An analytic approximation of the likelihood function for the Heston model volatility estimation problem

2 authors, including:

Amir Atiya
Cairo University
206 PUBLICATIONS 5,141 CITATIONS

Some of the authors of this publication are also working on these related projects:

Applied Intelligence View project
Sparse Coding Prediction View project
An analytic approximation of the likelihood function for the Heston model volatility estimation problem

AMIR F. ATIYA*† and STEVE WALL‡§

†Department of Computer Engineering, Cairo University, Giza, Egypt
‡Long View Research Associates, Greenwich, CT, USA

(Received 25 August 2006; in final form 8 January 2008)

Estimating the volatility from the underlying asset price history for the discrete observations case is a challenging inference problem. Yet it has attracted much research interest due to the key role of volatility in many areas of finance. In this paper we consider the Heston stochastic volatility model and propose an accurate analytic approximation for the volatility likelihood function. The model is based on considering the joint probability density of the asset and the volatility, and integrating out past volatility variables. The likelihood simplifies to a product of $T$ terms, where $T$ is the length of the past history considered. An extension to the problem of fixed parameter estimation is also presented. Simulation results indicate the effectiveness and accuracy of the proposed method.

Keywords: Volatility; Volatility estimation; Heston model; Stochastic volatility; Particle filter

1. Introduction

Volatility plays a central role in many areas of finance, such as derivatives pricing, risk management and portfolio optimization. This has led to significant research effort in the area of volatility modelling and estimation. There have been two competing views for the volatility modelling problem. The somewhat earlier approach, the autoregressive conditional heteroscedastic model (ARCH), developed by Engle (1982) and Bollerslev (1986), is an econometric approach that models the current volatility as a function of past asset observations. The second approach is the stochastic volatility (SV) model (see Heston 1993, Taylor 1994, Hull and White 1987). In this approach, the volatility is modelled as a latent state in a stochastic process that describes the evolution of the volatility and the asset processes. The stochastic volatility approach has become more popular recently, owing to its ability to incorporate some of the features empirically observed in the financial markets, such as mean reversion, volatility clustering, and the leverage effect. The two major stochastic volatility models are the Heston model (1993; also called the square-root volatility model) and the lognormal stochastic autoregressive volatility model (see Jacquier et al. 1994).

In spite of the flexibility and power of SV models, they pose significant computational challenges. The volatility and the model parameters are unobserved variables and they have to be estimated from the data. This is a challenging inference problem for the discrete observations formulation. If the process were to be observed continuously, then volatility estimation would be a simpler task. The reason for the difficulty of the problem for the discrete domain is that the past volatilities, being unknown, have to be integrated out in the formula of the likelihood function. This leads to a prohibitive multi-integral formula. However, there have been many advances in developing computational methods for that problem. The early literature utilized mainly methods of moment matching. Examples are the method of moments (MM) (Taylor 1986), the generalized method of moments (GMM) (Gourieroux et al. 1993), and the efficient method of moments (EMM) (Andersen et al. 1999). Recently, work has focused mainly on particle-filtering-type approaches. In this approach a number of 'particles', describing the posterior probability density of the latent state (e.g. the volatility), are propagated forward in time accounting for the new
observations as they arrive (see the monograph by Doucet et al. 2001 and the review by Arulampalam et al. 2002). Particle filtering approaches have been applied to the SV problem (see for example Jacquier et al. 1994, Kim et al. 1998, Sandmann and Koopman 1998, Javaheri et al. 2003, Stroud et al. 2004, and Johannes et al. 2006a). Recently a monograph has appeared that is devoted to the SV estimation problem with emphasis on particle filtering approaches (Javaheri 2005).

An alternative approach for the SV problem is to consider the integral formulae that describe the inference problem, and develop efficient numerical procedures for computing them. Essentially, the likelihood function can be obtained by integrating out the latent volatility process from the joint density of the observed and the latent process. This can be accomplished by a sequence of one-dimensional integrations. (They are one-dimensional provided there is a single latent state variable, which is the case for SV models.) Kitagawa (1987) has been the first to propose such an approach in the context of general filtering theory, and it has been introduced to finance by Watanabe (1999). Fridman and Harris (1998) used the more efficient Gauss–Legendre quadrature to evaluate the integrals. Shimada and Tsukuda (2005) proposed making use of Laplace’s approximation for evaluating the integrals. Essentially, this amounts to approximating the posterior at any time step as normal. It is unclear, however, how this approximation would fare, as it has been documented that the volatility posterior is far from normal.

