Statistical properties of share volume traded in financial markets

Parameswaran Gopikrishnan,[1] Vasiliki Plerou,[1,2] Xavier Gabaix,[3], and H. Eugene Stanley[1]
[1]Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215
[2]Department of Physics, Boston College, Chestnut Hill, Massachusetts 02164
[3]Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142

(Received 1 May 2000)

We quantitatively investigate the ideas behind the often-expressed adage ''it takes volume to move stock prices,'' and study the statistical properties of the number of shares traded $Q_{\Delta t}$ for a given stock in a fixed time interval Δt. We analyze transaction data for the largest 1000 stocks for the two-year period 1994–95, using a database that records every transaction for all securities in three major US stock markets. We find that the distribution $P(Q_{\Delta t})$ displays a power-law decay, and that the time correlations in $Q_{\Delta t}$ display long-range persistence. Further, we investigate the relation between $Q_{\Delta t}$ and the number of transactions $N_{\Delta t}$ in a time interval Δt, and find that the long-range correlations in $Q_{\Delta t}$ are largely due to those of $N_{\Delta t}$. Our results are consistent with the interpretation that the large equal-time correlation previously found between $Q_{\Delta t}$ and the absolute value of price change $|G_{\Delta t}|$ (related to volatility) are largely due to $N_{\Delta t}$.

PACS number(s): 05.40.Fb, 05.45.Tp, 89.90.+n

The distinctive statistical properties of financial time series are increasingly attracting the interest of physicists [1]. In particular, several empirical studies have determined the scale-invariant behavior of both the distribution of price changes [2] and the long-range correlations in the absolute values of price changes [3]. It is a common saying that ''it takes volume to move stock prices.'' This adage is exemplified by the market crash of 19 October 1987, when the Dow Jones Industrial Average dropped 22.6% accompanied by an estimated 6×10^8 shares that changed hands on the New York Stock Exchange alone. Indeed, an important quantity that characterizes the dynamics of price movements is the number of shares $Q_{\Delta t}$ traded (share volume) in a time interval Δt. Accordingly, in this Rapid Communication we quantify the statistical properties of $Q_{\Delta t}$ and the relation between $Q_{\Delta t}$ and the number of trades $N_{\Delta t}$ in Δt. To this end, we select 1000 largest stocks from a database [4] recording all transactions for all US stocks, and analyze transaction data for each stock for the two-year period 1994–95.

First, we consider the time series [15] of $Q_{\Delta t}$ for one stock, which shows large fluctuations that are strikingly non-Gaussian [Fig. 1(a)]. Figure 1(b) shows, for each of four actively traded stocks, the probability distributions $P(Q_{\Delta t})$ which are consistent with a power-law decay,

$$P(Q_{\Delta t})\sim\frac{1}{(Q_{\Delta t})^{1+\lambda}}. \qquad (1)$$

When we extend this analysis [16] to the each of the 1000 stocks [Figs. 1(c) and 1(d)], we obtain an average value for the exponent $\lambda=1.7\pm0.1$, within the Lévy stable domain $0<\lambda<2$.

We next analyze correlations in $Q_{\Delta t}$. We consider the family of autocorrelation functions $\langle[Q_{\Delta t}(t)]^a[Q_{\Delta t}(t+\tau)]^a\rangle$, where the parameter a ($<\lambda/2$) is required to ensure that the correlation function is well defined. Instead of analyzing the correlation function directly, we apply detrended fluctuation analysis [5], which has been successfully used to study long-range correlations in a wide range of complex systems [6]. We plot the detrended fluctuation function $F(\tau)$ as a function of the time scale τ. Absence of long-range correlations would imply $F(\tau)\sim\tau^{0.5}$, whereas $F(\tau)\sim\tau^{\delta}$ with $0.5<\delta\leqslant1$ implies power-law decay of the autocorrelation function,

$$\langle[Q_{\Delta t}(t)]^a[Q_{\Delta t}(t+\tau)]^a\rangle\sim\tau^{-\kappa}\ \ (\kappa=2-2\delta). \qquad (2)$$

For the parameter $a=0.5$, we obtain the average value $\delta=0.83\pm0.02$ for the 1000 stocks [Figs. 2(a) and 2(b)]; so from Eq. (2), $\kappa=0.34\pm0.04$ [7].

