Quantitative Finance, Vol. 5, No. ?, Month?? 2005, 1–13

Routledge
Taylor & Francis Group

Stochastic volatility and the goodness-of-fit of the Heston model

GILLES DANIEL†, NATHAN L. JOSEPH*‡ and DAVID S. BRÉE†

†School of Computer Science, University of Manchester, Oxford Road, Manchester, M13 9PL, UK
‡Aston Business School, Aston University, Aston Triangle,
Birmingham, B4 7ET, UK

(*Received* [redacted]*; in final form* [redacted]) 1

Recently, Drăgulescu and Yakovenko proposed an analytical formula for computing the probability density function of stock log returns, based on the Heston model, which they tested empirically. Their research design inadvertently favourably biased the fit of the data to the Heston model, thus overstating their empirical results. Furthermore, Drăgulescu and Yakovenko did not perform any goodness-of-fit statistical tests. This study employs a research design that facilitates statistical tests of the goodness-of-fit of the Heston model to empirical returns. Robustness checks are also performed. In brief, the Heston model outperformed the Gaussian model only at high frequencies and even so does not provide a statistically acceptable fit to the data. The Gaussian model performed (marginally) better at medium and low frequencies, at which points the extra parameters of the Heston model have adverse impacts on the test statistics.

1. Introduction

Theoretical pricing models of risky assets and contingent claims make important assumptions regarding the underlying distribution of the associated asset's returns. Many theoretical models of investor behaviour predict that risky asset returns follow a log-normal distribution. For example, under the Markowitz model agents are said to select portfolios in which their preferences depend on the mean and variance of returns. This in turn implies a normal distribution for the asset returns contained in the portfolio. The normality assumption is also required for asset pricing models, e.g. Black–Scholes option pricing, when investors are not assumed to have quadratic utility functions. However, empirical investigations consistently report that risky asset returns are leptokurtic with more peakedness and fat tails than the log-normal distribution, particularly at high frequency.

Since most theoretical models of investors' behaviour assume a normal distribution, many empirical studies have sought to describe the empirical distribution of asset returns and the extent to which their statistical properties matched those that are predicted by theoretical models.† To capture the fat-tailed behaviour of asset returns, several statistical distributions have been suggested. One distribution that has featured prominently is a general class of stable Paretian distributions (see Mandelbrot 1963, Fama 1965). An important feature of this family of distributions is that it includes empirical distributions that have fat tails, albeit with infinite variance, when the characteristic exponent (α) is less than 2.0.‡ The stable Paretian distribution with $\alpha < 2.0$ appears to provide a *better* fit for stock returns compared with the normal distribution (see, e.g. Fama 1965, p. 92). However, empirical evidence suggests that α varies over time (see Blattberg and Gonedes 1974) thus violating

*Corresponding author. Email: n.l.joseph@aston.ac.uk

†The normality assumption relies on the theory of random walks, beginning with Bachelier (1900). Since the 1960s, many empirical studies have sought to describe the empirical distribution of risky asset returns (see Mandelbrot 1963, Fama 1965, Bookstaber and McDonald 1987).

‡An α in the interval $0 < \alpha < 2.0$ results in a stable Paretian distribution with tails that are much higher than those of the normal distribution. A distribution with an α of 2 is normally distributed, while one with an α of 1 depicts a Cauchy distribution. Beyond those special cases the probability density function (PDF) is not known.

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online © 2005 Taylor & Francis Group Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/14697680500148521

the constancy assumption of sums of *independent and identically distributed* (*iid*) stable Paretian variables under addition. The Paretian distributions have fairly limited application in finance as a description of security returns.

Another statistical model that has featured well in empirical work is the Student *t* distribution. Blattberg and Gonedes (1974) show that the Student *t* distribution provides a better fit than the stable Paretian distribution if the variance of a continuous mixture of underlying Gaussian distributions is itself random and follows an inverted Gamma distribution.

Empirical work has also evaluated the asymmetry, i.e. skewness, and volatility clustering in asset returns, first noted by Mandelbrot (1963). Initial empirical work focused on the unconditional density distribution of returns. It turns out that this approach might not be useful for capturing riskiness. There is now a substantial body of theoretical and empirical work concerned with the conditional variance and volatility of returns.† One important class of such models is the ARCH and GARCH family introduced by Engle (1982) and Bollerslev (1986), respectively. This family of models allows normally distributed disturbances to have conditionally time-varying variances in ways that account for certain stylized facts, e.g. kurtosis, fat tails, etc., regarding security price behaviour. Indeed, the conditional models of Nelson (1991) and Engle and Ng (1993) seek to capture the asymmetry in the variance of returns—often called the leverage effect.‡ Standard GARCH models are not able to accommodate all of the kurtosis in asset returns (see Franses and Ghijsels 1999) and indeed other important features, such as the time scaling of the probability density function (PDF) of returns (e.g. Serva *et al.* 2002). Harvey and Siddique (1999) also show that the inclusion of conditional skewness in a GARCH framework reduces the persistence in the conditional volatility of those returns.

An alternative but related set of models to the GARCH takes the form of multiplicative Brownian motion, with stochastic volatility. This set of models assumes that there is a constant drift in the mean asset value but that the variability is not constant; the volatility can be disturbed, but it reverts to its own mean value after a disturbance. The exact form that the disturbance takes varies in theoretical work. For example, Hull and White (1987) assume that the variability is subject to random shocks that are proportional to the variance while Heston (1993) assumes that random shocks to the variance are proportional to the square root of the variance. Miccichè *et al.* (2002) have shown that Hull and White's (1987) model describes better the PDF of stock volatility for high (relative to low) values of volatility whiles the log-normal model describes the PDF better for low (relative to high) volatility values. Recently, Drăgulescu and Yakovenko (2002)—hereafter D-Y—derived a closed-form analytical solution for the PDF of stock index returns based on the Heston model, which they fitted to empirical data. D-Y show that the Heston (1993) model adequately describes the PDF of empirical security returns and the model can therefore assist in the valuation of path-dependent options.

In this study, we raise several important objections to D-Y's research design and the robustness of their empirical results (see below). Indeed, we show that both of those considerations would have led D-Y to different empirical conclusions. We analyse the returns for three stock indices including the one analysed by D-Y and evaluate the empirical fit of the returns relative to both the Heston and the Gaussian models. We show that a better distributional fit can be found when using a neural network albeit not for all frequencies; but this underlying distribution is unspecified and not driven by economic theory. The Heston model is primarily concerned with derivatives pricing while our analysis focuses on spot returns.§ However, we follow Heston's (1993, p. 336; p. 340) argument that '... stochastic volatility model can conveniently explain properties of the option prices in terms of the underlying distribution of spot returns'. Also, the majority of the options on the market are usually traded at or near-the-money. The use of the returns is likely to give some indication of the distribution of those option prices.

The remaining sections of this study are as follows. Section 2 presents the motivation for the study whiles section 3 briefly describes D-Y's analytical solution to the Heston model. Section 4 presents the data sets and we replicate D-Y's empirical results in section 5 to provide a basis for our empirical work. Our own research design and main empirical results are presented in sections 6 and 7, respectively, and we present some conclusions in section 8.

2. Motivation for use of stochastic volatility models

The research interests in stochastic volatility models and the Heston model in particular, are motivated by the following considerations. Firstly, there is a general belief among researchers that alternatives to or extensions on the basic Black–Scholes (B-S) model can *better* accommodate the volatility clustering and fat tailedness of returns. This belief is evidenced by the '... explosion of new models that each relax some of the restrictive

†The early empirical studies (see Mandelbrot 1963, Fama 1965) focused on the parametric unconditional density of asset returns. Since then, several other statistical models that assume unconditional (e.g. Theodossiou 1998) or conditional (e.g. Engle 1982) statistical properties for asset price behaviour and returns have been put forward.

‡It is expected that empirical models that accommodate both skewness and kurtosis will account more fully for the riskiness of returns, even if kurtosis is more pronounced than skewness in asset returns. Skewness is also an important variable in portfolio selection decisions (see Chunhachinda *et al.* 1997).

§This is a valid point made by an anonymous referee of this journal.

Black–Scholes . . . assumptions' (Bakshi *et al.* 1997, p. 2003).† Indeed, the biased parameter estimates generated by the B-S model have been attributed to the failure of the normality assumption to hold (see, e.g. Rubinstein 1985). If higher moments, e.g. skewness and kurtosis, statistically relate to the riskiness of asset prices (see, e.g. Harvey and Siddique 1999), taking those parameters into account is likely to improve on one's ability to price those assets.

