Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Physica A 344 (2004) 227–235
PHYSICA A
www.elsevier.com/locate/physa

Exponential distribution of financial returns at mesoscopic time lags: a new stylized fact

A. Christian Silva, Richard E. Prange, Victor M. Yakovenko*

Department of Physics, University of Maryland, College Park, MD 20742-4111, USA

Received 16 December 2003
Available online 22 July 2004

Abstract

We study the probability distribution of stock returns at mesoscopic time lags (return horizons) ranging from about an hour to about a month. While at shorter microscopic time lags the distribution has power-law tails, for mesoscopic times the bulk of the distribution (more than 99% of the probability) follows an exponential law. The slope of the exponential function is determined by the variance of returns, which increases proportionally to the time lag. At longer times, the exponential law continuously evolves into Gaussian distribution. The exponential-to-Gaussian crossover is well described by the analytical solution of the Heston model with stochastic volatility.

PACS: 02.50.−r; 89.65.−s

Keywords: Econophysics; Exponential distribution; Stylized facts; Stochastic volatility; Heston model; Stock market returns; Empirical characteristic function

*Corresponding author. Tel.: +1-301-405-6151; fax: +1-301-314-9465.
E-mail address: yakovenk@physics.umd.edu (V.M. Yakovenko).
*URL:*http://www2.physics.umd.edu/˜yakovenk/econophysics.html.

0378-4371/$ - see front matter
doi:10.1016/j.physa.2004.06.122

1. Introduction

The empirical probability distribution functions (EDFs) for different assets have been extensively studied by the econophysics community in recent years [1–10]. Stock and stock-index returns have received special attention. We focus here on the EDFs of the returns of individual large American companies from 1993 to 1999, a period without major market disturbances. By 'return' we always mean 'log-return', the difference of the logarithms of prices at two times separated by a time lag t.

The time lag t is an important parameter: the EDFs evolve with this parameter. At micro lags (typically shorter than 1 h), effects such as the discreteness of prices and transaction times, correlations between successive transactions, and fluctuations in trading rates become important. Power-law tails of EDFs in this regime have been much discussed in the literature before [2,3]. At 'meso' time lags (typically from an hour to a month), continuum approximations can be made, and some sort of diffusion process is plausible, eventually leading to a normal Gaussian distribution. On the other hand, at 'macro' time lags, the changes in the mean market drifts and macroeconomic 'convection' effects can become important, so simple results are less likely to be obtained. The boundaries between these domains to an extent depend on the stock, the market where it is traded, and the epoch. The micro–meso boundary can be defined as the time lag above which power-law tails constitute a very small part of the EDF. The meso–macro boundary is more tentative, since statistical data at long time lags become sparse.

The first result is that we extend to meso time lags a stylized fact known since the 19th century [11] (quoted in Ref. [12]): with a careful definition of time lag t, the variance of returns is proportional to t.

The second result is that log-linear plots of the EDFs show prominent straight-line (tent-shape) character, i.e., the bulk (about 99%) of the probability distribution of log-return follows an exponential law. The exponential law applies to the central part of EDFs, i.e., not too big log-returns. For the far tails of EDFs, usually associated with power laws at micro time lags, we do not have enough statistically reliable data points at meso lags to make a definite conclusion. Exponential distributions have been reported for some world markets [4–10] and briefly mentioned in the book [1, see Fig. 2.12]. However, the exponential law has not yet achieved the status of a stylized fact. Perhaps this is because influential work [2,3] has been interpreted as finding that the individual returns of all the major US stocks for micro to macro time lags have the same power-law EDFs, if they are rescaled by the volatility.

The Heston model is a plausible diffusion model with stochastic volatility, which reproduces the timelag-variance proportionality and the crossover from exponential distribution to Gaussian. This model was first introduced by Heston, who studied option prices [13]. Later Drăgulescu and Yakovenko (DY) derived a convenient closed-form expression for the probability distribution of returns in this model and applied it to stock indexes from 1 day to 1 year [4]. The third result is that the DY formula with three lag-independent parameters reasonably fits the time evolution of EDFs at meso lags.

