A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

August 2008

EPL, **83** (2008) 30003
doi: 10.1209/0295-5075/83/30003

www.epljournal.org

Measurable inhomogeneities in stock trading volume flow

A. A. G. CORTINES, R. RIERA and C. ANTENEODO[(a)]

Departamento de Física, Pontifícia Universidade Católica do Rio de Janeiro
CP 38097, 22453-900, Rio de Janeiro, Brazil

received 23 May 2008; accepted in final form 16 June 2008
published online 1 July 2008

PACS 05.40.-a – Fluctuation phenomena, random processes, noise, and Brownian motion
PACS 02.50.Ey – Stochastic processes
PACS 89.65.Gh – Economics; econophysics, financial markets, business and management

Abstract – We investigate the statistics of volumes of shares traded in stock markets. We show that the stochastic process of trading volumes can be understood on the basis of a mixed Poisson process at the microscopic time level. The beta distribution of the second kind (also known as q-gamma distribution), that has been proposed to describe empirical volume histograms, naturally results from our analysis. In particular, the shape of the distribution at small volumes is governed by the degree of granularity in the trading process, while the exponent controlling the tail is a measure of the inhomogeneities in market activity. Furthermore, the present case furnishes empirical evidence of how power law probability distributions can arise as a consequence of a fluctuating intrinsic parameter.

Introduction. – In recent years, stock-index returns and volatility have received special attention in the econophysics community. Meanwhile, there are other market quantities that also provide useful information for evaluating trade strategies, such as the *trading volume*, defined as the number of shares, bonds or contracts traded during a given period Δt. In fact, the trading volume is a measure of asset liquidity and a key determinant of prices and their changes [1]. In this work, we discuss the statistical properties of the share trading volume V in stock markets.

The probability density function (pdf) of high-frequency share volumes of individual stocks was first shown to exhibit a power law decay [2]. Later, statistical analysis including the low-volume region evidenced that the pdf of normalized trading volumes υ could be fitted by the *ansatz* [3]

$$P(\upsilon)=\frac{1}{Z}\left(\frac{\upsilon}{\theta}\right)^{\beta-1}\left[1+(q-1)\,\frac{\upsilon}{\theta}\right]^{-\frac{1}{q-1}}, \qquad (1)$$

where θ is a scale parameter, $\beta>1$, $q\geqslant 1$, and $Z\equiv Z(\beta,q)$ is a normalization constant. Equation (1), defined for positive argument, corresponds to the beta distribution of the second kind. It will be referred to as q-gamma distribution, as soon as it generalizes the regular gamma probability law when $q\neq 1$.

The q-gamma shape was observed for diverse markets in the high-frequency scale: the 10 top volume stocks in NYSE and Nasdaq in 2001, for 1, 2 and 3 minute ticks [3], the 30 stocks of Dow Jones Industrial Average in 2004, for 1 minute ticks [4,5], and the 1 minute Citigroup stocks at NYSE in 2004 [6].

One possible origin of the q-gamma function stands in the realm of mixed distributions [7]. In physics, mixture models have recently been used to describe non-equilibrium systems with a long-term stationary state specified by a spatiotemporal fluctuating intensive quantity [8,9]. In particular, one may find the q-gamma function from the composition of two regular gamma statistics. Based on this property, a plausible scenario has been conjectured to describe the statistics of volumes [4]. However, a consistent empirical support is still lacking. The aim of the present work is to fill that lacuna and model the statistics of volumes from observations and fundamental arguments.

Descriptive statistics of trading volumes. – We take as a representative example the intraday trading volumes of Brazilian stock market Bovespa, from January 3, 2005 to September 13, 2007. The data were obtained at 30 minutes ticks, consisting of 9970 points. Time is counted over trading days only, skipping weekends and holidays. The original trading volume time series, $V(t)$, has been rescaled by the empirical mean value $\langle V\rangle$ and is displayed in fig. 1, which shows the intermittent pattern of trading activity.

