Stock return distributions: Tests of scaling and universality from three distinct stock markets

Vasiliki Plerou and H. Eugene Stanley
Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA
(Received 8 July 2007; revised manuscript received 3 February 2008; published 19 March 2008)

We examine the validity of the power-law tails of the distributions of stock returns $P\{R>x\}\sim x^{-\zeta_R}$ using trade-by-trade data from three *distinct* markets. We find that both the negative as well as the positive tails of the distributions of returns display power-law tails, with mutually consistent values of $\zeta_R\approx 3$ for all three markets. We perform similar analyses of the related microstructural variable, the number of trades $N\equiv N_{\Delta t}$ over time interval Δt, and find a power-law tail for the cumulative distribution $P\{N>x\}\sim x^{-\zeta_N}$, with values of ζ_N that are consistent across all three markets analyzed. Our analysis of U.S. stocks shows that the exponent values ζ_R and ζ_N do not display systematic variations with market capitalization or industry sector. Moreover, since ζ_R and ζ_N are remarkably similar for all three markets, our results support the possibility that the exponents ζ_R and ζ_N are universal.

DOI: 10.1103/PhysRevE.77.037101 PACS number(s): 89.65.−s, 05.45.Tp

Define the "log return" of stock price $S(t)$ over a time interval Δt: $R\equiv R_{\Delta t}(t)=\log S(t+\Delta t)-\log S(t)$ [1]. Analyses of returns of individual stocks [2,3] and stock indices [4] have shown that the cumulative distribution of returns is well fit by a power-law asymptotic behavior with tail exponent ζ_R,

$$P\{R>x\}\sim x^{-\zeta_R}. \tag{1}$$

Based on analysis of tick data for the 1000 largest USA stocks, Refs. [3,4] report values of $\zeta_R\approx 3$ for both the positive and negative tails for time scales $\Delta t<1$ day up to a few weeks. Qualitatively similar results for currency exchange fluctuations as well as stocks can be found in Refs. [5–13].

The tail exponent $\zeta_R\approx 3$ [Eq. (1)] implies that $P(R>x)$ does not belong to the family of Lévy stable distributions which requires $\zeta_R<2$ [14]. Thus it is particularly interesting that the values of ζ_R seem similar for a wide range of stocks and time scales.

Our goal is to understand whether the dispersion in the measured exponent ζ_R across different stocks reflects statistical variations around a "universal" value or genuine variations from stock to stock and market to market. We show that the exponent ζ_R is universal in the following respects: (a) We find that ζ_R does not show significant dependence on market capitalization and (b) ζ_R does not show any systematic variations with industry sector. We further extend our analysis to two quite different markets—the London Stock Exchange (LSE) and the Paris Bourse—that show the validity of the power-law distribution Eq. (1) for these markets with similar estimates for the exponent ζ_R. Moreover, performing the same analysis on a related and equally important variable, the number of trades $N\equiv N_{\Delta t}(t)$ in the time interval Δt, we show that the exponent ζ_N describing the tails of the distribution $P\{N>x\}\sim x^{-\zeta_N}$ is universal in the same way as ζ_R. This work complements recent work which reports universal behavior of the power-law exponents of the distribution of trade sizes and volume [15].

We analyze detailed trade-by-trade data from three *distinct* markets (same data set analyzed in Ref. [15]): (a) 1000 major USA stocks for the 2-yr period 1994–1995 ($\approx 10^8$ records), (b) 85 major stocks traded on the London Stock Exchange for the 2-yr period 2001–2002 which form part of the FTSE 100 index ($\approx 4\times 10^7$ records), and (c) 13 major stocks traded on the Paris Bourse that form part of the CAC 40 index for the 4.7-yr period 3 Jan 1995–22 Oct 1999 ($\approx 2\times 10^7$ records). To examine the behavior of the distributions over a larger time horizon, we analyze daily data from (d) the CRSP database for 422 stocks for the 35-yr period Jan 1962–Dec 1996.

