International Journal of Theoretical and Applied Finance
Vol. 3, No. 3 (2000) 365–369

THE DISTRIBUTION OF RETURNS OF STOCK PRICES

LUÍS A. N. AMARAL*, VASILIKI PLEROU, PARAMESWARAN GOPIKRISHNAN, MARTIN MEYER and H. EUGENE STANLEY

Center for Polymer Studies and Department of Physics,
Boston University, Boston, MA 02215, USA
E-mail: amaral@buphy.bu.edu

We perform a phenomenological study of stock price fluctuations of individual companies. We systematically analyze two different databases covering securities from the three major US stock markets. We consider (i) the trades and quotes (TAQ) database, for which we analyze 40 million records for 1000 US companies for the 2-year period 1994–95, and (ii) the Center for Research and Security Prices (CRSP) database, for which we analyze 35 million daily records for approximately 16,000 companies in the 35-year period 1962–96. We study the probability distribution of returns over varying time scales — from 5 min up to 4 years. For time scales from 5 min up to approximately 16 days, we find that the tails of the distributions can be well described by a power-law decay, characterized by an exponent $\alpha \approx 3$ — well outside the stable Lévy regime $0 < \alpha < 2$. For time scales greater than 16 days, we observe results consistent with a slow convergence to Gaussian behavior.

Keywords: Scaling, distribution, power-laws, returns, Lévy, stock prices.

1. Introduction

The study of financial markets poses many challenging questions. For example, how can one understand a strongly fluctuating system that is constantly driven by external information? And, how can one account for the role of the feedback between the markets and the outside world, or of the complex interactions between traders and assets? An advantage for the researcher trying to answer these questions is the availability of very large quantities of data for analysis. Indeed, the activities at financial markets result in several observables, such as the values of different market indices, the prices of the different stocks, trading volumes, etc.

Some of the most widely studied market observables are the values of market indices. Previous empirical studies [1] show that the distribution of fluctuations — measured by the returns — of market indices has slow decaying tails and that the distributions apparently retain the same functional form for a range of time scales. Here, we focus on a more "microscopic" quantity: individual companies. We analyze every transaction for the 1000 publicly-traded US companies with the

*http://polymer.bu.edu/~amaral.

largest market capitalizations in Jan 1994 and systematically study the statistical properties of their stock price fluctuations. In a preliminary study [2], we reported that the distribution of the 5 min returns for 1000 individual companies and the S&P 500 index decays as a power-law with an exponent $\alpha \approx 3$ — well outside the stable Lévy regime ($\alpha < 2$). Earlier independent studies on stock returns on longer time scales — but with fewer companies — also yield $\alpha \approx 3$ [3].

We investigate how the nature of the distribution of individual stock returns change with an increasing time scale Δt. Namely, does the distribution retain its power-law functional form for longer time scales, or does it converge to a Gaussian, as found for market indices? If the distribution indeed converges to Gaussian behavior, how fast does this convergence occur? For the S&P 500 index, for example, one finds the distribution of returns to be consistent with a *non-stable* power-law functional form ($\alpha \approx 3$) until approximately 4 days, after which an onset of convergence to Gaussian behavior is found [4]. To answer these questions, we perform an empirical analysis of individual company returns for a wide range of time scales. Using two distinct databases, we find that the cumulative distribution of individual-company returns is consistent with a power-law asymptotic behavior with exponent $\alpha \approx 3$, which is outside the stable Lévy regime. We also find that these distributions appear to retain the same functional form for time scales up to approximately 16 days. For longer time scales, we observe results consistent with a slow convergence to Gaussian behavior.

2. Empirical Results

The basic quantity studied for individual companies — $i = 1, 2, \ldots, 1000$ — is the market capitalization $S_i(t)$, defined as the share price multiplied by the number of outstanding shares. The time t runs over the working hours of the stock exchange.[a] For each company, we analyze the return $G_i(t, \Delta t) \equiv \ln S_i(t + \Delta t) - \ln S_i(t)$. For time scales shorter than 1 day, we analyze the data from the TAQ database. We consider the largest 1000 companies, in decreasing order of values of their market capitalization on 3 January 1994. We sample the price of these 1000 companies at 5 min intervals and filter the data to remove spurious events.[b] Each of the 1000 time series has approximately 40,000 data points, or about 40 million data points in total. For each time series, we compute the 5 min returns.

