RESEARCH PAPER
quant.iop.org
QUANTITATIVE FINANCE VOLUME 1 (2001) 262–269
INSTITUTE OF PHYSICS PUBLISHING

Price fluctuations, market activity and trading volume

Vasiliki Plerou[1,2], Parameswaran Gopikrishnan[1], Xavier Gabaix[3], Luís A Nunes Amaral[1] and H Eugene Stanley[1]

[1] Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA
[2] Department of Physics, Boston College, Chestnut Hill, MA 02164, USA
[3] Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

Received 10 October 2000

Abstract
We investigate the relation between trading activity—measured by the number of trades $N_{\Delta t}$—and the price change $G_{\Delta t}$ for a given stock over a time interval $[t,\ t + \Delta t]$. We relate the time-dependent standard deviation of price changes—volatility—to two microscopic quantities: the number of transactions $N_{\Delta t}$ in Δt and the variance $W^2_{\Delta t}$ of the price changes for all transactions in Δt. We find that $N_{\Delta t}$ displays power-law decaying time correlations whereas $W_{\Delta t}$ displays only weak time correlations, indicating that the long-range correlations previously found in $|G_{\Delta t}|$ are largely due to those of $N_{\Delta t}$. Further, we analyse the distribution $P\{N_{\Delta t} > x\}$ and find an asymptotic behaviour consistent with a power-law decay. We then argue that the tail-exponent of $P\{N_{\Delta t} > x\}$ is insufficient to account for the tail-exponent of $P\{G_{\Delta t} > x\}$. Since $N_{\Delta t}$ and $W_{\Delta t}$ display only weak interdependence, we argue that the fat tails of the distribution $P\{G_{\Delta t} > x\}$ arise from $W_{\Delta t}$, which has a distribution with power-law tail exponent consistent with our estimates for $G_{\Delta t}$. Further, we analyse the statistical properties of the number of shares $Q_{\Delta t}$ traded in Δt, and find that the distribution of $Q_{\Delta t}$ is consistent with a Lévy-stable distribution. We also quantify the relationship between $Q_{\Delta t}$ and $N_{\Delta t}$, which provides one explanation for the previously observed volume–volatility co-movement.

1. Introduction

This paper summarizes the results of a few recent empirical studies focused on quantifying the statistical features of stock price fluctuations, market activity, and share-volume traded. Our results are based on the analysis of transaction data for 1000 US stocks for the two-year period 1994–1995.

The first result we shall discuss concerns the asymptotic nature of the distribution of stock returns. Our analysis shows that the tails of the distribution of stock returns display a power-law decay with exponents α larger than the upper bound for a Lévy-stable distribution. Our estimates of the exponents indicate that although the second moment exists, moments larger than the third (in particular, the kurtosis) are divergent. Our analysis also shows that, although not statistically stable, the estimates of exponents characterizing these power laws are quite similar for time scales Δt up to approximately 16 trading days, indicating the slow *onset* of convergence to Gaussian behaviour.

The second result concerns the statistical relationship between returns and market activity. We relate the statistical properties of the time-dependent standard deviation of returns—volatility—to those of two quantities: the number of trades $N_{\Delta t}$ in Δt and the variance $W^2_{\Delta t}$ of the price changes for all trades in Δt. We find that returns scaled by the volatility have a distribution with tails that are consistent with those of

1469-7688/01/020262+08$30.00 PII: S1469-7688(01)21452-3

a Gaussian. We next relate the non-Gaussian properties of returns to those of $N_{\Delta t}$ and $W_{\Delta t}$. We find that while the long-range correlations found in $|G_{\Delta t}|$ are largely due to $N_{\Delta t}$, the pronounced tails in the distribution of price fluctuations arise largely from $W_{\Delta t}$.

The third result concerns the statistical properties of the number of shares $Q_{\Delta t}$ traded in Δt and its relation to the number of trades $N_{\Delta t}$. Our analysis of the number of shares q_i traded in each trade i shows that the distribution $P\{q > x\}$ has a tail that decays as a power law with an exponent within the Lévy-stable domain. Since, the number of shares traded in Δt is $Q_{\Delta t} = \sum_{i=1}^{N_{\Delta t}} q_i$, in the limit of sufficiently large $N_{\Delta t}$, if q_i are i.i.d., the sum implies a statistical relationship between $Q_{\Delta t}$ and $N_{\Delta t}$. We test the data for such a relationship and find good agreement. We argue that the Q–N relationship, together with the relationship between volatility and the number of trades, provides an explanation for the previously observed volume–volatility co-movement.

2. Databases analysed

Our empirical results are based on the analysis of different databases covering securities traded in the three major US stock exchanges, namely (i) the New York Stock Exchange (NYSE), (ii) the American Stock Exchange (AMEX), and (iii) the National Association of Securities Dealers Automated Quotation (Nasdaq).

For studying short-time scale dynamics, we analyse the Trades and Quotes (TAQ) database, from which we select the two-year period January 1994 to December 1995. Nasdaq and AMEX merged in October 1998, after the end of the period studied in this work. The TAQ database, which has been published by NYSE since 1993, covers *all* trades at the three major US stock markets. This huge database is available in the form of CD-ROMs. The rate of publication was one CD-ROM per month for the period studied, but recently has increased to two–four CD-ROMs per month. The total number of transactions for the largest 1000 stocks is of the order of 10^9 in the two-year period studied. We analyse the largest 1000 stocks, by capitalization on 3 January 1994, which survived through to 31 December 1995.

The data are adjusted for stock splits and dividends. The data are also filtered to remove spurious events, such as occur due to the inevitable recording errors. The most common error is missing digits which appears as a large spike in the time series of returns. These are much larger than the usual fluctuations and can be removed by choosing an appropriate threshold. We tested a range of thresholds and found no effect on the results.

To study the dynamics at longer time horizons, we analyse the Center for Research and Security Prices (CRSP) database. The CRSP Stock Files cover common stocks listed on NYSE beginning in 1925, the AMEX beginning in 1962, and the Nasdaq Stock Market beginning in 1972. The files provide complete historical descriptive information and market data including comprehensive distribution information, high, low and closing prices, trading volumes, shares outstanding, and total returns. In addition to adjusting for stock splits and dividends, we have also detrended the data for inflation.