For each of the 85 UK stocks, we examine the cumulative distribution of returns for $\Delta t=5$ min [Fig. 1(a)]. For each stock i, we find that the cumulative distribution is consistent with a power law [Eq. (1)] with exponent ζ_{R_i}. Figure 1(b) shows the exponent estimates obtained using Hill's estimator for the positive and negative tails for each of the 85 stocks. We obtain mean values

$$\zeta_R^{\mathrm{LSE}}=\begin{cases}2.96\pm 0.05 & [\text{positive tail}],\\ 2.88\pm 0.04 & [\text{negative tail}].\end{cases} \tag{2}$$

Note that these results are consistent with previous results for USA stocks [3] and indices [4].

We find similar results for each of the 13 Paris Bourse stocks. Figure 1(c) shows that the cumulative distribution $P\{R>x\}$ displays power-law tails as in Eq. (1). We obtain the mean values

$$\zeta_R^{\mathrm{Bourse}}=\begin{cases}3.13\pm 0.08 & [\text{positive tail}],\\ 3.03\pm 0.06 & [\text{negative tail}],\end{cases} \tag{3}$$

which are consistent with previous findings for the USA data [3] and the above results for the LSE data.

One of the striking features of the exponent ζ_R in Eq. (1) is that the estimates of ζ_R seem to be similar for all 1000 stocks analyzed in Ref. [3], with some dispersion around $\zeta_R\approx 3$ over intraday time scales. Our analysis extends these results and shows consistent values among all stocks for both the LSE and the Paris Bourse data, raising the interesting possibility that the actual value of the exponent ζ_R is "universal" and the dispersion in exponent values may be purely statistical around a "true" value. To investigate this possibility further, we examine for the USA data the dependence of ζ_R estimates on stock-specific factors such as industry sector and market capitalization.

Figure 2 shows for each of the 1000 USA stocks the estimates of ζ_R for the positive and negative tails plotted as functions of the average market capitalization of the stock in

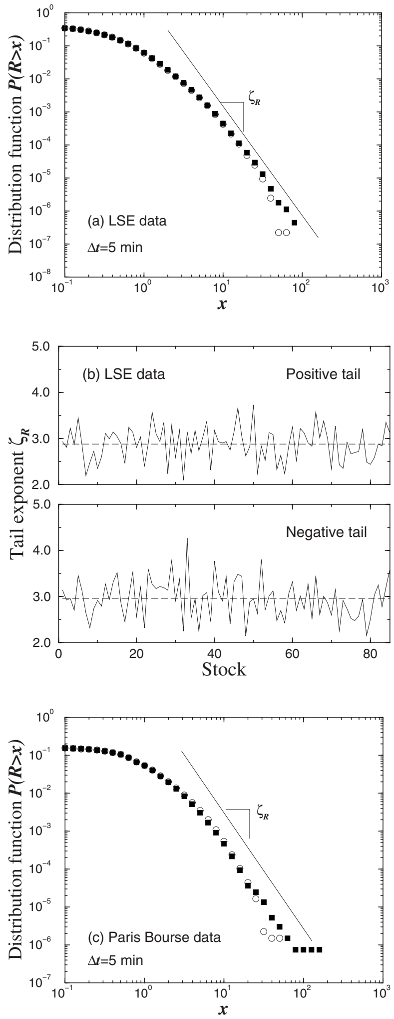

FIG. 1. (a) Cumulative distribution function $P\{R>x\}$ for the 85 largest stocks that form part of the FTSE 100 index and survived through the 2-yr period 2001–2002. Here the $\Delta t=5$ min returns of each stock have been normalized to zero mean and unit variance. (b) Estimates of the exponent ζ_R obtained using Hill's method. Exponent estimates for the positive tail (top panel) and negative tail (bottom panel) for which we obtain mean values $\zeta_R=2.96\pm0.05$ and $\zeta_R=2.88\pm0.04$ for the positive and negative tails, respectively. (c) Cumulative distribution function $P(R>x)$ for the 13 Paris Bourse stocks that form part of the CAC40 index for the 4-yr period 1995–1999. Here the $\Delta t=5$ min returns of each stock have been normalized to zero mean and unit variance.

the 2-yr period. We find only a statistical dispersion around the mean value with no systematic dependence on market capitalization [cf. caption of Fig. 2].