Secondly, since stochastic volatility models appear to capture certain stylized facts of stock returns their use is likely to improve on one's ability to price assets, beyond the benchmarks provided by B-S. The Heston model accommodates a volatility parameter which can increase with the level of kurtosis. So we would expect the Heston model to improve on the fit provided by the Gaussian model, particularly at higher frequency, since it is supposed to capture both the volatility of price fluctuations and skewness which are more pronounced in high frequency data. Under the Heston model, the joint PDF of returns and their variance are a function conditioned on the initial variance value. Integrating the joint PDF over the variance results in the marginal PDF of returns which is unconditional on the variance.

Fiorentini *et al.*'s (2002, p. 227) empirical results indicate, however, that although the Heston model offers '. . . a slight overall improvement over the performance of . . .' the B-S model *in option pricing*, it has a tendency to overprice out-of-the-money (underprice in-the-money) calls for daily data. Indeed, Bouchaud and Potters (2004, pp. 141–144), show that the model seriously fails to capture adequately important empirical properties of price returns. In particular, they show that the model predicts a Gamma distribution for volatility, whereas empirically volatility is closer to an inverse Gamma distribution. Furthermore, for long time lags, the decay in volatility is much slower than the model predicts while the decay of kurtosis asymptotically is much faster empirically than the model predicts. These considerations suggest that D-Y's result that the Heston model provides a good fit to empirical data is therefore suspect as the Heston model would not appear to capture adequately certain properties of returns. Furthermore, D-Y did not perform any goodness-of-fit statistical tests. To relate aspects of the fit of the normal distribution, D-Y's figure 4 (p. 449) illustrates to a limited extent the decay in the non-Gaussian portion of the curve in relation to time. Our empirical work explores the empirical fit of returns in much greater detail and in particular considers the goodness-of-fit of both the Heston model and the normal distribution (which is assumed for B-S) and the conditions, if any, under which those models perform well. These are the key aims of our empirical study.

3. Heston's model and D-Y formula

The model starts from a geometric Brownian motion stochastic differential equation for the price S_t

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

where μ is the market trend (drift), σ_t is the volatility at time t and $W_t^{(1)}$ is a standard Weiner process. D-Y use $r_t = \log(S_t/S_0)$, the continuously compounded log return on the asset, which is preferable for empirical estimation. This is because, unlike S_t, r_t tends to be stationary—a feature which is important for most equilibrium models.

In a simple Brownian motion, r_t is then governed by

$$\mathrm{d}r_t = \left(\mu - \frac{\sigma^2}{2}\right)\mathrm{d}t + \sigma\,\mathrm{d}W_t \tag{2}$$

and is normally distributed. Unlike the Bachelier–Osborne model, a constant volatility σ is not assumed in the Heston model. So, setting $\sigma = v_t^{1/2}$, equation (2) becomes

$$\mathrm{d}r_t = \left(\mu - \frac{v_t}{2}\right)\mathrm{d}t + v_t^{1/2}\,\mathrm{d}W_t^{(1)}. \tag{3}$$

Centred log returns $x_t = r_t - \mu t$ are then introduced, so that

$$\mathrm{d}x_t = -\frac{v_t}{2}\,\mathrm{d}t + v_t^{1/2}\,\mathrm{d}W_t^{(1)}. \tag{4}$$

Heston assumes that v_t obeys the following mean-reverting stochastic differential equation

$$\mathrm{d}v_t = -\gamma(v_t - \theta)\,\mathrm{d}t + \kappa v_t^{1/2}\,\mathrm{d}W_t^{(2)}, \tag{5}$$

where θ is the long-run mean of v_t; γ is the rate of relaxation to the long-run mean; κ is a constant parameter called variance noise; $W_t^{(2)}$ is another standard Wiener process, not necessarily correlated with $W_t^{(1)}$. The mean-reverting attribute of v_t assumes that there exists a normal level of volatility to which the volatility itself will return, even if it has long memory. This is a widely agreed stylized fact in empirical finance (see Engle and Patton 2001). The forward Kolmogorov equation can be derived for the transition joint probability $P_t(x, v|v_{\mathrm{i}})$, where v_{i} is the initial value of the variance. This is to obtain the log return x_t and variance v_{i} at time t, for initial $x = 0$ and variance v_{i} at time $t = 0$:

$$\frac{\partial}{\partial t}P = \gamma\frac{\partial}{\partial v}[(v-\theta)P] + \frac{1}{2}\frac{\partial}{\partial x}(vP) + \rho k\frac{\partial^2}{\partial x\partial v}(vP) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(vP) + \frac{k^2}{2}\frac{\partial^2}{\partial v^2}(vP). \tag{6}$$

Solving equation (6) analytically, D-Y's final result in Fourier integral is

$$P_t(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{d}p_x\,\exp\left[\mathrm{i}p_x x + F_t(p_x)\right], \tag{7}$$

†Bakshi *et al.* (1997) provide a useful set of citations for some of the more prominent deterministic and stochastic volatility models in this area. Stochastic volatility models are considered more appropriate for empirical examination since they can more readily accommodate certain risk factors, e.g. jumps (see also Fiorentini *et al.* 2002).

which is an expression for the probability distribution of the centred log returns x for a frequency t where

$$F_t(p_x) = \frac{\gamma\theta}{k^2}\Gamma t - \frac{2\gamma\theta}{k^2}\ln\left[\cosh\frac{\Omega t}{2} + \frac{\Omega^2 - \Gamma^2 + 2\gamma\Gamma}{2\gamma\Omega}\sinh\frac{\Omega t}{2}\right] \tag{8}$$

in which: $\Gamma = \gamma + i\rho k p_x$; $\Omega = [\Gamma^2 + k^2(p_x^2 - ip_x)]^{1/2}$; ρ is the correlation between the two Weiner processes $W_t^{(1)}$ and $W_t^{(2)}$; and γ, θ, k and ν are the parameters of the model. Equation (7) is central to the model. For a given t, it gives the expected probability density of the centred log returns x. A full derivation of the model's parameters can be found in Drăgulescu and Yakovenko (2002).

4. Data sets

The datasets that we use to test the models are: (i) the Dow Jones Industrial Average index (hereafter DJIA8201); (ii) the Standard and Poor's (SP) 500 index (hereafter SP8201); and (iii) the Financial Times Stock Exchange (FTSE) 100 index (hereafter FTSE8401). Both the DJIA8201 and SP8201 span the period 1 January 1982 to 31 December 2001.† The FTSE8401 spans the period 2 January 1984 to 31 December 2001. The DJIA8201 and FTSE8401 stock indices were obtained from http://finance.yahoo.com while the SP8201 stock index was taken from http://economy.com. The DJIA8201 spans the same period and is from the same data source as the data set used by D-Y. All the series are trimmed of bank holiday observations but, to be consistent with D-Y's study, pre- and post-bank holiday observations are retained.

5. D-Y's research design and their results

To fit the empirical distribution of the DJIA8201 returns, D-Y generated a single log-return time series of overlapping return frequencies for each of $t = 1$, 5, 20, 40 and 250 days where t is the frequency in trading days for the price time series S_t. These frequencies are arbitrary but consistent with the need to price security at different intervals. Consider an index I, at time τ for a frequency of (say) $t = 5$ days. D-Y computed the log returns from the closing prices as $\{r_\tau | \tau \in [1, n-t]\}$, where n is the number of trading days in the data set and $r_t = \log(S_{\tau+t}/S_\tau)$, $\forall\tau \in [1, n-t]$. Thus for the DJIA8201, with $n = 5050$ trading days, D-Y computed *logreturns*

$$\begin{aligned} &\{r_1, r_2, \ldots, r_{n-t}\} \\ &= \{\log(S_{1+t}/S_1), \log(S_{2+t}/S_2), \ldots, \log(S_n/S_{n-t})\} \\ &= \{\log(S_6/S_1), \log(S_7/S_2), \ldots, \log(S_{5050}/S_{5045})\} \end{aligned} \tag{9}$$

for a given t (say 5 days).‡ This approach generates a single dataset of $n - t$ data points for each t.

