasoning is the basis of Clark's Mixture of Distribution Hypothesis [33], that is intimately related to the rate of arrival of information in the market. For effects of modelling the hypothesis was later generalised to consider the direct influence of trading volume on the volatility and vice-versa [39]. Other modelling proposals were introduced in a (G)ARCH context with integer or fractional exponents[60]. In spite of the case that in these models the kernels of integration of volatility and volume [62,63] are close there is some controversy on the conclusions. On

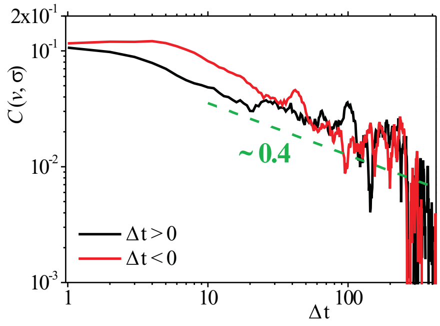

Fig. 12. Correlation function between the trading volume, v, and the volatility (absolute price fluctuations), σ, for 1-min sampling rate of Alcoa stocks traded at NYSE in the 2nd semester 2004. First presented in [45].

the one hand some authors suggest that the volume can be widely used as tool to improve volatility estimates [64], on the other hand some other authors [65] claim that the similarities only run in the short-term because trading volume presents little multiscaling features (it is less clustered).

For high-frequency data (5-min sampling rate) it is pointed that there is no same time correlation between trading volume and volatility [59], a property that would go against MDH. Moreover the existence of lead–lag relations and the tails in the volume-volatility correlation favour SIAH where the arrival new information to traders at different times creates some inertia in the market and a transient towards new steady states which would explain both. However, as I show in Fig. 12 are correlation at the same time are actually significant, although a maximum for 4-min (less than the sampling rate of [59]). The Mixture of Distributions Hypothesis was afterwards changed to take into consideration microstructural details of the market and centering its focus on the basic idea of a joint dependence of the trading volume and volatility on a latent event or information flow, first in Ref. [66] and subsequently in [67] result bolsters the Mixture of Distribution Hypothesis. That framework was more recently refined assuming jumps and leverage [68] with significant results. Complementary other line of action [69] consists of heeding that agents have heterogeneous degrees of sensitivity to information and each affecting in volatility and volume in different ways, a result that matches empirical findings [65].

4.3. Trading value

Alternatively to looking at financial markets observables a bivariate quantity, some authors have studied the cash flow generated by the trading of a given asset in a given period which was named *trading value* [74] or *price-volume* [75],

$$C_{\Delta}(t) \equiv S(t)\, V_{\Delta}(t), \tag{18}$$

where Δ represents the sample rating. The work by Eisler & Kértesz [74] provided systematic evidence about the finiteness of the second order moment and nonuniversality of the values describing the scaling of fluctuating moments, namely that these quantities dependent strongly on the size of the company that they defined by means of the average cash flow per minute, which they related to the market capitalisation as well. Explicitly, they found that a shorter scales the companies have got similar memory scaling described by a scaling exponent $\hat{\alpha}$ between 1 and 6/5.[12] However, one larger time scales are analysed it was verified a significant change in the qualitative behaviour; for very little market cash flow it can be found slightly antipersistent behaviour, $\hat{\alpha} \lesssim 1/2$ and for companies with high market cash flow the exponent denotes strong persistency near ballistic behaviour with $\hat{\alpha}$ around 0.9. This time and market cash flow dependence was also verified in multiscaling. Allowing for the fact that price fluctuations are uncorrelated this behaviour is in alignment with the two-scale dynamical approach. The increasing of the scaling exponents with the market cash flow is not that surprising, liquid companies have by definition large market depths are therefore they tend to have smoother dynamics.

Other results have analysed the cash flow with a 10-min sampling rate where they have assumed a log-Normal distribution of trading value [13] and introduce stochastic differential equations for describing the dynamics of the log-Normal parameters.