To analyze the dependence on the industry sector, we examine the exponent estimates for each stock as functions of the corresponding industry sector. To categorize each stock by industry sector, we use the Standard Industry Classification (SIC) code [25]. Figure 3 shows that ζ_R as a function of the first two digits of the SIC code (which shows major industry sectors) displays only a narrow scatter around the dashed lines which represent the mean value—consistent with the possibility that all individual distributions are characterized by the same power-law exponent.

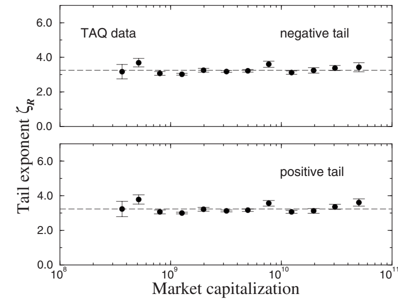

FIG. 2. Estimates of exponent ζ_R that describe the tail behaviors of the $\Delta t=15$ min returns for 1000 largest U.S. stocks from the TAQ database shows no clear dependence on market capitalization. Each point shows the average value of ζ_R for each market capitalization group, and the groups are spaced uniformly in logarithmic scale. The regression $A+B\log x$ gives an estimate of $B=-0.04\pm0.04$ (positive tail) and -0.01 ± 0.04 (negative tail) with negligible values of R^2.

Next we focus on the statistical properties of the number of trades $N\equiv N_{\Delta t}(t)$ in the interval Δt. The statistics of N is important for understanding the behavior of returns and share volume [17–24]. Analysis [16] of the statistics of N for the 1000 largest U.S. stocks (same as the USA data in our analysis) shows that the simplistic view of describing the dynamics of N by a Poisson process is not consistent in the following two respects: (a) The cumulative distribution $P\{N>x\}$ is found to display a power-law tail and (b) N displays long-range correlations that decay as a power law. Here we analyze the LSE and Paris Bourse data. Moreover, for the USA data, we analyze the dependence of the tail exponent ζ_N on market capitalization and industry sector.

Previous work [16] reports that the number of trades in Δt displays a power-law asymptotic behavior,

$$P\{N>x\} \sim x^{-\zeta_N}, \tag{4}$$

with $\zeta_N^{\mathrm{USA}}=3.40\pm0.05$.

To test whether the power-law distribution of N holds for other markets, we first examine the tick-by-tick data for the

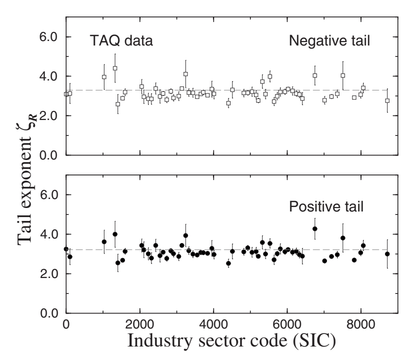

FIG. 3. Tail exponent as a function of the SIC code shows no clear dependence on the industry sector. Here we have binned using the first two digits of the SIC code [25] which shows major industry sectors. Points farthest from the mean have large standard errors and occur when only a few stocks contribute. The points at SIC code 0 show the 73 stocks in our sample of 1000 for which we did not have the corresponding SIC codes.