In this paper, we consider the latter ‘integration-based’ approach, and develop a latent state and fixed parameter estimation method for the Heston SV model. (The proposed approach applies specifically to the Heston model, but not to the other competing lognormal stochastic autoregressive volatility model, or to multivariate extensions.) Using an approximation that is valid when the time step is short (e.g. in case of daily observations), we obtain a semi-analytic approximation for the volatility likelihood function. The formula is roughly in the form of a product of $T$ terms, where $T$ is the number of time steps considered. To obtain the full posterior for all volatility values (say $K$ values), this formula has to be evaluated $K$ times leading to a computational complexity of $O(KT)$. In comparison, the approaches described above (Kitagawa and derivatives thereof) lead to a complexity of $O(K^2T)$. Concerning the fixed parameters, the corresponding likelihood is evaluated with almost no extra computation. However, to obtain maximum likely estimates of these, the expression has to be evaluated several times using an optimization algorithm.

A recently appearing paper by Ait-Sahalia and Kimmel (2007) presents an analytic approach that describes the likelihood function in terms of a partial differential equation. Ait-Sahalia and Kimmel’s approach is general in that it takes into account any volatility/asset processes and also incorporates options observables as well as asset price observables. On the other hand, our approach follows a different methodology, and leads to a simpler and more direct formula. This is partly due to its applicability to the Heston model only.

The paper is organized as follows. The next section describes the statement of the volatility estimation problem and the details of the proposed method. Section 3 presents simulation results for testing the proposed approach. The last section presents the conclusion of this work.

2. The proposed method

2.1. Statement of the problem

Consider the Heston stochastic volatility model, given by

$$\begin{align*}
    ds_i &= \mu s_i dt + \sqrt{\nu_i} s_i dw_i, \\
    d\nu_i &= (\alpha - \beta \nu_i) dt + \sigma \sqrt{\nu_i} dz_i.
\end{align*}$$

The first equation expresses the evolution of the asset price process, while the second one expresses the evolution of the volatility. We use here for $\nu_i$ the term volatility rather than the term variance, as some authors do. Here the term $\mu$ represents the drift parameter. For the volatility process, the term $\beta$ is the speed of reverting to the mean, the quantity $\alpha/\beta$ represents the long-run mean, and $\sigma$ is a parameter expressing the variability of the volatility. The terms $dw_i$ and $dz_i$ are the standard Wiener increments. They could possibly be correlated, i.e.

$$E(dw_i, dz_i) = \rho dt,$$

with $\rho$ being the correlation coefficient. Note that it is well known that the SV model does not yield to a riskless hedge (see Heston et al. 2007).

It is more convenient to perform a log transformation for the asset. Using Ito’s lemma, the log-asset process $y_i$ is given by

$$dy_i = \left(\mu - \frac{1}{2} \nu_i \right) dt + \sqrt{\nu_i} dw_i.$$

The volatility estimation problem is essentially the problem of inferring the current volatility from current and past asset prices. The conventional way is to evaluate the likelihood function, and from the likelihood function the maximum likely estimate of the volatility can be obtained. In addition to the volatility, the fixed parameters $\mu$, $\alpha$, $\beta$, $\sigma$ and $\rho$ have to be estimated. Also, the maximum likelihood framework is usually employed for this problem. The former problem – the estimation of the volatility – is often called the ‘filtering problem’, while the latter problem is simply called the ‘parameter estimation’ problem. As it turns out, the parameter estimation problem is the harder of the two because of the higher dimensionality of the parameter vector. The fact that the parameters are constant with time is not helping much in simplifying the problem (or at least there are not many algorithms in the literature that make use of this potentially simplifying fact).

