Evaluation of a two-step estimation procedure for a functional model of volatility

Suad Elezović
Department of Statistics, Umeå University

Abstract

A two-step procedure for volatility estimation is evaluated by a simulation study intended to mimic estimation from the Swedish limit order book. To simulate data with varying volatility the Heston stochastic volatility model is used. From the simulated data, the time series of realized quadratic variation (RQV) for a given relative quantity of shares are obtained. These time series are modeled in a functional time series context by fitting an autoregressive moving average model. This model may be estimated in two ways, either by obtaining the raw estimates of the coefficient functions (naive approach) or by smoothing the fitted coefficient functions (two-step approach). Our results show that the risk measures of the smooth coefficient functions are indeed smaller than the corresponding risk measures of the coefficient functions of raw estimates. Consequently, the two-step estimation procedure is considered to be more efficient than the naive approach within this framework.

Keywords: Financial volatility; realized quadratic variation; functional time series; Heston model; two-step estimation procedure; smoothing.

1 Introduction

A functional time series model is introduced in Elezović (2008) to model the financial volatility in the Swedish limit order book (LOB).
The model can be estimated in a naive manner (raw estimates) or more efficiently (two-step procedure by Fan and Zhang (2000)). The purpose of this paper is to study whether the method by Fan and Zhang (2000) is more efficient than the naive one. To evaluate the estimation procedure, data are simulated from a stochastic volatility (SV) model (Heston model) mimicking LOB data.

The SV models allow the asset volatility to be time varying, which is particularly appropriate to capture different aspects of the distribution of financial asset returns. Such an aspect is volatility clustering, commonly observed in empirical studies as a result of positive autocorrelations of volatility measures over periods of time. In particular, the Heston SV model, outlined later, is likely to capture deviations from normality assumption for asset returns, such as excess kurtosis due to fatter tails and more peaky distribution, as well as excess skewness. These features are typical for high-frequency financial data studied here (see, e.g., Mariani et al., 2007).

Essentially, a measure of variation of asset prices called realized quadratic variation (RQV), as a function of a given relative quantity of shares $w$ to be bought or sold, is modeled by a functional time series model. The ordinary RQV is defined as the sum of squared high-frequency financial returns sampled over equidistant time intervals over each period of time. As the number of these time intervals increases, RQV converges in probability to the increments of integrated variance (IV) of the efficient (true) prices. For more information about RQV see, e.g., Andersen et al. (2003) or Barndorff-Nielsen and Shephard (2004a). The concept of IV, which is essential to the underlying theory of SV models, is defined later in the paper (see Appendix A).

The two-step estimation procedure is performed by fitting an autoregressive moving average (ARMA) time series model to the data (RQV time series), obtaining the raw estimates of the coefficient functions. The raw estimates are then smoothed by a non-parametric smoothing method to obtain the smooth estimates of the coefficient functions. The performance of this procedure is then evaluated by comparing some risk measures (mean squared errors and some related measures) of the raw coefficient functions with the corresponding mea-
sures of the smooth estimates.

Our results show that the risk measures of the smooth coefficient function estimates are generally smaller than the corresponding measures for the ordinary raw estimates. As a consequence, we find that the two-step estimation procedure may be considered as an improvement in volatility estimation within this framework.

The plan of the paper is as follows. After this introductory section, in Section 2, a description of the Heston SV model is presented together with some discretizations for this model. Section 3 describes the two-step estimation procedure for volatility modelling. Afterwards, in Section 4, the simulation study with the purpose of evaluating the mentioned procedure is described. The main results from the simulation analysis are then presented. Finally, the paper is briefly summarized in Section 5.

2 Heston model for varying volatility

2.1 Basic properties

In the Heston SV continuous time model the instantaneous squared volatility follows a square-root mean-reverting diffusion correlated with the stock price process represented by a modified geometric Brownian motion. This modification allows volatility of the return process to vary over time in contrast with basic geometric Brownian motion with constant volatility.