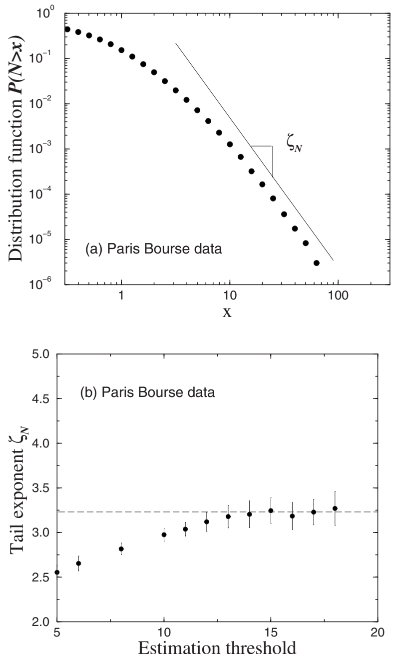

FIG. 4. (a) Cumulative distribution $P\{N>x\}$ for 13 stocks which form part of the CAC40 index of the Paris Bourse for the 4-yr period (1995–99). Power-law regression gives the value $\zeta_N = 3.24 \pm 0.06$ ($x>10$). (b) Tail exponent obtained using Hill estimator as a function of estimation threshold shows an increase followed by a plateau behavior around $\zeta_N = 3.25$.

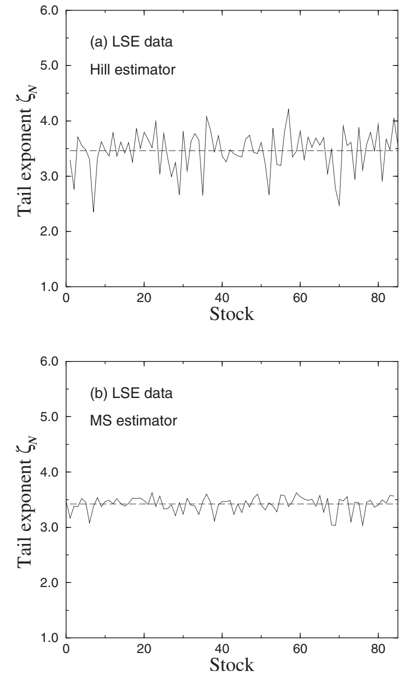

FIG. 5. (a) Estimates of the exponent ζ_N for each stock in our sample using Hill's estimator with an estimation threshold of 4 yields an average exponent of $\zeta_N = 3.46 \pm 0.04$. For each stock N has been normalized by its first centered moment. Performing the same estimation for larger thresholds gives consistently larger exponents $\zeta_N = 3.90 \pm 0.05$ and $\zeta_N = 4.41 \pm 0.09$ for five and seven times the average, respectively, although this increase is an artifact of the estimator [15]. (b) To obtain a reliable estimate of the exponent ζ_N, we use the MS estimator which does not rely on an estimation threshold and obtain a mean value $\zeta_N = 3.42 \pm 0.02$.

Paris Bourse stocks. We use data for the 13 largest stocks that are part of the CAC 40 index and survived through the ≈4.7-yr period 4 January 1995–22 October 1999.

For each stock, we find that the cumulative distribution $P\{N>x\}$ displays a power-law functional form for large x. The exponents characterizing the distribution are consistent across all stocks and scaling N by its first centered moment we find good data collapse. Under the assumption that the underlying distributions are identical, we use the scaled data to improve the tail statistics. Figure 4(a) shows that the distribution is consistent with a power law as in Eq. (4) with an exponent $\zeta_N^{\mathrm{Bourse}} = 3.24 \pm 0.06$.

Analyzing the Hill exponent estimate as a function of the number of tail events used in the estimation procedure, we find a region around $\zeta_N \approx 3.25$ around the tail, showing that the exponent estimate is robust with the number of tail events. Figure 4(d) shows the analogous plot where the Hill exponent estimate is plotted as a function of the estimation threshold instead of the number of tail events. We find an increase of the exponent estimate ζ_N followed by a plateau around

$$\zeta_N^{\mathrm{Bourse}} = 3.23 \pm 0.12. \tag{5}$$

We use the threshold-independent Meerschaert-Scheffler (MS) estimator [26] to obtain a separate estimate of the tail exponent and find $\zeta_N^{\mathrm{Bourse}} = 2.99 \pm 0.03$.

