Available online at www.sciencedirect.com
SCIENCE @ DIRECT®
Physica A 324 (2003) 303–310
PHYSICA A
www.elsevier.com/locate/physa

Comparison between the probability distribution of returns in the Heston model and empirical data for stock indexes

A. Christian Silva, Victor M. Yakovenko*

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

Abstract

We compare the probability distribution of returns for the three major stock-market indexes (Nasdaq, S&P500, and Dow-Jones) with an analytical formula recently derived by Drăgulescu and Yakovenko for the Heston model with stochastic variance. For the period of 1982–1999, we find a very good agreement between the theory and the data for a wide range of time lags from 1 to 250 days. On the other hand, deviations start to appear when the data for 2000–2002 are included. We interpret this as a statistical evidence of the major change in the market from a positive growth rate in 1980s and 1990s to a negative rate in 2000s.

PACS: 02.50.−r; 89.65.−s

Keywords: Econophysics; Stochastic volatility; Heston; Stock-market returns

1. Introduction

Models of multiplicative Brownian motion with stochastic volatility have been a subject of extensive studies in finance, particularly in relation with option pricing [1]. One of the popular models is the so-called Heston model [2], for which many exact mathematical results can be obtained. Recently, Drăgulescu and Yakovenko (DY) [3]

* 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

doi:10.1016/S0378-4371(02)01903-9

derived a closed analytical formula for the probability distribution function (PDF) of log-returns in the Heston model. They found an excellent agreement between the formula and the empirical data for the Dow-Jones index for the period of 1982–2001. (Discussion of other work on returns distribution and references can be found in Ref. [3].)

In the present paper, we extend the comparison by including the data for Nasdaq and S&P500. We find that the DY formula agrees very well with the data for the period of 1982–1999. However, when the data for 2000–2002 are included, systematic deviations are observed, which reflect a switch of the market from upward to downward trend around 2000.

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

In this section, we briefly summarize the results of the DY paper [3]. Let us consider a stock, whose price S_t, as a function of time t, obeys the stochastic differential equation of multiplicative Brownian motion

$$\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma_t S_t\,\mathrm{d}W_t^{(1)} \,. \tag{1}$$

Here the subscript t indicates time dependence, μ is the drift parameter, $W_t^{(1)}$ is a standard random Wiener process, and σ_t is the time-dependent volatility. Changing the variable in (1) from price S_t to log-return $r_t = \ln(S_t/S_0)$ and eliminating the drift by introducing $x_t = r_t - \mu t$, we find

$$\mathrm{d}x_t = -\frac{v_t}{2}\,\mathrm{d}t + \sqrt{v_t}\,\mathrm{d}W_t^{(1)} \tag{2}$$

where $v_t = \sigma_t^2$ is the variance.

Let us assume that the variance v_t obeys the following mean-reverting stochastic differential equation

$$\mathrm{d}v_t = -\gamma(v_t - \theta)\,\mathrm{d}t + \kappa\sqrt{v_t}\,\mathrm{d}W_t^{(2)} \,. \tag{3}$$

Here θ is the long-time mean of v, γ is the rate of relaxation to this mean, $W_t^{(2)}$ is a standard Wiener process, and κ is the variance noise. In general, the Wiener process in (3) may be correlated with the Wiener process in (1):

$$\mathrm{d}W_t^{(2)} = \rho\,\mathrm{d}W_t^{(1)} + \sqrt{1-\rho^2}\,\mathrm{d}Z_t \,, \tag{4}$$

where Z_t is a Wiener process independent of $W_t^{(1)}$, and $\rho \in [-1, 1]$ is the correlation coefficient.