5. Remarks and perspectives

In this paper I have reviewed results on trading volume obtained within the context of applications of physics to financial analysis and modelling. Empirical analysis of high-frequency trading volume from different markets have shown features that are typically associated with complexity, namely fat tailed (asymptotic power-law) stationary distribution functions, long lasting autocorrelations & dependence and multiscaling. In addition, I have reviewed dynamical scenarii that associate the emergence of tails in the trading volume distribution with fluctuations in the number of active agents in the market that induce an evolution of the (local) average trading volume. This hypothesis of local steady state (or equilibrium in financial parlance) has statistical significance and it was found to have relation to state of agitation/nervousness in the market. At this point we can start assembling the elements I have given an account of as pieces of a jigsaw puzzle in order to have to insight into the big picture that we can represent with theses to questions: Does it really takes volume to make prices move? Which is the right description for the input of information in a market, the MDH or the SIAH? In respect of the first question, Section 4.1 indicates that there is an obvious relation between trading volume and price fluctuations, especially in what regards price fluctuations around its mode. However, this approach fails by a factor 10 for large price fluctuations. Therefore, the fluctuations of the parameters the impact function, $\mathcal{I}(v)$ that are dismissed might be relevant in proving a better relation between trading volume and price fluctuations. In [48] it

[12] I consider it similar because in order to have a change from the minimal Hurst exponent to the maximal one must increase the average cash flow by a factor one million.

[13] The authors do not test the q-Gamma distribution.

was shown that mean values are independent of the size of the segments, but no further analysis, namely its relation to volatility, imbalance or bid-ask gap; these issues are worth studying are can be matched with the analysis of mechanical and informational contributions to market impact analysed for order books [70].[14] I recall that after a systematic analysis of order books, it was assigned to volume imbalance collaborators [6] and other microstructure factors of order placement/cancellation [71] the origin of fat tails in price fluctuations. For the second question the blending scenario of the two main hypothesis can also be asserted since both pictures contain valid elements. Although physicist tend to be fond of simple arguments, the simple option for one or another approach reveals quite restrictive. Being a complex system, a financial market is characterised by relations between microscopical behaviour and macroscopical relations; this relations are not static they evolve in time. We are living an era where information spreads at the speed of light in optical fibre cables from news agencies to trading robots and brokers on the trading floor. These changes can be comprehended from the analysis of the autocorrelation of price fluctuation. For instance: a) data of US markets from the 1990s showed significant autocorrelations of the price fluctuations up to 20-min [72], but recent data from the 2000s shows that price fluctuations autocorrelations are already at noise level for lags of 1-min; b) the intraday $\cup$-shape profile is getting flatter; these suggest that steady states are achieved very swiftly, a trait that is closer to MDH than SIAH. On the other hand, the positive correlation between — which is a key element of the former — can be also included in the latter taking into considerations on the nature of agents (informed, uninformed and market makers) and the degree of sensitivity to price that changes it is influenced by the performance of each investor's portfolio. Notwithstanding, one can retrieve [56] the results and set the hypothesis that the from positive to negative return-volatility correlations is not only due but to a superdiffusion of information among agents because SIAH asserts that the simultaneous receiving of information by all the agents implies negative correlations whereas sequential arrival leads to positive correlations. Nevertheless, one can intuitively reckon that the sequential arrival of information can be adequate for cases of private information [73] or cases of markets (or assets) marked by short liquidity. On the other hand, we can compare the major points established by MDH with empirical results obtained from high-frequency data. Consequently, one can observe that in opposition to Clark [33] the distribution is not log-Normal but it is assumed to follow a q-Gamma distribution with statistical significance. Yet this is a minor point since there are MDH proposals assuming other distributions for trading volume, e.g., in Ref. [67] a Poisson distribution. Yet, the power-law relation between the size of the price fluctuations and the trading volume is consistent with Epps & Epps formulation [35]-a); the result that $f(r|v)$ is independent of v at the tick (order size) [35]-c), but for a given clock time there is a positive correlation between volatility and trading volume that can be appraised by the dependence of $f(r|v)$ on v as shown in Fig. 10; Eq. 14 agrees with the emergence of excess kurtosis (non-Gaussianity) of the price fluctuations PDF.