We next analyze the statistics of the number of trades, using the UK LSE database. For each of the 85 stocks, we construct a time series of $N_{\Delta t}$ for $\Delta t = 5$ min. We find that individual distribution functions $P\{N>x\}$ show power-law asymptotic behavior consistent with Eq. (4). We apply Hill's estimator to the 85 stocks in our database and find similar exponent values for each stock [Fig. 5(a)]. We obtain a mean exponent value

$$\zeta_N^{\mathrm{LSE}} = 3.46 \pm 0.04, \tag{6}$$

which is consistent with the behavior of ζ_N for the USA data in Eq. (6). Since Hill's estimator displays a dependence on the estimation threshold, we apply the MS estimator to the data to obtain a threshold-independent estimate for the exponent ζ_N. Figure 5(b) shows the MS estimate of ζ_N for the same set of stocks and finds a mean value $\zeta_N^{\mathrm{LSE}} = 3.42 \pm 0.02$.

We next test whether the estimates of the exponent ζ_N show any systematic variations with stock-specific variables such as market capitalization and industry sector. Figure 6(a) shows that the exponent estimates of ζ_N for $\Delta t = 15$ min obtained from the TAQ database shows no statistically significant dependence on the market capitalization. Although a log-linear regression $y = A \log x + B$ gives a slope $A = 0.10 \pm 0.02$, the statistical significance of this relation is small as indicated by a negligible $R^2 \approx 0.02$. We perform the same regression using exponent estimates obtained from the MS estimator and find similar results: A slightly positive slope with little statistical significance. Based on our statis-

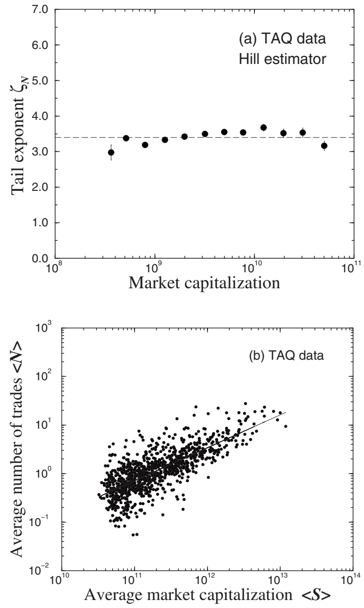

FIG. 6. (a) Hill estimates of the tail exponents ζ_N plotted against the average market capitalization for the TAQ data. Here we have binned logarithmically in market capitalization so that each point represents the average ζ_N for stocks belonging to that group. A logarithmic regression $y=A \log x+B$ on the data without any binning gives $B=0.10\pm0.02$ with negligible $R^2=0.02$. We repeat this test using ζ_N from the MS estimator and find no dependence on market capitalization. (b) Average number of trades over the interval $\Delta t=5$ min for USA stocks as a function of average market capitalization in the 2-yr period 1994–95.

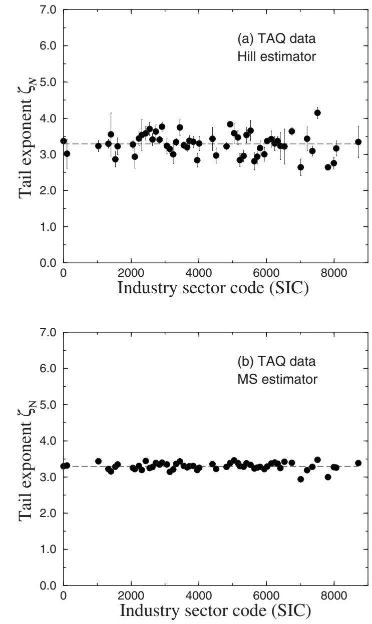

FIG. 7. (a) Estimates of the tail exponent ζ_N obtained using Hill's estimator as a function of the SIC codes which denote industry sector. Here we have grouped stocks by the first two digits of the SIC code which shows major industry sectors. (b) Same as (a) but using the MS [26] estimator shows a less dispersed plot confirming the lack of dependence on industry sector as seen in (a).

tics, we conclude that ζ_N does not show any systematic variations with market capitalization.

