On statistical properties of traded volume in financial markets

J. de Souza,* L. G. Moyano,† and S. M. Duarte Queirós‡

Centro Brasileiro de Pesquisas Físicas, 150, 22290-180, Rio de Janeiro - RJ, Brazil

(Dated: May 2, 2006)

In this article we study the dependence degree of the traded volume of the Dow Jones 30 constituent equities by using a nonextensive generalised form of the Kullback-Leibler information measure. Our results show a slow decay of the dependence degree as a function of the lag. This feature is compatible with the existence of non-linearities in this type time series. In addition, we introduce a dynamical mechanism whose associated stationary probability density function (PDF) presents a good agreement with the empirical results.

I. INTRODUCTION

The study of complexity, in particular within financial systems, has become one of the main focus of interest in statistical physics [1]. In fact, several statistical properties verified in financial observables, *e.g.*, relative price changes (*the return*) and returns standard deviation (*the volatility*), have enabled the establishment of new models which characterise systems ever better [2]. Along with the previous two quantities, another key observable is the number of stocks of a certain company traded in a given period of time, *the traded volume*, v. In this article we analyse the dependence degree of 1-minute traded volume time series, $V(t)$, of the constituents of the Dow Jones Industrial Average 30 index (DJ30), between the 1st of July 2004 and the 31st of December 2004. We introduce also a dynamical mechanism that provides the same stationary PDF [3, 4, 5]. In order to avoid spurious features, we have removed intra-day pattern of the original time series and normalised each element of the series by its mean value defining the normalised traded volume, $v(t) = \frac{V'(t)}{\langle V'(t) \rangle}$, where $V'(t) = \frac{V(t)}{\Xi(t')}$, $\Xi(t') = \frac{\sum_{i=1}^{N} V(t'_i)}{N}$ and $\langle \ldots \rangle$ is defined as the average over time (t' represents the intra-day time and i the day).

II. DEPENDENCE DEGREE

Discrimination between two hypothesis, *consistent testing*, is ubiquitous in science. Examples are the stationary/non-stationary character of time series or the dependence degree between its elements. Concerning the latter, the most widely applied measure of "dependence" between variables is the correlation function mathematically defined as,

$$C[v(t), v(t+\tau)] = \frac{\langle v(t) \, v(t+\tau) \rangle - \langle v(t) \rangle^2}{\langle v(t)^2 \rangle - \langle v(t) \rangle^2}.$$

Since the correlation function is basically a normalised covariance (or the second cumulant of the stochastic process), it will only be a suitable statistical procedure for linear correlations or correlations that can be written in a linear way. In other words, the correlation function is not able to determine conveniently non-linearities in a given group of data. Aiming to consistently test the dependence or independence of stochastic variables it was recently defined a dependence measure that has been able to evaluate non-linearities, for instance, in daily return time series [6] and GARCH processes [7] for which the correlation function gives zero value.

---

*e-mail address: jeferson@cbpf.br

†e-mail address: moyano@cbpf.br

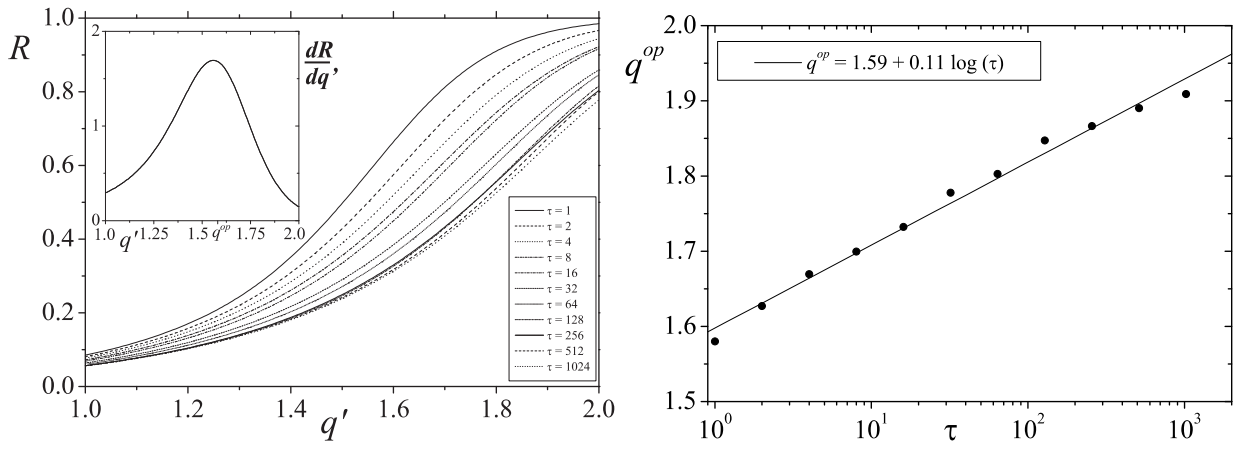‡e-mail address (Corresponding author): sdqueiro@cbpf.br