To investigate the reasons for the observed power-law tails of $P(Q_{\Delta t})$ and the long-range correlations in $Q_{\Delta t}$, we first note that

$$Q_{\Delta t}\equiv\sum_{i=1}^{N_{\Delta t}} q_i \qquad (3)$$

is the sum of the number of shares q_i traded for all $i=1,\dots,N_{\Delta t}$ transactions in Δt. Hence, we next analyze the statistical properties of q_i. Figure 3(a) shows that the distribution $P(q)$ for the same four stocks displays a power-law decay $P(q)\sim1/q^{1+\zeta}$. When we extend this analysis to each of the 1000 stocks, we obtain the average value $\zeta=1.53\pm0.07$ [Fig. 3(b)].

Note that ζ is within the stable Lévy domain $0<\zeta<2$, suggesting that $P(q)$ is a positive (or one-sided) Lévy stable distribution [8,9]. Therefore, the reason why the distribution $P(Q_{\Delta t})$ has similar asymptotic behavior to $P(q)$, is that $P(q)$ is Lévy stable, and $Q_{\Delta t}$ is related to q through Eq. (3). Indeed, our estimate of ζ is comparable within error bounds to our estimate of λ. We also investigate if the q_i are correlated in ''transaction time,'' defined by i, and we find only ''weak'' correlations (the analog of δ has a value $=0.57\pm0.04$, close to 0.5).

To confirm that $P(q)$ is Lévy stable, we also examine the behavior of $Q_n\equiv\Sigma_{i=1}^{n}q_i$. We first analyze the asymptotic behavior of $P(Q_n)$ for increasing n. For a Lévy stable distribution, $n^{1/\zeta}P([Q_n-\langle Q_n\rangle]/n^{1/\zeta})$ should have the same functional form as $P(q)$, where $\langle Q_n\rangle=n\langle q\rangle$ and $\langle\cdots\rangle$ de-