The coupled stochastic processes (2) and (3) constitute the Heston model [2]. In a standard manner [4], the Fokker–Planck equation can be derived for the transition probability $P_t(x, v \,|\, v_i)$ to have log-return x and variance v at time t given the initial log-return $x=0$ and variance v_i at $t=0$:

$$\frac{\partial}{\partial t}P=\gamma\frac{\partial}{\partial v}[(v-\theta)P]+\frac{1}{2}\frac{\partial}{\partial x}(vP)+\rho\kappa\frac{\partial^2}{\partial x\,\partial v}(vP)+\frac{1}{2}\frac{\partial^2}{\partial x^2}(vP)+\frac{\kappa^2}{2}\frac{\partial^2}{\partial v^2}(vP)\,. \tag{5}$$

A general analytical solution of Eq. (5) for $P_t(x,v\,|\,v_i)$ was obtained in Ref. [3]. Then $P_t(x,v\,|\,v_i)$ was integrated over the final variance v and averaged over the stationary distribution $\Pi_*(v_i)$ of the initial variance v_i:

$$P_t(x)=\int_0^\infty \mathrm{d}v_i\int_0^\infty \mathrm{d}v\,P_t(x,v\,|\,v_i)\,\Pi_*(v_i)\,. \tag{6}$$

The function $P_t(x)$ in Eq. (6) is the PDF of log-returns x after the time lag t. It can be directly compared with financial data. It was found in Ref. [3] that data fits are not very sensitive to the parameter ρ, so below we consider only the case $\rho=0$ for simplicity.

The final expression for $P_t(x)$ at $\rho=0$ (the DY formula [3]) has the form of a Fourier integral

$$P_t(x)=\frac{\mathrm{e}^{-x/2}}{x_0}\int_{-\infty}^{+\infty}\frac{\mathrm{d}\tilde{p}}{2\pi}\,\mathrm{e}^{\mathrm{i}\tilde{p}\tilde{x}+F_{\tilde{t}}(\tilde{p})}\,, \tag{7}$$

$$F_{\tilde{t}}(\tilde{p})=\frac{\alpha\tilde{t}}{2}-\alpha\ln\left[\cosh\frac{\tilde{\Omega}\tilde{t}}{2}+\frac{\tilde{\Omega}^2+1}{2\tilde{\Omega}}\sinh\frac{\tilde{\Omega}\tilde{t}}{2}\right]\,, \tag{8}$$

$$\tilde{\Omega}=\sqrt{1+\tilde{p}^2},\quad \tilde{t}=\gamma t,\quad \tilde{x}=x/x_0,\quad x_0=\kappa/\gamma,\quad \alpha=2\gamma\theta/\kappa^2\,. \tag{9}$$

In the long-time limit $\tilde{t}\gg 2$, Eqs. (7) and (8) exhibit scaling behavior, i.e., $P_t(x)$ becomes a function of a single combination z of the two variable x and t (up to the trivial normalization factor N_t and unimportant factor $\mathrm{e}^{-x/2}$):

$$P_t(x)=N_t\,\mathrm{e}^{-x/2}P_*(z),\quad P_*(z)=K_1(z)/z,\quad z=\sqrt{\tilde{x}^2+\bar{t}^2}\,, \tag{10}$$

$$\bar{t}=\alpha\tilde{t}/2=t\theta/x_0^2,\quad N_t=\bar{t}\mathrm{e}^{\bar{t}}/\pi x_0\,, \tag{11}$$

where $K_1(z)$ is the first-order modified Bessel function.

3. Comparison between the DY theory and the data

We analyzed the data for the three major stock-market indexes: Dow-Jones, S&P500, and Nasdaq. From the Yahoo Web site [5], we downloaded the daily closing values of Dow-Jones and S&P500 from 1 January 1982 to 22 October 2002 and all available data for Nasdaq from 4 October 1984 to 22 October 2002. The downloaded time series