We verified that our empirical trading volume histograms, at different timeframes Δt, can be well approximated by the q-gamma pdf, as illustrated in fig. 2

(a) E-mail: celia@fis.puc-rio.br

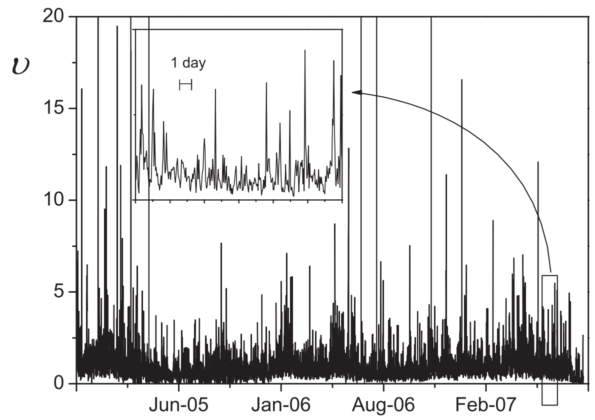

Fig. 1: Trading volume time series of Bovespa from January 3, 2005 to September 13, 2007. Ticks represent the volume traded in intervals $\Delta t = 30$ minutes, normalized by the mean value ($\upsilon = V/\langle V\rangle$).

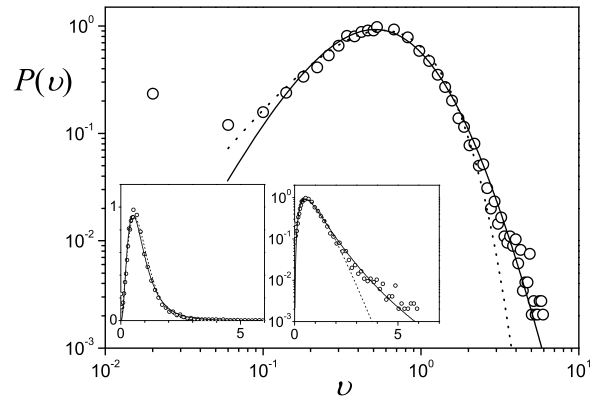

Fig. 2: Log-log plot of the pdf of 30-minute trading volumes in Bovespa during 2005–2007. Symbols represent real data and solid lines are a least-squares fit of eq. (1) with parameters shown in fig. 3. Dotted lines correspond to the fit of eq. (1) with $q = 1$, drawn for comparison. Insets represent linear-linear and log-linear plots of the same data.