Such data does not allow an evaluation of the robustness of statistical tests because there is only one series for each frequency. Also, the resulting series is unsuitable for statistical testing because the data is reused. For example, if a rare event, such as a stock market crash, occurred on just one day, say τ^*, D-Y's approach will take this event into account exactly t times in the log returns $\{r_{\tau^*-t}, r_{\tau^*-t+1}, \ldots, r_{\tau^*-t+2}\}$, thereby resulting in a return series with fatter tails for this rare event whenever $t > 1$. D-Y have also trimmed their data of extreme values, even though their aim was to model the fat tails.§ Trimming out extreme data points is controversial and its application is likely to reduce the extent of skewness and kurtosis. Instead, we follow Fama's (1965, p. 42) argument that 'Unlike the statistician, … the investor cannot ignore the possibility of large price changes …'. Excluding outliers which are a feature of the data, throws away important information regarding the data generating process of the series (Mandelbrot 1963). Trimming the data is likely to favourably bias the distributional fit of the Heston model.

Following D-Y, the four parameters γ, θ, k and ν (see equation (5)) were trained in order to fit the empirical returns according to equation (9). The experiment was implemented by minimizing the mean square deviation of

$$E = \sum\nolimits_{x,t} \left|\log P_t^*(x) - \log P_t(x)\right|^2 \tag{10}$$

summed over all available centred log returns x and over all frequencies $t = 1$, 5, 20, 40 and 250 days. Here, $P_t^*(x)$ is the empirical probability mass for a centred return of x and $P_t(x)$ is the probability mass predicted by the D-Y formula in equation (7). We also set the correlation coefficient between the two Weiner processes, ρ, to zero, since D-Y report that the empirical fit does not depend on the value of ρ.

We precisely replicated D-Y's results for their trimmed and reused DJIA8201 returns. We found that the Heston model generally fitted the trimmed empirical returns well but there was some variation in the fit at some frequencies as indicated by D-Y. We also found that the distributional fit was indeed much better than the Gaussian distribution, especially in the tails (see Daniel *et al.* 2003).¶

†The actual data series begins on 4 January 1982 as the first three days of January 1982 were all non-trading days. This would also apply to D-Y's data set, although they did not explicitly state this.
‡D-Y say they had 5049 data points. After accounting for bank holidays between the dates they gave, there are actually 5050 *prices*, giving a maximum of 5049 returns.
§D-Y did not indicate that their data was trimmed. However, we were alerted to this fact on observing strange points in our results using their research design. They kindly provided us with the numerical boundaries that they had used for trimming their data.
¶Those empirical results are reported in Daniel *et al.* (2003) together with the replicated plots for D-Y's results.

2005 style (V 7.51g) [27.5.2005-8:01pm] First Proof [1-13] [Page No. 4] {TandF}Rquf/RQUF-114835.3d (RQUF) RQUF-114835

D-Y judged the fit of the model visually so they could not take account of the greater number of parameters in the Heston model that were estimated from the data (4 rather than 2). It is noteworthy that Silva and Yakovenko (2003) used D-Y's methodology but a different sample period and did not find a good fit for the Heston model.

Both trimming and reusing the data can (favourably) bias the empirical fit of the distribution. Indeed, statistical goodness-of-fit tests are particularly sensitive to the fit of the empirical distribution. Below, we present empirical results where the data is neither trimmed nor reused, and we statistically test the goodness-of-fit of the various models. To do this we put forward our own research design below.

6. Modified research design

To circumvent the potential bias caused by trimming and reusing the data, we modified D-Y's research design as follows. Instead of working with a single time series of $n-t$ centred log returns for each frequency, t, we created t time series, called *paths*, each of length $m=\lfloor(n-t)/t\rfloor$, for each value of $t=1$, 5, 20, 40, 80, 100, 250 days:

$$r_{k,\tau}=\log(S_{k+\tau+t}/S_{k+\tau}),\ \forall k\in[0,t-1],\ \forall\tau\in[1,m]. \tag{11}$$

Each of the t paths for each value of t was then centred, by subtracting from each $r_{k,\tau}$, the mean value of all the $r_{k,\tau}$'s in the corresponding path. Unlike D-Y, no data was trimmed up to this point. As a consequence, the level of kurtosis in our returns is considerably greater.†

As our approach does not reuse the data, each test statistic is repeated on all t paths for each frequency t. We assume that each of the t paths is drawn from one and the same distribution and that the system is ergodic, which we test below. We believe that our research design is acceptable since when financial economists refer to weekly returns, for example, they mean the returns from Monday to Monday, Tuesday to Tuesday and so on, rather than the average of the returns over five consecutive days. Thus assuming no bank holidays, a frequency of say $t=5$ days gives a set of five individual return series—paths—over each of the five days within that frequency. The set of paths with the same frequency also facilitate a robustness check on the statistical analysis. We can therefore compute the average of the test statistics and their standard deviations, based on the combined statistics for the individual paths within the same frequency for any $t\geq 5$ days. We then obtained the frequency partition of log-returns, which we also term the empirical PDF, *empPDF*. We performed goodness-of-fit tests under the assumption that the empirical model follows not only the Heston model (*hestonPDF*) but also the Gaussian (*normPDF*) distribution. Finally, we also fitted a neural network (*nnPDF*) to generate an unknown distribution which we use as a benchmark.‡

7. Experimental results

7.1. *Preliminary results*

Table 1 indicates that the log returns contain much more kurtosis at higher frequency ($t=1$ and 5 days) compared with lower frequencies ($t=200$ and 250 days). This finding was expected. The extent of kurtosis varies across frequencies, and so the empirical fit of the returns will also vary.

The kurtosis measure exhibits a comparatively high variance across the different paths with the same frequency, t. This suggests that the samples might not be ergodic across paths with the same frequency.§

†While the exclusion of observations outside certain ranges is not uncommon in empirical work (see, e.g. Theodossiou 1998, p. 1658), the impact of the trimming undertaken by D-Y which seems extensive, can be dramatically seen by comparing the level of kurtosis for the trimmed and untrimmed DJIA8201 dataset, thus:

t days	1	5	20	40	250
k_{trimmed}	1.40	0.72	0.43	0.56	−0.53
$k_{\text{untrimmed}}$	69.26	19.68	7.80	6.02	−0.33

The (excess) kurtosis values for k_{trimmed} do not appear to be statistically significant under the normal distribution. The corresponding values for $k_{\text{untrimmed}}$ are statistically significant for all frequencies except $t=250$ days.

‡Theoretically, a neural network can closely approximate any unknown continuous function to any desired degree of accuracy (see Conti *et al.* 1994). So we would expect the *unspecified* model generated by the neural network to exhibit the best overall fit and as such it would serve as a useful benchmark for comparison with the other models. The chosen neural network structure was a feed-forward back-propagation (B-P) network, with a five node hidden layer and a single node output layer. The transfer functions are respectively *tansig* and *purelin*, where $\text{tansig}(n)=2/(1+\exp(-2n))-1$ and $\text{purelin}(n)=n$. The B-P function used is *trainscg* where the weight and bias values are updated according to Levenberg–Marquardt optimization. This optimization method minimizes a combination of squared errors and weights and then determines the correct combination so as to produce a network which generalizes well.

§This finding has been noted elsewhere, albeit using alternative sampling procedures (see, e.g. Fama 1965). Variation in the extent of kurtosis suggests that the system is not ergodic. This finding has important implications for the constancy of the parameter estimates and the experimental design. If the system is ergodic we would expect the shape parameters to be almost constant from one path to the next, within the same frequency. So our test for ergodicity for all t paths at each t seeks to test for such variation. In general, both the mean and standard deviation of the returns increase as the frequency decreases, albeit not proportionally. The $\sigma_{k,t}$ appears to stabilize beyond $t>80$ days. This might be because the distribution moves towards the normal distribution beyond this point, whereas for (say) $t<80$ days, kurtosis will be less well defined if the distribution is non-normal.