Although this storyline over the relevance of trading volume has got information as one of its main characters, so far I have said little about information. News are often assumed as a synonymous of information. Within a financial context, the first attempt to describe the impact of news in high-frequency stock prices is attributed to Berry & Howe in 1994 [76], whose work was followed by studies relating news to volatility and intra-day dynamics [77]. More recent indicated that news tend to play a minor role in large price fluctuations [78]. Although this is apparently at odds with previous analysis [5] where it was found evidence of the role of scheduled and unscheduled information in the dynamics of a market, the explanation of the apparent “contradiction” lies in the fact that the authors in use order book and high-frequency data on which stationary state treatment in applied these data can hardly capture the impact of news because, as the authors mention, the most relevant news even some decisions with political content tend to be publicised after the markets are closed, especially before weekends. Under these assumptions, the most probable span to find dependence on news would be the pre-market or the very first minutes of the session with their major impact going to the filtered component of the price fluctuation/volatility/trading volume. Nonetheless, the results [78] are challenged by the results in Ref. [79] where the authors found that news correlate with returns, volatility, trading volume and the bid-ask spread. The mismatch between both conclusions might be assigned to at least one of three different factors: the data (market and spell) and the definition of information. Reference [78] uses NYSE and NASDAQ data from 2004 to 2006 whereas in Ref. [79] uses LSE data from 2006 to 2008. Although both markets present strong liquidity, LSE is slightly less inefficient than NYSE.[15] Moreover, the spell of the former is a bullish period in opposition to the spell of the latter work the last six months of which are strongly bearish. Regarding the third ingredient it is important to note that information is defined differently because in Ref. [79] the sentiment as set forth by the Reuters News Scope Sentiment Engine is taken into consideration. In a recent work on commodities (crude oil and gold) futures [80] traded at NYMEX and COMEX, respectively, it was found that the sentiment enclosed in news have an impact on trading dynamics. In a different perspective to information, activity in social media and its relation to market dynamics has been studied as early as the early 2000s [82,83]. First focussing on the investor's fora (e.g. seekingalpha, rangingbull, etc.) [87], it was already found some relation between message activity and price fluctuations and trading volume. However, this relation, at least for

[14] In his seminal work [37], Copeland proposes that the trading volume is a logarithmic function of the severity of the information described by shift in the demand curve.

[15] Using forecasting methods based on generalised entropic forms, it is possible to forecast the residuals of NYSE stocks with a performance of 52% whereas for LSE stocks that performance increases to 56%. As a means of comparison, a little liquid market like São Paulo has a forecasting rate of 64% [81].

technological/internet companies, was found in a reverse direction, *i.e.*, (positive) opinions and abnormal activity are preceded by significant returns. More recent results using social networks [84], where sentiment was taken into consideration, and generic internet tools [85] pointed to a possibility of forecasting price movements. With those results in mind it is worth recalling that financial markets are a for-profit industry, and therefore it is important to understand whether such information contained in social media provides a means of beating the market. The response to this question has been partially given in [86] where it was found that apparently such trends contain basically the same information that is conveyed by the market price, which ultimately could be seen as point in favour of one of the sacred cows of econometrics, the martingale nature of stock prices.

An interesting challenge it is therefore to combine all these ingredients considering that in order to have ever better representations of the market dynamics and specifically of the relations between volatility, volume, price fluctuations and information — which can be read entropy — either using coarse grained techniques like differential/discrete stochastic equations for those macroscopic quantities or using agent-based models.

Acknowledgements

It is my pleasure to acknowledge C. Anteneodo, L. Borland, S. Camargo, Z. Eisler, M.B. Graczyk, P.G. Lind, R.N. Mantegna, T. di Matteo, L.G. Moyano, R. Riera, J. de Souza, S. Thurner, C. Tsallis and W.-X. Zhou either for their direct and invaluable collaboration in some of the work herein reviewed or discussions on matters related to the quantitative description of financial markets. I would like to thank Olsen Financial Services for having provided data used in several papers as well. This review was prepared whilst benefiting from financial support from FAPERJ [Contract nos. APQ1-110.635/2014, Apoio a Projetos Temáticos no Estado do Rio 210.958/1025 and Jovem Cientista do Nosso Estado 202.881/2015], and CNPq [Contract no. 308737/2013-0].