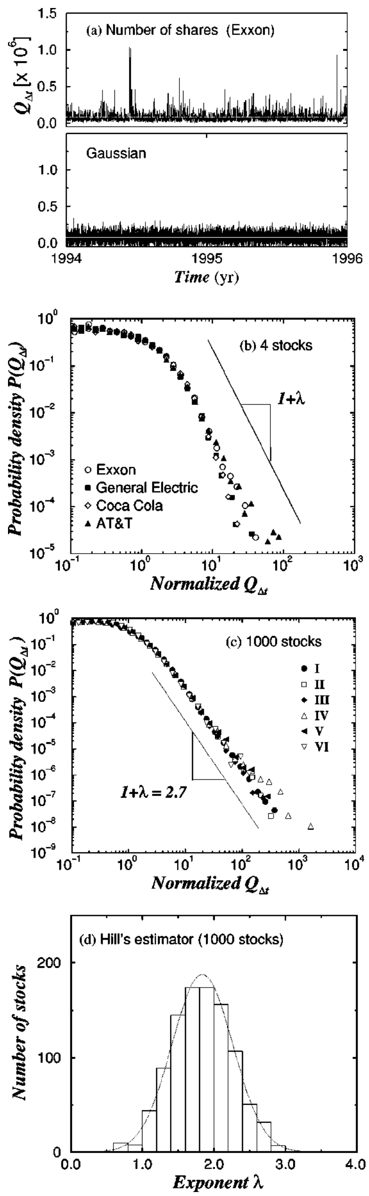

FIG. 1. (a) Number of shares traded [15] for Exxon Corporation (upper panel) for an interval $\Delta t = 15$ min compared to a series of Gaussian random numbers with the same mean and variance (lower panel). (b) Probability density function $P(Q_{\Delta t})$ for four actively traded stocks, Exxon Corp., General Electric Co., Coca Cola Corp., and AT&T Corp., shows an asymptotic power-law behavior characterized by an exponent $1+\lambda$. Hill's method [16] gives $\lambda = 1.87 \pm 0.14$, 2.10 ± 0.17, 1.91 ± 0.20, and 1.71 ± 0.09, respectively. (c) $P(Q_{\Delta t})$ for 1000 stocks on a log-log scale. To choose compatible sampling time intervals Δt, we first partition the 1000 companies studied into six groups [12] denoted I–VI, based upon the average time interval between trades δt. For each group, we choose $\Delta t > 10\delta t$, to ensure that each interval has a sufficient $N_{\Delta t}$. Thus we choose $\Delta t = 15$, 39, 65, 78, 130, and 390 min for groups I–VI respectively, each containing ≈ 150 companies. Since the average value of $Q_{\Delta t}$ differs from one company to the other, we normalize $Q_{\Delta t}$ by its median. Each symbol shows the probability density function of normalized $Q_{\Delta t}$ for all companies that belong to each group. Power-law regressions on the density functions of each group yield the mean value $\lambda = 1.78 \pm 0.07$. (d) Histogram of exponents λ_i for $i = 1, \ldots, 1000$ stocks obtained using Hill's estimator [16], shows an approximately Gaussian spread around the average value $\lambda = 1.7 \pm 0.1$.

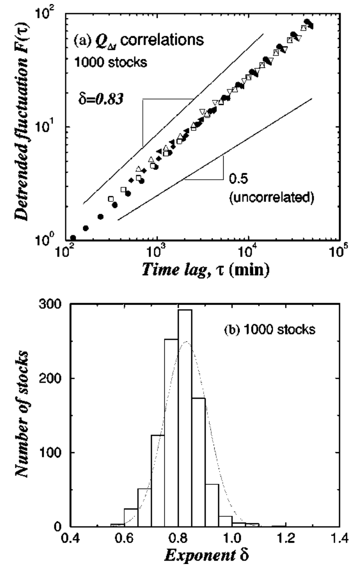

FIG. 2. (a) Detrended fluctuation function $F(\tau)$ for $(Q_{\Delta t})^a$ for $a = 0.5$ [7], averaged for all stocks within each group (I–VI) as a function of the time lag τ. $F(\tau)$ for a time series is defined as the χ^2 deviation of a linear fit to the integrated time series in a box of size τ [5]. An uncorrelated time series displays to $F(\tau) \sim \tau^{\delta}$, where $\delta = 0.5$, whereas long-range correlated time series display values of exponent in the range $0.5 < \delta \leqslant 1$. In order to detect genuine long-range correlations, the U-shaped intraday pattern for $Q_{\Delta t}$ is removed by dividing each $Q_{\Delta t}$ by the intraday pattern [3]. (b) Histogram of δ obtained by fitting $F(\tau)$ with a power-law for each of the 1000 companies. We obtain a mean value of the exponent 0.83 ± 0.02.

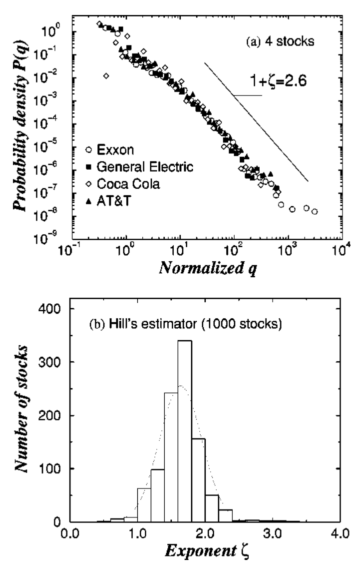

FIG. 3. (a) Probability density function of the number of shares q_i traded, normalized by the average value, for all transactions for the same four actively traded stocks. We find an asymptotic power-law behavior characterized by an exponent ζ. Fits yield values $\zeta = 1.87 \pm 0.13, 1.61 \pm 0.08, 1.66 \pm 0.05, 1.47 \pm 0.04$, respectively for each of the four stocks. (b) Histogram of the values of ζ obtained for each of the 1000 stocks using Hill's estimator [16], whereby we find the average value $\zeta = 1.53 \pm 0.07$.