Although the functional form Eq. (4) and the exponent values do not show significant dependence on market capitalization $\langle S\rangle$, the average number of trades $\langle N\rangle$ for each stock displays a power-law dependence on market capitalization $\langle N\rangle \sim S^{\beta}$ with an exponent $\beta=0.68\pm0.02$. Similar results can be found in Ref. [27].

Figure 7(a) shows that ζ_N obtained using Hill's estimator [28] does not display any significant dependence on industry sectors. Similar results are found using the MS estimator [Fig. 7(b)].

[1] *Proceedings of The Third Nikkei Econophysics Symposium*, edited by H. Takayasu (Springer, New York, 2006).
[2] T. Lux, Appl. Financ. Econ. **6**, 463 (1996).
[3] V. Plerou, P. Gopikrishnan, L. A. NunesAmaral, M. Meyer, and H. E. Stanley, Phys. Rev. E **60**, 6519 (1999); P. Gopikrishnan *et al.*, Eur. Phys. J. B **3**, 139 (1998).
[4] P. Gopikrishnan, V. Plerou, L. A. NunesAmaral, M. Meyer, and H. E. Stanley, Phys. Rev. E **60**, 5305 (1999).
[5] R. R. Officer, J. Am. Stat. Assoc. **67**, 807 (1972).
[6] P. D. Praetz, J. Business **45**, 49 (1972).
[7] R. C. Blattberg *et al.*, J. Business **47**, 244 (1974).
[8] D. Jansen and C. de Vries, Rev. Econ. Stat. **73**, 18 (1991).
[9] M. Loretan and P. C. B. Phillips, J. Empirical Finance **1**, 211 (1994).
[10] A. Pagan, J. Empirical Finance **3**, 15 (1996).
[11] U. A. Muller *et al.*, in *A Practical Guide to Heavy Tails*, edited by R. J. Adler *et al.* (Birkhäuser Publishers, Boston, 1998), p. 83.
[12] S. Sinha and R. K. Pan, e-print arXiv:physics/0605247.
[13] X. Gabaix *et al.*, Nature (London) **423**, 267 (2003); Quart. J. Econom. **121**, 461 (2006).
[14] B. B. Mandelbrot, J. Business **36**, 394 (1963).
[15] V. Plerou and H. E. Stanley, Phys. Rev. E **76**, 046109 (2007).
[16] V. Plerou, P. Gopikrishnan, L. A. NunesAmaral, X. Gabaix, and H. E. Stanley, Phys. Rev. E **62**, R3023 (2000).
[17] B. B. Mandelbrot and H. Taylor, Oper. Res. **15**, 1057 (1967).
[18] P. K. Clark, Econometrica **41**, 135 (1973).
[19] T. W. Epps and M. L. Epps, Econometrica **44**, 305 (1976).
[20] T. Ane and H. Geman, J. Financ. **55**, 2259 (2000).
[21] P. Gopikrishnan, V. Plerou, X. Gabaix, and H. E. Stanley, Phys. Rev. E **62**, R4493 (2000).
[22] G. Tauchen and M. Pitts, Econometrica **51**, 485 (1983).
[23] J. Stock, J. Am. Stat. Assoc. **83**, 77 (1988).
[24] C. Hopman, "Are Demand and Supply Driving Stock Price?," working paper, MIT Dept. of Economics, 2002 (unpublished).
[25] http://www.osha.gov/pls/imis/sic-manual.html.
[26] M. Meerschaert and H. Scheffler, J. Stat. Plan. Infer. **71**, 19 (1998).
[27] G. Zumbach, Quant. Finance **4**, 441 (2004); X. Gabaix *et al.* (unpublished).
[28] B. M. Hill, Ann. Stat. **3**, 1163 (1975).