Article

Volatility is log-normal — but not for the reason you think

Martin Tegnér 1 and Rolf Poulsen 2,*

1 University of Oxford, Department of Engineering Science; martin.tegner@eng.ox.ac.uk
2 University of Copenhagen, Department of Mathematical Sciences; rolf@math.ku.dk
* Correspondence to Poulsen

Version February 16, 2018 submitted to Risks

Abstract: Despite first appearances, it is impossible to discriminate between the commonly used continuous-time stochastic volatility models Heston, log-normal, and 3-over-2 on the basis of exponentially weighted averages of squared daily rates of returns. However, with 5-minute sampling frequency, the models can be told apart and empirical evidence overwhelmingly favours a fast mean-reverting log-normal model.

Keywords: Volatility, estimation, Heston, log-normal, 3-over-2, fast mean-reversion

1. Introduction: Seeing and believing

A widespread method for measuring the instantaneous return variance1 of a price process $S$ is by an exponentially weighted average of past squared rates of returns,

$$\sigma_t^2 = \frac{1 - \lambda}{1 - \lambda N + 1} \sum_{j=0}^{N} \left( \ln \left( \frac{S_{t-j}}{S_{t-(j+1)}} \right) \right)^2 \lambda^j,$$

see for instance RiskMetrics (1996, Table 5.1), Wilmott (1998, Section 45.3), or Hull (2009, Chapter 17).

In Figure 1 we apply this to 20 years of daily rates of return on the S&P500 index with the suggested values of $\lambda = 0.94$ and $N = 20$ and plot the non-parametrically estimated density. We also show the best-fitting log-normal and gamma densities. The log-normal distribution gives a decent fit, the gamma distribution does not. Ipso facto: Clear empirical evidence against the Heston stochastic volatility model – whose stationary distribution is gamma – and for a log-normal model of volatility.

But there are problems with this reasoning. The first comes immediately when we want to go from an eyeball test to a quantitative test. The Kolmogorov-Smirnov (K-S) test $p$-values for, respectively, log-normal and gamma are $7.8 \cdot 10^{-5}$ and $2.2 \cdot 10^{-16}$; so even though the log-normal is better both distributions are deemed way off. However, the K-S test assumes independence, and the observations used to create the density in Figure 1 are far from that; their first order auto-correlation is about 0.99. This makes differences look more significant than they are. So the empirical density might be log-normal. But continuing that line of thought, how do we know the gamma density doesn’t

1 Often, we will use the term (instantaneous) volatility. This means taking the square root of the instantaneous variance and annualizing it by multiplying it by $\sqrt{252}$, where 252 represents the (approximate) no. trading days in a year. The annualization is a convention, not a theorem; when returns are not iid — which they are definitely not in mean-reverting stochastic volatility models — we do not get the standard deviation of yearly returns by multiplying the standard deviation daily returns by $\sqrt{252}$. Notice further that log-normal models are stable to roots and squares, so it is not misleading to say “volatility is log-normal” even though we model variance. The Heston and the 3-over-2 models do not posses this stability.
pass muster as well? The answer lies in realizing that a diffusion model makes much stronger statements about the behaviour of volatility than just its unconditional (or stationary or asymptotic) distribution. The model determines conditional distributions at any time-horizon and for any initial value of volatility. In the 1990’s a sound theory for statistical inference for diffusion processes was established; we’ll just use that. Then comes another problem. Equation (1) does not give us the true continuous-time object that Heston’s and other stochastic volatility models describe; it is a measured quantity and as such contaminated by noise.

In this paper we attack these problems in order to discover, what we can and cannot say about volatility with statistical certainty. The answers are in the title. More specifically, we look at three continuous-time models (Heston, log-normal, and 3-over-2) and show that with daily measurements, equation (1) cannot be used to separate the models. One-day-head conditional distributions of volatility are too noisy to capture the fine structure of the models, and an analysis like in Figure 1 may not only be inconclusive, but outright misleading; for all models the log-normal density performs best on the stationary distribution of measured variance. However, shifting to a 5-minute observation frequency, which is quite feasible in liquid markets, it is possible to discriminate between the models both in controlled simulation studies and on market data. We find overwhelming support for a fast mean-reverting log-normal model, both for the S&P500 index and at individual stock level. This is not just a case of the one-eyed man being king in the land of the blind. We find that the log-normal model passes goodness-of-fit tests convincingly and is robust to jump corrections.

The rest of the paper is organized as follows: Section 2 reviews theory of models, inference, and measurement, Section 3 is a comprehensive simulation study, Section 4 reports our empirical results.

---

2 A further complication is that for the models we consider, not all parameters are identified in the stationary distribution.
for US data, and finally the epilogue in Section 5 discusses the limitations of our analysis and where we might be heading from here.

2. Theory: Three Pieces

2.1. Continuous-time stochastic volatility models

We model the price of a given financial asset, \( S = (S(t))_{t \geq 0} \), by a stochastic differential equation

\[
dS(t) = \mu S(t)dt + \sqrt{V(t)}S(t)dZ(t),
\]

(2)

where \( Z \) is a Wiener process, \( \mu \) is the drift (rate) coefficient, and the (stochastic) instantaneous variance process \( V = (V(t))_{t \geq 0} \) is our main modelling target.