We first calculate the cumulative distributions of the 5 min returns for 10 randomly-chosen companies [Fig. 1(a)]. For each company i, the asymptotic

[a] The New York Stock Exchange is open from Monday through Friday 9:30 a.m. to 4:00 p.m. The time runs over the working hours only. Nights, week-ends, and holidays are removed.

[b] The analyzed data are affected by several types of recording errors. The most common error is missing digits which appears as a large spike in the time series of returns. Additionally we checked *individually* that the removed events correspond to missing digits in entering the data. There are also stock splits and take-overs which always occur overnight. To account for these, we take to be zero *all* the returns that happen overnight that are merely due to change in the number of outstanding shares.

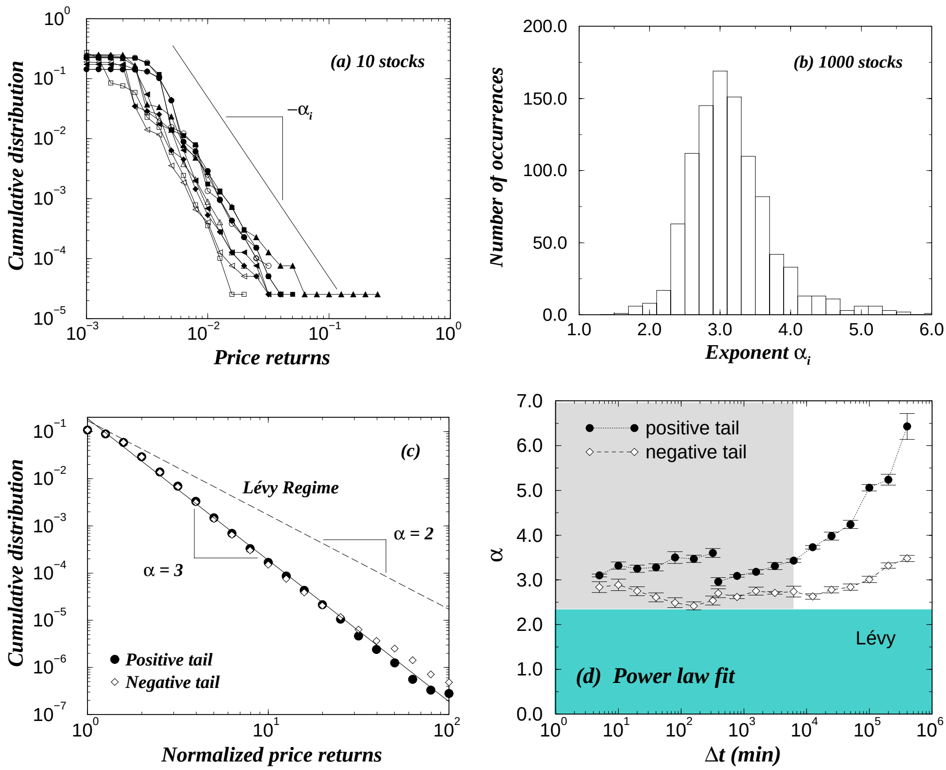

Fig. 1. (a) Cumulative distributions of the positive 5 min returns for 10 randomly-selected companies. (b) Histogram of the power-law exponents obtained by power-law regression fits to the individual cumulative distribution functions, where the fit is for all x larger than 2 standard deviations. (c) Cumulative distributions of the positive and negative tails of the normalized returns of the 1000 largest companies in the TAQ database. A power-law regression fit in the region $2 \leq x \leq 80$ (solid line) yields $\alpha = 3.10 \pm 0.03$ for the positive tail and $\alpha = 2.84 \pm 0.12$ for the negative tail. (d) Values of the exponent α as a function of the time scale Δt obtained using a power-law fit. The values of α for Δt <1 day are calculated from the TAQ database while for $\Delta t \geq$1 day they are calculated from the CRSP database. The unshaded region indicates the range of time scales where we find results consistent with a slow convergence to Gaussian behavior.

behavior of the functional form of the cumulative distribution is "visually" consistent with a power-law,

$$P(G_i > x) \sim x^{-\alpha_i}\,, \tag{2.1}$$

where α_i is the exponent characterizing the power-law decay. We perform power-law regression fits to the individual cumulative distributions of all 1000 companies studied. The histogram has most probable value $\alpha_{MP} = 3$; cf. Fig. 1(b).