3. The distribution of stock returns

The nature of the distribution of stock returns has been a topic of interest for over 100 years [1]. A reasonable *a priori* assumption, motivated by the central limit theorem, is that the returns are independent, identically Gaussian distributed (i.i.d.) random variables, or equivalently, a random walk in the logarithm of price [2].

Empirical studies (for a review, see [3–6]) show that the distribution of returns has pronounced tails, in striking contrast to those of a Gaussian. To account for the fat tails, Mandelbrot [7] proposed a simple generalization of the central limit theorem to variables which lack a finite second moment, which leads to Lévy-stable distributed returns [7, 8]. This particularly simple way of obtaining fat-tailed distributions is however shown to be inconsistent by empirical studies on the decay of the tails of the return distribution [9–12]. In particular, alternative hypotheses for modelling the return distribution were proposed, which include a log-normal mixture of Gaussians [12], Student *t* distributions [9–11], and exponentially truncated Lévy distributions [13, 14].

Conclusive results on the distribution of returns are difficult to obtain, and require a large number of data to study the rare events that give rise to the tails. In our study [15–17], we analyse approximately 40 million records of stock prices sampled at 5 min intervals for the 1000 leading US stocks for the two-year period 1994–1995 and 30 million records of daily returns for 6000 US stocks for the 35-year period 1962–1996. For time scales shorter than one day, we analyse the data from the TAQ database.

The basic quantity studied for individual companies is the price $S_i(t)$. The time t runs over the working hours of the stock exchange—excluding nights, weekends and holidays. For each company, we calculate the return

$$G_i \equiv G_i(t, \Delta t) \equiv \ln S_i(t + \Delta t) - \ln S_i(t) . \qquad (1)$$

We then calculate the cumulative distributions—the probability of a return larger than or equal to a threshold—of returns G_i for $\Delta t = 5$ min. For each stock $i = 1, \ldots, 1000$, the asymptotic behaviour of the functional form of the cumulative distribution is consistent with a power law,

$$P\{G_i > x\} \sim \frac{1}{x^{\alpha_i}} , \qquad (2)$$

where α_i is the exponent characterizing the power-law decay. In order to compare the returns of different stocks with different volatilities, we define the normalized return $g_i \equiv (G_i - \langle G_i \rangle_T)/v_i$, where $\langle \ldots \rangle_T$ denotes a time average over the 40 000 data points of each time series, for the two-year period studied, and the time-averaged volatility v_i of company i is the standard deviation of the returns over the two-year period $v_i{}^2 \equiv \langle G_i{}^2 \rangle_T - \langle G_i \rangle_T{}^2$. Values of the exponent α_i can be estimated by a power-law regression on each of these distributions $P\{g_i > x\} \sim x^{-\alpha}$, whereby we obtain the average value for the 1000 stocks,

$$\alpha = \begin{cases} 3.10 \pm 0.03 & \text{(positive tail)} \\ 2.84 \pm 0.12 & \text{(negative tail)} \end{cases} \qquad (3)$$

where the fits are performed in the region $2 \leqslant g \leqslant 80$. In figure 1(*a*) we show the histogram for α_i, obtained from power-law regression-fits to the positive tails of the individual cumulative distributions of all 1000 companies studied, which shows an approximate Gaussian spread around the mean value $\alpha = 3.10 \pm 0.03$. These estimates of the exponent α are well outside the Lévy-stable range (which requires $0 < \alpha < 2$), and is therefore consistent with a finite variance for returns. However, moments larger than 3, in particular the kurtosis, seem to be divergent [3, 18]. Our results are similar to the results of the analysis of the daily returns of 30 German stocks comprising the DAX index [19], daily CRSP returns [3], and foreign exchange rates [20].

In order to obtain an alternative estimate for α, we use the estimator of Hill [3, 21]. We calculate the inverse local slope of the cumulative distribution function $P(g)$, $\gamma \equiv -(\mathrm{d}\log P(g)/\mathrm{d}\log g)^{-1}$ for the negative and the positive tail. We obtain an estimator for γ, by sorting the normalized increments by their size, $g^{(1)} > g^{(2)} > \ldots > g^{(N)}$. The cumulative distribution can then be written as $P(g^{(k)}) = k/N$, and we obtain for the local slope

$$\gamma = \left[(N-1)\sum_{i=1}^{N-1} \log g^{(i)}\right] - \log g^{(N)}, \qquad (4)$$

where N is the number of tail events used. We use the criterion that N does not exceed 10% of the sample size, simultaneously ensuring that the sample is restricted to the tail events [3]. Thus, we obtain the average estimates for 1000 stocks,

$$\alpha = \begin{cases} 2.84 \pm 0.12 & \text{(positive tail)} \\ 2.73 \pm 0.13 & \text{(negative tail)} \end{cases}. \qquad (5)$$

Removing overnight events yields the average values of $\alpha = 3.11 \pm 0.15$ for the positive tail and $\alpha = 3.03 \pm 0.21$ for the negative tail.

Motivated by scale-free phenomena observed in a wide range of complex physical systems that have a large number of interacting units, it was believed that the power-law behaviour of the tails of the return distribution must apply to a wide range of observations. However, economists [18, 19, 22] have emphasized that power-law tails may apply only to a small fraction of extreme events. In our case, we find the power-law distribution of returns to hold for events larger than approximately 2 standard deviations, below which the nature of the distribution is affected by the discreteness of prices.

A parallel analysis on the distribution of S&P 500 index returns shows consistent asymptotic behaviour [16], although the central part of the distribution seems to display Lévy behaviour for short time scales (< 30 min) [13]. One reason for a different behaviour at the central part of the distribution of S&P 500 returns is the discreteness [2, 23] of the prices of individual stocks (which causes a cut-off for low values of returns) that comprise the S&P 500 index.