2. Probability distribution of log-returns in the Heston model

In this section, the Heston model [13] and the DY formula [4] are briefly summarized. The price S_t of a model stock obeys the stochastic differential equation of multiplicative Brownian motion: $\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sqrt{v_t} S_t\,\mathrm{d}W_t^{(1)}$. Here the subscript t indicates time dependence, μ is the drift parameter, $W_t^{(1)}$ is a standard random Wiener process, and v_t is the fluctuating time-dependent variance. The detrended log-return is defined as $x_t = \ln(S_t/S_0) - \mu t$, although detrending is a minor correction at meso lags. In the Heston model, the variance v_t obeys the mean-reverting stochastic differential equation:

$$\mathrm{d}v_t = -\gamma(v_t - \theta)\,\mathrm{d}t + \kappa\sqrt{v_t}\,\mathrm{d}W_t^{(2)} . \qquad (1)$$

Here θ is the long-time mean of v, γ is the rate of relaxation to this mean, and κ is the variance noise. We take the Wiener processes $W_t^{(1,2)}$ to be uncorrelated.

The DY formula [4] for the probability density function (PDF) $P_t(x)$ is

$$P_t(x) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}k}{2\pi}\, \mathrm{e}^{\mathrm{i}kx + F_{\tilde{t}}(k)},\ F_{\tilde{t}}(k) = \frac{\alpha\tilde{t}}{2} - \alpha \ln\left[\cosh\frac{\Omega\tilde{t}}{2} + \frac{\Omega^2+1}{2\Omega}\sinh\frac{\Omega\tilde{t}}{2}\right] , \qquad (2)$$

$$\tilde{t} = \gamma t, \quad \alpha = 2\gamma\theta/\kappa^2, \quad \Omega = \sqrt{1 + (k\kappa/\gamma)^2}, \quad \sigma_t^2 \equiv \langle x_t^2 \rangle = \theta t . \qquad (3)$$

The variance $\sigma_t^2 \equiv \langle x_t^2 \rangle$ (3) of the PDF (2) increases linearly in time, while $\langle x_t \rangle = 0$. The three parameters of the model are γ, θ and α. At short and long time lags, the PDF (2) reduces to exponential (if $\alpha = 1$) and Gaussian [4]:

$$P_t(x) \propto \begin{cases} \exp(-|x|\sqrt{2/\theta t}), & \tilde{t} = \gamma t \ll 1 , \\ \exp(-x^2/2\theta t), & \tilde{t} = \gamma t \gg 1 . \end{cases} \qquad (4)$$

In both limits, it scales with the volatility: $P_t(x) = f\left(x/\sqrt{\langle x_t^2 \rangle}\right) = f(x/\sqrt{\theta t})$, where f is the exponential or the Gaussian function.

3. Data analysis and discussion

We analyzed the data from Jan/1993 to Jan/2000 for 27 Dow companies, but show results only for four large cap companies: Intel (INTC) and Microsoft (MSFT) traded at NASDAQ, and IBM and Merck (MRK) traded at NYSE. We use two databases, TAQ to construct the intraday returns and Yahoo database for the interday returns. The intraday time lags were chosen at multiples of 5 min, which divide exactly the 6.5 h (390 min) of the trading day. The interday returns are as described in Refs. [4,5] for time lags from 1 day to 1 month = 20 trading days.

In order to connect the interday and intraday data, we have to introduce an effective overnight time lag T_n. Without this correction, the open-to-close and close-to-close variances would have a discontinuous jump at 1 day, as shown in the inset of Fig. 1a. By taking the open-to-close time to be 6.5 h, and the close-to-close time to be