*** 2005 style (V 7.51g) [27.5.2005–8:01pm] First Proof [1–13] [Page No. 5] {TandF}Rquf/RQUF-114835.3d (RQUF) RQUF-114835

The mean, $\mu_{k,t}$, of the returns and the standard deviation, $\sigma_{k,t}$, of each path k, t were used to test for ergodicity. If the samples are ergodic, these parameters are expected to remain constant across paths with the same frequency. The Kruskal–Wallis (K-W) statistic clearly did not allow us to reject the null hypothesis that the time series formed by the different paths with the same frequency came from a population with the same median (p-value >0.98). So there is support for the equivalence of the different paths with the same frequency in terms of their medians. The K-W statistic, however, does not capture information contained in the sample variance. There was a much stronger rate of variation in $\sigma_{k,t}$ compared with $\mu_{k,t}$ across paths with the same frequency t. Box plots of the different paths (not shown) indicated some degree of variability in the paths with similar frequency. The variability in the time paths might not directly relate to seasonal effects since our paths are based on *trading* rather than calendar days.† Overall, the findings suggest that the log returns for the different time paths are not equivalent. So, for both practical and statistical reasons, the returns in the different paths should not be combined as D-Y did.

Clear evidence of departure from normality is provided by the normal cumulative probability plots for the series at various frequencies (not shown). These did not cluster around the fitted straight line; particularly at higher frequencies ($t=1$ and 5 days), there was the usual elongated *S*-like pattern associated with fat tails. The departure from normality appeared to be primarily associated with excess kurtosis in the returns rather than skewness (see also Badrinath and Chatterjee 1988).

To discount any possibility that the distributions might be normally distributed, both the Jarque-Bera and the Lilliefors statistics were applied to each path separately.‡ Table 2 shows that both test statistics rejected the null hypothesis of normality for high frequency returns. These results were expected and are also consistent with those in table 1. The normality hypothesis, however, cannot be rejected for $t \geq 200$ days.

Table 1. Mean and standard deviation of path kurtosis at various frequencies.The data is non-overlapping and is not trimmed. The kurtosis measure is for all t paths for each frequency t. DJIA8201 and SP8201 represent the returns for the Dow Jones Industrial Average index and the Standard and Poor's 500 index returns, respectively, for the period 1 January 1982 to 31 December 2001. FTSE8401 represents the returns for the Financial Times Stock Exchange 100 index for the period 1 January 1984 to 31 December 2001. All data points excluded bank holiday observations.

Frequency t days	DJIA8201	SP8201	FTSE8401
1	69.27	50.12	12.72
5	16.87 ± 16.20	13.25 ± 12.21	12.34 ± 5.46
20	7.75 ± 4.77	6.77 ± 4.49	10.67 ± 5.99
40	5.69 ± 2.40	4.97 ± 2.14	5.48 ± 2.44
80	2.75 ± 1.21	2.20 ± 1.09	2.70 ± 2.08
100	1.68 ± 0.99	1.30 ± 0.98	2.10 ± 1.70
200	-0.06 ± 0.57	0.28 ± 0.57	0.14 ± 0.86
250	-0.52 ± 0.37	-0.18 ± 0.78	-0.04 ± 0.74

Table 2. Proportion of paths rejecting the null hypothesis.The tests were run on each of the t paths of each frequency, t. If the null hypothesis was rejected at a 5% significance level, the output was given a value of one. Otherwise, i.e. if the hypothesis was accepted, the output was given a value of zero.

Frequency t days	DJIA8201		SP8201		FTSE8401	
	Jarque-Bera	Lilliefors	Jarque-Bera	Lilliefors	Jarque-Bera	Lilliefors
1	1	1	1	1	1	1
5	1	1	1	1	1	1
20	1	0.85	1	0.70	1	0.65
40	1	0.72	1	0.65	0.97	0.42
80	0.90	0.19	0.82	0.14	0.80	0.16
100	0.60	0.06	0.43	0.11	0.54	0.08
200	0	0.04	0.04	0.09	0	0.03
250	0	0.02	0.04	0.01	0	0.11

†This finding might be thought to be due to seasonal effects in the index returns, as depicted in the finance literature (see, e.g. Keim 1989). We believe this to be unlikely in our case. Consider a path beginning on a Monday. A frequency of 5 days will shift to Tuesday after the first public holiday, to Wednesday after the second holiday and so on, returning to Monday after 5 public holidays. There are about 10 public holidays in each year under consideration, so in each year any one path is forwarded about 10 times. So a path for $t = 5$ will have data from all the days of the week within about 6 months, although not in equal proportions. Thus, while we cannot rule out a seasonal day effect in our paths covering 18 or 20 years, it is not likely to be present.

‡These standard statistical tests typically reject the normality hypothesis for high frequency data. The (asymptotic) Jarque-Bera statistic tests a composite normality hypothesis which means that it assumes that the unknown parameters for computing the test statistic can be estimated from the data. The Lilliefors statistic also tests a composite hypothesis but, unlike the Jarque-Bera statistic, emphasizes the maximum departure of the empirical distribution from the normal distribution. When the sample size is large almost any goodness-of-fit tests would reject normality. However, the magnitude of the test statistics shown here does not suggest that approximate normality is a reasonable conclusion. Since our goodness-of-fit tests are performed on untrimmed data sets, we do not expect as good an empirical fit as D-Y have claimed. Indeed, in an application of both the Jarque-Bera and Lilliefors tests, the trimmed data exhibited a much better fit than the untrimmed data.

7.2. *Graphical plots*

Figure 1 graphically shows the plots for the empirical data. These are the centred log returns for frequencies $t=1$, 5, 20, 250 days for the untrimmed DJAI8201, SP8201 and FTSE8401, together with the PDF plots of the fitted models.† To aide clarity, only the first path for each frequency is shown. The plots show that the Heston model captures the fat tails much better than the Gaussian. The fit of the DJIA8201 in figure 1 is much worse than the fit shown in D-Y's figure 2 (p. 446). We attribute this difference to the bias in D-Y's research design. However, figure 1 suggests that the Heston model fits the empirical cumulative distribution better than the Gaussian model, even when the data is not trimmed.‡

To further examine the adequacy of the distributional fit, figure 2 shows the cumulative density functions for the statistical models at $t=5$. Clearly, the adequacy of the empirical fit depends on the theoretical distribution. If we zoom in around the lower tails of the fitted models, we can see at the extreme tail that both models underestimate the empirical data. This finding would suggest that the Heston model would underprice or overprice an option depending on the option's moneyness. The plots also show that, at values less extreme than the lowest 2% of the data, the Gaussian model overestimates the empirical data while the Heston model slightly underestimates the empirical data in the lowest 35%. We illustrate this for $t=5$ in figure 3.

Appearances, however, can be misleading. The plots do not give us much confidence in making strong