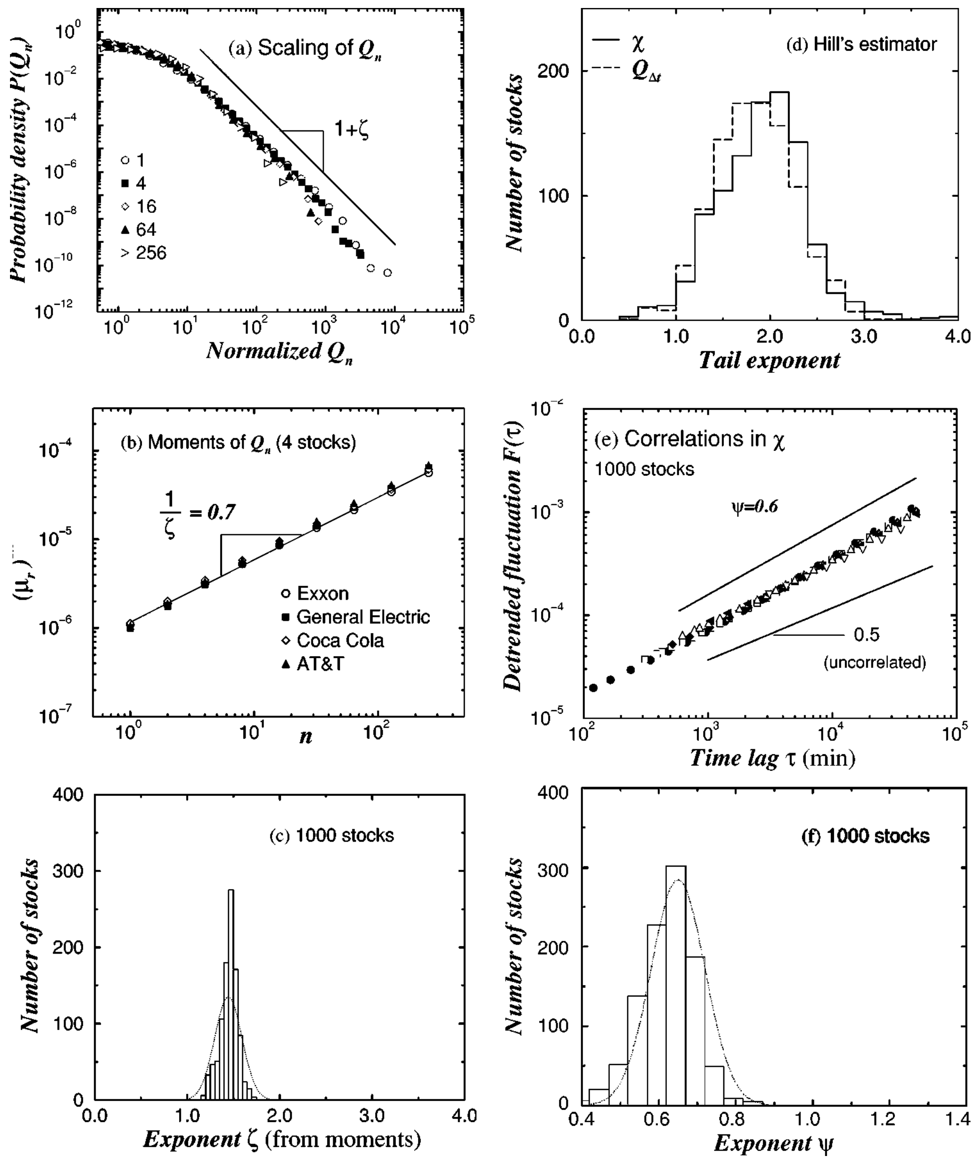

FIG. 4. (a) Probability distribution of Q_n as a function of increasing $n=1,\ldots,256$ apparently retains the same asymptotic behavior. (b) Scaling of the rth root of the rth moments $[\mu_r]^{1/r}$ with increasing n for the same four stocks. The inverse slope of this line yields an independent estimate of the exponent ζ. We obtain $\zeta=1.43\pm0.02,1.35\pm0.03,1.42\pm0.01,1.41\pm0.02$, respectively. (c) Histogram of exponents ζ obtained by fitting a power-law to the equivalent of part (b) for all 1000 stocks studied. We thus obtain a value $\zeta=1.45\pm0.03$ consistent with our previous estimate using Hill's estimator. (d) Histogram of slopes estimated using Hill's estimator for the scaled variable $\chi\equiv[Q_{\Delta t}-\langle q\rangle N_{\Delta t}]/N_{\Delta t}^{1/\zeta}$ compared to that of $Q_{\Delta t}$. We obtain a mean value 1.7 ± 0.1 for the tail exponent of χ, consistent with our estimate of the tail exponent λ for $Q_{\Delta t}$. (e) Detrended fluctuation function $F(\tau)$ for χ, where each symbol denotes an average of $F(\tau)$ for all stocks within each group (I–VI as in Fig. 1). (f) Histogram of detrended fluctuation exponents for χ. We obtain an average value for the exponent 0.61 ± 0.03 which indicates only weak correlations compared to the value of the exponent $\delta=0.83\pm0.03$ for $Q_{\Delta t}$.