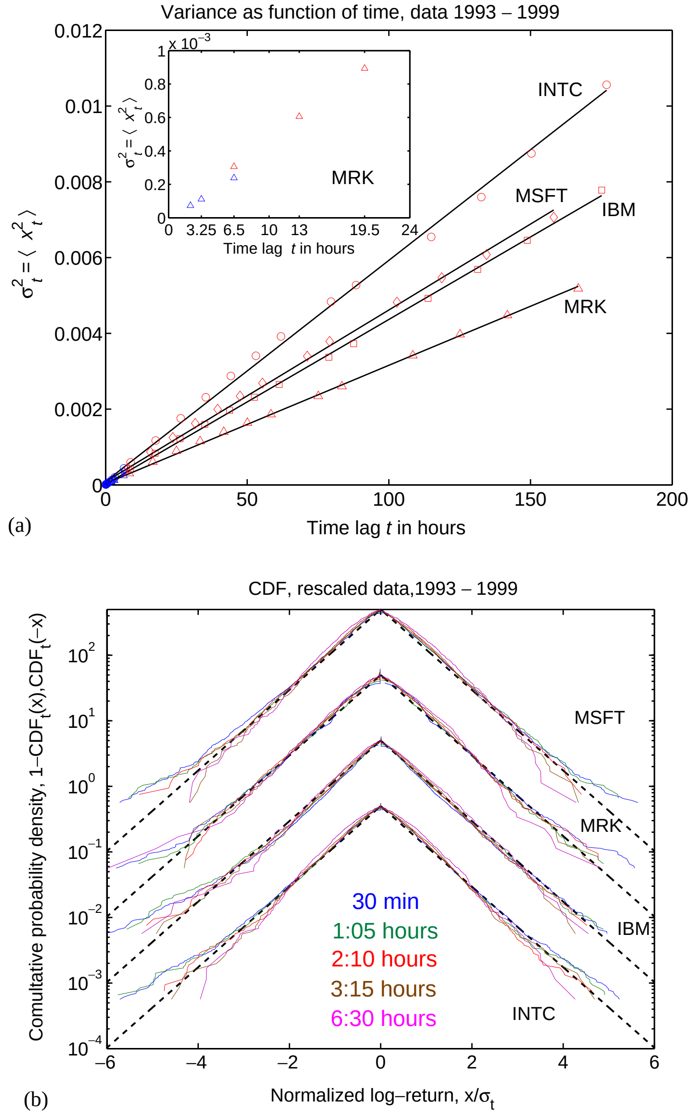


Fig. 1. (a) variance $\langle x_t^2 \rangle$ vs. time lag t. Solid lines: linear fits $\langle x_t^2 \rangle = \theta t$. Inset: variances for MRK before adjustment for the effective overnight time T_n. (b) log-linear plots of CDFs vs. $x/\sqrt{\theta t}$. Straight dashed lines $-|x|\sqrt{2/\theta t}$ are predicted by the DY formula (4) in the short-time limit. The curves are offset by a factor of 10.

Table 1
Fitting parameters of the Heston model with $\alpha = 1$ for the 1993–1999 data

	$\gamma\left(\frac{1}{\mathrm{h}}\right)$	$1/\gamma$ (h)	$\theta\left(\frac{1}{\mathrm{year}}\right)$	$\mu\left(\frac{1}{\mathrm{year}}\right)$	T_n (h)
INTC	1.029	0:58	13.04%	39.8%	2:21
IBM	0.096	10:25	9.63%	35.3%	2:16
MRK	0.554	1:48	6.57%	29.4%	1:51
MSFT	1.284	0:47	9.06%	48.3%	1:25

$6.5\,\mathrm{h} + T_n$, we find that variance $\langle x_t^2 \rangle$ is proportional to time t, as shown in Fig. 1a. The slope gives us the Heston parameter θ in Eq. (3). T_n is about 2 h (see Table 1).

In Fig. 1b, we show the log-linear plots of the cumulative distribution functions (CDFs) vs. normalized return $x/\sqrt{\theta t}$. The $\mathrm{CDF}_t(x)$ is defined as $\int_{-\infty}^{x} P_t(x')\,\mathrm{d}x'$, and we show $\mathrm{CDF}_t(x)$ for $x<0$ and $1 - \mathrm{CDF}_t(x)$ for $x>0$. We observe that CDFs for different time lags t collapse on a single straight line without any further fitting (the parameter θ is taken from the fit in Fig. 1a). More than 99% of the probability in the central part of the tent-shape distribution function is well described by the exponential function. Moreover, the collapsed CDF curves agree with the DY formula (4) $P_t(x) \propto \exp(-|x|\sqrt{2/\theta t})$ in the short-time limit for $\alpha = 1$ [4], which is shown by the dashed lines.