4. Scaling of the distributions of returns and correlations in the volatility

Since the values of α we find are inconsistent with a statistically stable law, we expect the distribution of returns $P(G)$ on larger time scales to converge to Gaussian. In contrast, our analysis of daily returns from the CRSP database shows that the distributions of returns retain the same functional form for a wide range of time scales Δt, varying over three orders of magnitude, 5 min $\leqslant \Delta t \leqslant$ 6240 min = 16 days (figure 1(*b*)). The *onset* of convergence to a Gaussian starts to occur only for $\Delta t >$ 16 days [16, 17]. In contrast, n-partial sums of computer-simulated time series of the same length generated with a probability distribution with the same

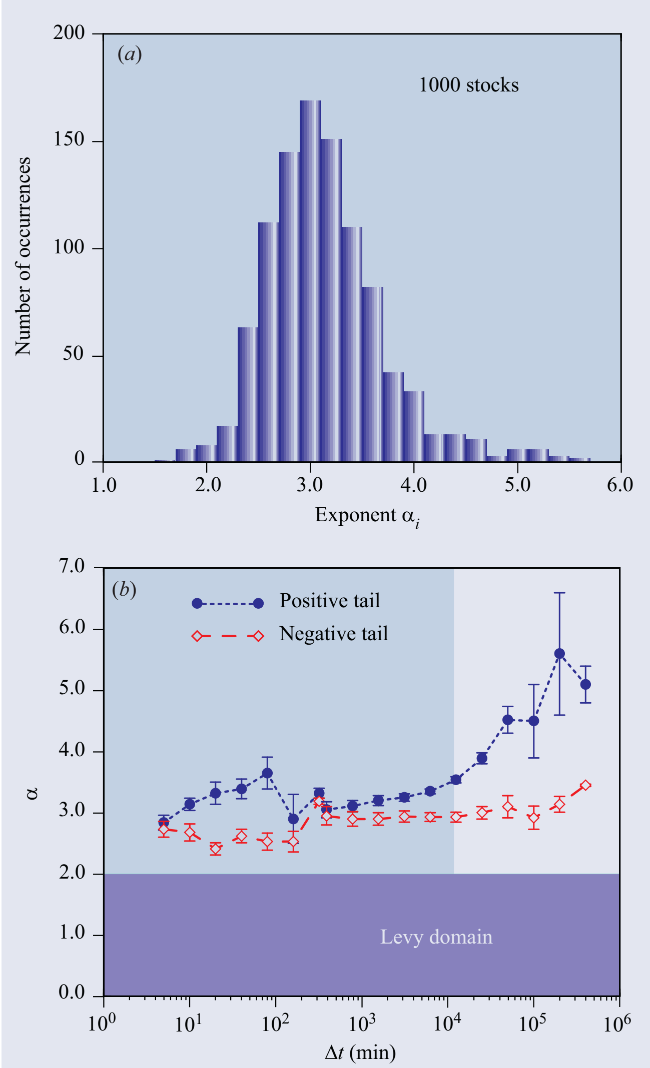

Figure 1. (*a*) The histogram of the tail exponents of the distribution of returns obtained by power-law regressions to the cumulative distribution functions of each stock, where the fit is for all x larger than 2 standard deviations. This histogram is not normalized—the y-axis indicates the number of occurrences of the exponent. (*b*) The average values of the exponent α characterizing the asymptotic power-law behaviour of the distribution of returns as a function of the time scale Δt obtained using Hill's estimator. The values of α for $\Delta t < 1$ day are calculated from the 1000 stocks from the TAQ database while for $\Delta t \geqslant 1$ day they are calculated from 6000 stocks from the CRSP database. The unshaded region, corresponding to time scales larger than $(\Delta t)_\times \approx 16$ days (6240 min), indicates the range of time scales where we find results consistent with the onset of a slow convergence to Gaussian behaviour. The rate of convergence to Gaussian appears to be slower for the negative tail.

asymptotic behaviour agree well with Gaussian behaviour for $n \geqslant 256$ [6, 16]. Thus, the rate of convergence of the distribution $P\{G > x\}$ to a Gaussian is remarkably slow, indicative of time dependences [2, 24] which violate the conditions necessary for the central limit theorem to apply.

To test for time dependences, we analysed the autocorrelation function of returns, which we denote $\langle G(t)G(t+\tau)\rangle$, using 5 min returns of 1000 stocks. Our results show pronounced short-time (< 30 min) anti-correlations due to the bid–ask bounce [2, 25]. For larger time scales, the correlation function is at the level of noise, consistent with the efficient market hypothesis [2, 26, 27]. However, lack of linear correlation does not imply independent returns, since there may exist higher-order correlations. Our recent studies [28–30] show that the amplitude of the returns measured by the absolute value or the square has long-range autocorrelations with persistence [31] up to several months,

$$\langle |G(t)|\,|G(t+\tau)|\rangle \sim \tau^{-a}\,, \tag{6}$$

where $\langle \ldots \rangle$ denotes the autocorrelation function. We obtain the average value for the exponent $a = 0.34 \pm 0.09$ for the 1000 stocks studied. In order to detect genuine long-range correlations, the effects of the U-shaped intraday pattern [32,33] for $|G|$ has been removed [30]. This result is consistent with previous studies [2, 34–36] which also report long-range correlations. In addition to analysing the correlation function directly, we apply power spectrum analysis and the recently developed detrended fluctuation analysis [30, 37]. Both of these methods yield consistent estimates of the exponent a.

5. Analysis of trading activity

5.1. Statistics of trading activity

In order to understand the reasons for the pronounced tails of the return distribution and long-range correlations in volatility, we follow an approach in the spirit of models of time deformation proposed by Clark [12], Tauchen and Pitts [38], Stock [39], Lamoureux and Lastrapes [40], Ghysels and Jasiak [41], and Engle and Russell [42].

Returns G over a time interval Δt can be expressed as the sum of several changes δp_i due to the $i = 1, \ldots, N_{\Delta t}$ trades in the interval $[t, t + \Delta t]$,

$$G_{\Delta t} = \sum_{i=1}^{N_{\Delta t}} \delta p_i\,. \tag{7}$$

If Δt is such that $N_{\Delta t} \gg 1$, and δp_i have finite variance, then one can apply the central limit theorem, whereby one would obtain the result that the unconditional distribution $P(G)$ is Gaussian [12, 43]. It is implicitly assumed in this description that $N_{\Delta t}$ has only *narrow* Gaussian fluctuations, i.e. has a standard deviation much smaller than a large mean value $\langle N_{\Delta t}\rangle$.