2.2. The proposed method

Let the size of the time step be $\Delta$. To simplify notation, let us redefine $y_i$ and $\nu_i$ to denote the $i$-th value of the
sampled log-asset and volatility process, respectively. The transition probability density is given by

\[
p(y_{t+1}, v_{t+1} | y_t, v_t) = \mathcal{N}(\mu_{t+1}, \Sigma_{t+1}),
\]

(5)

where \(\mathcal{N}\) denotes the normal density of mean \(\mu_{t+1}\) and covariance matrix \(\Sigma_{t+1}\). The mean and covariance matrix are given by

\[
\mu_{t+1} = \left( y_t + \left[ \mu - \frac{v_t}{2} \right] \Delta \right)
\]

\[
v_{t+1} = \left[ \alpha - \beta v_t \right] \Delta
\]

(6)

\[
\Sigma_{t+1} = v_t \Delta \left( \frac{\rho \sigma \sigma^2}{\sigma^2} \right).
\]

(7)

Let \(y_{1:T}\) and \(v_{0:T}\) denote, respectively, the past asset price observables from time 0 to time \(T\) and the past volatilities from time 0 to time \(T\). The filtering problem can be formulated as that of obtaining the likelihood of the volatility given the past observations

\[
L_T(v_T) = p(v_T | y_{1:T})
\]

(8)

\[
\propto p(v_T, y_{1:T} | v_0)
\]

(9)

\[
= \int p(y_{1:T}, v_{0:T} | v_0) \, dv_{0:T-1},
\]

(10)

where \(L_T\) denotes the likelihood function and the proportionality symbol in (9) means that the likelihood function equals the RHS (which has the variable of interest, \(v_T\)) times a positive factor that is not a function of the volatility \(v_T\), and hence will not influence the position of the maximum of \(v_T\). We note that the likelihood here is only approximate. The reason is that, as we move from continuous time to discrete time, this entails some approximation in the expressions (5), (6) and (7), because the discrete-time formulation assumes constant process values between each sampling time and the next. However, since all subsequent work is based on assuming that the time step is short (\(\Delta\) is small), the approximation error will be small.

The maximum likely estimate of the volatility is then simply the maximum of the function \(p(v_T, y_{1:T} | v_0)\). Using the Markov property, we can evaluate the previous integral as follows:

\[
L_T(v_T) \propto \int_{v_{T-1}} \cdots \int_{v_0} \prod_{t=1}^T \left[p(y_t, v_t | y_{t-1}, v_{t-1})\right] \times p(v_0) \, dv_0 \cdots dv_{T-1},
\]

(11)

where the term \(p(v_0)\) represents the \textit{a priori} probability density for the volatility at time \(t = 0\). We have assumed that \(v_0\) does not depend on \(y_0\), which seems a reasonable assumption (even if it is not exactly true it is usually too remote from the current time \(T\) to have a noticeable effect). Note that the previous equation is somewhat of a deviation from the formulation typically used in the filtering literature. This is due to our desire to accommodate directly the possible dependence between the random components \(d_{w_t}\) and \(d_{z_t}\). From the previous equation, one can see that the likelihood at time \(t + 1\) can be obtained recursively in terms of that at time \(t\). The formulation becomes

\[
L_{t+1}(v_{t+1}) \propto \int_{v_t} p(y_{t+1}, v_{t+1} | y_t, v_t) L_t(v_t) \, dv_t,
\]

(12)

with the starting value \(L_0(v_0) = p(v_0)\). Substituting for \(p(y_{t+1}, v_{t+1} | y_t, v_t)\) using the normal density formula in (5), (6) and (7), we get

\[
L_{t+1}(v_{t+1}) \propto \int_0^\infty \left[ \frac{e^{-a v_t (b_t / v_t)}}{v_t} \right] \, L_t(v_t) \, dv_t,
\]

(13)

where

\[
a = \frac{\beta^2 + \rho \sigma \beta \Delta + \frac{\sigma^2 \Delta^2}{2}}{2 \sigma^2 (1 - \rho^2) \Delta}
\]

(14)

\[
\beta' = 1 - \beta \Delta \quad \text{and} \quad \Delta v_{t+1} = v_{t+1} - v_t.
\]

(15)

The function in square brackets in (13) achieves a maximum at the point

\[
v_{\text{peak}} = -1 + \sqrt{1 + 4 a b_1}
\]

(16)

\[
L_{t+1}(v_{t+1}) \propto d_t \int_0^\infty \left[ \frac{e^{-a v_t (b_t / v_t)}}{v_t} \right] \, L_t(v_{\text{peak}})
\]

(17)

\[
+ L_t'(v_{\text{peak}}) (v_t - v_{\text{peak}})
\]

(18)

\[
+ \frac{1}{2} L_t''(v_{\text{peak}}) (v_t - v_{\text{peak}})^2 + \cdots \] dv_t.