The Heston model is specified by the coupled two-dimensional stochastic differential equation (SDE) of the following form, (see, e.g., Glasserman, 2004, p. 121):

\[
\begin{align*}
\frac{dS(t)}{S(t)} &= \mu dt + \sqrt{V(t)}dW_1(t), \\
dV(t) &= \alpha (\theta - V(t)) dt + \sigma \sqrt{V(t)}dW_2(t),
\end{align*}
\]  

where \( S(t) \) is the price of an asset in continuous time; drift \( \mu \) is the expected instantaneous rate of return of the asset; the squared volatility
V(t) follows a mean-reverting square-root diffusion, converging to the positive long-run mean $\theta$ at the rate governed by a strictly positive parameter $\alpha$; the parameter $\sigma$, sometimes called volatility of volatility, is a strictly positive constant which, together with $V(t)$, builds the diffusion term $\sigma\sqrt{V(t)}$; $(W_1, W_2)$ is a two-dimensional Brownian motion ($W_1$ and $W_2$ are standard Brownian processes such that $dW_1(t)dW_2(t) = \rho dt$ and $\rho$ is a correlation constant in $[-1, 1]$).

The correlation $\rho$, which affects the skewness of an asset’s distribution, is usually assumed to take a negative value emphasizing the stylized fact that equity returns and the corresponding volatility are often negatively correlated. The volatility of volatility parameter $\sigma$ determines how peaky the distribution is and hence affects kurtosis. The diffusion term $\sigma\sqrt{V(t)}$ tends to zero as $V(t)$ approaches the origin preventing $V(t)$ from taking negative values. The drift $\mu$ is trivial and may be omitted for simplicity if necessary (often assumed as zero in practical applications). The mean reversion parameter $\alpha$ affects the level of volatility clustering, which is another stylized fact about properties of asset returns (see, e.g., Cont (2001) for more details about the stylized facts of asset returns).

The squared volatility $V(t)$ in (1) may be treated as the spot (or instantaneous) variance of relative changes of $S(t)$. One usually wants to estimate the quadratic variation (QV) of $dS(t)/S(t)$ over $[t, t + dt]$ which is expressed in terms of $V(t)dt$, as pointed out in Andersen (2007). See also Appendix A for a brief introduction to the role of QV as a measure of volatility.

When the focus is put on log-returns, it is necessary to transform the process in (1a) by using Itô’s formula from stochastic calculus (e.g. Glasserman, 2004, pp. 94, 545-547), obtaining:

$$d\log S(t) = \left(\mu - \frac{1}{2} V(t)\right) dt + \sqrt{V(t)}dW_1(t).$$

(2)

Observe that $V(t)$, defined by the square-root diffusion process in (1b), will never be negative provided that $V(0) > 0$. Furthermore, the process remains strictly positive for all $t$ when $V(0) > 0$ and $2\alpha\theta \geq \sigma^2$. When $2\alpha\theta < \sigma^2$, zero is an attainable boundary for the SDE in (1b)
Any SV model, including the Heston model, relies on an assumption that the underlying volatility is an unobservable (latent) factor but essential in asset return process generation. Many other theoretical models for asset returns are built upon the same assumption (Andersen and Sørensen, 1996). As pointed out in Aït-Sahalia and Kimmel (2007), such a model may be more effective in capturing important empirical properties of the joint behavior of stock and option prices than some more restricted models which assume constant volatility. Further motivation for this model may be found in Daniel et al. (2005), who advocates that the Heston model is likely to outperform the Gaussian model in terms of sensitivity to price fluctuations.

Although the Heston model provides a closed-form solution for the price of a European call option, there is no explicit solution for the SDE that defines the instantaneous squared volatility process. As a consequence, some approximation or numerical discretization method has to be used to simulate stock prices and instantaneous volatilities from this model.