Heston (1993) suggests a model where \( V \) follows a square-root process (also known as a Feller or a Cox-Ingersoll-Ross process),

\[
dV(t) = \kappa (\theta - V(t))dt + \epsilon \sqrt{V(t)}dW(t)
\]

(3)

with mean reversion speed \( \kappa \), long-term mean \( \theta \), diffusion \( \epsilon \), and a Wiener process \( W \) correlated to \( Z \).

Applying the Ito formula to \( e^{\kappa t}V(t) \), known as using integrating factor, we get that

\[
X(t) = X(0)e^{-\kappa t} + \theta (1 - e^{-\kappa t}) + \sigma \int_0^t e^{\kappa(s-t)} \sqrt{X(s)}dW(s).
\]

(4)

From this we immediately conclude that

\[
E(V(t+h)|V(t)) = V(t)e^{-\kappa h} + \theta (1 - e^{-\kappa h}).
\]

(5)

By taking conditional variance on equation (4), using the Ito isometry, and equation (5) for the conditional mean, elementary integration shows that

\[
\text{Var}(V(t+h)|V(t)) = V(t)\frac{\epsilon^2}{\kappa} \left( e^{-\kappa h} - e^{-2\kappa h} \right) + \theta \frac{\epsilon^2}{2\kappa} \left( 1 - e^{-\kappa h} \right)^2.
\]

(6)

It can be shown\(^3\) that the conditional distribution of \( V \) in Heston’s model is a non-central chi-squared distribution with \( df = 4\kappa \theta / \epsilon^2 \) degrees of freedom, non-centrality parameter \( \nu(t) = V(t) \cdot 4\kappa e^{-\kappa h} / (\epsilon^2(1 - e^{-\kappa h})) \), and scaling \( c = 4\kappa / (\epsilon^2(1 - e^{-\kappa h})) \); or shortly written

\[
c \cdot V(t+h)|V(t) \sim \chi^2(df, \nu(t)).
\]

If the parameters satisfy the Feller condition \( 2\kappa \theta > \epsilon^2 \), then \( V \) has a stationary gamma distribution with shape parameter \( df/2 \) and scale parameter \( \epsilon^2 / (2\kappa) \).\(^4\) Arguably, our treatment of the Heston model in this paper does not fairly reflect the seminal nature of Heston’s work. He demonstrated that by working with characteristic functions (or in Fourier space), almost-closed-form expressions could be derived for option prices. This transform method that has since been shown to work in considerably more general so-called affine models, see Duffie et al. (2000).

---

\(^3\) A detailed derivation of this is in Cairns (2004, Appendix B.2).

\(^4\) Notice that stationary distribution depends on \( \kappa \) and \( \epsilon \) only through \( \epsilon^2 / \kappa \). This was what we meant, when we talked about identification problems in an earlier footnote.
Table 1. Impact of stochastic volatility research measured by no. citations in Google Scholar on February 15, 2018.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Source</th>
<th># citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affine, basic</td>
<td>Heston (1993)</td>
<td>7637</td>
</tr>
<tr>
<td>Affine, fancy</td>
<td>Duffie, Pan, and Singleton (2000)</td>
<td>3009</td>
</tr>
<tr>
<td>Log-normal, no κ</td>
<td>Hagan, Kumar, Lesniewski, and Woodward (2002)</td>
<td>974</td>
</tr>
<tr>
<td>Log-normal, high κ</td>
<td>Fouque, Papanicolaou, and Sircar (2000)</td>
<td>175</td>
</tr>
<tr>
<td>3-over-2</td>
<td>Carr and Sun (2007)</td>
<td>94</td>
</tr>
<tr>
<td>All of the above</td>
<td>Lewis (2000)</td>
<td>766</td>
</tr>
</tbody>
</table>

We get a log-normal model by letting $V(t) = \exp(X(t))$, where $X$ is a mean-reverting Gaussian process (also known as an Ornstein-Uhlenbeck or a Vasicek process), i.e.

$$dX(t) = \kappa(\theta - X(t))dt + \varepsilon dW(t)$$

with mean reversion speed $\kappa$, long-term mean $\theta$, diffusion parameter $\varepsilon$, and a Wiener process $W$ correlated with $Z$ with correlation coefficient $\rho$. Using an integrating factor again and the fact that something deterministic integrated with respect to a Wiener process is normally distributed, we conclude that $X$ is a Gaussian process (making $V$ log-normal) with conditional distribution

$$X(t+h)|X(t) \sim N\left(X(t)e^{-\kappa h} + \theta(1-e^{-\kappa h}), \frac{\varepsilon^2(1-e^{-2\kappa h})}{2\kappa}\right)$$

and a stationary distribution (let $h \to \infty$) that is normal with mean $\theta$ and variance $\varepsilon^2/(2\kappa)$. When modelling something inherently positive, the log-normal distribution is a natural choice. It is therefore not surprising that it was used in early stochastic volatility models; for a taxonomy see Table 1 in Poulsen et al. (2009). However, for option pricing, the quantity $\int_0^T V(u)du$ plays a central role and in log-normal models its distribution isn’t known in closed form, basically because a sum of log-normals isn’t log-normal. So no closed-form expression is known for option prices. This has diminished the academic appeal of log-normal models as revealed by Table 1, which shows the impact (measured by citation count) of different stochastic volatility papers. Option price approximations have been developed. The two most popular ones treat opposite ends of the parametric spectrum; Hagan et al. (2002) work with the so-called SABR model where volatility has no mean-reversion, while Fouque et al. (2000) develop approximations that work well if the mean-reversion parameter $\kappa$ is high (here, 10 is a high number).