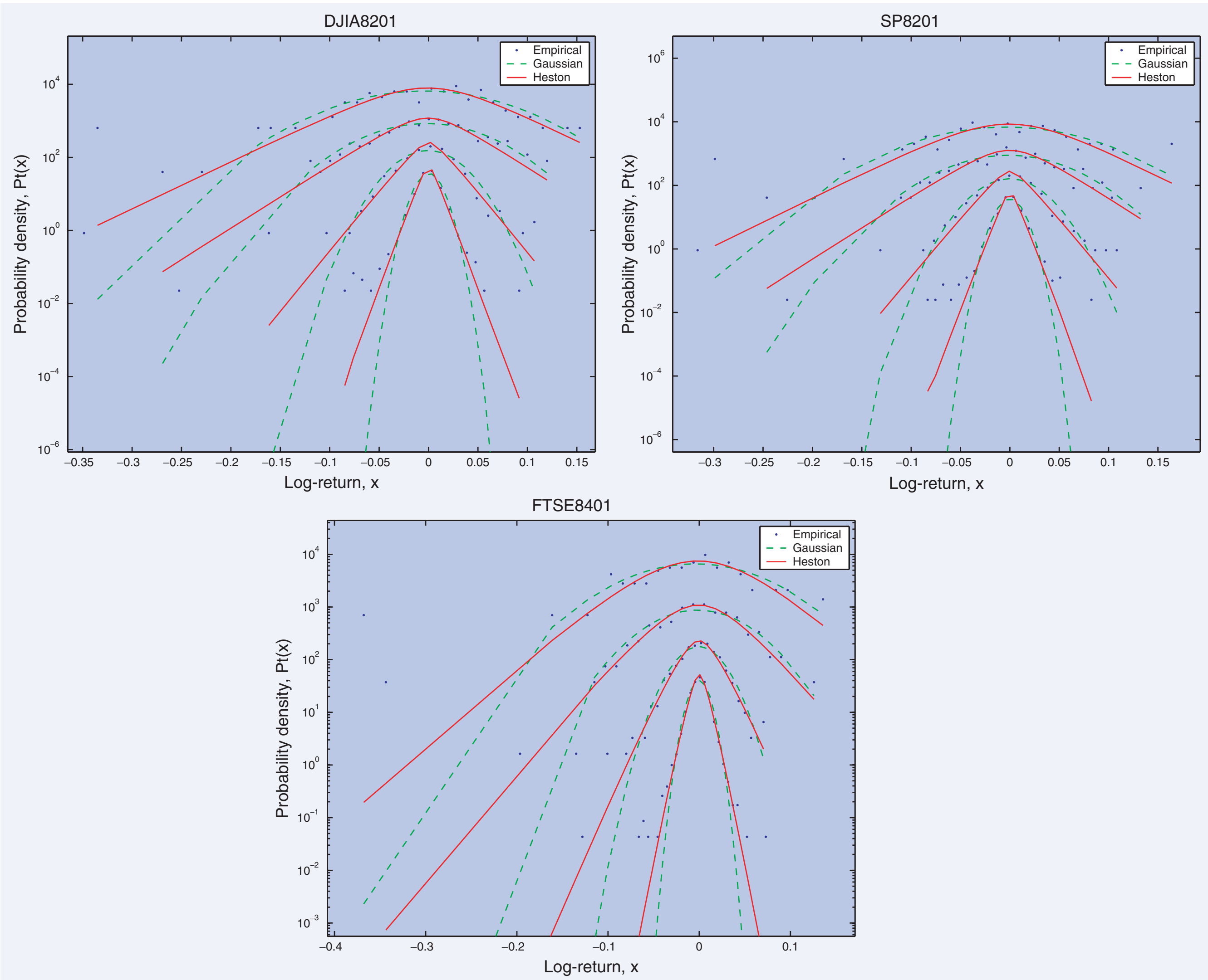

Figure 1. Probability plots of the empirical distribution (dots), the fitted Gaussian (dash line) and the Heston model (straight line) at various frequencies.

†The data for different frequencies have been shifted up by a multiple of 10 in order to separate them out in the figure.
‡We also fitted the *nnPDF* to the *empPDF*. Those plots are not shown due to the need to retain clarity in both the plots and the focus of the paper. These results can be obtained from the authors. The neural network provides the best *overall* fit, particularly at high frequencies. The overall performance of the neural network in terms of statistical goodness-of-fit is discussed in subsequent subsections.

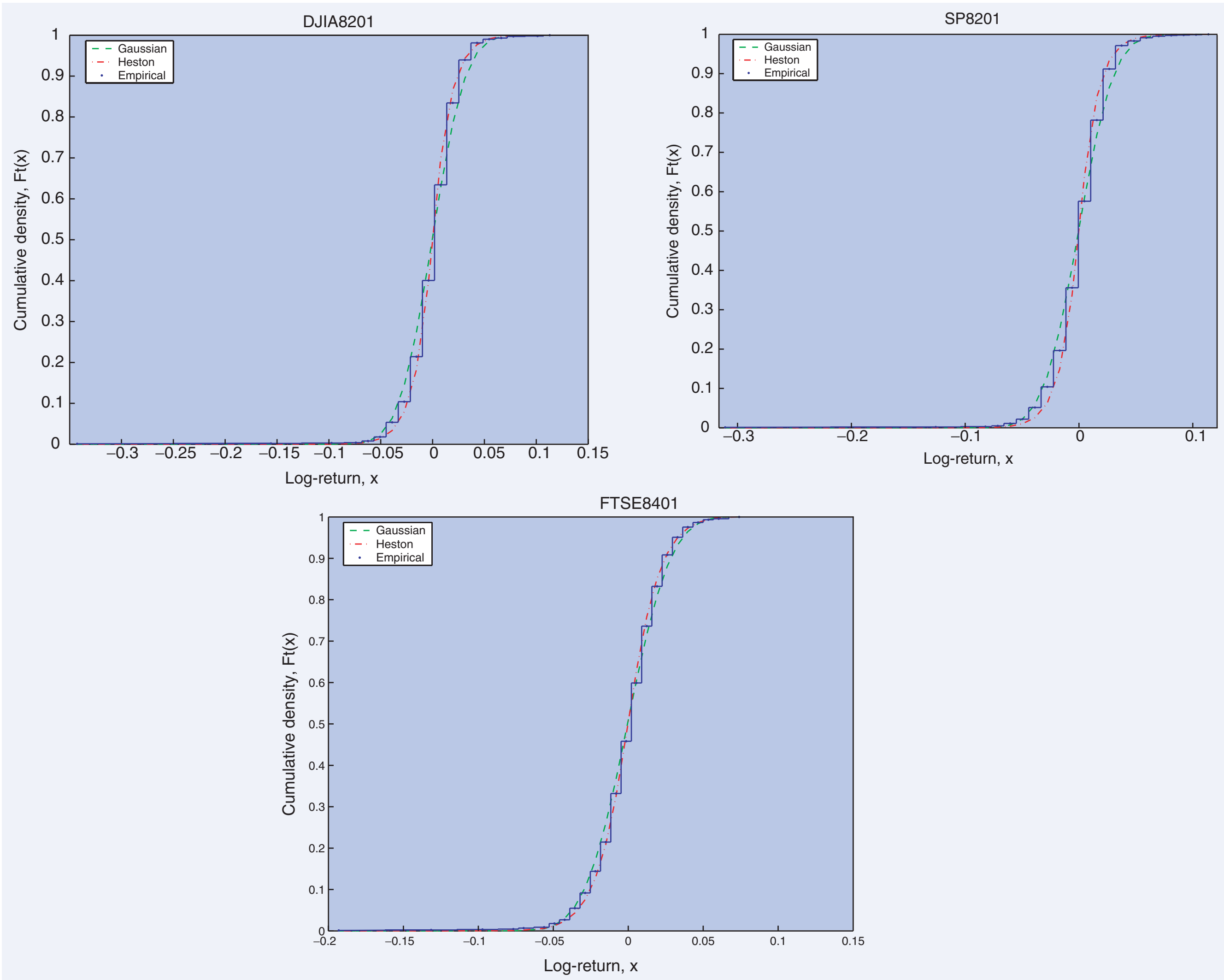

Figure 2. Cumulative density plots of the empirical distribution (dots), the fitted Gaussian (dash line) and the Heston model (straight line) for five-day returns.

inferences regarding the fit of the Heston model or the Gaussian model despite the differences in our research design. Therefore, we statistically tested the goodness-of-fit of the models.

7.3. *Statistical goodness-of-fit tests*

Both the chi-square χ^2 and the Kolmogorov–Smirnov Z statistics were used to test the goodness-of-fit of the empirical distributions to the models. While the parameter estimates for the χ^2 statistic can be derived from the empirical dataset, unfortunately, the Z statistic is based on a simple hypothesis. However, whatever the test statistic, we derived the parameters: (i) $\mu_{k,t}$ and $\sigma_{k,t}$ for *normPDF*; (ii) γ, θ, k and μ for *hestonPDF*; and (iii) the weights and biases for *nnPDF*, all from the empirical dataset. The associated results are presented below.

7.3.1. Chi-square statistic. To compute the χ^2 statistic the observations for the series were partitioned into bins. We split the log return axis into equal expected occupancy ($\geq$5) bins so that none of the log returns would be discarded or re-combined on finding bins that contained an expected occupancy of less than 5. The χ^2 statistic could not be computed for the log returns at lower frequency ($t > 80$ days) due to a loss of degrees of freedom (DF). The number of DF is: *noBins-1-noParas*, where *noParas* is the number of parameters of the model. *noParas* = 2 for the Gaussian, *noParas* = 4 for Heston model and *noParas* = 11 for the neural network. For each path and each of the three datasets we built the empirical cumulative density function *empCDF* and also the expected cumulative density functions (CDFs) for each of the three models (*normCDF*, *hestonCDF* and *nnCDF*).

Table 3 shows the mean χ^2 value, $\hat{\chi}^2$, and its standard deviation, $\hat{\sigma}_{\chi^2}$, for each path for each of the models. The standard deviation seems relatively large in relation to $\hat{\chi}^2$, especially at low frequencies. This suggests that the different paths for the same frequency are not equivalent, which legitimizes, *a posteriori*, our method of creating *t different paths of length* $\lfloor (n-t)/t \rfloor$ *for each frequency t, instead of using just one time series of length* $n-t$.