(19)

and most support of that function lies sharply in a very small interval around the peak, due to the existence of \(\Delta\) (which is very small, for example, in the case of daily observations) in the denominator of \(a\) and \(b_1\) and hence in the denominator of the exponent. Expanding the likelihood function \(L_t(v_t)\) using a Taylor series\(^\dagger\) around \(v_{\text{peak}}\), we obtain

\[
L_{t+1}(v_{t+1}) \propto d_t \int_0^\infty \left[ \frac{e^{-a v_t (b_t / v_t)}}{v_t} \right] \, L_t(v_{\text{peak}})
\]

\[
+ L_t'(v_{\text{peak}}) (v_t - v_{\text{peak}})
\]

\[
+ \frac{1}{2} L_t''(v_{\text{peak}}) (v_t - v_{\text{peak}})^2 + \cdots \] dv_t.

(20)

\(^\dagger\)Note that the domain of convergence of the series cannot be easily specified because the functional form of \(L_t(v_t)\) is not explicitly available.
Let us consider the general $j$th term in the Taylor series, and let us denote it by $H_j$, i.e.

$$
H_j = \frac{1}{j!} L_j^{(0)}(v_{\text{peak}}) \int_0^{\infty} (v_t - v_{\text{peak}})^j \left[ e^{-av_t - (b_t/a)} \right] dv_t. 
$$

This term is first expanded as

$$
H_j = \frac{1}{j!} L_j^{(0)}(v_{\text{peak}}) \sum_{i=0}^{j} \binom{j}{i} (-v_{\text{peak}})^{j-i} \int_0^{\infty} v_t^{-1} e^{-av_t - (b_t/a)} dv_t.
$$

The integral on the right-hand side can be obtained as (see Gradshteyn and Ryzhik 1998)

$$
\int_0^{\infty} v_t^{-1} e^{-av_t - (b_t/a)} dv_t = 2 \left( \frac{b_t}{a} \right)^{i/2} K_{i/2}(2\sqrt{ab_t}),
$$

where $K_{i/2}(z)$ denotes the modified Bessel function of the second kind. For large $z$, the following series is a good approximation (Gradshteyn and Ryzhik 1998):

$$
K_{i/2}(z) = \frac{\pi}{\sqrt{2z}} \sum_{k=0}^{\infty} \frac{1}{(2\pi)^k k!} (v + k + \frac{1}{2}) (v - k + \frac{1}{2}) + O(z^{-n}).
$$

In our case $z = 2\sqrt{ab_t}$ is typically very large because of the existence of $\Delta$ in the denominator of each of $a$ and $b_t$ (for daily price observations, $\Delta = 1/250$ and hence it is small). For typical examples such as examples 1 and 2 in the simulations section, we find $z^{-1} \approx 10^{-4}$ to around $10^{-3}$. For these reasons, we take only the first term in the series (25). This series truncation gave an error only of the order of around 0.01–0.1%. Taking more than one term would not increase the complexity of the resulting formula by much, but we just did not find that necessary.

The formula in (23) becomes

$$
H_j \approx \frac{1}{j!} L_j^{(0)}(v_{\text{peak}}) \sqrt{\pi \frac{ab_t}{b_t}} e^{-2\sqrt{ab_t} \sum_{i=0}^{j} \binom{j}{i} \left( \frac{b_t}{a} \right)^{i/2} \left( -v_{\text{peak}} \right)^{j-i}}.
$$

Substituting in (21), and recognizing that the resulting expression is the Taylor series, evaluated at $v_t = \sqrt{b_t/a}$, we obtain the final expression for the likelihood:

$$
L_{t+1}(v_{t+1}) \propto d_t(ab_t)^{-1/4} e^{-2\sqrt{ab_t} L_t \left( \frac{b_t}{a} \right)}.
$$

Using the derived formula (28), one could design a way to obtain the likelihood $L_{t+1}$ at any value $v_{t+1}$ as follows. We first evaluate $v_t = \sqrt{b_t/a}$, which depends on $v_{t+1}$ (since $b_t$ is a function of $v_{t+1}$). Then, we evaluate the likelihood for that particular value, i.e. $L_t(\sqrt{b_t/a})$. This in turn depends on the likelihood $L_{t-1}(\sqrt{b_{t-1}/a})$, which has to be obtained first. We continue going this way until time 0. This solution, however, is not of a sequential nature and is therefore not suitable. For every time step we have to go back all the way until $t = 0$ to evaluate the likelihood.

