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 $\Delta t$. 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 $\Delta t$, 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 numbers: 05.45.Tp, 89.90.+n, 05.40.-a, 05.40.Fb

The distinctive statistical properties of financial time series are increasingly attracting the interest of physicists \cite{Gopikrishnan96}. In particular, several empirical studies have determined the scale-invariant behavior of both the distribution of price changes \cite{Gopikrishnan98} and the long-range correlations in the absolute values of price changes \cite{Gopikrishnan98}. 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 $6 \times 10^8$ 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 traded (share volume) in a time interval $\Delta t$. Accordingly, in this paper 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 $\Delta t$. To this end, we select 1000 largest stocks from a database \cite{Gopikrishnan96} recording all transactions for all US-stocks, and analyze transaction data for each stock for the 2-year period 1994-95.

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

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

When we extend this analysis to the each of the 1000 stocks [Fig. 1c,d], we obtain an average value for the exponent $\lambda = 1.7 \pm 0.1$, within the Lévy stable domain $0 < \lambda < 2$.

We next analyze correlations in $Q_{\Delta t}$. We consider the family of correlation 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 \cite{Gopikrishnan98}, which has been successfully used to study long-range correlations in a wide range of complex systems \cite{Gopikrishnan98}. We plot the detrended fluctuation function $F(\tau)$ as a function of the time scale $\tau$. Absence of long-range correlations would imply $F(\tau) \sim \tau^{1/2}$, whereas $F(\tau) \sim \tau^\delta$ with $0 < \delta < 1$ implies power-law decay of the correlation function,

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

For the parameter $a = 0.5$, we obtain the average value $\delta = 0.83 \pm 0.02$ for the 1000 stocks [Fig. 2a,b]; so from Eq. (2), $\kappa = 0.34 \pm 0.04$ \cite{Gopikrishnan96}.

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

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

is the sum of the number of shares $q_i$ traded for all $i = 1, \ldots, N_{\Delta t}$, transactions in $\Delta t$. Hence, we next analyze the statistical properties of $q_i$. Figure 3a shows that the distribution $P(q)$ for the same four stocks displays a power-law decay $P(q) \sim 1/q^{1+\zeta}$. When we extend this analysis to each of the 1000 stocks, we obtain the average value $\zeta = 1.53 \pm 0.07$ [Fig. 3b].

Note that $\zeta$ is within the stable Lévy domain $0 < \zeta < 2$, suggesting that $P(q)$ is a positive (or one-sided) Lévy stable distribution \cite{Gopikrishnan98}. 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. (2). 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 ± 0.04, close to 0.5).

To confirm that P(q) is Lévy stable, we also examine the behavior of Q_n ≡ ∑_{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/ζ} P([Q_n − ⟨Q_n⟩]/n^{1/ζ}) should have the same functional form as P(q), where ⟨Q_n⟩ = n ⟨q⟩ and ⟨...⟩ denotes average values. Figure 4a 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 µ_r(n) ≡ ⟨|Q_n − ⟨Q_n⟩|^r⟩, where r < ζ [11]. For a Lévy stable distribution [µ_r(n)]^{1/r} ∼ n^{1/ζ}. Hence, we plot [µ_r(n)]^{1/r} as a function of n [Fig. 4b,c] and obtain an inverse slope of ζ = 1.45 ± 0.03 — 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_{∆t} ≡ ∑_{i=1}^{N_{∆t}} q_i can show much stronger correlations (δ = 0.83). To address this question, we note that (i) N_{∆t} is long-range correlated [4], and (ii) P(q) is consistent with a Lévy stable distribution with exponent ζ, and therefore, N_{∆t}^{1/ζ} P([Q_{∆t} − ⟨q⟩ N_{∆t}]/{N_{∆t}^{1/ζ}}) should, from Eq. [3], have the same distribution as any of the q_i. Thus, we hypothesize that the dependence of Q_{∆t} on N_{∆t} can be separated by defining η ≡ [Q_{∆t} − ⟨q⟩ N_{∆t}]/N_{∆t}^{1/ζ}, where η is a one-sided Lévy-distributed variable with zero mean and exponent ζ [11]. To test this hypothesis, we first analyze P(η) and find similar asymptotic behavior to P(Q_{∆t}) [Fig. 4d]. Next, we analyze correlations in η and find only weak correlations [Fig. 4e,f] — implying that the correlations in Q_{∆t} are largely due to those of N_{∆t}.

An interesting implication is an explanation for the previously-observed [2,13] equal-time correlations between Q_{∆t} and volatility V_{∆t}, which is the local standard deviation of price changes G_{∆t}. Now V_{∆t} = W_{∆t} / √N_{∆t}, since G_{∆t} depends on N_{∆t} through the relation G_{∆t} = W_{∆t} / √N_{∆t} ε, where ε is a Gaussian-distributed variable with zero mean and unit variance and W_{∆t} is the variance of price changes due to all N_{∆t} transactions in ∆t [11]. Consider the equal-time correlation, ⟨Q_{∆t} V_{∆t}⟩, where the means are subtracted from Q_{∆t} and V_{∆t}. Since Q_{∆t} depends on N_{∆t} through Q_{∆t} = ⟨q⟩ N_{∆t} + N_{∆t}^{1/ζ} η, and the equal-time correlations ⟨N_{∆t} W_{∆t}⟩, ⟨N_{∆t} η⟩, and ⟨W_{∆t} η⟩ are small (correlation coefficient of the order of ≈ 0.1), it follows that the equal-time correlation ⟨Q_{∆t} V_{∆t}⟩ ∝ ⟨N_{∆t}^{3/2}⟩ − ⟨N_{∆t}⟩^{1/2} ⟨N_{∆t}^{1/2}⟩, which is positive due to the Cauchy-Schwartz inequality. Therefore, ⟨Q_{∆t} V_{∆t}⟩ is large because of N_{∆t}. [11].