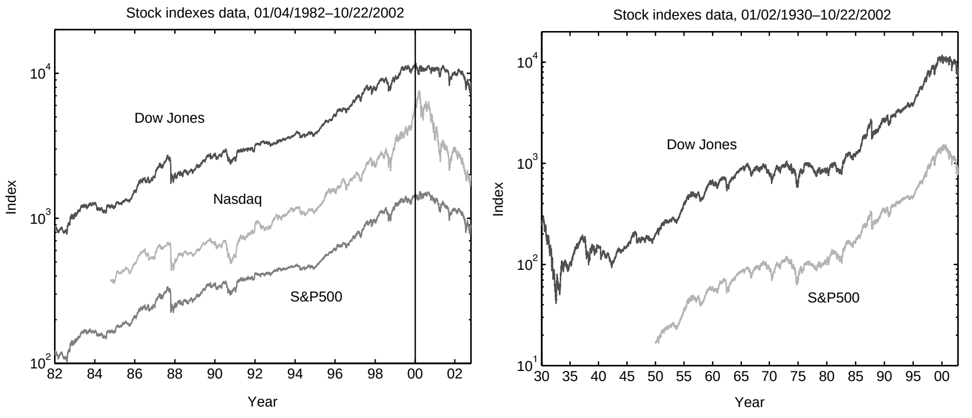

Fig. 1. Historical evolution of the three major stock-market indexes, shown in the log-linear scale. The Nasdaq curve is shifted up by the factor of 1.5 for clarity. The vertical line separates the regions with the average positive and negative growth rates.

$\{S_\tau\}$ are shown in the left panel of Fig. 1. It is clear that during 1980s and 1990s all three indexes had positive exponential growth rates, followed by negative rates in 2000s. For comparison, in the right panel of Fig. 1, we show the time series from 1930 to 2002. Contrary to the mutual-funds propaganda, stock market does not always increase. During 1930s (Great Depression) and 1960s–1970s (Stagnation), the average growth rate was zero or negative. One may notice that such fundamental changes of the market trend occur on a very long time scale of the order of 15–20 years.

Using the procedure described in Ref. [3], we extract the PDFs $P_t^{(\mathrm{data})}(r)$ of log-returns r for different time lags t from the time series $\{S_\tau\}$ for all three indexes. In the DY theory [3], the actual (empirically observed) growth rate $\bar{\mu}$ is related to the bare parameter μ by the following relation: $\bar{\mu} = \mu - \theta/2$, and $P_t^{(\mathrm{data})}(x)$ is obtained by replacing the argument $r \to x + \mu t$. The parameters $\bar{\mu}$ were found by fitting the time series in the left panel of Fig. 1 to straight lines. With the constraint $\mu = \bar{\mu} + \theta/2$, the other parameters of the Heston model (γ, θ, κ) were obtained by minimizing the mean-square deviation $\sum_{x,t} |\ln P_t^{(\mathrm{data})}(x) - \ln P_t(x)|^2$ between the empirical data and the DY formula (7) and (8), with the sum taken over all available x and the time lags $t = 1$, 5, 20, 40, and 250 days. This procedure was applied to the data from 1982 (1984 for Nasdaq) to 31 December 1999, and the values of the obtained parameters are shown in Table 1. The model parameters for Dow-Jones and S&P500 are similar, whereas some parameters for Nasdaq are significantly different. Namely, the variance relaxation time $1/\gamma$ is much shorter, the variance noise κ is much bigger, and the parameter α is much smaller for Nasdaq. All of this is consistent with the general notion that Nasdaq is more volatile than Dow-Jones and S&P500. On the other hand, the average growth rates $\bar{\mu}$ of all three indices are about the same, so the greater risk in Nasdaq does not result in a higher average return.

Fig. 2 compares the 1984–1999 data for Nasdaq (points) with the DY theory (curves). The left panel shows the PDFs $P_t(x)$ (7) for several time lags t, and the

Table 1
Parameters of the Heston model obtained from the fits of the Nasdaq, S&P500, and Dow-Jones data from 1982 to 1999 using $\rho = 0$ for the correlation coefficient

	γ $\frac{1}{\text{year}}$	$1/\gamma$ day	θ $\frac{1}{\text{year}}$	κ $\frac{1}{\text{year}}$	μ $\frac{1}{\text{year}}$	$\bar{\mu}$ $\frac{1}{\text{year}}$	α	x_0
Nasdaq	114	2.2	3.6%	5.3	16%	14%	0.3	4.7%
S&P500	17	15	1.8%	0.67	13%	12%	1.36	4.0%
Dow-Jones	24	10	2%	0.94	14%	13%	1.1	3.9%