notes average values. Figure 4(a) shows that the distribution $P(Q_n)$ retains its asymptotic behavior for a range of n, consistent with a Lévy stable distribution. We obtain an independent estimate of the exponent ζ by analyzing the scaling behavior of the moments $\mu_r(n)\equiv\langle|Q_n-\langle Q_n\rangle|^r\rangle$, where $r<\lambda$ [10]. For a Lévy stable distribution $[\mu_r(n)]^{1/r}\sim n^{1/\zeta}$. Hence, we plot $[\mu_r(n)]^{1/r}$ as a function of n [Figs. 4(b) and 4(c)] and obtain an inverse slope of $\zeta=1.45\pm0.03$, consistent with our previous estimate of ζ [11].

Since the q_i have only weak correlations (the analog of δ has the value $=0.57$), we ask how $Q_{\Delta t}=\Sigma_{i=1}^{\mathrm{Nat}}q_i$ can show much stronger correlations ($\delta=0.83$). To address this question, we note that (i) $N_{\Delta t}$ is long-range correlated [12], and (ii) $P(q)$ is consistent with a Lévy stable distribution with exponent ζ, and therefore, $N_{\Delta t}^{1/\zeta}P([Q_{\Delta t}-\langle q\rangle N_{\Delta t}]/N_{\Delta t}^{1/\zeta})$ should, from Eq. (3), have the same distribution as any of the q_i. Thus, we hypothesize that the dependence of $Q_{\Delta t}$ on $N_{\Delta t}$ can be separated by defining $\chi\equiv[Q_{\Delta t}-\langle q\rangle N_{\Delta t}]/N_{\Delta t}^{1/\zeta}$, where χ is a one-sided Lévy-distributed variable with zero mean and exponent ζ [8,9]. To test this hypothesis, we first analyze $P(\chi)$ and find similar asymptotic behavior to $P(Q_{\Delta t})$ [Fig. 4(d)]. Next, we analyze correlations in χ and find only weak correlations [Figs. 4(e) and 4(f)], implying that the correlations in $Q_{\Delta t}$ are largely due to those of $N_{\Delta t}$.

An interesting implication is an explanation for the previously observed [13,14] equal-time correlations between $Q_{\Delta t}$ and volatility $V_{\Delta t}$, which is the local standard deviation of price changes $G_{\Delta t}$. Now $V_{\Delta t} = W_{\Delta t}\sqrt{N_{\Delta t}}$, since $G_{\Delta t}$ depends on $N_{\Delta t}$ through the relation $G_{\Delta t} = W_{\Delta t}\sqrt{N_{\Delta t}}\epsilon$, where ϵ is a Gaussian-distributed variable with zero mean and unit variance and $W_{\Delta t}^2$ is the variance of price changes due to all $N_{\Delta t}$ transactions in Δt [12]. Consider the equal-time correlation, $\langle Q_{\Delta t} V_{\Delta t}\rangle$, where the means are subtracted from $Q_{\Delta t}$ and $V_{\Delta t}$. Since $Q_{\Delta t}$ depends on $N_{\Delta t}$ through $Q_{\Delta t} = \langle q\rangle N_{\Delta t} + N_{\Delta t}^{1/\zeta}\chi$, and the equal-time correlations $\langle N_{\Delta t} W_{\Delta t}\rangle$, $\langle N_{\Delta t}\chi\rangle$, and $\langle W_{\Delta t}\chi\rangle$ are small (correlation coefficient of the order of ≈ 0.1), it follows that the equal-time correlation $\langle Q_{\Delta t} V_{\Delta t}\rangle \propto \langle N_{\Delta t}^{3/2}\rangle - \langle N_{\Delta t}\rangle\langle N_{\Delta t}^{1/2}\rangle$, which is positive due to the Cauchy-Schwartz inequality. Therefore, $\langle Q_{\Delta t} V_{\Delta t}\rangle$ is large because of $N_{\Delta t}$.