In order to ensure that the sampling time interval Δt for each stock contains a sufficient number of transactions $N_{\Delta t}$, we partition the set of 1000 stocks into six groups (I–VI) based on the average time between trades $\langle \delta t\rangle$. Each group contains approximately 150 stocks. For a specific group, we choose a sampling time Δt at least 10 times larger than the average value of $\langle \delta t\rangle$ for that group. We choose the sampling time intervals $\Delta t = 15$, 39, 65, 78, 130 and 390 min respectively for groups I–VI. Overnight returns have been excluded.

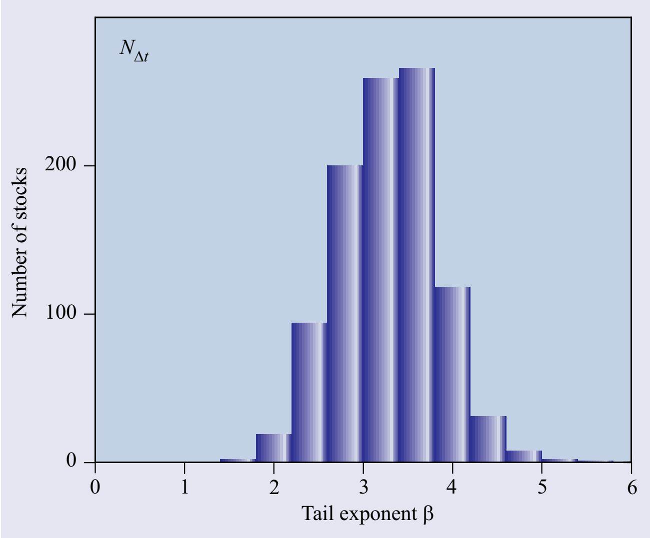

Figure 2. Histogram of tail exponents β characterizing the decay of the cumulative distributions $P\{N_{\Delta t} > x\}$, estimated using Hill's estimator for each of the 1000 stocks. We obtain an average value $\beta = 3.40 \pm 0.05$.

Our investigation of $N_{\Delta t}$ [44] shows stark contrast with a Gaussian time series with the same mean and variance—there are several events of the magnitude of tens of standard deviations which are inconsistent with Gaussian statistics [42, 45–48]. Specifically, we find that the distribution of $N_{\Delta t}$ displays an asymptotic power-law decay [44]

$$P\{N_{\Delta t} > x\} \sim x^{-\beta} \quad (x \gg 1)\,. \tag{8}$$

For the 1000 stocks that we analyse, we estimate β using Hill's method [21] and obtain a mean value $\beta = 3.40 \pm 0.05$ (figure 2). Note that $\beta > 2$ is outside the Lévy-stable domain $0 < \beta < 2$ and is inconsistent with a stable distribution for $N_{\Delta t}$ [49].

5.2. Price fluctuations and trading activity

Since we find that $P\{G_{\Delta t} > x\} \sim x^{-\alpha}$, we can ask whether the value of β we find for $P\{N_{\Delta t} > x\}$ is sufficient to give rise to the fat tails of returns. To test this possibility, we implement, for each stock, the least-squares regression

$$\ln |G_{\Delta t}(t)| = a + b \ln N_{\Delta t}(t) + \psi(t)\,, \tag{9}$$

where $\psi(t)$ has mean zero and the equal time covariance $\langle N_{\Delta t}\psi(t)\rangle = 0$. Only those intervals having $N_{\Delta t} \geqslant 20$ are considered for this regression. For 30 actively traded stocks, we find the average value $b = 0.55 \pm 0.04$.

Values of $b \approx 0.5$ are consistent with what we would expect from equation (7), if δp_i are i.i.d. with finite variance. In other words, suppose δp_i are chosen *only from the interval* $[t,\ t + \Delta t]$, and let us hypothesize that *these* δp_i are mutually independent, with a common distribution $P(\delta p_i | t \in [t, t+\Delta t])$

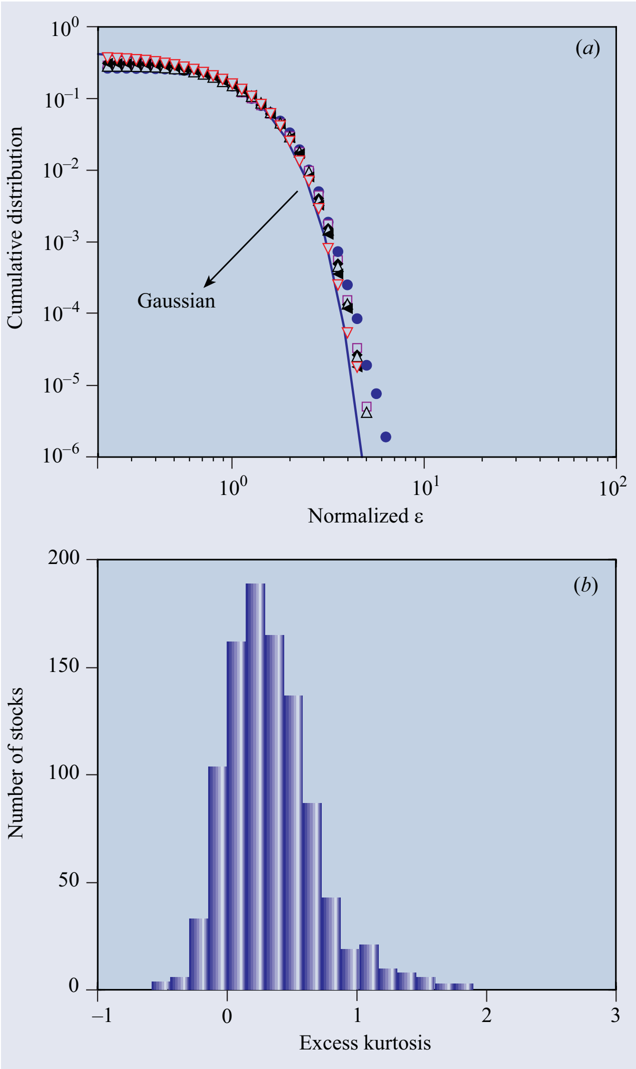

Figure 3. (*a*) Positive tail of the cumulative distribution of ϵ averaged over all stocks within each of the six groups I–VI (based on the average time between trades). Each symbol shows the average behaviour of the cumulative distributions of the scaled variable ϵ for all stocks in each of the six groups. The negative tail (not shown) displays similar behaviour. (*b*) Histogram of excess kurtosis of ϵ. For the 1000 stocks studied, we obtain the average value of kurtosis 0.46 ± 0.03 and skewness 0.018 ± 0.002.