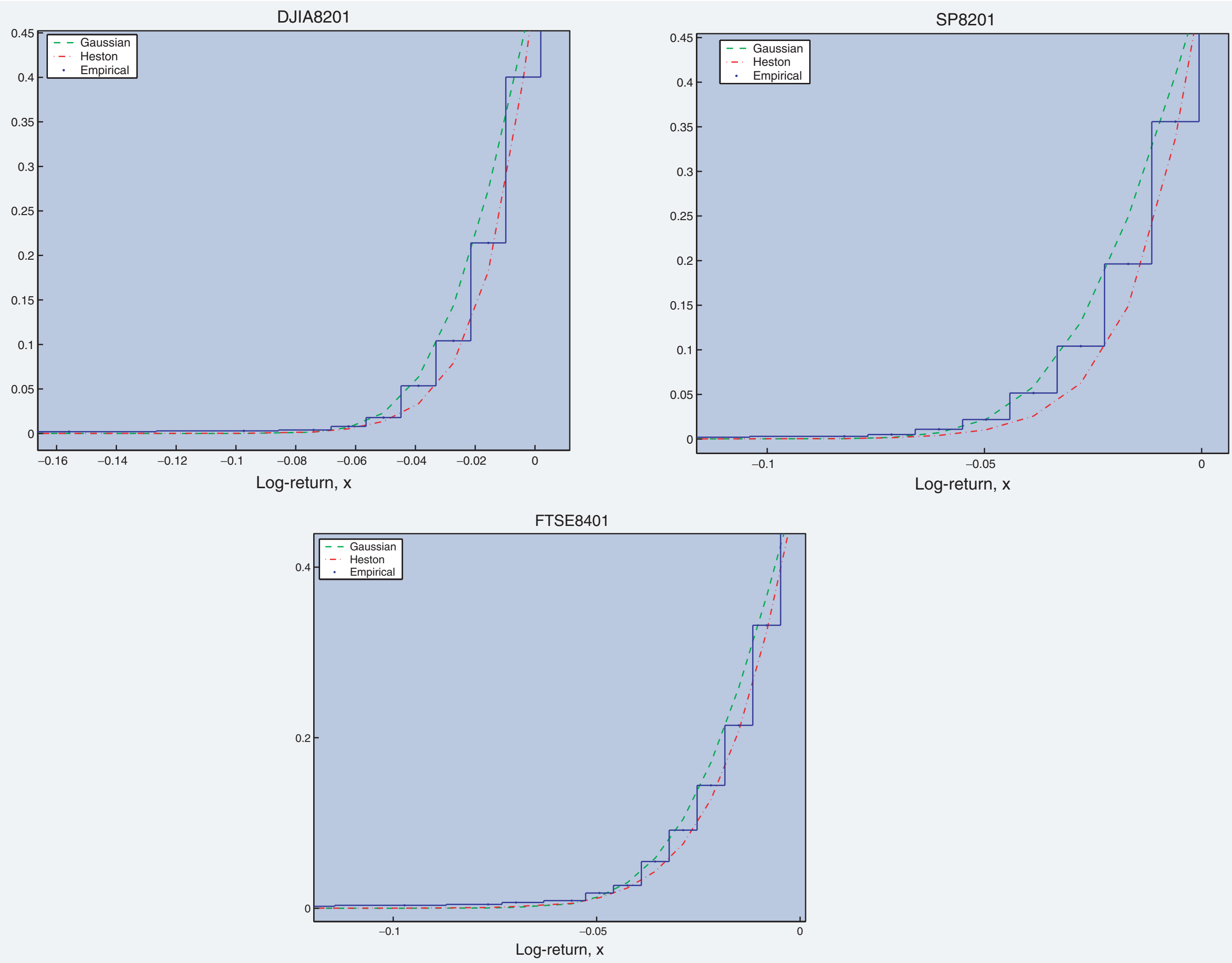

Figure 3. Zoom around the lower tail of the cumulative density plots of the empirical distribution (dots), the fitted Gaussian (dash line) and the Heston model (straight line) for five-day returns.

The mean $\hat{\chi}^2$ values indicate that the Heston model outperforms the Gaussian model for all but one frequency. This comparison using $\hat{\chi}^2$ alone takes no account of the two extra parameters to be estimated for the Heston model; inferences based on the magnitude of $\hat{\chi}^2$ are affected by the number of DF. Since the p-values associated with the estimates are more reliable, these are given for $\hat{\chi}^2 - \hat{\sigma}_{\chi^2}$ and $\hat{\chi}^2 + \hat{\sigma}_{\chi^2}$ (i.e. ± 1 from $\hat{\chi}^2$) in addition to $\hat{\chi}^2$, for all frequencies except $t = 1$ day.

At $t = 1$ day, the fit of the Heston model is unsatisfactory for all three datasets, although it is better than the fit provided by the Gaussian model. Notice that at $t = 1$ day, we have only one path and therefore the test statistic would lack power relative to those for $t \geq 5$ days. The Heston model also outperforms the Gaussian model at $t = 5$ and 20 days, but not significantly; at $t = 5$ it is, however, only marginally satisfactory. While both the Heston and the Gaussian models provide a satisfactory fit for $t = 40$ and 80 days, importantly, the Gaussian model marginally outperforms the Heston model on these frequencies.

As expected, the neural network outperforms both models at $t = 1$ and 5 days, but, contrary to expectations, not at lower frequencies. Also at $t = 1$ day, the fit of the neural network is marginally unsatisfactory (p-value $\geq$ 0.084). So the neural network does not consistently generate a distribution that fits at all frequencies well because of its relatively large number of parameters.

Thus no model consistently provides a superior fit at all frequencies. For example, at $t = 1$ and 5 days, the p-values generated for the neural network are better than those of the other models. At $t = 20$ and 40 days, the Heston model marginally outperforms all the other models. The Gaussian model generally provides a better fit at lower frequencies, partly due to its reduced complexity having a proportionally greater effect on the DF and partly to the central limit theorem. Overall, according to the χ^2 statistic the Heston model provides a better (albeit unsatisfactory) fit than the Gaussian model for high frequencies ($t \leq 20$ days), but the Gaussian model is better for low frequency returns ($t = 40$, 80 days).

7.3.2. Kolmogorov–Smirnov statistic. The goodness-of-fit of the models was also assessed using the Kolmogorov–Smirnov Z (see table 4). This statistic is

Table 3. Results of chi-square χ^2 goodness-of-fit statistic on models with parameters estimated from the data.

	Gaussian			Heston model			Neural network		
Frequency t days	$\hat{\chi}^2 \pm \hat{\sigma}_{\chi 2}$	DF	p-Values at: $\hat{\chi}^2 - \hat{\sigma}_{\chi 2}, \hat{\chi}^2, \hat{\chi}^2 + \hat{\sigma}_{\chi 2}$	$\hat{\chi}^2 \pm \hat{\sigma}_{\chi 2}$	DF	p-Values at: $\hat{\chi}^2 - \hat{\sigma}_{\chi 2}, \hat{\chi}^2, \hat{\chi}^2 + \hat{\sigma}_{\chi 2}$	$\hat{\chi}^2 \pm \hat{\sigma}_{\chi 2}$	DF	p-Values at: $\hat{\chi}^2 - \hat{\sigma}_{\chi 2}, \hat{\chi}^2, \hat{\chi}^2 + \hat{\sigma}_{\chi 2}$
Panel A: DJIA8201									
1	1790.0	1010	6.29e−11	1420.0	1000	1.16e−04	2230.0	997	0.084
5	255.0 ± 30.0	198	5.38e−05, 4.07e−03, 0.093	244.0 ± 26.0	196	332e−04, 0.011, 0.133	232.0 ± 38.0	189	0.056, 0.346, 0.817
20	61.0 ± 12.0	47	7.99e−03, 0.082, 0.409	48.50 ± 11.5	45	0.066, 0.333, 0.796	45.9 ± 11.1	38	0.047, 0.250, 0.688
40	29.1 ± 7.0	22	0.029, 0.141, 0.451	27.30 ± 6.1	20	0.030, 0.126, 0.385	21.5 ± 6.3	13	0.006, 0.055, 0.333
80	10.4 ± 4.6	9	0.091, 0.320, 0.760	9.70 ± 4.4	7	0.049, 0.206, 0.624	7.6 ± 6.3	0	N/A
Panel B: SP8201									
1	1401.0	1002	7.77e−16	1137.0	1000	1.16e−03	1035.0	993	0.170
5	242.4 ± 24.0	198	8.16e−04, 0.017, 0.152	219.0 ± 22.0	196	0.014, 0.120, 0.471	217.0 ± 36.7	189	1.16e−03, 0.079, 0.662
20	55.6 ± 10.1	47	0.036, 0.181, 0.536	47.9 ± 8.9	45	0.110, 0.354, 0.722	45.4 ± 7.1	38	0.058, 0.191, 0.458
40	27.5 ± 6.9	22	0.044, 0.192, 0.548	23.9 ± 7.2	20	0.053, 0.245, 0.673	20.9 ± 6.8	13	9.56e−03, 0.074, 0.369
80	9.20 ± 3.6	9	0.167, 0.410, 0.772	8.0 ± 3.2	7	0.124, 0.326, 0.684	7.5 ± 3.9	−1	N/A
Panel C: FTSE8401									
1	1083.0	890	8.70e−06	974.0	888	0.023	906.0	881	0.269
5	235.4 ± 15.9	184	7e−04, 6.27e−03, 0.038	212.3 ± 23.5	182	0.004, 0.061, 0.348	199.5 ± 10.9	175	0.035, 0.099, 0.228
20	50.8 ± 13.6	43	0.019, 0.193, 0.718	44.4 ± 11.6	41	0.059, 0.330, 0.815	38.9 ± 10.1	34	0.046, 0.258, 0.720
40	22.9 ± 5.0	20	0.112, 0.295, 0.597	21.2 ± 5.5	18	0.086, 0.271, 0.614	17.2 ± 5.1	11	0.022, 0.101, 0.356
80	9.69 ± 3.3	8	0.114, 0.287, 0.599	9.9 ± 3.8	6	0.032, 0.128, 0.416	7.5 ± 4.3	−1	N/A

