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 number(s): 05.40.Fb, 05.45.Tp, 89.90.\(\ast\)

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 6×10^10 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 $\Delta t$. 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 $\Delta t$. 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,

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

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 $\lambda = 1.7\pm 0.1$, within the Lévy stable domain 0 < $\lambda$ < 2.

We next analyze correlations in $Q_{\Delta t}$. We consider the family of autocorrelation functions $\langle [Q_{\Delta t}(t)][Q_{\Delta t}(t+\tau)]^{\alpha}\rangle$, where the parameter $\alpha$ (<< $\lambda$/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 $\tau$. Absence of long-range correlations would imply $F(\tau) \sim \tau^{0.5}$, whereas $F(\tau) \sim \tau^{\delta}$ with $0.5 < \delta \leq 1$ implies power-law decay of the autocorrelation function,

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

For the parameter $\alpha = 0.5$, we obtain the average value $\delta = 0.83 \pm 0.02$ for the 1000 stocks [Figs. 2(a) and 2(b)], so from Eq. (2), $\kappa = 0.34 \pm 0.04$ [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

$$Q_{\Delta t} = \sum_{i=1}^{N_{\Delta t}} q_i \tag{3}$$

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 3(a) shows that the distribution $P(q)$ for the same four stocks displays a power-law decay $P(q) \sim 1/q^{1+\xi}$. When we extend this analysis to each of the 1000 stocks, we obtain the average value $\xi = 1.53 \pm 0.07$ [Fig. 3(b)].

Note that $\xi$ is within the stable Lévy domain 0 < $\xi$ < 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 $\xi$ is comparable within error bounds to our estimate of $\lambda$. We also investigate if the $q_i$ are correlated in “transaction time,” defined by $i$, and we find only “weak” correlations (the analog of $\delta$ 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 $Q_n = \sum_{i=1}^{n} q_i$. We first analyze the asymptotic behavior of $P(Q_n)$ for increasing $n$. For a Lévy stable distribution, $n^{1/\xi} P ([Q_n - \langle Q_n \rangle]/n^{1/\xi})$ should have the same functional form as $P(q)$, where $(Q_n) = \langle q \rangle$ and (\ldots)}
FIG. 1. (a) Number of shares traded \([Q_{\Delta t}]\) 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 \(\lambda=1.87\pm 0.14, 2.10\pm 0.17, 1.91\pm 0.20,\) 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 [12] 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 \(\Delta t=15, 39, 78, 130,\) 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 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 \(\lambda_i\) 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 \(\tau\). \(F(\tau)\) 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 \(\tau\) [5]. An uncorrelated time series displays to \(F(\tau) \sim \tau^d\), where \(d=0.5\), whereas long-range correlated time series display values of exponent in the range \(0.5<d<1.0\). 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 \(d\) 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 four stocks. (b) Histogram of the values of \(\zeta\) obtained for each of the 1000 stocks using Hill’s estimator [16], whereby we find the average value \(\zeta = 1.53\pm 0.07\).
Hence, we plot much stronger correlations with our previous estimate of the exponent $\zeta$. We obtain $\zeta=1.43\pm0.02,1.35\pm0.03,1.42\pm0.01,1.41\pm0.02$, respectively. Figure 4 shows scaling of the moments $\mu_r(n)$ as a function of increasing $n$ for 1000 stocks studied. We thus obtain a value $\zeta=1.45\pm0.03$ consistent with our previous estimate using Hill’s estimator. Histograms of slopes estimated using Hill’s estimator for the scaled variable $X^r=[Q_{\Delta t}-\langle q \rangle N_{\Delta t}]/N_{\Delta t}^{1/r}$ compared to that of $Q_{\Delta t}$. We obtain an average value 1.7$\pm$0.1 for the tail exponent of $\chi$, consistent with our estimate of the tail exponent $\lambda$ for $Q_{\Delta t}$. 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). Histogram of detrended fluctuation exponents for $\chi$. We obtain an average value 1.7$\pm$0.1 for the tail exponent 1.7$\pm$0.1 for the tail exponent $\delta$ of $Q_{\Delta t}$.

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 $\zeta$ by analyzing the scaling behavior of the moments $\mu_r(n)=[Q_n-\langle Q_n \rangle]^{1/r}$, where $r<\lambda$ [10]. For a Lévy stable distribution $[\mu_r(n)]^{1/r}=n^{1/r \zeta}$. 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 $\zeta$ [11].

Since the $q_i$ have only weak correlations (the analog of $\delta$ has the value $\approx 0.57$), we ask how $Q_{\Delta t}=[\sum q_i N_{\Delta t}]$ 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 $\zeta$, and therefore, $N_{\Delta t}^{1/\zeta}P([Q_{\Delta t}-\langle q \rangle N_{\Delta t}]/N_{\Delta t})$ 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 $\chi=[Q_{\Delta t}-\langle q \rangle N_{\Delta t}]/N_{\Delta t}^{1/\zeta}$, where $\chi$ is a one-sided Lévy-distributed variable with zero mean and exponent $\zeta$ [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 $\chi$ 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 previ-
ously 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 $\epsilon$ 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 $\Delta t$ [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 $Q_{\Delta t} = \langle q \rangle N_{\Delta t} + \langle N_{\Delta t}^{1/2} \rangle \chi$, 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 $\approx 0.1$), it follows that the equal-time correlation $\langle Q_{\Delta t}V_{\Delta t} \rangle \propto \langle N_{\Delta t}^{1/2} \rangle - \langle N_{\Delta t} \rangle \langle N_{\Delta t}^{1/2} \rangle$, which is positive due to the Cauchy-Schwarz inequality. Therefore, $\langle Q_{\Delta t}V_{\Delta t} \rangle$ is large because of $N_{\Delta t}$.

[7] Here, $\kappa$ is the exponent characterizing the decay of the autocorrelation function, compactly denoted $\langle [Q_{\Delta t}(t)]^\alpha [Q_{\Delta t}(t + \tau)]^\beta \rangle$. Values of $\alpha$ in the range $0.1 < \alpha < 1$ yield $\delta$ in the range $0.75 < \delta < 0.88$, consistent with long-range correlations in $Q_{\Delta t}$.
[9] The general form of a characteristic function of a Lévy stable distribution is $\varphi(x) = i\mu x - \gamma |x|^{\alpha}[1 + i\beta |x|^{\alpha} \tan(\pi \alpha/2)]$ [\(\alpha \neq 1\)], where the tail exponent $\alpha$ is in the domain $0 < \alpha < 2$, $\gamma$ is a positive number, $\mu$ is the mean, and $\beta$ is an asymmetry parameter. The case where the parameter $\beta = 1$ gives a positive or one-sided Lévy stable distribution.
[10] The values of $\xi$ 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 $\xi$, the moments $\langle \mu(n) \rangle$ are constructed after randomizing each time series of $q_i$. Without randomizing, the same procedure gives an estimate of $\xi = 1.31 \pm 0.03$.
[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.