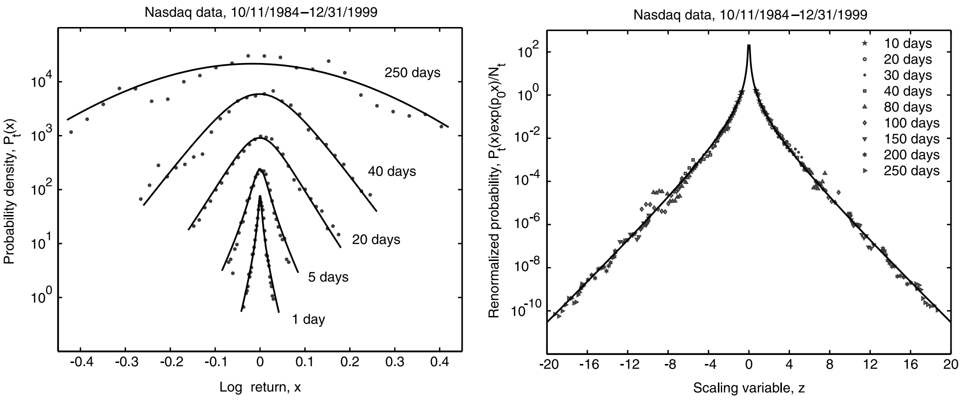

Fig. 2. Comparison between the 1984 and 1999 Nasdaq data (points) and the Drăgulescu–Yakovenko theory [3] (curves). Left panel: PDFs $P_t(x)$ of log-returns x for different time lags t shifted up by the factor of 10 each for clarity. Right panel: Renormalized PDF $P_t(x)e^{x/2}/N_t$ plotted as a function of the scaling argument z given in Eq. (10). The solid line is the scaling function $P_*(z) = K_1(z)/z$ from Eq. (10), where K_1 is the first-order modified Bessel function.

right panel demonstrates the scaling behavior (10). The overall agreement is quite good. Particularly impressive is the scaling plot, where the points for different time lags collapse on a single nontrivial scaling curve spanning 10 (!) orders of magnitude. On the other hand, when we include the data up to 22 October 2002, the points visibly run off the theoretical curves, as shown in Fig. 3. We use the same values of the parameters (μ, γ, θ, κ) in Fig. 3 as in Fig. 2, because attempts to adjust the parameters do not reduce the discrepancy between theory and data. The origin of the discrepancy is discussed in Section 4.

Similarly to Nasdaq, the S&P500 data for 1982–1999 agree well with the theory, as shown in Fig. 4. However, when the data up to 2002 are added (Fig. 5), deviations occur, albeit not as strong as for Nasdaq. For Dow-Jones 1982–1999 (Fig. 6), the data agrees very well with the theory. The PDFs for 1982–2002, shown in the left panel of Fig. 7, still agree with the theory, but deviations are visible in the scaling plot in the right panel of Fig. 7. They come from the time lags between 40 and 150 days not shown in the left panel.

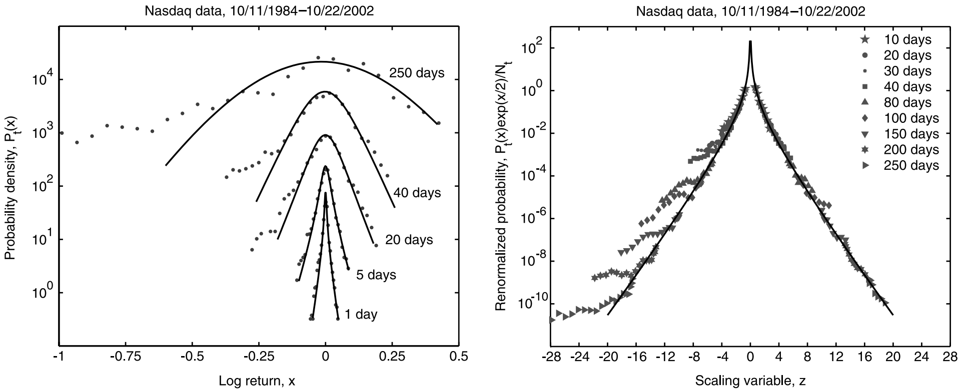

Fig. 3. The same as in Fig. 2 for 1984–2002.