The dynamics of the instantaneous variance in the 3-over-2 model is given by

$$dV(t) = V(t)\kappa(\theta - V(t))dt + \varepsilon V(t)^{3/2}dW(t), \quad (7)$$

where $\kappa$, $\theta$ and $\varepsilon$ are parameters. Notice that the speed of mean-reversion depends on $V$ itself; the drift function is quadratic, not affine. Applying the Ito formula to the reciprocal variance process $Y(t) = 1/V(t)$, we see that

$$dY(t) = \kappa \theta \left(\frac{\kappa + \varepsilon^2}{\kappa \theta} - Y(t)\right) - \varepsilon \sqrt{Y(t)}dW(t), \quad (8)$$

so $Y$ follows a square-root process with mean reversion speed $\bar{\kappa} = \kappa \theta$ and long-term mean $\bar{\theta} = (\kappa + \varepsilon^2)/(\kappa \theta)$. We will refer to the model specified by equation (8) as the reciprocal 3-over-2 model with reciprocal parameters ($\bar{\kappa}, \bar{\theta}, \varepsilon$). This implies that $V$ will have a stationary reciprocal gamma distribution.
with shape parameter \(5 \tilde{2}\tilde{k}/\varepsilon^2\) and rate parameter \(2\tilde{k}/\varepsilon^2\). The brilliant book Lewis (2000) treats option pricing in the 3-over-2 model, although that flew somewhat under the radar. The paper Carr and Sun (2007) renewed interest in the model. The 3-over-2 model is “wild”, which is good in some contexts, and much pricing can be done by transform methods, although in our experience almost everything requires very delicate numerical treatment.

### 2.2. Estimation of discretely observed diffusion processes

Suppose for now that we have observed the instantaneous variance process or some non-parameter-dependent transformation of it; generically \(X\). The observations are made at discrete time-points \(t_0, \ldots, t_N\) with spacings are \(h_i = t_{i+1} - t_i\), typically one day or a few minutes for high-frequency data. We thus have an observed path \(x = (x_0, x_1, \ldots, x_N)\).

For the log-normal model with parameters \(\psi = (\kappa, \theta, \epsilon)\), the role of \(X\) is played by \(\ln(V)\), and we can write the log-likelihood as

\[
I(\psi; x) = \sum_{i=0}^{N-1} \log \phi(x_{i+1}; m_{i+1|j}, \sigma^2_{i+1|j})
\]

where \(\phi\) denotes the normal density, here with (conditional) mean \(m_{i+1|j} = x_i e^{-\kappa h_i} + \theta (1 - e^{-\kappa h_i})\) and variance \(\sigma^2_{i+1|j} = \varepsilon^2 (1 - e^{-2\kappa h_i}) / (2\varepsilon)\). The maximum likelihood estimate \(\hat{\psi} = (\hat{\kappa}, \hat{\theta}, \hat{\epsilon})\) is the argument that maximizes the log-likelihood (9). It may be found in closed form if observations are equidistant (and easily by a numerical optimizer if they are not). Calculating the observed information matrix

\[
I_0 = -\frac{\partial^2I}{\partial \psi^T \partial \psi}|_{\psi=\hat{\psi}}
\]

by numerical differentiation at estimated values gives an estimated standard error of the \(j\)th parameter as \(\sqrt{(I_0^{-1})_{ij}}\) where \((A^{-1})_{ij}\) denotes element \(i, j\) of the inverse of a matrix \(A\).

For the Heston model, the role of \(X\) is played by \(V\) itself. Optimization on the non-central chi-squared distribution that enters the Heston model’s log-likelihood is not possible in closed form, and numerically it is delicate and slow. A simple way to obtain a consistent estimator of \(\hat{\psi} = (\hat{\kappa}, \hat{\theta}, \hat{\epsilon})\) is by plugging the conditional moments into a Gaussian likelihood and maximizing. In other words this means using the approximate log-likelihood

\[
I^a(\psi; v) = \sum_{i=0}^{N-1} \log \phi(v_{i+1}; m_{i+1|j}, \sigma^2_{i+1|j}),
\]

where the conditional mean and variance are given by equations (5)-(6). The reason this does indeed give consistent estimators is that the first order conditions for optimization of (10) are a set of so-called martingale estimating equations, see Sørensen (1999). We approximate standard errors of the estimated parameters from the observed information matrix of the approximated log-likelihood function by numerical differentiation.

In a similar fashion for the (reciprocal) 3-over-2 model, the approximate log-likelihood (10) with \(1/V\) in the role of \(X\), can be optimized to obtain maximum likelihood estimates of the reciprocal parameters \((\hat{\tilde{k}}, \hat{\theta}, \hat{\epsilon})\). These estimates may then be transformed to estimates of the original parameters through \(\kappa = \hat{\tilde{k}} \tilde{k} - \tilde{\varepsilon}^2\) and \(\theta = \tilde{k} / (\hat{\tilde{k}} - \tilde{\varepsilon}^2)\) while \(\epsilon = \hat{\epsilon}\). As well, standard errors of the reciprocal

---

5 The first moment of the inverse-gamma distribution is defined if the shape parameter is greater than one, a condition certified by the Feller condition. In this case \(\mathbb{E}[V(t)] = 2\varepsilon / (2\tilde{\epsilon} - \varepsilon^2) = 2\tilde{\varepsilon}^2 / (2\varepsilon + \varepsilon^2)\).