having a finite variance $W^2_{\Delta t}$. Under this hypothesis, the central limit theorem, applied to the sum of δp_i in equation (7), implies that the ratio

$$\epsilon \equiv \frac{G_{\Delta t}}{W_{\Delta t}\sqrt{N_{\Delta t}}} \qquad (10)$$

must be a Gaussian-distributed random variable with zero mean and unit variance [38, 43]. We test this hypothesis by analysing the distribution $P(\epsilon)$ and the correlations in ϵ. Our results indicate that the distribution $P(\epsilon)$ is consistent with a Gaussian (figures 3(*a*) and (*b*)), with mean values of excess kurtosis ≈ 0.46 for all 1000 stocks [44]. This is noteworthy since, for the unconditional distribution $P(G_{\Delta t})$, the kurtosis is divergent (empirical estimates yield mean values ≈ 80 for 1000 stocks). Similar results have also been reported by recent independent studies [46–48].

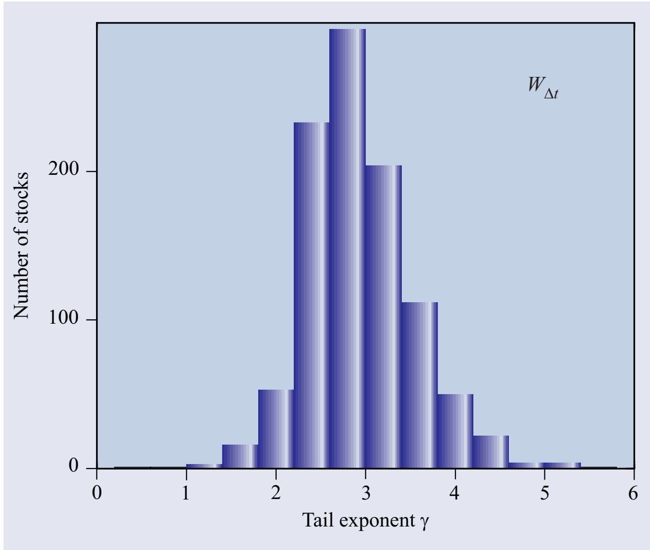

Figure 4. Histogram of power-law exponents γ for the cumulative distribution $P\{W_{\Delta t} > x\}$, obtained using Hill's estimator for each of the 1000 stocks separately. We obtain an average value $\gamma = 2.9 \pm 0.1$.

One implication of our result is that the fat tails of $P\{G_{\Delta t} > x\} \sim x^{-\alpha}$ *cannot* be caused solely due to $P\{N_{\Delta t} > x\} \sim x^{-\beta}$, because by conservation of probabilities $P\{\sqrt{N_{\Delta t}} > x\} \sim x^{-2\beta}$ with $2\beta \approx 6.8$. Equation (10) then implies that $N_{\Delta t}$ alone cannot explain the value $\alpha \approx 3$. Since $N_{\Delta t}$ is not sufficient to account for the fat tails in $G_{\Delta t}$, one other possibility is that it arises from the local variance $W^2_{\Delta t}$. By definition $W^2_{\Delta t}$ is the variance of all δp_i in Δt. We investigate the statistics of $W_{\Delta t}$ and examine if the distribution of $W_{\Delta t}$ is sufficient to explain the value of α found for $P\{G_{\Delta t} > x\}$. We find that the distribution of $W_{\Delta t}$ displays power-law asymptotic behaviour, $P\{W_{\Delta t} > x\} \sim x^{-\gamma}$. For the 1000 stocks analysed, we obtain the average value $\gamma = 2.9 \pm 0.1$ [44], consistent with the estimates of α for the same stocks (figure 4). Thus, the pronounced tails of the distribution of returns are largely due to those of $W_{\Delta t}$.

5.3. Volatility correlations and trading activity

Thus far we discussed equation (10) from the point of view of distributions. Next, we analyse time correlations in $N_{\Delta t}$, and relate them to the time correlations of $|G_{\Delta t}|$ [44]. To detect genuine long-range correlations, we first remove the marked U-shaped intradaily pattern [32, 33] in $N_{\Delta t}$ using the procedure of [30]. We find that the autocorrelation function $\langle N_{\Delta t}(t)N_{\Delta t}(t+\tau)\rangle \sim \tau^{\nu}$. We obtain the mean value of the exponents $\nu = 0.30 \pm 0.02$ for all 1000 stocks (figure 5) using the detrended fluctuation analysis method [37].

In addition, we investigate how the exponent ν of the autocorrelation function of $N_{\Delta t}$ is related to that of $|G_{\Delta t}|$. To this end, we also estimate the time correlations in $W_{\Delta t}$ and $|\epsilon|$ and find only short-range correlations. Thus, the long-range correlations in volatility ($W_{\Delta t}\sqrt{N_{\Delta t}}$) arise due to those of $N_{\Delta t}$. Indeed, our estimate of the average value of the exponent $a = 0.34 \pm 0.09$ is consistent with our estimate