[1] J.D. Farmer, Computing in Science & Engineering 1, 26 (1999).
[2] T. Lux, Applied Financial Economics 6, 463 (1996); P. Gopikrishnan et al. Phys. Rev. E 60, 5305 (1999); V. Plerou et al., ibid. 60, 6519 (1999).
[3] Y. Liu et al., Phys. Rev. E 60, 1390 (1999); M. Lundin et al., in Financial Markets Tick by Tick, P. Lequeux (ed), p.91 (J. Wiley, New York 1999); Z. Ding et al., J. Empirical Finance 1, 83 (1993).
[4] The Trades and Quotes Database, 24 CD-ROMs for 1994-95, published by the New York Stock Exchange.
[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, ζ is the exponent characterizing the decay of the autocorrelation function, compactly denoted ⟨[Q_{∆t}(t)]^r][Q_{∆t}(t + τ)]⟩. Values of ζ in the range 0.1 < ζ < 1 yield η in the range 0.75 < δ < 0.88 — consistent with long-range correlations in Q_{∆t}.
[8] M. F. Shlesinger et al. (eds.), Lévy Flights and Related Topics in Physics (Springer, Berlin, 1995); C. Tsallis, Physics World 10, 42 (1997); J.-F. Bouchaud and A. Georges, Phys. Rep. 195, 127 (1990).
[9] The general form of a characteristic function of a Lévy stable distribution is ln ϕ(x) ≡ μx − γ|x|^α [1 + iβ(π/2)sgn(x)] [α ≠ 1], where 0 < α < 2, γ is a positive number, μ is the mean, and β is an asymmetry parameter. The case where the parameter β = 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 [µ_r(n)]^{1/r} are constructed after randomizing each time series of q_i. Without randomizing, the same procedure gives an estimate of ζ = 1.31 ± 0.03.
[12] J. Karpoff, Journal of Financial and Quantitative Analysis 22, 109 (1987); C. Jones et al., Reviews of Financial Studies 7, 631 (1994); A. R. Gallant et al., ibid. 5, 199 (1992).
[13] 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).
[14] V. Plerou et al., cond-mat/9912051.
[15] Opening trades are not shown in this plot. For all calculations, we have normalized Q_{∆t} by the total number of outstanding shares in order to account for stock splits.
[16] B. M. Hill, Ann. Stat. 3, 1163 (1975).
FIG. 1. (a) Number of shares traded \( I \) 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 4 actively-traded stocks Exxon Corp., General Electric Co., Coca Cola Corp., and A T & T Corp., shows an asymptotic power-law behavior characterized by an exponent \( 1 + \lambda \). Hill’s method [5] gives \( \lambda = 1.87 \pm 0.14, 2.10 \pm 0.17, 1.91 \pm 0.20, \) and \( 1.71 \pm 0.09 \) respectively. (c) \( P(Q_{\Delta t}) \) for 1000 stocks on a log-log scale. To choose compatible sampling time intervals \( \Delta t \), we first partition the 1000 companies studied into six groups [4] denoted I - VI, based upon the average time interval between trades \( \delta t \). 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 \( \approx 150 \) companies. Since the average value of \( Q_{\Delta t} \) differs from one company to the other, we normalize \( Q_{\Delta t} \) by its average. 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 \( \lambda_i \) for \( i = 1, \ldots, 1000 \) stocks obtained using Hill’s estimator [4].

FIG. 2. (a) Detrended fluctuation function \( F(\tau) \) for \( (Q_{\Delta t})^a \) for \( a = 0.5 \), averaged for all stocks within each group (I-VI) as a function of the time lag \( \tau \). \( F(\tau) \) for a time series is defined as the \( \chi^2 \) deviation of a linear fit to the integrated time series in a box of size \( \tau \). An uncorrelated time series displays to \( F(\tau) \sim \tau^0 \), whereas long-range correlated time series display values of exponent in the range \( 0.5 < \delta \leq 1 \). 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 \( \bar{\chi} \). (b) Histogram of \( \delta \) obtained by fitting \( F(\tau) \) with a power-law for each of the 1000 companies. We obtain a mean value of the exponent \( 0.83 \pm 0.02 \).
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 $\zeta$. 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 4 stocks. (b) Histogram of the values of $\zeta$ obtained for each of the 1000 stocks using Hill’s estimator \cite{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 $r^{th}$ moments $\mu_r$ with increasing $n$ for the same four stocks. The inverse slope of this line yields an independent estimate of the exponent $\zeta$. We obtain $\zeta = 1.43 \pm 0.02, 1.35 \pm 0.03, 1.42 \pm 0.01, 1.41 \pm 0.02$ respectively. (c) Histogram of exponents $\zeta$ obtained by fitting a power-law to the equivalent of part (b) for all 1000 stocks studied. We thus obtain a value $\zeta = 1.45 \pm 0.03$ consistent with our previous estimate using Hill’s estimator. (d) Histogram of slopes estimated using Hill’s estimator for the scaled variable $\chi \equiv [Q_{n \Delta t} - \langle q \rangle_{n \Delta t}]N_{n \Delta t}^{1/\zeta}$ compared to that of $Q_{n \Delta t}$. We obtain a mean value $1.7 \pm 0.1$ for the tail exponent of $\chi$, consistent with our estimate of the tail exponent $\lambda$ for $Q_{n \Delta t}$. (e) Detrended fluctuation function $F(\tau)$ for $\chi$, 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 $\chi$. We obtain an average value for the exponent $0.61 \pm 0.03$ which indicates only weak correlations compared to the value of the exponent $\delta = 0.83 \pm 0.03$ for $Q_{n \Delta t}$.