[1] J. D. Farmer, Comput. Sci. Eng. **1**, 26 (1999).

[2] T. Lux, Appl. Finan. Econom. **6**, 463 (1996); P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer, and H. E. Stanley, Phys. Rev. E **60**, 5305 (1999); V. Plerou, P. Gopikrishnan, L. A. N. Amaral, M. Meyer, and H. E. Stanley, *ibid.* **60**, 6519 (1999).

[3] Y. Liu, P. Gopikrishn, P. Cizeau, C.-K. Peng, M. Meyer, and H. E. Stanley, Phys. Rev. E **60**, 1390 (1999); M. Lundin *et al.*, in *Financial Markets Tick by Tick*, edited by P. Lequeux (Wiley, New York 1999), p. 91; Z. Ding *et al.*, J. Empir. Fin. **1**, 83 (1993); R. A. Wood *et al.*, J. Finance **40**, 723 (1985).

[4] *The Trades and Quotes Database*, 24 CD-ROMs for 1994–95 (New York Stock Exchange, New York, 1994).

[5] C.-K. Peng *et al.*, Phys. Rev. E **49**, 1685 (1994).

[6] E. Koscielny-Bunde *et al.*, Phys. Rev. Lett. **81**, 729 (1998); C.-K. Peng *et al.*, *ibid.* **70**, 1343 (1993); Nature (London) **356**, 168 (1992).

[7] Here, κ is the exponent characterizing the decay of the autocorrelation function, compactly denoted $\langle [Q_{\Delta t}(t)]^a [Q_{\Delta t}(t+\tau)]^a\rangle$. Values of a in the range $0.1 < a < 1$ yield δ in the range $0.75 < \delta < 0.88$, consistent with long-range correlations in $Q_{\Delta t}$.

[8] *Lévy Flights and Related Topics in Physics*, edited by M. F. Schlesinger *et al.* (Springer, Berlin, 1995); C. Tsallis, Phys. World **10**, 42 (1997); J.-P. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990).

[9] The general form of a characteristic function of a Lévy stable distribution is $\ln \varphi(x) \equiv i\mu x - \gamma|x|^{\alpha}\{1 + i\beta(x/|x|) tg[(\pi/2)\alpha]\}$ [$\alpha \neq 1$], where the tail exponent α is in the domain $0 < \alpha < 2$, γ is a positive number, μ is the mean, and β is an asymmetry parameter. The case where the parameter $\beta = 1$ gives a positive or one-sided Lévy stable distribution.

[10] The values of ζ reported are using $r = 0.5$. Varying r in the range $0.2 < r < 1$ yields similar values.

[11] To avoid the effect of weak correlations in q on the estimate of ζ, the moments $[\mu_r(n)]$ are constructed after randomizing each time series of q_i. Without randomizing, the same procedure gives an estimate of $\zeta = 1.31 \pm 0.03$.

[12] V. Plerou, P. Gopikrishnan, L. A. N. Amaral, X. Gabaix, and H. E. Stanley, Phys. Rev. E **62**, R3023 (2000) (e-print cond-mat/9912051).

[13] J. Karpoff, J. Finan. Quantitat. Anal. **22**, 109 (1987); C. Jones *et al.*, Rev. Finan. Stud. **7**, 631 (1994); A. R. Gallant *et al.*, *ibid.* **5**, 199 (1992).

[14] G. Tauchen and M. Pitts, Econometrica **57**, 485 (1983); T. W. Epps and M. L. Epps, *ibid.* **44**, 305 (1976); P. K. Clark, *ibid.* **41**, 135 (1973).

[15] Opening trades are not shown in this plot. For all calculations, we have normalized $Q_{\Delta t}$ by the total number of outstanding shares in order to account for stock splits.

[16] B. M. Hill, Ann. Math. Stat. **3**, 1163 (1975).