We propose here an adaptation that is of a sequential nature. Assume that we have evaluated the likelihood $L_t$ for all values $v_t$. When we go forward one time step to $t + 1$, we use this likelihood curve $L_t$ to obtain the new one $L_{t+1}$. To design the algorithm this way we first start at time $t = 0$ with $v_0$ chosen as a uniform grid, and evaluating $L_0(v_0) = p(v_0)$. Then, we move to time $t = 1$. The values of the grid points $v_1$ are selected such that the corresponding points $v_0$ of the previous grid equal $\sqrt{b_0/a}$ (remember that $b_t$ is a function of $v_t$). Then we move to time $t = 2$ and choose the $v_2$ grid such that the corresponding $v_1$ points equal $\sqrt{b_1/a}$, and continue in this manner.

To evaluate the $v_{t+1}$ grid point from the corresponding $v_t$ grid point, we note that

$$
v_{t+1} = \sqrt{b_t/a},
$$

where $b_t$ is a function of $v_{t+1}$ as given in (15). All that we need is to invert (29). That leads to the following relation:

$$
v_{t+1} = \sqrt{B^2 - C - B},
$$

where

$$
\begin{align*}
B &= -\alpha \Delta - \rho \sigma (\Delta v_{t+1} - \mu \Delta) \\
C &= \alpha^2 \Delta^2 + 2 \rho \sigma \Delta (\Delta v_{t+1} - \mu \Delta) \\
&\quad + \sigma^2 (\Delta v_{t+1} - \mu \Delta)^2 - 2 \sqrt{\sigma^2 \Delta^2 - (1 - \rho^2) \Delta}.
\end{align*}
$$

We note that the $v_{t+1}$ grid is very close to the $v_t$ grid. The difference is only second order, owing to the fact that to a first-order approximation (by observing formulae 14 and 15):

$$
\sqrt{b_t/a} \approx v_{t+1}.
$$

However, it is recommended to use the exact solution in (30), because it is these second-order differences that, when accumulating over time, make a difference.

The following is a summary of the proposed algorithm:

- Start at time 0 with a uniform grid for the volatility $v_0$, and with $L(v_0) = p(v_0)$.
- For $t = 0$ to $T - 1$,
  - obtain grid points for $v_{t+1}$ according to equation (30);
  - compute the likelihood $L_{t+1}$ for the grid points $v_{t+1}$ according to the equation

$$
L_{t+1}(v_{t+1}) \propto d_t(ab_t)^{-1/4} e^{-2\sqrt{ab_t} L_t(v_t)}.
$$

A very important point to note is that when evaluating (28), it is imperative to combine the exponent in $d_t$ (equation 16) and the exponent $-2\sqrt{ab_t}$ in (28) before exponentiating, in order to avoid numerical errors or possibly overflow. The reason is that both exponents are large and almost equal in magnitude but with opposite signs. Please also note that the sources of the error in the developed formula are threefold. The first one is the space
discretization of $v_{t+1}$. The third one is in the error that exists in the previous time instant’s likelihood estimate, $L_t(y_t)$ in the RHS of equation (34). An empirical examination of the overall error will be presented in the simulation results section.

2.3. The parameter estimation issue

Let us arrange the fixed parameters in a vector: 
\[ \theta = (\mu, \alpha, \beta, \sigma, \rho)^T. \]

The maximum likelihood estimate of the parameters is the value of $\theta$ that maximizes

\[ L_T^* = p(y_1:T|y_0, \theta) \]

\[ = \int_{v_T} p(y_T, v_T|y_0, \theta) dv_T. \]  

But, the expression $p(v_T, y_1:T|y_0, \theta)$ has already been obtained when deriving the volatility likelihood, see (9). In fact the integration of this expression should also have been evaluated when normalizing the likelihood so that it integrates to 1. So essentially the likelihood of the parameters can be obtained by applying the same procedure detailed for the volatility estimation problem. The likelihood should then be optimized using some optimization algorithm, in order to obtain the most likely parameters. In the simulation experiments we used the Adaptive Simulated Annealing (ASA) optimization procedure (Ingber 1989). For small-to-medium scale this is one of the best global optimization methods (see the comparison in Mongeau et al. 2000). It is important to have the search range of the ASA algorithm chosen so that the parameters sampled are in the admissible range, for example do not lead to zero volatility (which could happen when $\alpha \leq 0$).