6 Older readers may remember that Chan et al. (1992), a very influential paper in the 90’s, estimated the volatility of the short-term interest rate, \(r\) to be of the form \(1.18^{1.48}\), i.e. almost exactly \(3/2\), but their model had an affine drift, which (7) doesn’t.
parameters may be calculated from the observed information matrix and standard errors of the original parameters by the delta method.

2.2.1. Goodness of fit and uniform residuals

A goodness-of-fit analysis that takes into consideration the full conditional structure of a diffusion process \(X\) (and not just its stationary distribution) uses the so-called uniform residuals, see Pedersen (1994). With \(F_{i+1}(x|\psi)\) denoting the conditional distribution function of \(X(t_{i+1})|X(t_i)\), the uniform residuals are defined by the

\[ U_{i+1} = F_{i+1}(X(t_{i+1})|\psi), \quad i = 0, \ldots, N - 1. \]

The uniform residuals \(U_1, \ldots, U_N\) will be independent and identically \(U(0,1)\) distributed if the true distribution of \(X\) is that of the family \(\{F_{i+1}, i = 0, \ldots, N - 1\}\). The proof of this result, that goes back at least to Rosenblatt (1952), is straightforward: If the random variable \(X\) has a strictly increasing, continuous distribution function \(F\), then \(F(X)\) is \(U(0,1)\). Now combine hat with the Markov property and use iterated expectations to get the independence.

For the Heston model (and the reciprocal 3-over-2 model) we use the true conditional non-central chi-squared distribution find uniform residuals – not the Gaussian density used for the parameter estimation.

2.3. Measuring volatility

Contrary to the simplifying assumption in the previous subsection, the instantaneous variance process \(V\) is not directly observable. In practice we must measure it from observations of the asset price process \(S\). One way to do that is by equation (1). Another way, and one that enables us to make quantitative statements about the accuracy of the measurement, comes by following Andersen and Benzoni (2009) in a study of so-called realized volatility. To this end denote the logarithmic asset price by \(Y(t) = \ln S(t)\) and the continuously compounded return over a measurement interval \([t - k, t]\) by \(r(t, k) = Y(t) - Y(t - k)\). Applying the Ito formula to equation (2) and putting \(\alpha(t) = \mu - \frac{V(t)}{2}\), the continuously compounded return over \([t - k, t]\) is

\[ r(t, k) = \int_{t-k}^{t} \alpha(u) du + \int_{t-k}^{t} \sqrt{V(u)} dZ(u) \]

and its quadratic variation is

\[ QV(t, k) = \int_{t-k}^{t} V(u) du. \]

For a partition of \([t - k, t]\) of the form \(\{t - k + \frac{j}{n}, j = 1, \ldots, n\}\), the realized variance defined by given by

\[ RV(t, k; n) = \sum_{j=1}^{n} r\left(t - k + \frac{j}{n}, \frac{1}{n}\right)^2. \]

General semimartingale theory ensures convergence in probability

\[ RV(t, k; n) \rightarrow QV(t, k) \text{ as } n \rightarrow \infty. \] (11)

So for high-frequency returns (a large \(n\)) \(RV\) gives a good approximation of \(QV\) and we may use \(\int_{t-k}^{t} V(u) du \approx V(I) \cdot k\) for some \(I \in [t - k, t]\) to obtain our measured variance path. Thus, to adopt our

\[ \text{The convergence is holds for any partition whose mesh size goes to 0.} \]
notation, for an asset log-price $Y$ observed at time points $t_0, \ldots, t_N$ and a measurement interval (say one day) $[t_{i-n}, t_i]$, i.e. observation frequency set by $n$ (say every 5 minutes), calculate
\begin{equation}
\nu_i = \frac{1}{t_i - t_{i-n}} \sum_{j=1}^{n} \left( y_{i-n+j} - y_{i-n+(j-1)} \right)^2 \tag{12}
\end{equation}
for $i = n, 2n, \ldots, N$ as the measured variance path $\nu = (\nu_n, \nu_{2n}, \ldots, \nu_N)$ of the non-observable variance $V(t_n), V(t_{2n}), \ldots$. Note that we obtain daily variance measurements while we require asset price observations of a much higher frequency. Of course, we may reduce $n$ to get more frequent variance measurements but for a cost of poorer approximation with respect to the convergence in (11).

We have described a simple-to-implement, model independent, and theoretically sound way to measure realized and instantaneous variance. But it is far from the only one. For instance, Reno (2008) describes the use of so-called $\delta$-sequences and kernels in volatility measurement and Andreasen (2016, equation (2), page 29) suggests an estimator based on absolute, not squared, returns. A different, global time-series approach is offered by the numerous variants of GARCH models, see Engle et al. (2012). We refrain from analyzing these techniques further. Partly in the interest of space, partly because specific estimation choices might stack the deck for or against one of our three models.