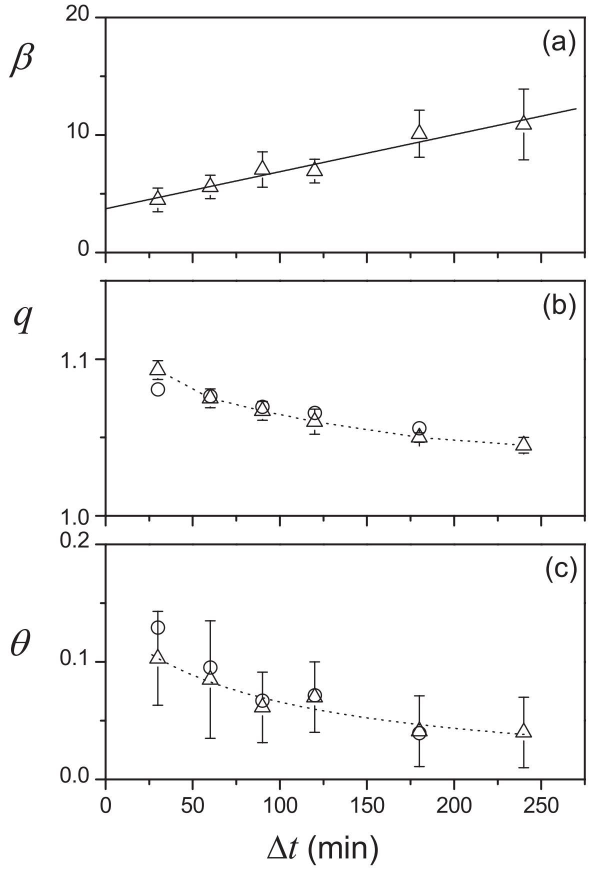

Fig. 3: q-gamma fitting parameters β (a), q (b) and θ (c), for normalized volume pdfs, as a function of Δt (triangles). Circles correspond to theoretical values obtained from eqs. (3). Dotted lines are a guide to the eye. The solid line in (a) corresponds to a linear fit.

(solid lines). A deviation between them is noticeable only for very low volumes, specially for increasing data frequency (see also refs. [3,4]). The excess of the empirical frequency may be due to a lower bound for market transactions, whose statistics is not in the scope of the present analysis. Nevertheless, there is a very good agreement for the wide range encompassing the most significant volumes. Least-squares fits of eq. (1) were performed over the range of sample data, discarding the very small volume region (typically below $\upsilon \simeq 10^{-1}$). Statistical weights $1/P(\upsilon)$ were considered in order to ponder extreme volumes. The resulting values of the fitting parameters are displayed in fig. 3. Consistently with a unitary average rescaled volume, while the three parameters vary, the theoretical q-gamma average, $\beta\theta/((1-q)\beta + 2 - q)$, remains close to one (within a discrepancy of at most 6%).

We also calculated first- and second-order autocorrelation functions and observed (not shown) that they decay exponentially fast with the time lag, in accord with previous results [5]. Moreover, they fall down almost to the noise level at the time lag of about 30 min. This means that memory effects at this scale of data aggregation are very weak.

Modeling. – In this section we model the process that yields the empirical distributions of trading volumes. We will argue in terms of point or arrival processes.

At a given level of data aggregation, each tick of width Δt contains the overall volume negotiated in that interval. Within Δt, a number of negotiations, each one involving a given volume of stocks, occurs. Since, our database does not give any direct information about the individual volume at each negotiation, we will think, conversely, in the arrival of ticks throughout the flow of cumulative volume. The interval of cumulative volume between successive ticks is $\Delta\upsilon_c = \upsilon$. Then, we will deduce the statistics of υ in analogy to that of time intervals (or waiting times) in usual arrival processes.

In that framework, the statistics of volumes constitutes itself a Poisson arrival process. In fact, it is a valid approximation to consider: i) the number of events (ticks) in non-overlapping volume increments as independent and stationary (in the analyzed intraday timescales), ii) that, for sufficiently small $\delta\upsilon_c$, the probability of occurrence of one tick is proportional to $\delta\upsilon_c$ and iii) that the probability of more than one tick arriving in $\delta\upsilon_c$ is negligible. In such case, if the rate of occurrence (average number of ticks per cumulative volume) is η, then the pdf of intervals υ will follow the single exponential law $P(\upsilon) = \eta\exp(-\eta\upsilon)$, akin waiting times between successive occurrences in Poisson arrivals [10]. However, market transactions are complex activities involving several decision steps at the microscopic level of individual events, and volumes on any longer scale are built out of volumes on the event scale. Then, it is reasonable to assume that, instead of being a simple (single step) Poissonian process, the stochastic process of volumes is described by β stages, each one associated to the simple exponential statistics, with similar rates. For independent stages, occurring at a rate η, the volume distribution will be given by the Erlangian pdf [11], namely,

$$P(\upsilon) = \frac{\eta}{\Gamma(\beta)}(\eta\upsilon)^{\beta-1}\exp(-\eta\upsilon)\,. \tag{2}$$

If β is no longer required to be an integer, eq. (2) is known as the gamma distribution $\Gamma_{\beta,1/\eta}$. In this case, the shape parameter β can be interpreted as an effective number of stages of the combined process. Notice also that eq. (2) is equivalent to eq. (1) when $q \to 1$.