Because the parameter γ drops out of the asymptotic Eq. (4), it can be determined only from the crossover regime between short and long times, which is illustrated in Fig. 2a. We determine γ by fitting the characteristic function $\tilde{P}_t(k)$, a Fourier transform of $P_t(x)$ with respect to x. The theoretical characteristic function of the Heston model is $\tilde{P}_t(k) = \mathrm{e}^{F_t(k)}$ (2). The empirical characteristic functions (ECFs) can be constructed from the data series by taking the sum $\tilde{P}_t(k) = \mathrm{Re}\sum_{x_t} \exp(-\mathrm{i}kx_t)$ over all returns x_t for a given t [14]. Fits of ECFs to the DY formula (2) are shown in Fig. 2b. The parameters determined from the fits are given in Table 1.

In Fig. 3a we compare the empirical PDF $P_t(x)$ with the DY formula (2). The agreement is quite good, except for the very short time lag of 5 min, where the tails are visibly fatter than exponential. In order to make a more detailed comparison, we show the empirical CDFs (points) with the theoretical DY formula (lines) in Fig. 3b. We see that, for micro time lags of the order of 5 min, the power-law tails are significant. However, for meso time lags, the CDFs fall onto straight lines in the log-linear plot, indicating exponential law. For even longer time lags, they evolve into the Gaussian distribution in agreement with the DY formula (2) for the Heston model. To illustrate the point further, we compare empirical and theoretical data for several other companies in Fig. 4.

In the empirical CDF plots, we actually show the ranking plots of log-returns x_t for a given t. So, each point in the plot represents a single instance of price change. Thus, the last one or two dozens of the points at the far tail of each plot constitute a statistically small group and show large amount of noise. Statistically reliable conclusions can be made only about the central part of the distribution, where the points are dense, but not about the far tails.

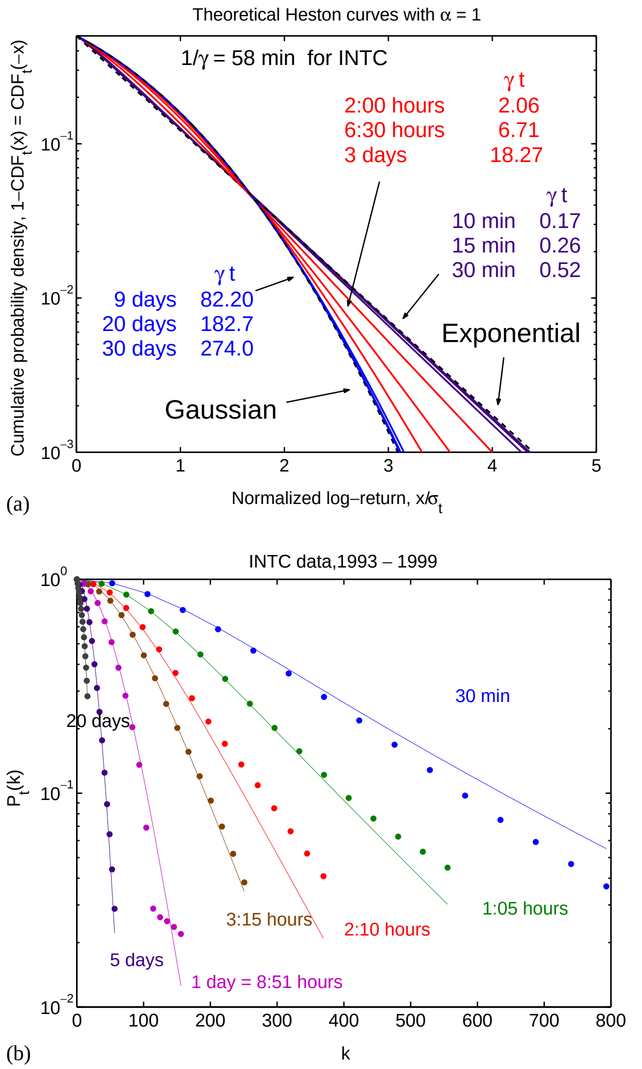
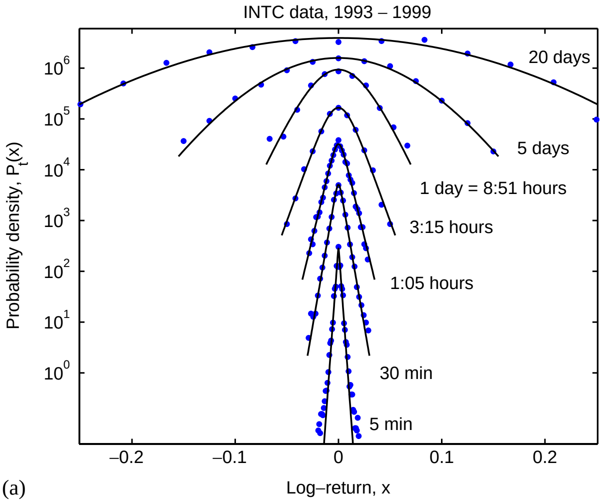