2.3.1. Realized volatility for an asset price with jumps

The asset price process in equation (2) has continuous paths; arguably a restrictive constraint on market data of asset prices.\(^8\) Andersen and Benzoni (2009) suggest adding a compound Poisson-type, finite activity jump part to the price dynamics,
\[
dS(t) = \mu S(t)dt + \sqrt{V(t)S(t)}dZ(t) + (e^{\tilde{\xi}(t)} - 1)S(t^-)dN(t).
\]
where $N = (N(t))_{t \geq 0}$ is a Poisson process uncorrelated to $Z$, while $\tilde{\xi} = (\tilde{\xi}(t))_{t \geq 0}$ determines the magnitude of a jump $\Delta S(t) = S(t) - S(t^-) = S(t^-)e^{\tilde{\xi}(t)}$ that occurs at time $t$. The Ito formula for diffusions with jumps applied to $Y(t) = \log S(t)$ gives
\begin{equation}
\begin{aligned}
dY(t) &= \alpha(t)dt + \sqrt{V(t)}dZ(t) + \tilde{\xi}(t)dN(t). \tag{13}
\end{aligned}
\end{equation}
The jump term $\tilde{\xi}(t)dN(t)$ is non-zero only if a jump occurs at time $t$ and it should be understood in integral form
\[
\int_0^t \tilde{\xi}(u)dN(u) = \sum_{T_i \in [0,t]} \tilde{\xi}(T_i)
\]
where $T_1, \ldots, T_{N(t)}$ are jump times of $N$ in $[0,t]$. The quadratic variaton of $Y$ in (13) includes both the integrated variance and a jump part
\[
QV(t,k) = \int_{t-k}^t V(u)du + \sum_{T_i \in [t-k,t]} \tilde{\xi}^2(T_i),
\]
while the convergence
\[
RV(t,k;n) \longrightarrow QV(t,k) \text{ as } n \rightarrow \infty.
\]

\(^8\) Although, as Christensen et al. (2014) points out, perhaps not as restrictive as believed in some areas in the literature.
still holds. As an estimator for the integrated variance part we employ the \textit{realized bipower variation} introduced by Barndorff-Nielsen and Shephard (2004)

\[ BV(t,k;n) = \frac{\pi}{2} \sum_{j=2}^{n} |r(t - k + \frac{j}{n}, \frac{1}{n}) || r(t - k + \frac{j-1}{n}, \frac{1}{n}) | \]

which provides a consistent estimate of the variance part that is robust to jumps. The ratio \((RV(t,k;n) - BV(t,k;n))/RV(t,k;n)\) gives us a measure of how the cumulative variance that is due to jumps.

3. Simulation: Model discrimination

To assess the statistical properties of our model discrimination test, we look at simulated asset price and variance data.

The first port of call when simulating solutions to stochastic differential equations is the Euler scheme, see Kloeden and Platen (1992). It provides a simple method for the generation of a time-discrete approximations to sample paths. However, we quickly run into problems with Euler scheme for the models we consider; there are negative variance outcomes. We require methods that better preserve distributional properties, positivity specifically. For the log-normal model the solution is straightforward; simulate the log of the process, which is easy to do in an exact manner. Alas, the log-trick does not work for the Heston model; apply the Ito formula to see why. Therefore we use the simulation scheme from Broadie and Kaya (2006) with a modification that approximates the integrated variance with a trapezoidal rule, see Platen and Bruti-Liberati (2010). This scheme is (almost) exact, but slow. Faster methods have been suggested, see Andersen et al. (2010); these generally preserve positivity hard, but not the exact distributional characteristics. In a similar fashion, Baldeaux (2012) provides an exact simulation methods for the 3-over-2 model.

3.1. S&P 500 data revisited

To return to the question of credibility the analysis in the introduction, we perform a mimicking simulation where we apply the exponentially weighted moving average measure from equation (1) to 4,532 daily prices from a simulated model of either Heston or log-normal type with parameters given by the estimates in Table 3. The result for measured variance of Heston’s model in left-hand graph in Figure 2 shows that fitting the unconditional distribution of the Heston simulated data leads us the the same conclusion as with the S&P 500 data: The data look log-normal.

To overcome the correlation problem and to fully exploit conditional information, we run discriminatory tests between the three models based on uniform residuals when variance is measured by equation (1). The results are reported in the right-most column in Table 2, where we report proportion (out of 1000 repetitions) a given model is accepted. In every single simulation, the true underlying model ia rejected – as are the wrong models. Variance measured by equation (1) is simply too noisy to capture the fine structure of continuous-time stochastic volatility models.

We now change to simulation at 5-minute frequency and use equation (12) with \(n = 192\) (i.e. two-day aggregation in the measurement of instantaneous variance) and run the uniform residuals goodness-of-fit test on this data. That gives the second-to–right column in Table 2. Here we achieve model discrimination: In all three cases (across models and simulations), the wrong models are rejected, and the true model is typically accepted; 57% of the cases for for Heston, 78% for log-normal and 90% 3-over-2.

Finally, we ran the goodness-of-fit the on the true simulated variances; in this way we investigate if there is a gain in going beyond a 5-minute sampling frequency. The results in the second-to-left column in Table 2 show that there isn’t.
Figure 2. Fitted empirical, log-normal and gamma densities to measured variance with $\lambda = 0.94$ of 4,532 simulated prices with Heston’s (left figure) and the log-normal model (right figure).

Table 2. Discriminatory power of Kolmogorov-Smirnov tests on uniform residuals for different methods of variance measurement. The numbers (each based on 1000 simulations) indicate how often uniformity of residuals is accepted when the uniformization is based on the model indicated in the test alternative column.