FIG. 1: Left: Normalised generalised Kullback-Leibler measure, $R_{q'}$, vs. entropic index, q', for the International Business Machines (IBM). The inset shows, as mere illustration, the derivative of R in respect to q' for $\tau = 1$. The maximum corresponds to $q^{op} = 1.58$. Right: The symbols represent the dependence degree, q^{op}, vs. τ (in minutes) averaged over the 30 time series. The line represents a fitting logarithmic function ($q^{op} = 1.59 + 0.11 \log(\tau)$) (the correlation coefficient is 0.9944) pointing up the slow increase of q^{op}.

So, let us start by defining our dependence measure as the *non-extensive generalised mutual information measure*,

$$I_{q'} = - \int p(y) \ln_{q'} \frac{p'(y)}{p(y)} dy$$

where $\ln_{q'}(y) = \frac{y^{q'-1}-1}{q'-1}$ ($\ln_{q'}(y) = \ln_1(y)$), which emerged within the non-extensive formalism based on Tsallis entropy [8]. For $q' = 1$, it is equivalent to the Kullback-Leibler information gain [9, 10].

Let us now assume that y is a two-dimensional random variable $y = (x, z)$. We can quantify the degree dependence between x and z by computing $I_{q'}$ for $p(x, z)$ and $p'(x, z) = p_1(x) p_2(z)$, where $p_{...}(...)$ represents the marginal probability. For this case, $I_{q'}$ presents both a lower bound and an upper bound. The former, $I_{q'}^{MIN} = 0$, corresponds to *total independence* between x and z, i.e. $p(x, z) = p'(x, z)$. The latter, $I_{q'}^{MAX}$, represents a *one-to-one dependence* between variables and is given by,

$$I_{q'}^{MAX} = \begin{array}{c} - \int \int p(x, z) [\ln_{q'} p_1(x) + \\ (1-q) \ln_{q'} p_1(x) \ln_{q'} p_2(z)] dx \, dz. \end{array}$$

From these two extreme values, it is then possible to define a normalised measure,

$$R_{q'} = \frac{I_{q'}}{I_{q'}^{MAX}} \in [0, 1],$$

which has an *optimal* index, q^{op} (where the prime was suppressed for clarity).

This index is optimal in the sense that the gradient of the measure R is most sensitive and hence most capable of determine variations in the dependence among the variables. Moreover, it is optimal because its two extreme values are associated to full dependence and full independence between x and z. Analytically, it is determined by the inflection point of $R_{q'}$ vs q' curves. For one-to-one dependence we have $q^{op} = 0$, and $q^{op} = \infty$ for total independence (see reference [9] for a detailed discussion).