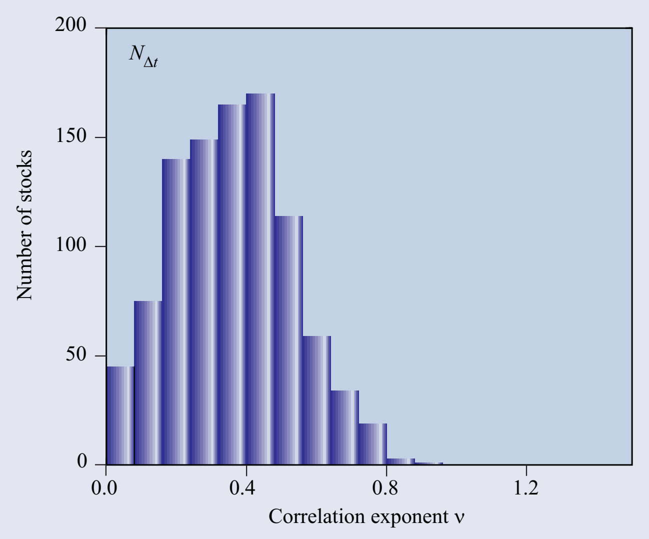

Figure 5. Histogram of exponents ν characterizing the power-law decay of the autocorrelation function $\langle N_{\Delta t}(t)N_{\Delta t}(t+\tau)\rangle$. We find the mean value for 1000 stocks $\nu = 0.30 \pm 0.02$, consistent with long memory. In order to detect genuine long-range correlations, the U-shaped intraday pattern for $N_{\Delta t}$ has been removed by dividing each $N_{\Delta t}$ by the intraday pattern [30].

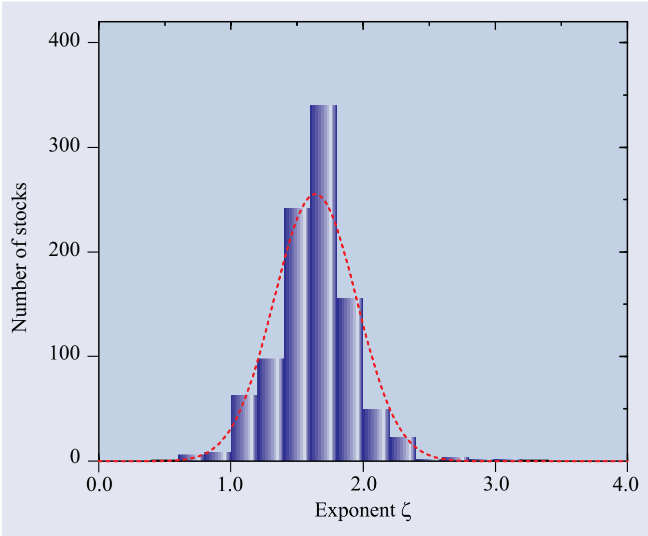

Figure 6. Histogram of the values of the tail exponent ζ of the distribution of the number of shares traded per trade $P\{q > x\}$ obtained for each of the 1000 stocks using Hill's estimator, whereby we find the average value $\zeta = 1.53 \pm 0.07$. The dotted curve shows a Gaussian distribution with the same mean and variance.

of $\nu = 0.30 \pm 0.02$. Together with the above discussion on distribution functions, these results suggest an interesting dichotomy—that the fat tails of the distribution of returns $G_{\Delta t}$ arise from $W_{\Delta t}$ and the long-range volatility correlations arise from trading activity $N_{\Delta t}$ [44].

6. Statistics of share volume traded

Understanding the equal-time correlations between volume and volatility and, more importantly, understanding how the number of shares traded impacts the price has long been a topic of active research [2, 12, 23, 38, 46, 50–53]. Using an approach similar to the above, we analyse the statistics of the number of shares $Q_{\Delta t}$ traded in a time interval Δt [54]. We find that the probability distributions $P\{Q_{\Delta t} > x\}$ display a power-law asymptotic behaviour

$$P\{Q_{\Delta t} > x\} \sim \frac{1}{x^{\lambda}} . \tag{11}$$

Using Hill's estimator, we obtain an average value $\lambda = 1.7 \pm 0.1$ [54], within the Lévy-stable domain $0 < \lambda < 2$.

We next analyse time 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 analysing the correlation function directly, we use the method of detrended fluctuation analysis [37]. We find a power-law decay of the autocorrelation function,

$$\langle [Q_{\Delta t}(t)]^a [Q_{\Delta t}(t+\tau)]^a \rangle \sim \tau^{-\kappa} . \tag{12}$$

For the parameter $a = 0.5$,[4] we obtain the average value $\kappa = 0.34 \pm 0.04$ [54]. We also investigate if the q_i are correlated in 'transaction time', defined by i, and we find only short-range correlations.

6.1. Share volume traded and number of trades

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

$$Q_{\Delta t} \equiv \sum_{i=1}^{N_{\Delta t}} q_i , \tag{13}$$

is the sum of the number of shares q_i traded for all $i = 1, \ldots, N_{\Delta t}$ transactions in Δt. Hence, we next analyse the statistical properties of the number of shares traded during each trade q_i and find that the distribution $P\{q > x\}$ displays a power-law decay

$$P\{q > x\} \sim \frac{1}{q^{\zeta}} , \tag{14}$$

where ζ has the average value $\zeta = 1.53 \pm 0.07$ (figure 6) for 1000 stocks [54].

Note that ζ is within the interval $0 < \zeta < 2$, indicating that $P\{q > x\}$ is a positive (or one-sided) distribution within the Lévy-stable domain [55].[5] Therefore, the reason why the distribution $P\{Q_{\Delta t} > x\}$ has similar asymptotic behaviour to $P\{q > x\}$, is that $P\{q > x\}$ lacks a finite second moment, and $Q_{\Delta t}$ is related to q through equation (13). Indeed, our estimate of ζ is comparable within error bounds to our estimate of λ.

[4] 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 $Q_{\Delta t}$.

[5] The general form of a characteristic function $\phi(x)$ of a Lévy-stable distribution $P(q)$ is $\ln \varphi(x) \equiv i\mu x - \gamma |x|^{\zeta} \left[1 + i\beta \frac{x}{|x|} \tan\left(\frac{\pi}{2}\zeta\right)\right]$ $[\zeta \neq 1]$, where $0 < \zeta < 2$, γ is a positive number, μ is the mean, and β is an asymmetry parameter. An inverse Fourier tranformation shows that these distributions have asymptotic behaviour described by a power-law decay $P\{q > y\} \sim y^{-\zeta}$ characterized by the exponent ζ. The case where the parameter $\beta = 1$ gives a positive or one-sided Lévy-stable distribution.