Next, we compute the time-averaged volatility $v_i \equiv v_i(\Delta t)$ of company i as the standard deviation of the returns over the 2-year period $v_i{}^2 \equiv \langle G_i{}^2\rangle_T - \langle G_i\rangle_T{}^2$, where $\langle \ldots \rangle_T$ denotes a time average over the 40,000 data points of each time series, for the 2-year period studied. In order to compare the returns of different companies with different volatilities, we define the normalized return $g_i(t, \Delta t) \equiv (G_i - \langle G_i \rangle_T)/v_i$. The distributions for all 1000 normalized returns g_i have similar functional forms. Hence, to obtain better statistics, we compute a *single* distribution of all the normalized returns. The cumulative distribution $P(g > x)$ shows a power-law decay [Fig 1(c)], and regression fits in the region $2 \leq g \leq 80$ yield $\alpha = 3.10 \pm 0.03$ for the positive tail, and $\alpha = 2.84 \pm 0.12$ for the negative tail. These estimates of α are well outside the stable Lévy range, $0 < \alpha < 2$.

Next, we compute the distribution of returns for longer time scales. For $\Delta t <$ 1 day, we use the TAQ database, and for time scales of 1 day or longer we analyze data from the CRSP database, which comprises approximately 3.5×10^7 daily records for about 16,000 companies for the 35-year period 1962–96. We study the cumulative distribution of the normalized returns for time scales of up to 1024 days. For $\Delta t < 16$ days, we observe good "data collapse" with consistent values of α which suggests that the distribution of returns appears to retain its functional form for these time scales. The scaling behavior of the distributions of returns appears to break down for $\Delta t \geq 16$ days, and we observe indications of a slow convergence to Gaussian behavior [Fig. 1(d)] [5].

3. Summary

To summarize, we find that (i) the distribution of normalized returns for individual companies is consistent with a power-law behavior characterized by an exponent $\alpha \approx 3$, (ii) the distributions of returns retain the same functional form for a wide range of time scales Δt, varying over 3 orders of magnitude, 5 min$\leq \Delta t \leq$ 6240 min = 16 days, and (iii) for $\Delta t > 16$ days, the distribution of returns appears to slowly converge to a Gaussian.

Acknowledgments

We thank J.-P. Bouchaud, M. Barthélemy, S. V. Buldyrev, A. Chessa, P. Cizeau, X. Gabaix, I. Grosse, S. Havlin, K. Illinski, C. King, C.-K. Peng, B. Rosenow, D. Sornette, D. Stauffer, S. Solomon, J. Voit and especially R. N. Mantegna for stimulating discussions and helpful suggestions. We thank X. Gabaix, C. King, J. Stein, R. Tomposki, and especially T. Lim for help with obtaining the data. We also thank DFG, FCT/Portugal and NSF for financial support.

References

[1] J. P. Bouchaud and M. Potters, *Theorie des Risques Financiéres* (Alea-Saclay, Eyrolles, 1998); R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge Univ. Press, Cambridge (1999); I. Kondor and J. Kertész (eds.), *Econophysics: An Emerging Science*, Kluwer, Dordrecht, (1999); R. N. Mantegna (ed.), *Proc. Int. Workshop on Econophysics and Statistical Finance*, *Physica* **A269** (1999); J.-P. Bouchaud, P. Alstrøm and K. B. Lauritsen

(eds.), *Application of Physics in Financial Analysis*, Int. J. Theor. Appl. Finance **3**, (2000); B. B. Mandelbrot, J. Business **36** (1963) 294; A. Pagan, J. Empirical Finance **3** (1999) 15.

[2] P. Gopikrishnan, M. Meyer, L. A. N. Amaral and H. E. Stanley, Eur. Phys. J. **B3** (1998) 139.

[3] T. Lux, Appl. Financial Economics **6** (1996) 463; M. M. Dacorogna, U. A. Muller, R. J. Nagler, R. B. Olsen and O. V. Pictet, J. Int. Money and Finance **12** (1993) 413; M. Loretan and P. C. B. Phillips, J. Empirical Finance **1** (1994) 211.

[4] P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer and H. E. Stanley, Phys. Rev. **E60** (1999) 5305.

[5] V. Plerou, P. Gopikrishnan, L. A. N. Amaral, M. Meyer and H. E. Stanley, Phys. Rev. **E60** (1999) 6519.