References

[1] Mantegna RN, Stanley HE. An Introduction to Econophysics: Correlations and Complexity in Finance. Cambrigde: Cambridge University Press; 1999.
[2] Bouchaud JP, Potters M. Theory of Financial Risks: From Statistical Physics to Risk Management, Cambridge: Cambridge University Press; 2000; Dacorogna M, Gençay R, Müller U, Olsen R, Pictet O. An Introduction to High-Frequency Finance, London: Academic Press; 2001; Empirical Science of Financial Fluctuations: The Advent of Econophysics, in: Takayasu H (Ed.), Berlin: Springer-Verlag; 2002; Voit J, The Statistical Mechanics of Financial Markets, Berlin: Springer-Verlag, 2003; Duarte Queirós SM, Anteneodo C, Tsallis C. Power-law distributions in economics: a nonextensive statistical approach, in: Noise and Fluctuations in Econophysics and Finance, in: Abbot D, Bouchaud JP, Gabaix X, McCauley JL (Eds.), Proc. of SPIE 5848 (2005) 151.
[3] Black F. Financ Anal J 1971;27:21.
[4] Karpoff JM. J Financ Quant Anal 1987;22:109.
[5] Chae J. J. Financ 2005;60:431.
[6] Farmer JD, Gillemot L, Lillo F, Mike S, Sen A. Quant. Financ. 4 (2004) 383; Weber P, Rosenow B, Quant. Financ. 925 6 (2006) 7; Farmer JD, Gillemot L, Lillo F, Mike S, Sen A, Quant. Financ. 6 (2006) 371.
[7] Admati A, Pfleiderer P. Rev Financ Stud 1988;1:3; Andersen T, Bollerslev T. J Empir Financ 1997;4:115.
[8] Wood RA, McInish TH, Ord JK. J Financ 1985;40:723.
[9] Jain PC, Joh G-H. J Financ Quant Anal 1988;23:269.
[10] Graczyk MB. Sazonalidades intra-diárias e não-estacionaridade de volume negociado em mercados financeiros: propriedades estatísticas e matriciais. M.sc. thesis; Centro Brasileiro de Pesquisas Físicas; Rio de Janeiro - RJ; 2016.
[11] Liu Y, Gopikrishnan P, Cizeau P, Meyer M, Peng C-K, Stanley HE. Phys Rev E 1999;60:1390.
[12] Pasquini M, Serva M. Econ Lett 1999;65:275.
[13] Duarte Queirós SM, Moyano LG. Physica A 2007;383:10.
[14] Beck C, Schlögl F. Thermodynamics of Chaotic Systems. Cambridge: Cambridge University Press; 1993.
[15] Gopikrishnan P, Plerou V, Gabaix X, Stanley HE. Phys Rev E 2000;62. R4493.
[16] Plerou V, Stanley HE. Phys Rev E 2007;76:046109.
[17] Osorio R, Borland L, Tsallis C. In: Gell-Mann M, Tsallis C, editors. Nonextensive Entropy - Interdisciplinary Applications. New York: Oxford University Press; 2004.
[18] Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL. Phys Rev E 1994;49:1685.
[19] de Souza J, Moyano LG, Duarte Queirós SM. Eur Phys J B 2006;50:165.
[20] Cortines AAG, Riera R, Anteneodo C. Europhys Lett 2008;83:30003.
[21] Anteneodo C, Duarte Queirós SM. J Stat Mech 2009. P10023.
[22] Lee KE, Lee JW. Physica A 2007;383:65.
[23] Mu G-H, Chen W, Kertész J, Zhou W-X. Eur Phys J B 2009;68:245.
[24] Gu G-F, Ren F, Ni X-H, Chen W, Zhou W-X. Physica A 2010;389:278.
[25] Qiu T, Zhong LX, Chen G, Wu XR. Physica A 2009;388:2427.
[26] Granger C, Lin J-L, Mendes DA. J Time Ser Anal 1994;15:371; Dionísio A, Menezes R. Physica A 2004;344:326; Syczewska EM, Struzik ZR. Acta Phys Pol A 2015;127:129.