To confirm that $P(q)$ is within the Lévy-stable domain, we also examine the behaviour of $Q_n \equiv \sum_{i=1}^{n} q_i$. We first analyse the asymptotic behaviour 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 $\langle \ldots \rangle$ denotes average values. We find that the distribution $P(Q_n)$ retains its asymptotic behaviour for a range of n—consistent with a Lévy-stable distribution [54]. We obtain an independent estimate of the exponent ζ by analysing the scaling behaviour of the moments $\mu_r(n) \equiv \langle |Q_n - \langle Q_n \rangle|^r \rangle$, where $r < \zeta$.[6] 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 and obtain an inverse slope of $\zeta = 1.45 \pm 0.03$—consistent with our previous estimate of ζ [54].[7]

6.2. Share volume traded and trading activity

Since the q_i have only weak correlations, we ask how $Q_{\Delta t} \equiv \sum_{i=1}^{N_{\Delta t}} q_i$ can show much stronger correlations. To address this question, we note that (i) $N_{\Delta t}$ is long-range correlated [44], and (ii) $P(q)$ is within the Lévy-stable domain 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 equation (13), 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 [54]

$$\chi \equiv \frac{Q_{\Delta t} - \langle q \rangle N_{\Delta t}}{N_{\Delta t}^{1/\lambda}}, \qquad (15)$$

where χ is a one-sided Lévy-distributed variable with zero mean and exponent ζ [55].[8] Since $N_{\Delta t}$ is a random variable, one crucial condition for the hypothesis (equation (15)) to be valid is that the distribution $P\{N_{\Delta t} > x\}$ must have a tail that decays much more rapidly than that of q_i [56], which is indeed what we have seen in the previous sections (our estimates of the tail exponent of $N_{\Delta t}$, β are significantly larger than those of ζ).[9] To test the hypothesis of equation (15), we first analyse $P(\chi)$ and find similar asymptotic behaviour to $P(Q_{\Delta t})$. Next, we analyse correlations in χ and find only weak correlations—implying that the correlations in $Q_{\Delta t}$ are largely due to those of $N_{\Delta t}$ [54]. Indeed, our estimates of the exponent $\kappa = 0.34 \pm 0.04$ is consistent with our estimate of the exponent $\nu = 0.30 \pm 0.02$ for $N_{\Delta t}$.

6.3. Price fluctuations and share volume traded

One interesting implication of these results is an explanation for the previously observed [12,50–53] 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, $V_{\Delta t} = W_{\Delta t} \sqrt{N_{\Delta t}}$ from equation (10). 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 if 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 coefficients ≈ 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. Thus, the previously observed volume–volatility co-movement is largely due to $N_{\Delta t}$—consistent with the results of [50] that, on average, the size of trades has no more information content than that contained in the number of transactions.

Another implication concerns the relationship between price fluctuations $G_{\Delta t}$ and order-imbalances $\phi \equiv q_b - q_s$, defined as the difference between the number of shares traded in buyer-initiated trades q_b and seller-initiated trades q_s [2]. Since the tail statistics of the distribution of returns $P\{G_{\Delta t} > x\}$ has a power-law asymptotic behaviour with exponent $\alpha \approx 3$, and that of the number of shares traded $P\{q > x\}$ has the exponent $\zeta \approx 1.7$, one can rule out a linear relationship between $G_{\Delta t}$ and ϕ. In order for the statistics to be consistent, the relationship has to be 'weaker' than linear, such as a square-root relationship [57]. This is consistent with empirical findings ([2] and references therein), which show a nonlinear relationship between price fluctuations and order imbalances.

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

[7] To avoid the effect of weak correlations in q on the estimate of ζ, the moments $[\mu_r(n)]^{1/r}$ are constructed after randomizing each time series of q_i. Without randomizing, the same procedure gives an estimate of $\zeta = 1.31 \pm 0.03$.

[8] See footnote 5.

[9] If $N_{\Delta t}$ were independent, it can be proved that this condition is sufficient for equation (15). However, in the presence of long-range correlations in $N_{\Delta t}$, although an exact proof is beyond the scope of this paper, based on numerical evidence, it seems reasonable to conjecture the sufficiency of this condition.

References

[1] Bachelier L 1900 Théorie de la spéculation *Annales Scientifiques de l'Ecole Normale Supérieure III-17* **21** *PhD Thesis in mathematics*

[2] Campbell J, Lo A W and MacKinlay A 1997 *The Econometrics of Financial Markets* (Princeton, NJ: Princeton University Press)

[3] Pagan A 1996 The econometrics of financial markets *J. Empirical Finance* **3** 15–102

[4] Farmer J D 1999 Physicists attempt to scale the ivory towers of finance *Comput. Sci. Engng* **1** 26–39

[5] Mantegna R N and Stanley H E 2000 *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge: Cambridge University Press)

[6] Bouchaud J-P and Potters M 2000 *Theory of Financial Risk* (Cambridge: Cambridge University Press)

[7] Mandelbrot B B 1963 The variation of certain speculative prices *J. Business* **36** 394–419

[8] Fama E and Roll R 1971 Parameter estimates for symmetric stable distributions *J. Am. Statist. Assoc.* **66** 331–38

[9] Officer R R 1972 The distribution of stock returns *J. Am. Statist. Assoc.* **67** 807–12

[10] Praetz P D 1972 The distribution of share price changes *J. Business* **45** 49–55

[11] Blattberg R C and Gonedes N 1974 A comparison of the stable Paretian and Student distributions as statistical models for prices *J. Business* **47** 244–80

[12] Clark P K 1973 A subordinated stochastic process model with finite variance for speculative prices *Econometrica* **41** 135–55

[13] Mantegna R N and Stanley H E 1995 Scaling behavior in the dynamics of an economic index *Nature* **376** 46–9

[14] Mantegna R N and Stanley H E 1994 Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight *Phys. Rev. Lett.* **73** 2946

[15] Gopikrishnan P, Meyer M, Amaral L A N and Stanley H E 1998 Inverse cubic law for the distribution of stock price variations *Eur. Phys. J.* B **3** 139–40