3. Simulation results

3.1. The filtering problem

The proposed model is tested on a number of simulated examples of the Heston model. To assess the comparative performance of the proposed model, we have compared it with the integration-based approach (i.e. the approach discussed in the introduction section, proposed by Kitagawa 1987, Fridman and Harris 1998, and Watanabe 1999), as well as the sampling importance resampling variant of the particle filtering (PF) approach (Arulampalam et al. 2002). To make the comparison fair, we have selected the integration discretization step size for the integration-based model and the number of particles for the particle filtering approach such that all three compared methods have the same computational cost, as measured by the CPU time. It is always possible for these competing methods to make an extra-fine volatility estimate by increasing the discretization size or the number of particles. But this of course is at the expense of significantly higher computational cost. For the integration method (INT), we used Simpson’s method of integration. It was hard to use more advanced methods such as quadrature methods, because one cannot apply them in a real-time or in a sequential nature. For any sampling instants indicated by the method at time $t$, all the sampling instants at times $t' < t$ have to be revised, leading to significant computation due to the need to redo all the computations.

For all three methods we computed the maximum of the likelihood function at each time step. For the PF method, to obtain the likelihood in terms of the generated particles, we used the kernel density estimator (we used MATLAB’s ksdensity program). Simpler methods like the histogram density estimators did not yield good results (the histogram method introduces a discretization error in the position of the density’s peak).

We compared the results of each method to the optimal volatility estimate, using the mean absolute percentage error (MAPE), defined as

\[ \text{MAPE} = \frac{1}{T} \sum_{t=1}^{T} \frac{|\hat{v}_t - v_t^{\text{opt}}|}{v_t^{\text{opt}}} \times 100\%, \]  

where $\hat{v}_t$ is the volatility estimate, $v_t^{\text{opt}}$ is the optimal volatility estimate, and $T$ is the length of the observation window. The optimal estimate is obtained using the integration approach (INT), but with a very fine discretization step of 0.002. Below 0.002 the results were almost the same, indicating that this choice was sufficient. However, we must mention that this optimal estimate is only approximately optimal, due to the fact that the likelihood is based on a discretization of the continuous time process and hence is only an approximation (as clarified at the beginning of section 2.2). However, because all methods considered are subject to the same approximation, the experiment can be considered a fair comparison.

For each parameter set, we generate $s_t$ and $v_t$ according to the Heston model equations. We assume that the observations are daily. Assuming around 250 trading days per year, we fix $\Delta = 1/250$. Moreover, we consider the size of the generated time series to be one year, i.e. $T = 250$ observations. For the same parameter set the estimation experiments are repeated 20 times, each time generating a different $(s_t, v_t)$ trajectory, applying the three candidate estimation models, and computing the estimation error (MAPE) for each method. The estimation errors are then averaged over the 20 trials. We assume that the initial volatility $v_0$ is distributed according to the one-sided exponential distribution with decay parameter equal to 1, i.e. $p(v_0) = e^{-v_0}I(v_0)$, where $I(v_0)$ denotes the indicator function. In the first experiment we considered the parameter set

Experiment 1: \[ \mu = 0.05, \quad \alpha = 0.15, \quad \beta = 0.08, \quad \sigma = 0.25, \quad \rho = -0.25. \]  

These parameter values are in the ballpark of those used by Johannes et al. (2006b). We used also another set of parameters, and repeated the same experiment. These are

Experiment 2: \[ \mu = 0.05, \quad \alpha = 0.02, \quad \beta = 0.02, \quad \sigma = 0.15, \quad \rho = -0.15. \]
Table 1. The estimation error (in MAPE, %) of the New Method (New Meth) versus the Integration (INT) method and the Particle Filtering (PF) method.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>MAPE (New Meth)</th>
<th>MAPE (INT)</th>
<th>MAPE (PF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>8.3</td>
<td>35.6</td>
<td>45.9</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>3.6</td>
<td>48.1</td>
<td>46.0</td>
</tr>
</tbody>
</table>