Fits of the histograms of trading volumes through eq. (2) are fairly good for intermediate volumes, as illustrated in fig. 2 (dotted lines). However, they present deviations at small and large volumes. The deviation at small values may be related to the fact that, since volumes below a threshold are absent, aggregation of small individual transactions occurs, overweighting the range immediately above the threshold. Meanwhile, the slightly non-exponential behavior of the tail may have a more fundamental origin. The analysis of correlations indicates that deviations are not due to the presence of strong correlations. Then, we proceeded to investigate other sources of the observed long tails. On one hand, trading is not homogeneous. On the other, it was shown by Wilk and Wlodarcsyk [8] that power law tail distributions emerge in a natural way from slow fluctuations of an inverse temperature-like parameter appearing in an originally exponential distribution. Hence, we scrutinized the possible stochastic character of the rate η in eq. (2), which quantifies the average number of events (ticks) per υ_c at each stage. For that purpose, we built a plot of the number of ticks as a function of the cumulative volume υ_c, which is exhibited in fig. 4. One observes that the effective local slope, let us call it σ, fluctuates at different scales, as shown in the inset. Notice that σ^{-1} is approximately a local measure of $\langle\upsilon\rangle$, which, according to eq. (2), satisfies $\langle\upsilon\rangle = \beta\eta^{-1}$. The local rate is then $\eta = \beta\sigma$ corresponding to the (hidden) rate of each stage. Indeed, the composite nature of the volume process implies that the observed rate is diminished as compared with the actual rate of single stages.

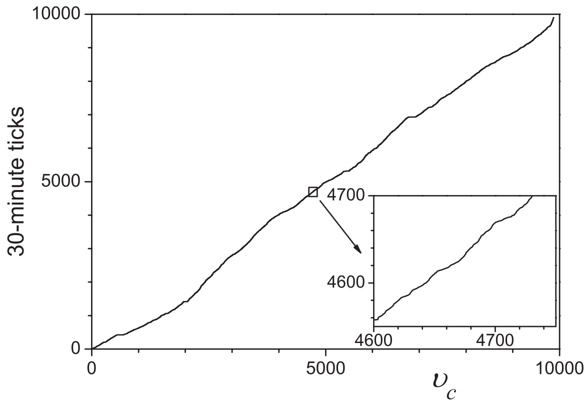

Fig. 4: Number of 30-minute ticks as function of the cumulative volume υ_c in Bovespa during the term 2005–2007. Lower inset: zoom of the curve.

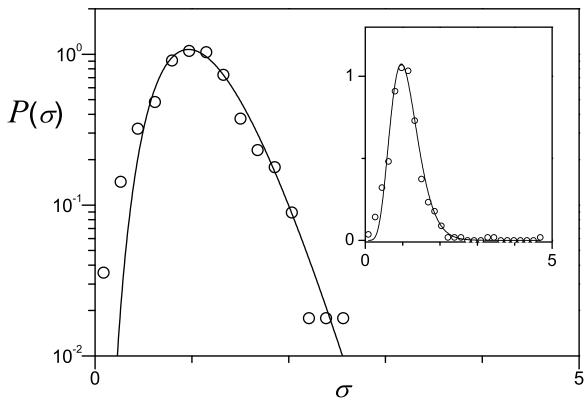

Fig. 5: Histogram of the local slopes of the plot given in fig. 4, calculated over 30 ticks (symbols). Solid lines correspond to the best least-squares fit of a gamma pdf. Inset: the same plot in linear scales.

In order to quantify rate fluctuations, the local slopes of fig. 4 were evaluated for windows of equal length υ_c. Small windows are too noisy while larger ones smear out the structure of fluctuations. Therefore, we considered intermediate-size windows corresponding typically to an interval of 30 ticks (about 2 trading days). The histogram of local slopes is presented in fig. 5. This figure shows that the statistics of slopes σ (hence that of local rates η) follows, in good approximation, a gamma distribution. By performing the Kolmogorov-Smirnoff goodness-of-fit test (KS-test), its result strongly supports the hypothesis of gamma distribution of σ. Contrarily, other two-parameter standard distributions were rejected, for instance the inverse gamma, with a p-value smaller than 0.035.