[27] Tsallis C. Phys Rev E 1998;58:1442; Borland L, Plastino AR, Tsallis C. J Math Phys 1998;39:6490.
[28] Duarte Queirós SM. Quant Financ 2005;5:475.
[29] Mandelbrot B. Fractal and multifractal finance: crashes and long-dependence (virtual selecta). http://users.math.yale.edu/~bbm3/webbooks/wb_fin.html. Calvet L, Fisher A. Multifractal Volatility: theory, Forecasting and Pricing, Academic Press, Burlington - MA, 2008; Arneodo A, Muzy J-F, Sornette D. Eur Phys J B 2(1998), 277; Vandewalle N, Ausloos M. Eur Phys J B 4(1998) 257; Muzy J-F, Delour J, Bacry E. Eur Phys J B 17(2000) 537; Lux T. Quant Financ 1(2001) 632; Matteo TD. Aste T, Dacorogna MM. J Bank Financ 29(2005) 827; Duarte Queirós SM, Moyano LG, de Souza J, Tsallis C. Eur Phys J B 55(2007) 161; Di Matteo T. Quant Financ 7(2007) 21; Duarte Queirós SM, Curado EMF, Nobre FD. Physica A 374 (2007) 715; Jiang Z-Q, Zhou W-X. Physica A 387(2008) 4881; Zhou W.-X. Europhys Lett 88(2009) 28004; de Souza J, Duarte Queirós S.M. Chaos Solitons Fractals 42(2009) 2512; Ł. Czarnecki D. Grech Acta Phys Pol A 117(2010) 623-629; Zhou W-X. Chaos Solitons Fractals 45(2012) 147; Buonocore RJ, Aste T, Di Matteo. Measuring multiscaling in financial time-series. doi: http://10.1016/j.chaos.2015.11.022 this issue (2015); Grech D. Alternative measure of multifractal content and its application in finance, arXiv:1309.5466 **[q-fin.ST]** this issue (2015); Borland L. Exploring the dynamics of financial markets: from stock prices to strategy returns this issue (2015).
[30] Kantelhardt JW, Zschiegner SA, Koscielny-Bunde E, Havlin S, Bunde A, Stanley HE. Physica A 2002;316:87.
[31] Moyano LG, de Souza J, Duarte Queirós SM. Physica A 2006;371:118.
[32] Mu G-H, Zhou W-X, Chen W, Kertész J. Phys Procedia 2010;3:1631.
[33] Clark PK. Econometrica 1973;41:135.
[34] Jensen HJ. Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems, Cambridge University Press, Cambridge, 1988; Bak P. How Nature Works: The Science of Self-organized Criticality, Copernicus New York, 1997; Newman MEJ. Proc. R. Soc. London B 263 (1996) 1605; A. Celikoglu, Tirnakli U, Duarte Queirós S.M. Phys. Rev. E 82 (2010) 021124.
[35] Epps TW, Epps ML. Econometrica 44 (1976) 305; Tauchen G, Pitts M. Econometrica 51 (1983) 485; Harris L. J. Finan. Quant. Anal. 22 (1987) 127; Jones CM, Kaul G, Lipson ML. Rev. Finan. Stud. 7 (1994) 631.
[36] Engle RF. Econometrica 50 (1982) 987; Bollerslev TJ. Econometrics 31 (1986) 307; Duarte Queirós SM, Tallis C. Europhys. Lett. 69 (2005) 893; Duarte Queirós SM. Physica A 344 (2004) 619; Duarte Queirós SM, Tallis C. Eur. Phys. J. B 48 (2005) 139; Duarte Queirós SM. EPL 80 (2007) 30005; Duarte Queirós SM, Tallis C. Eur. Phys. J. B 66 (2008) 137; Engle RF, Focardi SM, Fabozzi FJ. ARCH/GARCH Models in Applied Financial Econometrics, in Handbook of Finance, Volume III: Valuation, Financial Modeling, and Quantitative Tools, in: Fabozzi FJ (Ed.), John Wiley & Sons, New York, 2008.
[37] Copeland T. J Financ 1976;31:1149.