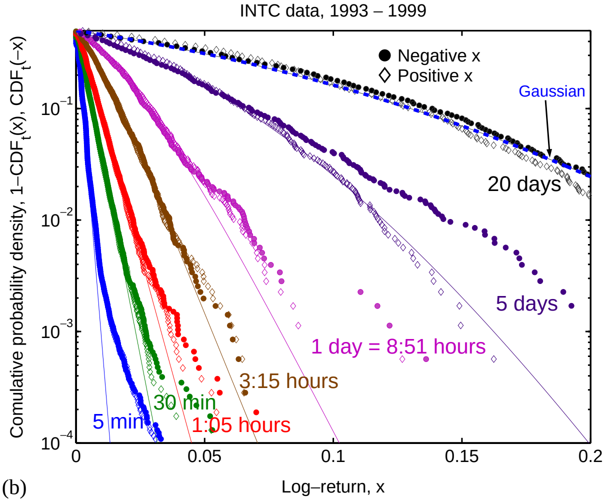
Fig. 2. (a) theoretical CDFs for the Heston model plotted vs. $x/\sqrt{\theta t}$. The curves interpolate between the short-time exponential and long-time Gaussian scalings. (b) comparison between empirical (points) and the DY theoretical (curves) characteristic functions $\tilde{P}_t(k)$.

4. Conclusions

We have shown that in the mesoscopic range of time lags, the probability distribution of financial returns interpolates between exponential and Gaussian law. The time range where the distribution is exponential depends on a particular company, but it is typically between an hour and few days. Similar



(a)



(b)

Fig. 3. Comparison between the 1993–1999 Intel data (points) and the DY formula (2) (curves) for PDF (a) and CDF (b).

exponential distributions have been reported for the Indian [6], Japanese [7], German [8], and Brazilian markets [9,10], as well as for the US market [4,5] (see also Fig. 2.12 in Ref. [1]). The DY formula [4] for the Heston model [13] captures the main features of the probability distribution of returns from an hour to a month with a single set of parameters. We believe that econophysicists should be aware of the presence of the exponential distribution and recognize it as another "stylized fact" in the set of analytical tools for financial data analysis.