<table>
<thead>
<tr>
<th>True model: Heston</th>
<th>Variance measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test alternative</td>
<td>true simulated</td>
</tr>
<tr>
<td>Heston</td>
<td>0.79</td>
</tr>
<tr>
<td>log-normal</td>
<td>0</td>
</tr>
<tr>
<td>3-over-2</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>True model: Log-normal</th>
<th>Variance measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test alternative</td>
<td>true simulated</td>
</tr>
<tr>
<td>Heston</td>
<td>0</td>
</tr>
<tr>
<td>log-normal</td>
<td>0.72</td>
</tr>
<tr>
<td>3-over-2</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>True model: 3-over-2</th>
<th>Variance measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test alternative</td>
<td>true simulated</td>
</tr>
<tr>
<td>Heston</td>
<td>0</td>
</tr>
<tr>
<td>log-normal</td>
<td>0</td>
</tr>
<tr>
<td>3-over-2</td>
<td>0.26</td>
</tr>
</tbody>
</table>

4. Empirics: Horseraces between models

The empirical analysis is based on data from the S&P 500 index (or, to be precise, an exchange traded fund (ETF) that tracks the S&P 500 index) and ten randomly picked securities listed on the New York Stock Exchange (NYSE). The data is retrieved from the TAQ database via Wharton Research Data Services. The raw data consists of intraday tick-by-tick trade data. Raw trade data is, however, not directly suited for analysis and we apply a series of preprocessing methods – with the help of the R-package highfrequency. Specifically, we use the step-by-step cleaning procedure proposed by Barndorff-Nielsen et al. (2009). First, entries with zero prices are deleted, as are entries with an abnormal sale condition as indicated from the sales condition code of the trade. The data is then restricted to trades from the exchange where the security is listed. Entries with the same time stamp
Figure 3. The left figure shows the S&P 500 ETF price from 1996-01-02 to 2013-12-31 with a 5 minute frequency and a total of 368,724 observations. Right figure: variance of the S&P 500 ETF price measured by realized volatility with frequency $n = 192$ and a total of 1,920 observations.

Table 3. S&P 500: Estimated parameters and standard errors from (approximate) maximum likelihood estimation of the Heston, log-normal, and 3-over-2 models.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\kappa}$</th>
<th>$\hat{\theta}$</th>
<th>$\hat{\epsilon}$</th>
<th>K-S p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heston</td>
<td>21.7 (3.63)</td>
<td>0.022 (0.0049)</td>
<td>2.98 (0.08)</td>
<td>0.00</td>
</tr>
<tr>
<td>log-normal</td>
<td>56.1 (3.36)</td>
<td>-3.65 (0.049)</td>
<td>11.5 (0.24)</td>
<td>0.68</td>
</tr>
<tr>
<td>3-over-2</td>
<td>2.7 (5.64)</td>
<td>30.0 (62.7)</td>
<td>98.2 (2.0)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

are replaced by the median price in the last step. The cleaned data will be on an irregular tick-by-tick time scale. The aggregation method of the highfrequency package aggregates prices to an equidistant 5 minute time grid by taking the last realized price before each grid point. Tick aggregation to 5 minutes is motivated by the fact that price processes are increasingly perturbed by micro-structure noise for higher frequencies. Finally, the data is restricted to exchange trading hours 9:30 a.m. to 4:00 p.m. from Monday to Friday.

The realized volatility measure is applied with a resolution of $n = 192$ prices per variance measurement for the first step of the test methodology and a plot of the resulting variance path of 1,920 observations is shown in right-hand graph in Figure 3. The measured variance is used in the second step for the maximum (approximate) likelihood estimation where a numerical optimization is performed on the likelihood function from each model. The resulting estimated parameters and standard errors are shown in Table 3. Note that even though each of parameters has the same qualitative interpretation across models ($\theta \sim$ level, $\kappa \sim$ speed of mean-reversion, $\epsilon$ volatility of volatility), their numerical values cannot be compared directly.

First sanity checks come from quick calculations of the typical levels of instantaneous volatility that the estimated parameters cause the models to produce. For the Heston model it is $\sqrt{0.022} = 0.148$ and for the log-normal model it is $e^{-3.65/2} = 0.161$; in line with what we would anticipate. However, we have to be careful here as non-linear functions abound and parameters are in a range where Jensen’s inequality matters; for instance in the log-normal model, the stationary distribution for $\sqrt{V} = e^{X/2}$ has mean $e^{(-3.65+11.5^2/2-56.1)/2} = 0.291$. The 3-over-2 parameter estimates look strange at first sight, but we can think of them this way: From equation (8) then the reciprocal parameters are $(\hat{\kappa}, \hat{\theta}, \hat{\epsilon}) = (81, 119, 98.2)$. From the right-hand panel of Figure 3 we see that 119 is a typical value for reciprocal variance, and that this process covers a large range of values leading to combined high volatility and speed of mean-reversion. The morale is that a non-linear and ill-fitting (see the right-most panel in Figure 4) model is hard to interpret.
To get a quantitative feeling for the speed of mean-reversion parameters, we use that for a process with linear mean-reversion, the time it takes to halve the deviation in expectation between current level and long-term level is

\[ X(t)e^{-\kappa t/2} + \theta(1 - e^{-\kappa t/2}) = (\theta + X(t))/2 \Rightarrow t_{1/2} = \ln(2)/\kappa. \]