[38] Jennings R, Starks L, Fellingham J. J Financ 1981;36:143; Smirlock M, Starks L. J Bank Financ 1988;12:31.
[39] Gallant AR, Rossi PE, Tauchen G. Rev Financ Stud 1992;5:199; Gallant AR, Rossi PE, Tauchen G. Econometrica 1993;61:871.
[40] Goodhart CAE, O'Hara M. Journal Empirical Finance 1997;4:73.
[41] Lux T, Marchesi M. Nature 1999;397:498; Bonanno G, Lillo F, Mantegna RN. Physica A 2000;280:136.
[42] Duarte Queirós SM. Europhys Lett 2005;71:339.
[43] Beck C, Cohen EGD. Physica A 2003;322:267.
[44] Gabaix X, Gopikrishnan P, Plerou V, Stanley HE. Nature (London) 2003;423:267.
[45] Duarte Queirós SM. Trading volume in financial markets: from order book placement to intra-day time series. tutorial talk presented at: econofis'10 – encontro de Econofísica, 25–26 March. 2010. São Paulo, Brazil. Available at: http://econofis10.fis.puc-rio.br/.
[46] Yamasaki K, Muchnik L, Havlin S, Bunde A, Stanley HE. Proc Natl Acad Sci USA 2005;102:9424; Wang F, Yamasaki K, Havlin S, Stanley HE. Phys Rev E 2006;73:026117.
[47] Camargo S, Duarte Queirós SM, Anteneodo C. Phys Rev E 2011;84:046702.
[48] Camargo S, Duarte Queirós SM, Anteneodo C. Eur Phys J B 2013;86:159.
[49] Friedrich R, Peinke J, Sahimi M, Tabar MRR. Phys Rep 2011;506:87; Anteneodo C, Duarte Queirós SM. Phys Rev E 2010;82:041122.
[50] Bouchaud JP, Farmer JD, Lillo F. How Markets Slowly Digest Changes in Supply and Demand. Handbook of Financial Markets: dynamics and Evolution. New York: Elsevier: Academic Press; 2008.
[51] Kyle AS. Econometrica 1985;53:1315.
[52] Lillo F, Farmer JD, Mantegna RN. Nature (London) 2003;421:129.
[53] Bouchaud JP, Gefen Y, Potters M, Wyart M. Quant Financ 2004;4:176.
[54] Campbell JY, Grossman SJ, Wang J. Q J Econ 1993;108:905.
[55] Glaser M, Weber M. J Financ Mark 2009;12:1.
[56] Chen G, Qiu T, Jiang X-F, Zhong L-X, Wu X-R. Physica A 2015;424:73.
[57] Bouchaud JP, Matacz A, Potters M. Phys Rev Lett 2001;87:228701.
[58] Graczyk MB, Duarte Queirós SM. Non-stationarities in the return - trading volume correlation. 2015. Work in progress.
[59] Darrat AF, Rahman S, Zhong M. J Bank Financ 2003;27:2035.
[60] Ding Z, Granger CWJ, Engle RF. J Econometrics 1996;73:61; Ding Z, Granger CWJ. J Empir Financ 1993;1:83.
[61] Lamoureux CG, Lastrapes W. J Financ 1990;45:221.
[62] Lamoureux CG, Lastrapes W. J Bus Econ Stat 1994;16:101.
[63] Bollerslev T, Jubinski D. J Bus Econ Sta 1999;17:9.
[64] Fleming J, Kirby C. J Bank Financ 2011;35:1714; Le V, Zurbruegg R. J Int Financ Mark Inst Money 2010;20:533.
[65] Anéa T, Ureche-Rangau L. J Int Finan Mark Inst Money 2008;18:216.
[66] Glosten LR, Milgrom PR. J Financ Econ 1985;14:71.