Parameters of the $\Gamma_{\alpha,1/\kappa}$ pdf, associated to the histograms of local η were obtained from the fitting of the gamma pdf to the histograms of $\sigma = \eta/\beta$, leading to $(\alpha, \kappa) = (7.93, 1.863)$. The distribution of slopes is expected to depend only marginally on the length of ticks Δt. In fact, we observed that α and $\kappa\beta$ are not significantly affected by the choice of Δt. Moreover, $\alpha/(\kappa\beta)$ corresponds to the average slope, $\langle\sigma\rangle$, which is unitary by construction.

If η fluctuates with a larger characteristic timescale than the underlying stochastic process, then eq. (2) becomes a conditional probability of v given η. For a gamma-distributed random variable η, with parameters α and κ, the unconditional (or mixed) pdf of volumes becomes a q-gamma [7]. The theoretical values of the parameters of the resulting generalized gamma pdf (1) are:

$$\beta_{th} = \beta\,, \quad q_{th} = 1 + \frac{1}{\alpha+\beta} \quad \text{and} \quad \theta_{th} = \frac{\kappa}{\alpha+\beta}. \qquad (3)$$

Notice that, while β controls the increase of the pdf at small volumes, α controls the tail: power law with exponent $\alpha + 1$ for finite α, exponential for $\alpha \to \infty$.

In sum, our analysis leads to the q-gamma pdf that describes, in very good approximation, empirical histograms (see fig. 2).

In fig. 3(a), one observes that β increases linearly with Δt from a value $\beta_0 > 1$. This implies the existence of an invariant effective number of stages β_0 associated with the most fundamental time level of description of the financial observables. Linearity accounts for the aggregation properties of the volume process. In fig. 3(b) one observes that q decays towards one when Δt increases. This implies that inhomogeneities smear out for large accumulation intervals Δt. It is noticeable that, for the ten high-volume stocks in Nasdaq and NYSE for 1, 2 and 3 minutes, similar behaviors are observed [3].

From eqs. (3), q is determined in a unique way from the effective number $\alpha + \beta$ of independent pieces of information (stages or degrees of freedom) which specify both gamma distributions. Then, on the one hand, q reflects the degree of granularity in the description of the trading process (through β). On the other, it also reflects the presence of inhomogeneities (through α), which are due to fluctuations in market environment, such as the unsteady flow of news and rumors. Both effects fade away for large accumulation intervals Δt.

Final remarks. – We have shown that a stochastic process given by the inhomogeneous occurrence of Poisson events describes the main statistical properties of an important market observable, the trading volume. The resulting pdf is the q-gamma, which is in very good accord with empirical histograms for the most significant volumes. Theoretical values of the parameters were obtained from eqs. (3), by drawing the effective number of stages β from the q-gamma fit to the histograms of volumes, while α and κ come from the gamma fit of the histogram of slopes. Theoretical and empirical values of the parameters are in excellent accord (see fig. 3), denoting the consistency of our approach. Being a measure of granularity, β affects the shape of the pdf at small volumes only. Meanwhile, the tail is determined solely by α, whose finiteness manifests a degree of inhomogeneity.

Let us note that the gamma pdf is observed in many phenomena involving multiple tasks or stages [11]. The correction of its exponential decay to a power law can be achieved by considering a combined Poisson process with non-homogeneous rate, which is itself random (drawn from a gamma distribution). This constitutes a special kind of non-homogeneous Poisson process in which the rate is not a smooth function of time but it evolves stochastically (Cox-like process), reflecting an additional source of random fluctuations. Similar models in which the rate parameter of the gamma pdf fluctuates with a gamma probability law have been proposed before in diverse contexts to explain the presence of heavy tails. Examples are: computer program failures [12], ecological disturbances [13] and ball touches during football matches [14]. This diversity shows the universality of the doubly stochastic gamma behavior. This may be due to the special character of the gamma family of pdfs, which is closed under convolutions. The frequent appearance of q-gamma pdfs, arising as a mixed pdf, suits within the frame of the approach introduced by Wilk and Wlodarczyk [8] and later generalized by Beck and Cohen in the so-called *superstatistics* [9], inspired in Tsallis proposal [15].