DF: degrees of freedom. N/A indicates that the appropriate test statistic could not be computed because of the loss of DF. For each frequency, t. the data in each dataset was divided into t non-overlapping *paths* and the χ^2 statistic was computed for each path. The p-values are for values of χ^2 of $\hat{\chi}^2 - \hat{\sigma}_{\chi 2}, \hat{\chi}^2, \hat{\chi}^2 + \hat{\sigma}_{\chi 2}$ where $\hat{\chi}^2$ is the mean and $\hat{\sigma}_{\chi^2}$ is the standard deviation of the χ^2 statistic over all the t paths with frequency t.

Table 4. Results of Kolmogorov–Smirnov goodness-of-fit Z statistic on models with parameters estimated from the data.

	Gaussian		Heston model		Neural network	
Frequency t days	$\hat{Z} \pm \hat{\sigma}_z$	p-Values at: $\hat{Z} - \hat{\sigma}_z, \hat{Z}, \hat{Z} + \hat{\sigma}_z$	$\hat{Z} \pm \hat{\sigma}_z$	p-Values at: $\hat{Z} - \hat{\sigma}_z, \hat{Z}, \hat{Z} + \hat{\sigma}_z$	$\hat{Z} \pm \hat{\sigma}_z$	p-Values at: $\hat{Z} - \hat{\sigma}_z, \hat{Z}, \hat{Z} + \hat{\sigma}_z$
Panel A: DJIA8201						
1	0.131	2.39e−75	0.109	1.2 e−53	0.106	1.2e−53
5	0.081 ± 0.020	1.75e−09, 2.98e−06, 9.88e−04	0.087 ± 0.019	2.08e−10, 3.64e−07, 1.48e−04	0.048 ± 0.014	1.64e−03, 0.026, 0.204
20	0.089 ± 0.013	9.51e−03, 0.036, 0.112	0.089 ± 0.014	8.75e−03, 0.033, 0.104	0.047 ± 0.009	0.430, 0.615, 0.778
40	0.104 ± 0.020	0.038, 0.122, 0.321	0.094 ± 0.010	0.125, 0.211, 0.337	0.071 ± 0.059	0.153, 0.746, 1.000
80	0.113 ± 0.021	0.199, 0.385, 0.649	0.116 ± 0.018	0.197, 0.355, 0.576	0.075 ± 0.061	0.729, 0.919, 0.995
100	0.113 ± 0.021	0.322, 0.533, 0.778	0.128 ± 0.019	0.215, 0.372, 0.585	0.076 ± 0.039	0.532, 0.871, 0.999
200	0.148 ± 0.038	0.339, 0.630, 0.917	0.163 ± 0.048	0.209, 0.512, 0.893	0.116 ± 0.034	0.126, 0.453, 0.932
250	0.170 ± 0.047	0.291, 0.598, 0.919	0.186 ± 0.046	0.224, 0.481, 0.816	0.137 ± 0.041	0.190, 0.502, 0.904
Panel B: SP820						
1	0.082	1.26e−29	0.029	3.80e−04	0.023	0.008
5	0.075 ± 0.014	2.13e−07, 2.18e−05, 0.001	0.087 ± 0.018	2.85e−10, 3.97e−07, 1.38e−04	0.042 ± 0.011	0.006, 0.056, 0.297
20	0.082 ± 0.017	0.013, 0.062, 0.226	0.084 ± 0.008	0.027, 0.055, 0.107	0.047 ± 0.006	0.456, 0.619, 0.786
40	0.107 ± 0.021	0.031, 0.108, 0.304	0.097 ± 0.020	0.060, 0.177, 0.421	0.055 ± 0.008	0.691, 0.853, 0.941
80	0.107 ± 0.020	0.259, 0.459, 0.715	0.113 ± 0.018	0.225, 0.391, 0.614	0.069 ± 0.018	0.722, 0.921, 0.996
100	0.115 ± 0.025	0.275, 0.516, 0.808	0.124 ± 0.021	0.231, 0.410, 0.649	0.079 ± 0.017	0.738, 0.908, 0.989
200	0.175 ± 0.054	0.140, 0.419, 0.850	0.192 ± 0.060	0.079, 0.306, 0.766	0.171 ± 0.062	0.126, 0.450, 0.927
250	0.182 ± 0.040	0.271, 0.509, 0.801	0.201 ± 0.040	0.188, 0.387, 0.674	0.189 ± 0.060	0.165, 0.466, 0.886
Panel C: FTSE8401						
1	0.046	1.03e−08	0.039	1.95e−06	0.047	3.59e−09
5	0.073 ± 0.003	6.9e−05, 1.45e−04, 2.97e−04	0.073 ± 0.005	3.48e−05, 1.23e−04, 4.01e−04	0.042 ± 0.005	0.038, 0.089, 0.188
20	0.086 ± 0.012	0.025, 0.068, 0.162	0.088 ± 0.014	0.017, 0.060, 0.174	0.049 ± 0.007	0.464, 0.632, 0.803
40	0.100 ± 0.016	0.088, 0.198, 0.392	0.107 ± 0.016	0.061, 0.144, 0.300	0.059 ± 0.010	0.639, 0.820, 0.949
80	0.124 ± 0.025	0.157, 0.340, 0.623	0.128 ± 0.021	0.157, 0.305, 0.529	0.081 ± 0.024	0.551, 0.844, 0.992
100	0.117 ± 0.025	0.323, 0.566, 0.836	0.127 ± 0.024	0.256, 0.465, 0.732	0.089 ± 0.030	0.546, 0.870, 0.998
200	0.169 ± 0.034	0.320, 0.549, 0.812	0.184 ± 0.036	0.227, 0.433, 0.711	0.173 ± 0.064	0.163, 0.515, 0.953
250	0.202 ± 0.046	0.238, 0.481, 0.796	0.225 ± 0.046	0.160, 0.350, 0.644	0.215 ± 0.082	0.096, 0.404, 0.922

The mean and standard deviation of the Z statistic were obtained in a similar manner as the χ^2 statistic in table 3. The p-values are for values of the values of Z of $\hat{Z} - \hat{\sigma}_z, \hat{Z}, \hat{Z} + \hat{\sigma}_z$, where $\hat{Z}$ is the mean and $\hat{\sigma}_Z$ is the standard deviation of the Z statistic, over the t paths for the time lag t. Note that these p-values are overestimated as no allowance could be made for the model parameters which were estimated from the data.

based on the maximal discrepancy between the expected and the observed cumulative distributions, for any return. So the test is sensitive to any difference in dispersion, including skewness and kurtosis. The Z statistic assumes that the distribution is continuous and that the parameters for the model are known. In any case, we expect the Z statistic to be large enough to reject the simple hypothesis and, *a fortiori*, the composite hypothesis (Brée 1975). If however, the value of Z is too small to reject the simple hypothesis, it does not mean that we can accept the composite hypothesis.

