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 $P(R > x) \sim x^{-\xi_R}$ 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 $\xi_R \approx 3$ for all three markets. We perform similar analyses of the related microstructural variable, the number of trades $N = N_\Delta$ over time interval $\Delta t$, and find a power-law tail for the cumulative distribution $P(N > x) \sim x^{-\xi_N}$, with values of $\xi_N$ that are consistent across all three markets analyzed. Our analysis of U.S. stocks shows that the exponent values $\xi_R$ and $\xi_N$ do not display systematic variations with market capitalization or industry sector. Moreover, since $\xi_R$ and $\xi_N$ are remarkably similar for all three markets, our results support the possibility that the exponents $\xi_R$ and $\xi_N$ are universal.

DOI: 10.1103/PhysRevE.77.037101

Define the “log return” of stock price $S(t)$ over a time interval $\Delta t$: $R = R_\Delta(t) = \log S(t+\Delta t) - \log S(t)$ [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 $\xi_R$.

\[ P(R > x) \sim x^{-\xi_R}. \]  

(1)

Based on analysis of tick data for the 1000 largest USA stocks, Refs. [3,4] report values of $\xi_R \approx 3$ 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 $\xi_R \approx 3$ [Eq. (1)] implies that $P(R > x)$ does not belong to the family of Lévy stable distributions which requires $\xi_R < 2$ [14]. Thus it is particularly interesting that the values of $\xi_R$ seem similar for a wide range of stocks and time scales.

Our goal is to understand whether the dispersion in the measured exponent $\xi_R$ 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 $\xi_R$ is universal in the following respects: (a) We find that $\xi_R$ does not show significant dependence on market capitalization and (b) $\xi_R$ 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 $\xi_R$. Moreover, performing the same analysis on a related and equally important variable, the number of trades $N = N_\Delta(t)$ in the time interval $\Delta t$, we show that the exponent $\xi_N$ describing the tails of the distribution $P(N > x) \sim x^{-\xi_N}$ is universal in the same way as $\xi_R$.

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 ($\approx 10^6$ 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).

We examine the validity of the power-law tails of the distributions 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 $\xi_R$. 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

\[ \xi_R^{\text{LSE}} = \begin{cases} 
2.96 \pm 0.05 & \text{[positive tail]}, \\
2.88 \pm 0.04 & \text{[negative tail]}. 
\end{cases} \]  

(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

\[ \xi_R^{\text{Bourse}} = \begin{cases} 
3.13 \pm 0.08 & \text{[positive tail]}, \\
3.03 \pm 0.06 & \text{[negative tail]}, 
\end{cases} \]  

(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 $\xi_R$ in Eq. (1) is that the estimates of $\xi_R$ seem to be similar for all 1000 stocks analyzed in Ref. [3], with some dispersion around $\xi_R \approx 3$ 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 $\xi_R$ 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 $\xi_R$ 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 $\xi_R$ for the positive and negative tails plotted as functions of the average market capitalization of the stock in

1539-3755/2008/77(3)/037101(4) 037101-1 ©2008 The American Physical Society
the first two digits of the SIC code industry sectors andponent estimates for the positive tail
BRIEF REPORTS PHYSICAL REVIEW E 77, 037101 (2008)
acterized by the same power-law exponent.
with the possibility that all individual distributions are char-
dashed lines which represent the mean value—consistent
the mean value with no systematic dependence on market
2-yr period. We find only a statistical dispersion around
the corresponding industry sector. To categorize each stock
amine the exponent estimates for each stock as functions of
Bourse stocks that form part of the CAC40 index for the 4-yr period
through the 2-yr period 2001–2002. Here the
cf. caption of Fig. 2
obtained using Hill’s method. Ex-
ponent estimates for the positive tail (top panel) and negative tail (bottom panel) for which we obtain mean values $\xi_R=2.96\pm0.05$ and $\xi_B=2.88\pm0.04$ 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 ex-
amine 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 $\xi_R$ 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 char-
acterized by the same power-law exponent.

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 $\xi_R$ obtained using Hill’s method. Exponent estimates for the positive tail (top panel) and negative tail (bottom panel) for which we obtain mean values $\xi_R=2.96\pm0.05$ and $\xi_B=2.88\pm0.04$ 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.

Next we focus on the statistical properties of the number of
trades $N=N_{\Delta t}(t)$ in the interval $\Delta t$. 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 $\xi_N$ on market capitalization and industry sector.

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

$$P(N>x) \sim x^{-\xi_N},$$

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

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

FIG. 2. Estimates of exponent $\xi_R$ 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 $\xi_R$ 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 $-0.01\pm0.04$ (negative tail) with negligible values of $R^2$.

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.
around find a region around number of tail events used in the estimation procedure, we the exponent estimate is robust with the number of tail increase of the exponent estimate threshold instead of the number of tail events. We find an exponent estimate is plotted as a function of the estimation construct a time series of using the UK LSE database. For each of the 85 stocks, we that are part of the CAC 40 index and survived through the that are part of the CAC 40 index and survived through the 4-yr period 4 January 1995–22 October 1999. We use the threshold-independent Meerschaert-Scheffler distribution is consistent with a power law as in Eq. MS estimator and find similar results: A slightly positive 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

\[ \zeta_{N,SE} = 3.46 \pm 0.04, \]  

which is consistent with the behavior of \( \zeta_N \) 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 \( \zeta_N \). Figure 5(b) shows the MS estimate of \( \zeta_N \) for the same set of stocks and finds a mean value \( \zeta_{N,MS} = 3.42 \pm 0.02 \).

We next test whether the estimates of the exponent \( \zeta_N \) show any systematic variations with stock-specific variables such as market capitalization and industry sector. Figure 6(a) shows that the exponent estimates of \( \zeta_N \) for \( \Delta t=5 \text{ min} \) 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 \( R^2=0.02 \). 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 $\zeta_N$ 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 $\zeta_N$ for stocks belonging to that group. A logarithmic regression $y = A \log x + B$ on the data without any binning gives $B = 0.10 \pm 0.02$ with negligible $R^2 \approx 0.02$. We repeat this test using $\zeta_N$ 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. Thus, we conclude that $\zeta_N$ 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 $\langle N \rangle = S^\beta$ with an exponent $\beta = 0.68 \pm 0.02$. Similar results can be found in Ref. [27]. Figure 7(a) shows that $\zeta_N$ 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)].