The gamma-gamma behavior has been discussed before also in connection with volume statistics [4,5]. There, the conditional gamma statistics derives from a stochastic dynamical equation whose associated stationary pdf is the gamma distribution. However, it has been proposed phenomenologically, while other equally plausible stochastic equations share the same property [16]. Moreover, it was dubious wether the gamma fluctuating intrinsic parameter was either the local mean trading volume or its inverse. Our results provide empirical support to the later conjecture against the former one.

Summarizing, we have investigated a real database of trading volumes in a mesoscopic timescale where both the granular nature of the transactions and aggregation properties are present. Besides furnishing new insights on trading dynamics, the present study provides a consistent real-data demonstration of how power law pdfs can emerge from observable fluctuation effects, in accordance with the superstatistical view. Our results, obtained for the timescale of the order of one hour, are also consistent with those observed for the time horizon of minutes [3]. However, an investigation along the present lines, analyzing the role of correlations in ultra-high-frequency data, should still be carried out.

* * *

We acknowledge Bovespa for providing the data and the Brazilian agency CNPq for financial support.

REFERENCES

[1] Cassidy D. L., *Trading on Volume: The Key to Identifying and Profiting from Stock Price Reversals* (McGraw-Hill, New York) 2002.

[2] Gopikrishnan P., Plerou V., Gabaix X. and Stanley H. E., *Phys. Rev. E*, **62** (2000) R4493; Gabaix X., Gopikrishnan P., Plerou V. and Stanley H. E., *Q. J. Econ.*, **121** (2006) 461.

[3] Osorio R., Borland L. and Tsallis C., in *Nonextensive Entropy - Interdisciplinary Applications*, edited by Gell-Mann M. and Tsallis C. (Oxford University Press, New York) 2004; see also Tsallis C., Anteneodo C., Borland L. and Osorio R., *Physica A*, **324** (2003) 89.

[4] Queirós S. M. D., *Europhys. Lett.*, **71** (2005) 339.

[5] de Souza J., Moyano L. G. and Queirós S. M. D., *Eur. Phys. J. B*, **50** (2006) 165; Queirós S. M. D. and Moyano L. G., *Physica A*, **383** (2007) 10.

[6] Queirós S. M. D., Moyano L. G., de Souza J. and Tsallis C., *Eur. Phys. J. B*, **55** (2007) 161.

[7] McDonald J. B. and Butler R. J., *Rev. Econ. Stat.*, **69** (1987) 232.

[8] Wilk G. and Wlodarczyk Z., *Phys. Rev. Lett.*, **84** (2000) 2770.

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

[10] Feller W., *An Introduction to Probability Theory and Its Applications* (Wiley, New York) 1968.

[11] Cooper R. B., *Introduction to Queuing Theory* (Elsevier North Holland, Inc., New York) 1981.

[12] Littlewood B. and Verall J. L., *Appl. Stat.*, **22** (1973) 332.

[13] Katz R. W., Brush G. S. and Parlange M. B., *Ecology*, **85** (2005) 1124.

[14] Mendes R. S., Malacarne L. C. and Anteneodo C., *Eur. Phys. J. B*, **57** (2007) 357.

[15] Tsallis C., *J. Stat. Phys.*, **52** (1988) 479; Tsallis C. and Souza A. M. C., *Phys. Rev. E*, **67** (2003) 026106.

[16] Anteneodo C. and Riera R., *Phys. Rev. E*, **72** (2005) 026106.