As before, table 4 shows the mean Z value, $\hat{Z}$, and its standard deviation, $\hat{\sigma}_Z$, for each frequency, together with their p-values. The magnitude of the p-values, under the (incorrect) assumption that the parameters were not estimated from the data, suggests that all the models are acceptable except for high frequencies, $t = 1$ and 5 days.

The table shows that the Gaussian generally has smaller Z statistics than the Heston model at almost all frequencies. However, these differences are not large and are well within one standard deviation of each other. As expected, the neural network outperforms the other models, except at the lowest frequencies, $t = 200$ and 250 days, where the Gaussian has a marginally better fit. Furthermore, there is not much difference between the fit provided by the Gaussian model and the Heston model.

In summary, the Z statistic accepts both the Heston and the Gaussian models for low frequency data ($t \geq 40$ days) but rejects them both for high frequency data ($t = 1$, 5 days). At low frequency the Gaussian model provides a better fit than the Heston, but not significantly so. For all frequencies except $t = 1$ day, the neural network provides a satisfactory fit and the best fit, as expected. The finding that the neural network is unable to find a satisfactory fit at $t = 1$ suggests that researchers will find it very difficult to specify a statistical distribution that provides a good fit for high frequency data.

8. Conclusions and implications

This empirical study focused on the Heston model as a statistical description of stock returns. We initially adopted D-Y's empirical research design and found it to be unreliable; they overstated their empirical results. We developed our own research design and then compared the empirical fit of the Heston model together with the fit of two other statistical models. Our goodness-of-fit tests confirm D-Y's claim that the Heston model fits the empirical data (p-value > 0.05), but this is so only for frequencies of 20 days or more. For returns at higher frequencies of 1 and 5 days the Heston model does not provide an acceptable fit to the data, but neither does the Gaussian model. The Heston model does, however, provide a better (albeit inadequate) fit than the Gaussian at these high frequencies. The Gaussian model, which is assumed for the B-S model, marginally outperformed the Heston model at low frequencies (40 to 250 days). Finally, the distribution generated by the neural network generally provides an acceptable fit to the returns, except at 1 day, as well as the best fit for higher frequencies.

Our results have important implications. The finding that the distributional fit of returns varies according to their frequency suggests that the extent of volatility can affect the distribution of returns (see also Engle and Patton 2001). Most simple statistical models, e.g. the normal distribution, perform reasonably well with low frequency data. The adequacy of the fit of the Heston model is therefore not encouraging given the complexity associated with its estimation. So some caution is called for when using the basic Heston model in (say) option pricing, as it does not adequately capture the volatility in returns. This basic result corroborates those of Fiorentini *et al.*'s (2002) who based their empirical results on option pricing data. If time-varying skewness has a stronger impact on volatility persistence than on kurtosis (see Harvey and Siddique 1999) then the Heston model might not fit the data well. The assumption of a non-time dependent drift parameter μ in the simple version of the model can also affect the empirical fit. Under those conditions, one might prefer the B-S model for practical purposes and particularly for low frequency data.

There are two important limitations of this empirical work which should be noted. Firstly, we have restricted the analysis to the use of the same parameters estimated by D-Y and have assumed that any variation in the magnitude of those parameters would not alter the reported results. As noted elsewhere, the inclusion of additional parameters in the Heston model such as random-jumps, (see, e.g. Bakshi *et al.* 1997), can improve on the distributional fit. Secondly, we have used spot returns to assess the distributional fit of stochastic volatility models. Future research should seek to explore those issues further.

Acknowledgements

We would like to thank Bin Yang for assisting in the data analysis. We are also grateful to the conference participants at the *9th Annual Meeting of the Society of Computational Economics* in July 2003 for their helpful comments on an earlier draft of part of this paper. We are very grateful to two anonymous referees of this journal for their invaluable comments on an earlier draft of this paper. The usual disclaimer applies.

References

Bachelier, L., Théorie de la speculation, *Ann. Sci. L'Ec. Norm. Supér.*, 1900, **17**, 21–88.

Badrinath, S. and Chatterjee, S., On measuring skewness and elongation in common stock return distributions: the case of the market index. *J. Bus.*, 1988, **61**, 451–472.

Bakshi, G., Cao, C. and Chen, Z., Empirical performance of alternative option pricing models. *J. Finance*, 1997, **52**, 2003–2049.

Blattberg, R. and Gonedes, N., A comparison of the Stable and Student distributions as statistical models for stock prices. *J. Bus.*, 1974, **47**, 244–280.

Bollerslev, T., Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 1986, **31**, 307–327.

Bookstaber, R. and McDonald, J., A general distribution for describing security price returns. *J. Bus.*, 1987, **60**, 401–424.

Bouchaud, J.-P. and Potters, M., *Theory of Financial Risk and Derivative Pricing: from Statistical Physics to Risk Management*, 2nd ed., pp. 141–144, 2004 (Cambridge University Press: Cambridge).

Brée, D., The distribution of problem-solving times: an examination of the stages model. *Br. J. Math. Stat. Psychol.*, 1975, **28**, 177–200.

Chunhachinda, P., Dandapani, K., Hamid, S. and Prakash, A., Portfolio selection and skewness: evidence from international stock markets. *J. Banking Finance*, 1997, **21**, 143–167.

Conti, M., Orcioni, S. and Turchetti, C., 1994, A class of neural networks based on approximate identity for analog IC's hardware implementation. *IEICE Trans. Fundam.*, **E77-A(6)**, 1069–1079.

Daniel, G., Joseph, N. and Brée, D., Goodness-of-fit of the Heston model, in *Ninth Annual Meeting of the Society of Computational Economics*, 2003, reprint: http://arXiv.org/abs/cs.CE/0305055.

Drăgulescu, A. and Yakovenko, V., Probability distribution of returns in the Heston model with stochastic volatility. *Quant. Finance*, 2002, **2**, 443–453.

Engle, R., Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, 1982, **50**, 987–1006.

Engle, R. and Ng, V., Measuring and testing the impact of news on volatility. *J. Finance*, 1993, **48**, 1749–1778.

Engle, R. and Patton, A., What good is a volatility model? *Quant. Finance*, 2001, **1**, 237–245.

Fama, E., The behavior of stock-market prices. *J. Bus.*, 1965, **38**, 34–105.

Fiorentini, G., León, A. and Rubio, G., Estimation and empirical performance of Heston's stochastic volatility model: the case of thinly traded market. *J. Empirical Finance*, 2002, **9**, 225–255.

Franses, P. and Ghijsels, H., Additive outliers, GARCH and forecasting volatility. *Int. J. Forecasting*, 1999, **15**, 1–9.

Harvey, C. and Siddique, A., Autoregressive conditional skewness. *J. Financ. Quant. Anal.*, 1999, **34**, 465–487.

Heston, S., A closed form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.*, 1993, **6**, 327–343.

Keim, D., Trading patterns, bid-ask spreads, and estimated security returns: the case of common stocks at calendar turning points. *J. Financ. Econ.*, 1989, **25**, 75–97.

Hull and White, The pricing of options on assets with stochastic volatilities. *J. Finance*, 1987, **42**, 281–300. 2

Mandelbrot, B., The variation of certain speculative prices. *J. Bus.*, 1963, **36**, 394–419.

Micciché, S., Bonanno, G., Lillo, F. and Mantegna, R.N., Volatility in financial markets: stochastic models and empirical results. *Physica* A, 2002, **314**, 756–761.

Nelson, D., Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 1991, **59**, 347–370.

Rubinstein, M., Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. *J. Finance*, 1985, **40**, 455–480.

Serva, M., Fuloc, U., Lyra, M. and Viswanathan, G., Kinematics of stock prices. Preprint, 2002, http://lanl.arXiv.org/abs/cond-mat/0209103.

Silva, A. and Yakovenko, V., A comparison between the probability distribution of returns in the Heston model and the empirical data for stock indices. *Physica A*, 2003, **324**, 303–310.

Theodossiou, P., Financial data and the skewed generalization of T distribution. *Manag. Sci.*, 1998, **44**, 1650–1661.

AUTHOR QUERIES

Journal id: ***RQUF 114835***

Query number	Query
1	Please provide Received Date.
2	Initials for Hull White are missing?