[67] Andersen TG. J Financ 1996;51:169.
[68] Andersen TG, Bollerslev T, Frederiksen P, Nielsen M. J Appl Econ 2010;25:233.
[69] Liesenfeld R. J Econom 2001;104:141.
[70] Farmer JD, Zamni N. Eur Phys J B 2007;55:189.
[71] Zovko I, Farmer JD. Quant Financ 2002;2:387; Mike S, Farmer JD. J Econ Dyn Control 2008;32:200.
[72] Stanley HE, Amaral LAN, Gabaix X, Gopikrishnan P, Plerou V. Physica A 2001;299:1.
[73] Darrat AF, Zhong M, Cheng LTW. J Bank Financ 2007;31:2711.
[74] Eisler Z, Kértesz J. Phys Rev E 2006;73:046109; Eisler Z, Kértesz J. Eur Phys J B 2006;51:145; Eisler Z, Kértesz J. Europhys Lett 2007;77:28001.
[75] Rocha P, Raischel F, Boto JP, Lind PG. Uncovering the evolution of non-stationary stochastic variables: the example of asset volume-price fluctuations (2015) Preprint. arxiv:1510.07280 [q-fin.ST].
[76] Berry TD, Howe KM. J Financ 1994;49:1331.
[77] Kalev PS, Liu WS, Pham PK, Jarnecic E, Bank J. Finance 28 (2004), 1441; Ranaldo A. Intraday market dynamics around public information disclosures. In: Lhabitant F-S, Gregoriou G, editors. Stock Market Liquidity. New Jersey: John Wiley & Sons; 2008. p. 199.
[78] Wyart M, Bouchaud JP, Kockelkoren J. Marc Potters and Michele Vettorazzo, Quant. Financ. 8 (2008) 41; Joulin A, Lefevre A, Grunberg D, Bouchaud JP, Stock price jumps: news 1110 and volume play a minor role (2008) Preprint. arXiv:0803.1769 [q-fin.ST].
[79] Groß-Klußmann A, Hautsch N. J Empir Financ 2011;18:321.
[80] Clements AE, Todorova N. J Futur Mark 2016;36:88.
[81] Duarte Queirós SM, Tsallis C. 2003. Unpublished.
[82] Chevalier JA, Mayzlin D. J Mark Res 2006;43:345; Clarkson PM, Joyce D, Tuttici I. Account Financ 2006;46:31; Liu Y. J Mark 2006;70:74; Boyd DM, Ellison NB. J Comp Med Commun 2007;13:210; Chen Y, Xie J. Manag Sci 2008;54:477; Zhu F, Zhang XM. J Mark 2008;74:133; Sabherwal S, Sarkar S. J Bus Financ Account 2010;38:1209; Gang W, Wang T, Wang B, Sambasivan D, Zhang Z, Zheng H, Zhao BY. Crowds on Wall Street: extracting Value from Social Investing Platforms (2014) Preprint. arxiv: 1406.1137[cs.SI].
[83] Mitra G, Mitra L. editors. The Handbook of News Analytics in Finance. Sussex: John Wiley & Sons; 2011.
[84] Sprenger TO, Tumasjan A, Sandner PG, Welpe IM. Eur Financ Manag 2014;20:926; Zhedulev I, Smith R, Aste T. Sci Rep 2014;4:4213.
[85] Da Z, Engleberg J, Gao P. J Financ 2011;65:1461; Preis T, Moat HS, Stanley HE. Sci Rep 2013;3:1684.
[86] Challet D, Ayed ABH. Do google trend data contain more predictability than price returns? (2014) Preprint. arxiv:1403.1715 [q-fin.TR].
[87] Tumarkin R, Whitelaw RF. Financ Anal J 2001;57:41.