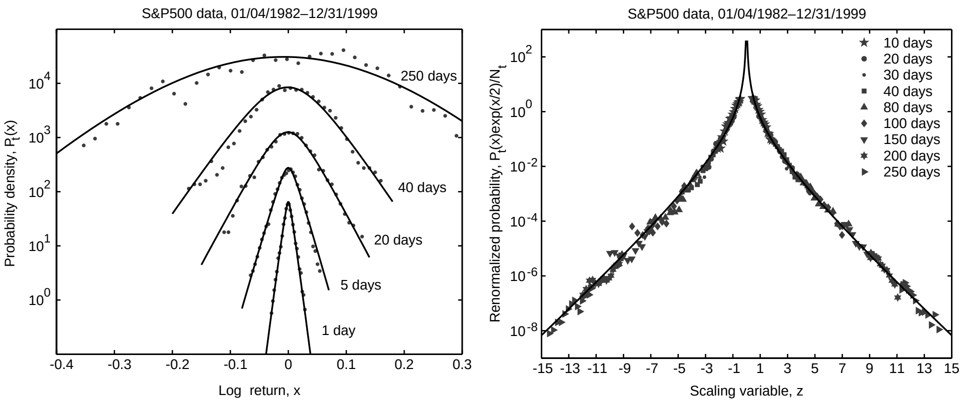

Fig. 4. The same as in Fig. 2 for S&P500 for 1982–1999.

4. Discussion and conclusions

We conclude that, overall, the PDFs of log-returns, $P_t(x)$, agree very well with the DY formula [3] for all three stock-market indices for 1982–1999. It is important to recognize that the single DY formula (7) and (8) fits the whole family of empirical PDFs for time lags t from one day to one year (equal to 252.5 trading days). The agreement with the nontrivial Bessel scaling function (10) extends over the astonishing *ten* orders of magnitude. These facts strongly support the notion that fluctuations of stock market are indeed described by the Heston stochastic process.

On the other hand, once the data for 2000s are included, deviations appear. They are the strongest for Nasdaq, intermediate for S&P500, and the smallest for Dow-Jones. The origin of the deviations can be recognized by looking in Fig. 1. Starting from 2000, Nasdaq has a very strong downward trend, yet we are trying to fit the data

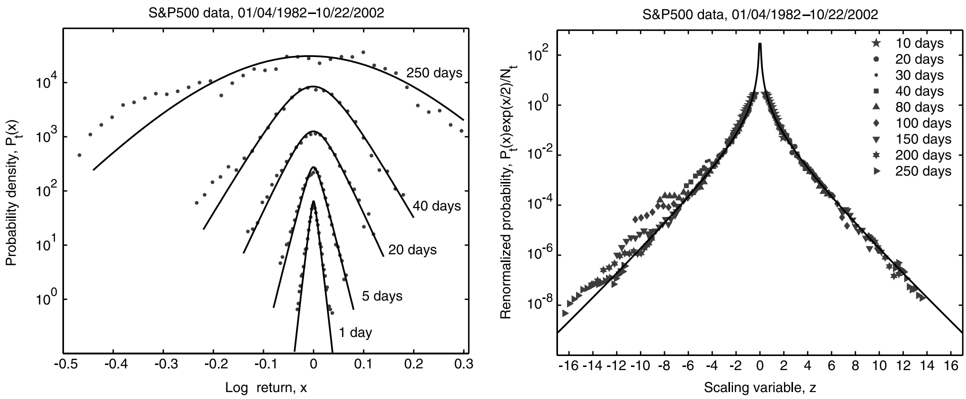

Fig. 5. The same as in Fig. 4 for 1982–2002.