The estimates imply that for Heston model the half-life of deviations is about 8 days, for the log-normal model the half-life (for log-variance) is just over 3 days. This is strong mean-reversion. Or differently put, mean-reversion is in effect short time-scales; in interest rate models typical \( \kappa \)-values are 0.1-0.25 which implies deviance half-life of 3-7 years. Such strong mean-reversion plays havoc with our usual “quantitative intuitive interpretation”, where it determines the standard deviation of the change (in absolute or relative terms) in the process over one year, i.e. a change of ± the volatility is “nothing out of the ordinary”. That interpretation is non-sensical here; if applied it would tell us that the instantaneous variance in Heston is typically in the range 0.022 ± 0.44, and that volatilities of 5000% and 0.05% are imminently plausible in the log-normal model. It is the combination of high speed of mean-reversion and high volatility of volatility that is needed to generated the decidedly spiky behaviour of volatility which is what our likelihood approach aims a capturing. For the Heston model we have \( 2\kappa\hat{\theta} = 0.95 < \ell^2 = 8.9 \), so the Feller condition is off by almost a factor 10. But what we think is the most striking results in \( 3 \) is the last column, which contains goodness-of-fit tests for the different models; \( p \)-values for Kolmogorov-Smirnov on the uniform residuals. The log-normal model provides a (surprisingly) good fit \( (p = 0.68) \), Heston and 3-over-2 do not \( (p < 10^{-12}) \); the quantile plots in Figure 4 is visual evidence.

Comparing our findings to the literature, our most clear ally in the recent empirical literature is Christoffersen et al. (2010) who also find strong support for log-normal(‘ish) volatility models compared to Heston and 3-over-2. They only use daily data, so their parameter values (see Table 1 in the paper) do not pick up the fast mean-reverting feature. In that respect, our results are are most in line with Fouque et al. (2000) who use a log-normal model (without explicit discriminatory goodness-of-fit test) and document strong mean-reversion (even stronger than ours; \( \kappa \sim 200 – 250 \) – although other work by the same authors suggests \( \kappa = 50 \)). Similarly to us Andersen et al. (2001)[Figures 1-3 for instance] report log-normality of variance measured on 5-minute sampling, but they do no find fast mean-reversion, but rather than volatility has long-term memory.

**Figure 4.** Quantile plots of the uniform \([0, 1]\) distribution from uniform residuals calculated with measured S&P 500 variance.
### Table 4. The number of asset price and measured variance observations for the stock data.

<table>
<thead>
<tr>
<th></th>
<th>IBM</th>
<th>MCD</th>
<th>CAT</th>
<th>MMM</th>
<th>MCO</th>
<th>XOM</th>
<th>AZN</th>
<th>GS</th>
<th>HPQ</th>
<th>FDX</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td>246,308</td>
<td>328,111</td>
<td>328,107</td>
<td>328,108</td>
<td>271,580</td>
<td>288,846</td>
<td>302,350</td>
<td>300,759</td>
<td>239,336</td>
<td>328,104</td>
</tr>
<tr>
<td>variance</td>
<td>1,282</td>
<td>1,708</td>
<td>1,708</td>
<td>1,708</td>
<td>1,414</td>
<td>1,504</td>
<td>1,504</td>
<td>1,574</td>
<td>1,574</td>
<td>1,246</td>
</tr>
</tbody>
</table>

### Table 5. Maximum likelihood estimated parameters and standard errors from the realized volatility measured variance with \( n = 192 \).