### 2.2 Discretizations of the Heston model

Two main problems with the discretization of the Heston model are evident from several empirical studies. First, the condition $2\alpha \theta \geq \sigma^2$ is typically not satisfied allowing $V(t)$ to hit zero, as pointed out in (Andersen, 2007), among others. Secondly, the simulations on a discrete time grid imply that the probability of obtaining a negative value for the discrete spot variance $V_t$ at the next time step is strictly greater than zero. This effect arises due to the presence of Brownian increments $dW(t)$ which are, when written in a discretized form as $\Delta W(t)$, normally distributed with mean zero and variance $\Delta t$. A vast number of methods are proposed both for discretization (approximation) of the continuous time processes in (1) and for fixing the negative variance, see Lord et al. (2006) for an overview.
2.2.1 The Euler-Maruyama method

A method called the Euler-Maruyama (EM) scheme (e.g. Kloeden and Platen, 1999) is often used as benchmark for other discretization methods due to its simplicity and speed. The EM method may be applied to model (1) as an approximation for the paths of the squared volatility process and the corresponding stock price process on a discrete time grid. By partitioning a time interval $T$ into $N$ segments of equal length $\Delta t = T/N$, $(0 = t_0 < t_1 < \ldots < t_N)$, this discretization is expressed as follows

$$S_{t_i} = S_{t_{i-1}} + \mu S_{t_{i-1}} \Delta t + \sqrt{V_{t_{i-1}}} \Delta W_{2,t_i}, \quad (3a)$$

$$V_{t_i} = V_{t_{i-1}} + \alpha (\theta - V_{t_{i-1}}) \Delta t + \sqrt{V_{t_{i-1}}} \sigma_v \Delta W_{1,t_i}, \quad (3b)$$

where $\Delta W_{j,t_i} = W_{j,t_i} - W_{j,t_{i-1}}$, $j = 1, 2$, are the Brownian increments, which are independent of each other and normally distributed with mean zero and variance $\Delta t$. These increments are simulated by

$$\Delta W_{1,t_i} \overset{d}{=} Z_V \sqrt{\Delta t},$$

$$\Delta W_{2,t_i} \overset{d}{=} Z_X \sqrt{\Delta t},$$

where $\overset{d}{=} \text{means } \text{“equal in distribution”}$. Here,

$$Z_X = \left( \rho Z_V + \sqrt{1 - \rho^2} Z_S \right),$$

assuming that $Z_V \sim N(0, 1)$ and $Z_S \sim N(0, 1)$ are independent. The correlation coefficient $\rho$ measures the level of linear dependence between the two Brownian processes $W_{1,t}$ and $W_{2,t}$.

When the log-prices are simulated according to (2), the discretization in (3a) is expressed as follows

$$\log(S_{t_i}) = \log(S_{t_{i-1}}) + \left( \mu - \frac{1}{2} V_{t_{i-1}} \right) \Delta t + \sqrt{V_{t_{i-1}}} \Delta W_{2,t_i}. \quad (4)$$

In practical applications, researchers commonly adopt a “fix” by either using absorption or reflection for the discrete process $V_t$ in (3b).
Absorption means that the values of $V_t$ are replaced with zeroes whenever they become negative. Reflection means taking the absolute values of $V_t$ under the square root in (3b), before advancing the recursion. Higham and Mao (2005) motivate the use of the reflection method by showing that an EM discretization provides correct approximations of the first and second moments.

Under certain conditions, the EM scheme converges to the true solution as the time step is made smaller and smaller, as pointed out in Higham (2001). However, the drift and diffusion terms in (1b) do not satisfy a linear growth condition and they are not globally Lipschitz (Lord et al., 2006). Hence, the standard convergence theory, as defined in Kloeden and Platen (e.g. 1999), cannot be applied for numerical simulations by using the EM procedure. Several other methods have been proposed, including the method in Kahl and Jäckel (2006) and the exact method in Broadie and Kaya (2006).