Table 1 shows the average MAPE’s (average over the 20 trials) for all three methods. Experiment 1 represents the experiment on parameter set (38), while Experiment 2 denotes the experiment on parameter set (39). For INT, the selected integration discretization step that would make it of similar computation time as the other two methods turned out to be 0.016 and 0.0145 for, respectively, Experiments 1 and 2. For PF, the number of particles (for similar computation time) turned out to be 25 for both experiments. One can see from the table that the new method considerably beats the two other competing methods. In Experiment 2, the outperformance is more than in Experiment 1. This is due to the fact that $\sigma$ is smaller in Experiment 2, making the approximations in deriving the method more accurate. Figure 1 shows the volatility estimate for all three methods versus the optimal estimate. To have an idea about how far the estimates are from the true volatility path (rather than the optimal estimate, as given above) we have computed the mean absolute percentage error (MAPE) w.r.t. the true volatility, as realized by the path simulations. For Experiment 1 this MAPE turned out to be 21.53, 26.10, 42.13 and 46.94% for, respectively, the optimal estimator (Opt Est), New Meth, INT and PF. For Experiment 2 this MAPE was 16.75, 17.21, 47.90 and 47.99% for, respectively, Opt Est, New Meth, INT and PF. Of course this measure is ‘noisier’ than the one with respect to the optimal estimator. Nevertheless, it is important to gauge its level because of its closer relation to the truly realized process.

3.2. The parameter estimation problem

We implemented the parameter estimation part of the proposed method, as described in section 2.3. Again we compared against INT and PF. As mentioned, we used the adaptive simulated annealing method (ASA) to find the most likely parameters, i.e. to optimize the likelihood function w.r.t. the five fixed parameters: $\alpha$, $\beta$, $\sigma$, $\mu$ and $\rho$. We used the following experimental set-up. We generated a one-year long (250 observations-long) volatility/asset price series, ‘the calibration series’, to be used for obtaining the maximum likely parameters. In addition, we generated another one-year long series, ‘the testing series’, to be used to test the estimated models. Of course these estimated models, whether INT, NEW or PF, use only estimated parameters. Hence the error on the testing series will incorporate the effects of the parameter estimation error, as well as the estimation error inherent in the volatility estimation or filtering step. The rationale of this set-up is that in typical practical situations the parameters have to be calibrated every few months. Then the parameters are fixed and the volatility is estimated going forward. So the ultimate goal is to obtain an accurate volatility estimate, taking into account that the parameters are just estimates.

We considered the two parameter sets in (38) and (39), and performed the experiment on each of them (Experiment 1 and Experiment 2). For each parameter set we performed five trials, where each trial represents a different trajectory generated using the same parameter set. The results are then averaged over the five trials. Again, the parameters such as the number of particles for PF and the integration discretization step for INT are fixed so that the computational cost is the same for all three methods. We also fixed the ASA parameters to be the same for all three methods. To focus the optimization into the most feasible portion of the parameter space we have added some bounds on the parameters. The bounds chosen are the following: $\sigma_{\text{min}} = 0$, $\sigma_{\text{max}} = 0.4$, $\beta_{\text{min}} = 0$, $\beta_{\text{max}} = 0.4$, $\sigma_{\text{min}} = 0.05$, $\sigma_{\text{max}} = 0.5$, $\mu_{\text{min}} = 0.01$, $\mu_{\text{max}} = 0.15$, $\rho_{\text{min}} = -0.4$ and $\rho_{\text{max}} = 0.4$. Beyond these ranges the parameter values are either inadmissible or are unrealistic for any type of market. The proposed method and the integration method follow the same formulation (as described in section 2.3), except for different ways to evaluate the integrals. However, for PF the situation is different, since what is essentially obtained in PF is the conditional probability density $p(v_{t+1}|y_{t+1})$ instead of the joint probability density $p(v_{t+1},y_{t+1})$. So the analysis in section 2.3 cannot apply. We used for PF the ‘joint filter’ (Javaheri 2005), in which we concatenate the volatility state and the parameter vector and apply one filter to this augmented state. The likelihood of the joint volatility state/parameter vector is then obtained from the particles using the kernel density estimator. We have used ASA to obtain the maximum of this distribution. As it turns out,
compute the computational cost is the same for all three methods.