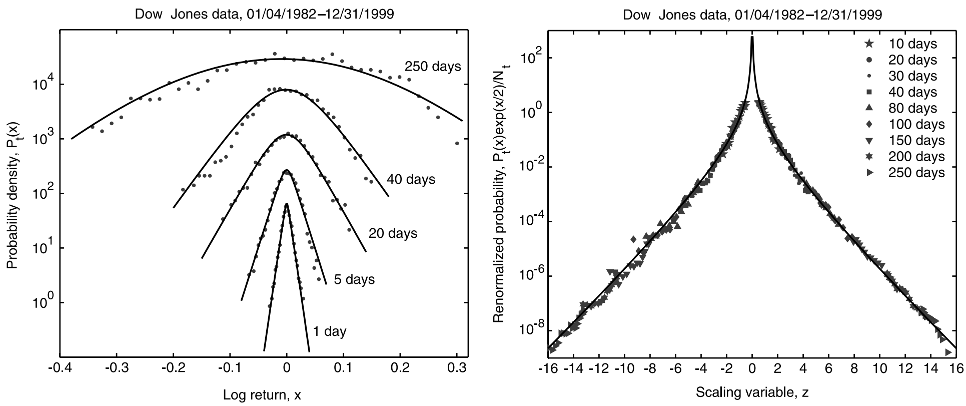

Fig. 6. The same as in Fig. 4 for Dow-Jones.

using a constant positive growth rate μ. Obviously, that would cause disagreement. For S&P500 and Dow-Jones, the declines in 2000s are intermediate and small, so are the deviations from the DY formula. We think these deviations are not an argument against the Heston model. They rather indicate the change of μ from a positive to a negative value around 2000. Our conclusion about the change of regime is based on the statistical properties of the data for the last 20 years. The situation is very different from the crash of 1987. As our plots show, the crash of 1987 did not have significant statistical impact on the PDFs of log-returns for 1980s and 1990s, because the market quickly recovered and resumed overall growth. Thus, the crash of 1987 was just a fluctuation, not a change of regime. To the contrary, the decline of 2000s (which is characterized by a gradual downward slide, not a dramatic crash on any particular day) represents a fundamental change of regime, because the statistical probability

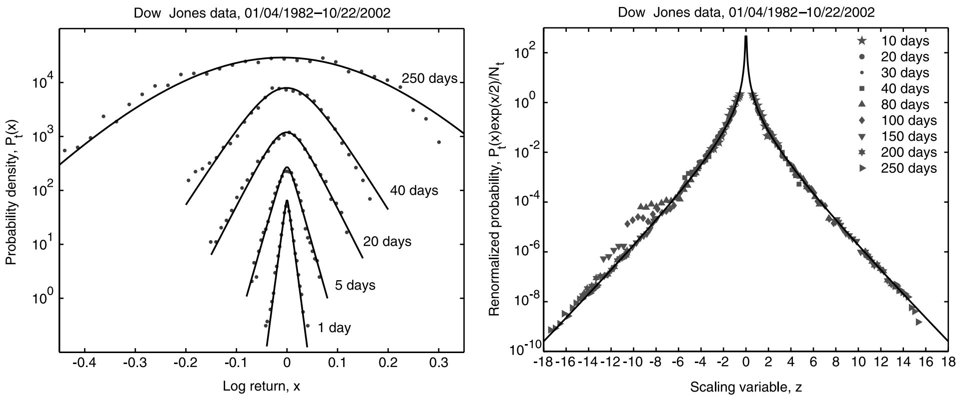

Fig. 7. The same as in Fig. 6 for 1982–2002.

distributions have changed. These conclusions are potentially important for investment decisions.

The average growth rate μ is an exogenous parameter in the Heston model and is taken to be constant only for simplicity. In a more sophisticated model, it could be a smooth function of time τ, reflecting the long-term trend of the market of the scale of 15–20 years. Using a properly selected function μ_τ, one could attempt to analyze the stock-market fluctuations on the scale of a century. That would be the subject of a future work.

Acknowledgements

We are grateful to Adrian Drăgulescu for help and sharing his computer codes for data processing and plotting.

References

[1] J.P. Fouque, G. Papanicolaou, K.R. Sircar, Derivatives in Financial Markets with Stochastic Volatility, Cambridge University Press, Cambridge, 2000; Int. J. Theor. Appl. Finance 3 (2000) 101.
[2] S.L. Heston, Rev. Financial Stud. 6 (1993) 327.
[3] A. Drăgulescu, V.M. Yakovenko, preprint http://lanl.arXiv.org/abs/cond-mat/0203046, Quantitative Finance 2 (2002) 443.
[4] C.W. Gardiner, Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences, Springer, Berlin, 1993.
[5] Yahoo Finance, http://finance.yahoo.com/. The data were downloaded for ˆDJI (Dow-Jones), ˆGSPC (S&P500), and ˆIXIC (Nasdaq).