<table>
<thead>
<tr>
<th></th>
<th>IBM</th>
<th>MCD</th>
<th>CAT</th>
<th>MMM</th>
<th>MCO</th>
<th>XOM</th>
<th>AZN</th>
<th>GS</th>
<th>HPQ</th>
<th>FDX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heston</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \kappa )</td>
<td>75.1 (7.2)</td>
<td>18.2 (4.0)</td>
<td>71.6 (5.3)</td>
<td>37.0 (4.52)</td>
<td>100.0 (6.85)</td>
<td>45.4 (4.35)</td>
<td>113.0 (7.92)</td>
<td>41.6 (5.14)</td>
<td>116.9 (12.1)</td>
<td>65.2 (6.79)</td>
</tr>
<tr>
<td>( \theta )</td>
<td>0.024 (0.004)</td>
<td>0.052 (0.012)</td>
<td>0.112 (0.009)</td>
<td>0.061 (0.006)</td>
<td>0.096 (0.009)</td>
<td>0.052 (0.005)</td>
<td>0.069 (0.005)</td>
<td>0.027 (0.009)</td>
<td>0.14 (0.012)</td>
<td>0.082 (0.008)</td>
</tr>
<tr>
<td>( \epsilon )</td>
<td>6.89 (0.4)</td>
<td>4.14 (0.1)</td>
<td>7.21 (0.2)</td>
<td>3.89 (0.1)</td>
<td>10.1 (0.4)</td>
<td>3.61 (0.1)</td>
<td>7.76 (0.3)</td>
<td>7.97 (0.3)</td>
<td>12.6 (0.6)</td>
<td>6.62 (0.2)</td>
</tr>
<tr>
<td>K/S p-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>IBM</th>
<th>MCD</th>
<th>CAT</th>
<th>MMM</th>
<th>MCO</th>
<th>XOM</th>
<th>AZN</th>
<th>GS</th>
<th>HPQ</th>
<th>FDX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \kappa )</td>
<td>86.6 (6.0)</td>
<td>61.1 (3.8)</td>
<td>83.9 (5.1)</td>
<td>76.9 (4.7)</td>
<td>71.6 (4.8)</td>
<td>72.3 (4.7)</td>
<td>126.2 (8.6)</td>
<td>58.3 (3.8)</td>
<td>102.2 (7.4)</td>
<td>88.0 (5.4)</td>
</tr>
<tr>
<td>( \theta )</td>
<td>-3.31 (0.05)</td>
<td>-2.96 (0.05)</td>
<td>-2.45 (0.04)</td>
<td>-3.08 (0.04)</td>
<td>-2.60 (0.05)</td>
<td>-3.10 (0.04)</td>
<td>-2.97 (0.04)</td>
<td>-2.49 (0.05)</td>
<td>-2.54 (0.05)</td>
<td>-2.63 (0.04)</td>
</tr>
<tr>
<td>( \epsilon )</td>
<td>13.5 (0.4)</td>
<td>11.4 (0.3)</td>
<td>12.7 (0.3)</td>
<td>12.1 (0.3)</td>
<td>43.8 (0.3)</td>
<td>11.3 (0.3)</td>
<td>16.6 (0.5)</td>
<td>11.6 (0.3)</td>
<td>14.9 (0.5)</td>
<td>12.8 (0.3)</td>
</tr>
<tr>
<td>K/S p-value</td>
<td>0.21</td>
<td>0.53</td>
<td>0.83</td>
<td>0.79</td>
<td>0.12</td>
<td>0.07</td>
<td>0.56</td>
<td>0.12</td>
<td>0.05</td>
<td>0.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>IBM</th>
<th>MCD</th>
<th>CAT</th>
<th>MMM</th>
<th>MCO</th>
<th>XOM</th>
<th>AZN</th>
<th>GS</th>
<th>HPQ</th>
<th>FDX</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-over-2(i)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \kappa )</td>
<td>105.0 (10.0)</td>
<td>27.3 (4.2)</td>
<td>77.7 (6.5)</td>
<td>77.6 (7.5)</td>
<td>48.8 (5.7)</td>
<td>60.9 (5.8)</td>
<td>152.0 (12.5)</td>
<td>45.6 (6.0)</td>
<td>105.8 (9.7)</td>
<td>91.6 (7.5)</td>
</tr>
<tr>
<td>( \theta )</td>
<td>40.0 (1.7)</td>
<td>27.2 (3.0)</td>
<td>16.3 (0.8)</td>
<td>32.0 (1.5)</td>
<td>19.2 (1.5)</td>
<td>31.1 (1.6)</td>
<td>30.6 (1.2)</td>
<td>15.7 (1.3)</td>
<td>18.7 (0.9)</td>
<td>19.4 (0.8)</td>
</tr>
<tr>
<td>( \epsilon )</td>
<td>95.2 (3.4)</td>
<td>67.8 (1.7)</td>
<td>60.7 (1.7)</td>
<td>82.8 (2.5)</td>
<td>61.9 (1.8)</td>
<td>67.2 (1.9)</td>
<td>122.5 (4.5)</td>
<td>56.6 (1.6)</td>
<td>71.9 (2.6)</td>
<td>67.9 (2.0)</td>
</tr>
<tr>
<td>K/S p-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\kappa} )</th>
<th>( \hat{\theta} )</th>
<th>( \hat{\xi} )</th>
<th>K-S p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heston</td>
<td>21.4 (3.8)</td>
<td>0.011 (0.004)</td>
<td>2.78 (0.08)</td>
<td>0.00</td>
</tr>
<tr>
<td>log-normal</td>
<td>55.3 (3.3)</td>
<td>-3.9 (0.05)</td>
<td>11.5 (0.2)</td>
<td>0.17</td>
</tr>
<tr>
<td>3-over-2</td>
<td>25.4 (4.0)</td>
<td>17.2 (8.8)</td>
<td>142.4 (4.3)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

4.1. Individual stocks

The Devil’s advocate might say: “The log-normality of the index is an aggregation effect; some stocks are Heston, some are 3-over-2, and when you mix them all up, it comes out as log-normal.” But it isn’t, as Table 5 shows. Heston and the 3-over-2 models are rejected with all \( p \)-values indistinguishable from 0, the log-normal model is accepted for all 10 stocks (although 2 are borderline), mean-reversion is strong (all \( \kappa \)'s > 50), and volatility of volatility is correspondingly high in order generate spikes.

4.2. Jump corrections

The Devil’s advocate might continue: “When you use equation (12) you pick up both jumps and diffusive volatility. It hink it’s the jumps that cause log-normality, for the diffusive part Heston will do.” (Conveniently, options can the be priced using transform methods from Duffie et al. (2000).) Table 6 shows that log-normality of variance prevails when it is measured through bipower variation.

5. Conclusion and future work

In this paper we have demonstrated that log-normal models provide a significantly better description of the empirical behaviour of instantaneous variance than the Heston and the 3-over-2 models do. However, at coarser time-resolutions it can be hard to discriminate the models. Our intention has been to give a reasonably self-contained treatment of the theory and the used models and methods than facilitate replication our our results.\(^9\)

We find that to replicate the spiky but stationary behaviour of volatility, we need a combination of strong mean-reversion and high volatility of volatility. From a modelling perspective this suggests that we look beyond diffusive model, towards models with higher degree of path roughness (as measured by degree of Hölder continuity) of volatility. A promising line of research are the rough volatility models suggested in Gatheral et al. (2014) (see also Andreasen (2017)) where volatility is driven fractional Brownian motion.

\(^9\) We are not allowed to make the high-frequency data public.


Gatheral, Jim, Thibault Jaisson, and Mathieu Rosenbaum. 2014. Volatility is rough. Workning paper available at SSRN.


© 2018 by the authors. Submitted to Risks for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).