We have computed $R_{q'}$ for all time series with $x = v(t)$, $z = v(t+\tau)$, where τ represents the lag. A typical example is presented in Fig. 1 (left panel). Analysing the behaviour of q^{op} as a function of τ, we have observed a slow increase of q^{op}, i.e., a slow decrease in the dependence degree between variables as it is visible in Fig. 1 (left panel). Our result reveals the existence of significant non-linear dependences which seem to be present even for large times. In Fig. 2 it is possible to see that the correlation value between $\tau = 1$ and $\tau = 1000$ diminishes around 80% while the q^{op} value between $\tau = 1024$ and $\tau = 1$ only reduces in 20% (approximately), i.e., a decrease in the dependence degree in the same amount.
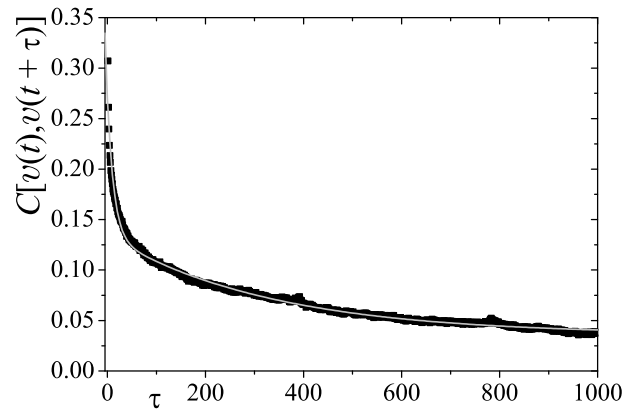
FIG. 2: Left: Symbols represent the average correlation function for the 30 time series analysed and the line represents a double exponential fit with characteristic times of $\gamma^{-1} = 13$ and $T = 332$ yielding a ratio about 25 between the two time scales Eq. (5) ($R^2 = 0.991$ and $\chi^2 = 9 \times 10^{-6}$ and time in minutes).

TABLE I: Obtained values from: PDF fitting (q, θ and α) and from correlation analysis (γT).

	q	θ	α	$T\gamma$
AA	1.19	8.81	2.67	29
AIG	1.22	4.32	1.84	34
AXP	1.21	6.51	2.06	26
BA	1.18	10.67	2.95	24
C	1.15	9.20	3.18	25
CAT	1.20	7.49	2.32	13
DD	1.20	7.33	2.26	53
DIS	1.21	7.29	2.19	20
GE	1.17	8.31	2.75	33
GM	1.21	8.14	2.46	29
HD	1.17	8.76	2.84	27
HON	1.19	9.06	2.67	70
HPQ	1.19	8.55	2.64	28
IBM	1.14	12.36	3.70	41
INTC	1.20	4.22	1.70	25
JNJ	1.17	8.55	2.91	11
JPM	1.17	9.14	2.92	22
KO	1.19	7.88	2.61	26
MCD	1.21	7.48	2.30	30
MMM	1.19	7.14	2.33	23
MO	1.18	7.73	2.66	12
MRK	1.25	1.24	0.61	21
MSFT	1.22	4.57	1.62	23
PFE	1.18	6.31	2.44	33
PG	1.16	8.94	2.99	23
SBC	1.19	8.62	2.57	25
UTX	1.14	18.47	4.71	32
VZ	1.17	8.83	2.84	34
WMT	1.16	10.24	3.23	30
XOM	1.15	11.45	3.50	31

III. A POSSIBLE DYNAMICAL MODEL FOR TRADED VOLUMES

The non-linear character of a time series manifests on the exhibition of (asymptotic) power-law behaviour of the stochastic variable (stationary) PDF. This power-law-like behaviour of the PDF was also verified for traded volume time series [3, 4]. In order to describe a possible dynamical mechanism for this observable, let us suppose that the traded volume of an equity is described by the following stochastic differential equation,

$$dv = -\gamma(v - \frac{\omega}{\alpha}) dt + \sqrt{2\frac{\gamma}{\alpha}v} \, dW_t, \tag{1}$$

where W_t is a regular Wiener process following a normal distribution and $v \geq 0$. The right-hand side of Eq. (1) may be interpreted as follows: the deterministic term represents a natural mechanism of the system which aims to keep the traded volume at some "normal" value, ω/α with a relaxation time of order of γ^{-1}. The stochastic term mimics the microscopic effects on the evolution of v, just like a multiplicative noise used to replicate intermittent processes. This dynamics and the corresponding Fokker-Planck equation [11] leads to an inverted Gamma stationary distribution,

$$f(v) = \frac{1}{\omega \, \Gamma[\alpha+1]} \left(\frac{v}{\omega}\right)^{-\alpha-2} \exp\left[-\frac{\omega}{v}\right]. \tag{2}$$