Figure 2. The estimated volatilities versus the optimal estimate (Opt Est) for the new method (New Meth), the Integration (INT) method, and the particle filtering (PF) approach, for a generated volatility path for Experiment 2 (see parameter set 39). All parameters of the methods are fixed so that the computational cost is the same for all three methods.

this formulation for PF is much more efficient than the one in section 2.3. The reason is that the function to be optimized is much simpler than the case of INT and New Meth. We fixed the number of particles \( N_p \) so that the overall computation time, including the optimization part, equals that of the other two methods. It yielded \( N_p = 4200 \) for each of Experiment 1 and Experiment 2. Even though it seems a large enough number, note that the augmented dimension is now six, so perhaps we do need a large number to cope with the large dimension. However, it was sufficient to beat the other two methods. Table 2 shows the MAPE of the three methods on the testing series. For INT, the selected integration discretization step that would make it of similar computation time as the other two methods turned out to be 0.016 and 0.0145 for, respectively, Experiments 1 and 2 (similar to the case of the filtering experiments presented in the last subsection). One can see from the table that PF yielded the best results followed by New Meth then by INT as distant third.

3.3. Real market application

We have applied the combined parameter estimation and volatility filter to the NASDAQ index time series. Specifically, we have used a set-up similar to the previous subsection. We considered the last two years of data, from 15 August 2004 to 14 August 2006. Again we used the first year to apply the parameter estimation method and calibrate the parameters. We then fixed the parameters and used the second year for the estimation of volatility, as a test for the proposed method. It is hard to assess the accuracy of the volatility estimate precisely, because we do not know the ‘ground truth’, i.e. the true volatility. However, to check visually that the results make sense, we have plotted the volatility estimate \( \nu_t \) of the test year against \((y_{t+1} - y_t - \mu \Delta)^2\). Figure 3 shows these results. One can see that in periods where \((y_{t+1} - y_t - \mu \Delta)^2\) was low, \( \nu_t \) tends to be low (e.g. the periods 1–20, 40–70, and 100–150). Conversely, the periods when \((y_{t+1} - y_t - \mu \Delta)^2\) was high, \( \nu_t \) tends to be high (e.g. the periods 20–40, 70–100 and 200–254). The period 150–200 had medium level square increments and its \( \nu_t \) transitioned from a low level to a high level.

3.4. Comments on the results

From the preceding simulations we can deduce the following observations.

- It can be seen that for the filtering problem the proposed method is clearly better than the other two by a large margin.
- The estimate of the proposed method follows the optimal estimate closely for the first 80 to 100 days, then it diverges a bit (probably due to the accumulation of error day after day). This suggests the need to perhaps recalibrate every few months with the optimal method (i.e. the integration method with sufficiently fine discretization).
- For the parameter estimation problem, PF was somewhat better than the proposed method. This suggests calibrating the parameters using

<table>
<thead>
<tr>
<th>Method</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE (New Meth)</td>
<td>15.4</td>
<td>18.2</td>
</tr>
<tr>
<td>MAPE (INT)</td>
<td>50.6</td>
<td>48.8</td>
</tr>
<tr>
<td>MAPE (PF)</td>
<td>13.6</td>
<td>11.4</td>
</tr>
</tbody>
</table>

Table 2. The estimation error (in MAPE, %) of the New Method (New Meth) versus the Integration (INT) method and the Particle Filtering (PF) method for the case when estimating the unknown parameters.
the PF method every few months or a year, but performing the day-to-day volatility estimation procedure using the proposed method.

- It was difficult for all methods including the optimal method to get close to the real parameter values for the parameter estimation problem. This is consistent with the findings of Javaheri (2005), who observed that the likelihood surface in parameter space is very flat. In order to get an estimate close to the real parameter values, a sample size of several tens of thousands of data points is needed. However, the fact that the parameter space is flat means that the error in the volatility estimate is somewhat tolerant to parameter estimation errors.

4. Conclusion

A new semi-analytic approximate formula is proposed in this paper for the volatility filtering problem for the Heston model. The formula reduces the likelihood at a particular point to a product of a number of terms. Such analytical approximations serve as one step closer to solving this hard problem, which is a central problem for many applications in finance. The challenge is to extend this analysis further to other volatility models, such as the Heston model with jumps. The extension is not as straightforward, but it could be an interesting future research project.

References