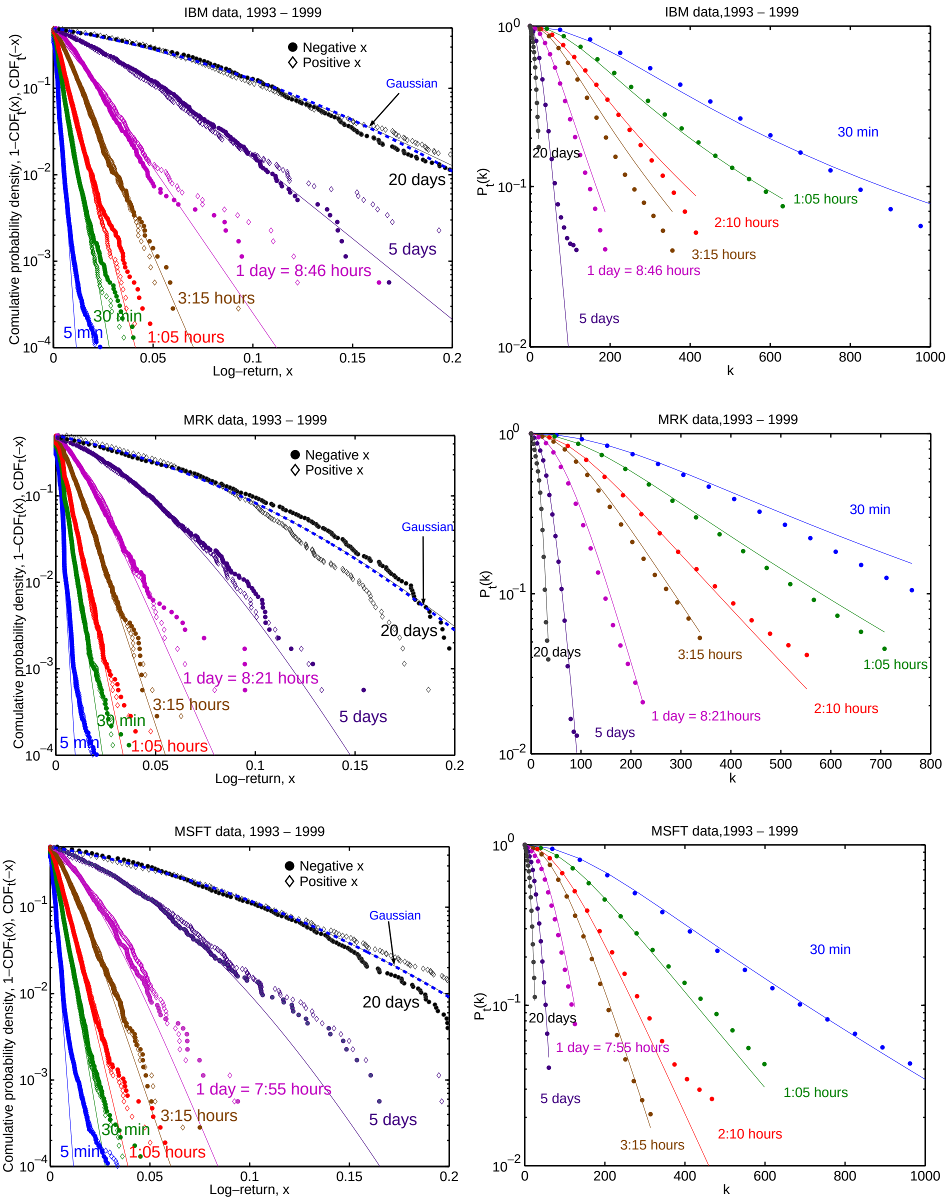


Fig. 4. Comparison between empirical data (symbols) and the DY formula (2) (lines) for CDF (left panels) and characteristic function (right panels).

Acknowledgements

We thank Chuck Lahaie from the Robert H. Smith School of Business at UMD for help with the TAQ database.

References

[1] J.-P. Bouchaud, M. Potters, Theory of Financial Risks, Cambridge University Press, Cambridge, 2000.
[2] V. Plerou, P. Gopikrishnan, L.N. Amaral, M. Meyer, H.E. Stanley, Phys. Rev. E 60 (1999) 6519.
[3] P. Gopikrishnan, V. Plerou, L.A.N. Amaral, M. Meyer, H.E. Stanley, Phys. Rev. E 60 (1999) 5305.
[4] A. Drăgulescu, V.M. Yakovenko, Quant. Finance 2 (2002) 443.
[5] A.C. Silva, V.M. Yakovenko, Physica A 324 (2003) 303.
[6] K. Matia, M. Pal, H. Salunkay, H.E. Stanley, Europhys. Lett. 66 (2004) 909.
[7] T. Kaizoji, M. Kaizoji, Adv. Complex Syst. 6 (2003) 303.
[8] R. Remer, R. Mahnke, Application of Heston model and its solution to German DAX data, talk presented at the APFA-4, 2003.
[9] R. Vicente, C.M. de Toledo, V.B.P. Leite, N. Caticha, preprint http://lanl.arXiv.org/abs/cond-mat/0402185.
[10] L.C. Miranda, R. Riera, Physica A 297 (2001) 509.
[11] J. Regnault, Calcul des Chances et Philosophie de la Bourse, Mallet-Bachelier et Castel, Paris, 1863.
[12] M. Taqqu, Paper #134, http://math.bu.edu/people/murad/articles.html.
[13] S.L. Heston, Rev. Financial Stud. 6 (1993) 327.
[14] N.G. Ushakov, Selected Topics in Characteristic Functions, VSP, Utrecht, 1999.