[16] Gopikrishnan P, Plerou V, Amaral L A N, Meyer M and Stanley H E 1999 Scaling of the distributions of fluctuations of financial market indices *Phys. Rev.* E **60** 5305
[17] Plerou V, Gopikrishnan P, Amaral L A N, Meyer M and Stanley H E 1999 Scaling of the distribution of price fluctuations of individual companies *Phys. Rev.* E **60** 6519–29
[18] Loretan M and Phillips P C B 1994 Testing the covariance stationarity of heavy-tailed time series *J. Empirical Finance* **1** 211–48
[19] Lux T 1996 The stable Paretian hypothesis and the frequency of large returns: an examination of major German stocks *Appl. Financial Economics* **6** 463–75
[20] Muller U A, Dacorogna M M and Pictet O V 1998 Heavy tails in high-frequency financial data *A Practical Guide to Heavy Tails* ed R J Adler, R E Feldman and M S Taqqu (Basle: Birkhäuser) pp 83–311
[21] Hill B M 1975 A robust estimator for the asymptotic behavior of certain time series *Ann. Math. Stat.* **3** 1163
[22] Jansen D and de Vries C 1991 On the frequency of large stock returns: putting booms and busts into perspective *Rev. Econ. Stat.* **73** 18–24
[23] Haussman J, Lo A and MacKinlay A C 1992 An ordered probit analysis of stock transaction prices *J. Finance Economics* **31** 319–79
[24] Lo A and MacKinlay A C 1988 Stock prices do not follow random walks: evidence from a simple specification test *Rev. Financial Studies* **1** 41–66
[25] Roll R 1984 A simple implicit measure of the effective bid–ask spread in an efficient market *J. Finance* **39** 1127–40
[26] Fama E F 1970 Efficient capital markets: a review of theory and empirical work *J. Finance* **25** 383–420
[27] Fama E F 1991 Efficient capital markets: II *J. Finance* **46** 1575–617
[28] Liu Y , Cizeau P, Meyer M, Peng C-K and Stanley H E 1997 Quantification of correlations in economic time series *Physica* A **245** 437–40
[29] Cizeau P, Liu Y, Meyer M, Peng C-K and Stanley H E 1997 Volatility distribution in the S&P 500 stock index *Physica* A **245** 441–5
[30] Liu Y, Gopikrishnan P, Cizeau P, Meyer M, Peng C-K and Stanley H E 1999 The statistical properties of the volatility of price fluctuations *Phys. Rev.* E **60** 1390–400
[31] Beran J 1994 *Statistics for Long-Memory Processes* (New York: Chapman & Hall)
[32] Wood R A, McInish T H and Ord J K 1985 An investigation of transactions data for NYSE stocks *J. Finance* **40** 723–739
[33] Admati A and Pfleiderer P 1988 A theory of intraday patterns: volume and price variability *Rev. Financial Studies* **1** 723–39
[34] Ding Z, Granger C W J and Engle R F 1993 A long memory property of stock market returns and a new model *J. Empirical Finance* **1** 83–105
[35] Granger C W J and Ding Z 1996 Varieties of long memory models *J. Econometrics* **73** 61
[36] Anderson T, Bollerslev T, Diebold F and Labys P 1999 The distribution of exchange rate volatility *NBER Working Paper WP6961*
[37] Peng C-K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L 1994 Mosaic organization of DNA nucleotides *Phys. Rev.* E **49** 1685–89
[38] Tauchen G and Pitts M 1983 The price variability–volume relationship on speculative markets *Econometrica* **57** 485–505
[39] Stock J 1988 Estimating continuous time processes subject to time deformation *J. Am. Statist. Assoc.* **83** 77–85
[40] Lamoureux C G and Lastrapes W D 1990 Heteroskedasticity in stock return data: volume versus GARCH effects *J. Finance* **45** 221–9
[41] Ghysels E and Jasiak J Stochastic volatility and time deformation: an application to trading volume and leverage effects *Preprint CRDE, Universite de Montreal*
[42] Engle R F and Russell J R 1999 Forecasting transaction rates: the autoregressive conditional duration model *Econometrica* **67** 387–402
[43] Feller W 1966 *An Introduction to Probability Theory and its Applications* (New York: Wiley)
[44] Plerou V, Gopikrishnan P, Amaral L A N, Gabaix X and Stanley H E 2000 Diffusion and economic fluctuations *Phys. Rev.* E **62** R3023
[45] Guillaume D M, Pictet O V, Muller U A and Dacorogna M M 1995 Unveiling non-linearities through time scale transformations *Preprint Olsen group OVP.1994-06-26* available at http://www.olsen.ch
[46] Ane T and Geman H 2000 Order flow, transaction clock and normality of asset returns *J. Finance* **55** 2259–84
[47] Anderson T, Bollerslev T, Diebold F and Ebens H 2000 The distribution of stock return volatility *NBER Working Paper WP7933*
[48] Anderson T G, Bollerslev T, Diebold F and Labys P 2000 Exchange rate returns standardized by realized volatility are (nearly) Gaussian *NBER Working Paper WP7488*
[49] Mandelbrot B B and Taylor H 1962 On the distribution of stock price differences *Operations Res.* **15** 1057–62
[50] Jones C, Gautam K and Lipson M 1994 Transactions, volumes and volatility *Rev. Financial Studies* **7** 631–51
[51] Gallant A R, Rossi and Tauchen 1992 Stock prices and volume *Rev. Financial Studies* **5** 199
[52] Karpoff J 1987 Price variability and volume: a review *J. Financial Quantitative Anal.* **22** 109
[53] Epps T W and Epps M L 1676 The stochastic dependence of security price changes and transaction volumes: implications of the mixture-of-distributions hypothesis *Econometrica* **44** 305–21
[54] Gopikrishnan P, Plerou V, Gabaix X and Stanley H E 2000 Statistical properties of share volume traded in financial markets *Phys. Rev.* E **62** 4493–6
[55] Shlesinger M F *et al* (ed) 1995 *Lévy Flights and Related Topics in Physics* (Berlin: Springer)
[56] Mittnik S and Rachev S 1993 Modeling asset returns with alternative stable distributions *Econometric Rev.* **12** 261–330
[57] Zhang Y C 1999 Towards a theory of marginally efficient markets *Physica* A **269** 30–9