Consider now, in the same lines of Beck and Cohen superstatistics [16], that instead of constant, ω is a time dependent quantity which evolves on a time scale T larger than the time scale γ^{-1} required by Eq. (1) to reach stationarity. This time dependence is, in the present model, associated to changes in the volume of activity (number of traders that performed transactions) [5]. Furthermore, if we assume that ω follows a Gamma PDF,

$$P(\omega) = \frac{1}{\lambda \Gamma[\delta]} \left(\frac{\omega}{\lambda}\right)^{\delta-1} \exp\left[-\frac{\omega}{\lambda}\right], \tag{3}$$

the long-term distribution of v will be given by $p(v) = \int f(v) P(\omega) \, d\omega$ which yields,

$$p(v) = \frac{1}{Z} \left(\frac{v}{\theta}\right)^{-\alpha-2} \exp_q\left[-\frac{\theta}{v}\right] \tag{4}$$

where $\lambda = \theta(q-1)$, $\delta = \frac{1}{q-1} - \alpha - 1$ and $\exp_q[x] \equiv [1+(1-q)x]^{1/(1-q)}$ is the q-exponential function, the inverse function of $\ln_q(y)$ ($\exp_1[x] = e^x$) [8], Z being the normalisation constant.

This approach is probabilistically equivalent to the one in [5, 12], but it is more realistic concerning the dependence on v of the Kramers-Moyal moments. In other words, this model is, *in principle*, a better dynamical approach. In regard of the measured values of q, θ, α in Tab. I, we verify that they are enclosed within a small interval in the q values, 1.19 ± 0.02 (close to $\frac{6}{5}$) and presents wider intervals for the other parameters, $\alpha = 2.63 \pm 0.48$ and $\theta = 8.31 \pm 1.86$. In Fig. 3 we present the best (Pfizer, PFE) and the worst (Du Pont, DD) fits.

With the α, θ and q fitting values in Tab. I we have generated a set of time series aiming to test the validity of our approach. For the evaluation of the time scales γ^{-1} and T, we have considered the simplest approach, i.e., the ratio between the two time scales which describe the CF for traded volume. See equity values of γT in Tab. I. As can be seen from Fig. 2, there is a fast decay of the CF, related to local equilibrium, and then a much slower decay for larger times that are due to a slow decay of correlations in ω, i.e.,

$$C[v(t), v(t+\tau)] = C_1 \, e^{-\gamma \tau} + C_2 \, e^{-\tau/T}. \tag{5}$$

This slow decay is consistent with a slow dynamics of ω, necessary condition for the appliance of a superstatistical model. In our numerical calculations we have defined time in γ^{-1} units and so $\gamma^{-1} = 1$. The ω values used to mimic the time series were obtained from stationary Feller processes [13] with a T_i relaxation for each i equity (see specific values of γT in Tab. I). Looking to Fig. 3 we have observed that our dynamical propose, using this simple approach, is able to provide good probabilistic description of the data.

IV. FINAL REMARKS

In this article we have analysed some statistical properties of the traded volume equities that constitute the DJ30 index, namely the dependence degree between time series elements and stationary PDF. For the dependence degree we have used a non-extensive generalised Kullback-Leibler information measure. With this procedure we have studied
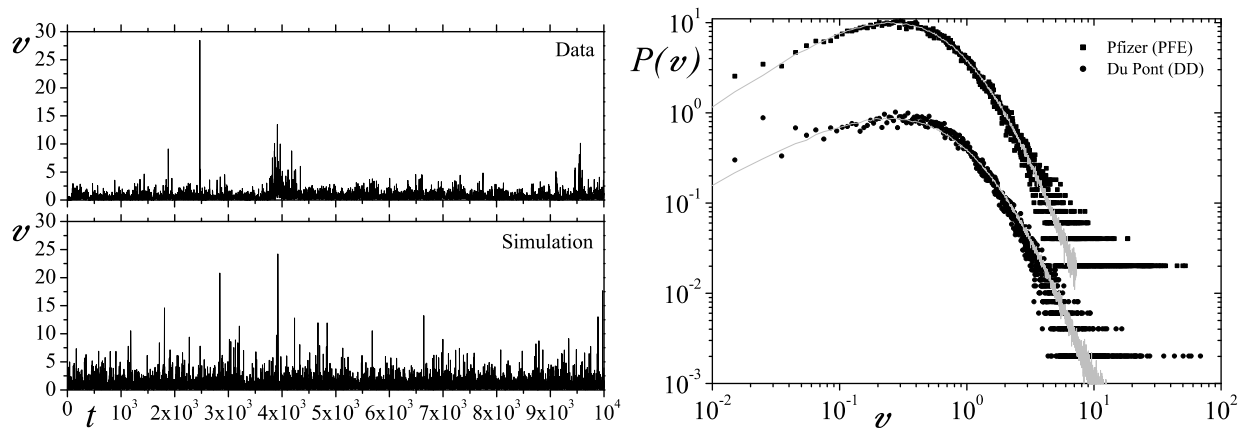
FIG. 3: Left: (Upper panel) Excerpt from the analysed Pfizer time series; (Lower panel) Excerpt from the time series generated to mimic Pfizer using the values presented in Tab. I. (t in minutes) Right: Symbols represent the empirical PDF for Pfizer (shifted by a factor of 10) and Du Pont normalised traded volume time series, which correspond to the best ($R^2 = 0.9953$ and $\chi^2 = 0.0002$) ($R^2 = 0.9763$ and $\chi^2 = 0.001$) and worst fits, respectively. The lines correspond to simulation using the values presented in Tab. I.

the dependence between variables which decreases on a logarithmic way with the lag. We have also verified that this decrease of the dependence is much slower than the one observed in the correlation function. This fact indicates that non-linearities are present in traded volume dynamics and that they may be important factors in other statistical features such as multi-fractality [14]. Analysing the stationary distribution we have verified that it fits well for a q-generalised inverted Gamma distribution presenting a q value around $\frac{6}{5}$ for all series. In addition, we developed a dynamical mechanism which has as stationary PDF the q-generalised inverted Gamma distribution. Further developments of these model may be achieved using perturbative calculus for a more accurate determination of γ [15] and determination of the ratio between the scale of local relaxation and the mean traded volume update [17].

[1] M. Gell-Mann and C. Tsallis (eds.), *Nonextensive Entropy - Interdisciplinary Applications* (Oxford Univiversity Press, New York) 2004.

[2] R.N. Mantegna and H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press) 1999; J.-P. Bouchaud and M. Potters, *Theory of Financial Risks: From Statistical Physics to Risk Management* (Cambridge University Press) 2000; J. Voit, *The Statistical Mechanics of Financial Markets* (Springer-Verlag, Berlin) 2003.

[3] P. Gopikrishnan, V. Plerou, X. Gabaix and H.E. Stanley, Phys. Rev. E **62**, R4493 (2000).

[4] R. Osorio, L. Borland and C. Tsallis, in Reference [1] pp. 321.

[5] S.M. Duarte Queirós, Europhys. Lett. **73**, 339 (2005).

[6] S.M. Duarte Queirós, Quantit. Finance **5**, 475 (2005).

[7] T. Bollerslev. J. Econometrics **31**, 307 (1986); S. J. Taylor, *Modelling Financial Time Series* (Wiley, New York, 1986); S.M. Duarte Queirós and C. Tsallis, Eur. Phys. J. B, in press (2005). [cond-mat/0501230]

[8] C. Tsallis, J. Stat. Phys. **52**, 479 (1988). Bibliography: http://tsallis.cat.cbpf.br/biblio.htm

[9] C. Tsallis, Phys. Rev. E **58**, 1442 (1998); L. Borland, A.R. Plastino and C. Tsallis, J. Math. Phys. **39**, 6490 (1998) and **40**, 2196 (1998).

[10] S. Kullback and R.A. Leibler, Ann. Math. Stat. **22**, 79 (1961).

[11] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, 2nd edition (Springer-Verlag, Berlin, 1989).

[12] As a matter of fact it can be shown that: $x^a e_q^{-\frac{x}{b}} = \left[\frac{b}{q-1}\right]^{\frac{1}{q-1}} x^{a-\frac{1}{q-1}} e_q^{-\frac{b/(q-1)^2}{x}}$ ($q > 1$). In other words, by a simple redefinition of parameters the inverted q-Gamma can be written as the q-Gamma PDF.

[13] W. Feller, Ann. of Math. **54**, 173 (1951).

[14] L.G. Moyano, J. de Souza and S.M. Duarte Queirós, in preparation.

[15] S.M. Duarte Queirós, C. Tsallis, in progress.

[16] C. Beck and E.G.D. Cohen, Physica A **322**, 267 (2003).

[17] C. Beck, E.G.D. Cohen and H.L. Swinney, `cond-mat/0507411` (preprint, 2005).