2.2.2 IJK-IMM discretization

Kahl and Jäckel (2006) consider discretization of the $V(t)$ process in (3b) using an implicit Milstein method (IMM)

$$V_{t_i} = V_{t_{i-1}} + \alpha (\theta - V_{t_{i}}) \Delta t + \sigma \sqrt{V_{t_{i-1}}} \Delta W_{1,t_{i}} + 0.25 \sigma^2 (\Delta^{2} W_{1,t_{i}} - \Delta t),$$

(5)

together with their implicit Jäckel-Kahl (IJK) discretization for the logarithms of stock prices $\log (S_t)$ in (4), as shown here

$$\log (S_{t_i}) = \log (S_{t_{i-1}}) + \mu \Delta t - 0.25 (V_{t_i} + V_{t_{i-1}}) \Delta t + \rho \sqrt{V_{t_{i-1}}} \Delta W_{1,t_{i}}$$

$$+ 0.5 \left( \sqrt{V_{t_i}} + \sqrt{V_{t_{i-1}}} \right) (\Delta W_{2,t_{i}} - \rho \Delta W_{1,t_{i}})$$

$$+ 0.25 \sigma \rho (\Delta W_{1,t_{i}}^2 - \Delta t).$$

(6)

This discretization scheme will result in positive paths for the spot variance process in (5) provided that $\sigma^2 < 4\alpha \theta$. As Kahl and Jäckel
(2006) do not provide a solution when this condition is not satisfied, Andersen (2007) recommends using $V_{t_i}^+$ and $V_{t_{i-1}}^+$ instead of $V_{t_i}$ and $V_{t_{i-1}}$ in (5) and (6), where the notation $x^+ = \max(0, x)$ is used. In Lord et al. (2006), this method is called “full truncation”.

Comparing several discretization procedures, Kahl and Jäckel (2006) claim that the IJK-IMM scheme yields the best results in terms of the strong convergence measure. Their method is particularly suited for the Heston model with a strong negative correlation $\rho$.

### 2.2.3 The Broadie-Kaya method

The BK method is probably the only existing procedure that simulate a solution to the SDE in (1b) without a bias. The algorithm is based on the result that the distribution of $V(t)$ given $V(u)$ is, up to a scale factor, a non-central chi-squared distribution, as shown here

$$V(t) \sim \frac{\sigma^2 (1 - e^{-\alpha(t-u)})}{4\alpha} \chi_d^2 \left\{ \frac{4\alpha e^{-\alpha(t-u)}}{\sigma^2 (1 - e^{-\alpha(t-u)})} V(u) \right\}, \quad t > u,$$

(7)

where $\chi_d^2(\xi)$ is a non-central chi-squared random variable with a non-centrality parameter $\xi$

$$\xi = \frac{4\alpha e^{-\alpha(t-u)}}{\sigma^2 (1 - e^{-\alpha(t-u)})}$$

(8)

and $d$ degrees of freedom

$$d = \frac{4\theta \alpha}{\sigma^2}.$$  

(9)

Hence, it will be possible to sample from the distribution of $V(t)$ exactly, provided that we can sample from a non-central chi-square distribution. The complete algorithm for simulation of the square root diffusion in (1b), including sampling from the non-central chi-squared distribution, is given in Glasserman (2004, pp. 122-124).

Worth noting, though, is the fact that the BK approximation scheme is highly time consuming and therefore not recommendable
in typical applications, as discussed in Andersen (2007). Particularly, simulation of \( \log(S(t)) \) in (2) within the framework of Broadie and Kaya (2006) is extremely complex allowing for noticeable biases. Furthermore, it is not straightforward how to introduce correlated Brownian processes in (1), as noted in Lord et al. (2006). Andersen (2007) proposes a discretization scheme for \( \log(S(t)) \) that will supposedly overcome problems with both the EM scheme and the BK scheme, as follows

\[
\log(S_t) = \log(S_{t-1}) + \mu \Delta t + K_0 + K_1 V_{t-1} + K_2 V_t + \sqrt{K_3 V_{t-1} + K_4 V_t} Z,
\]

where \( Z \) is the standardized Gaussian random variable independent of \( V_t \) and

\[
K_0 = -\frac{\rho \alpha \theta}{\xi} \Delta t, \quad K_1 = \eta_1 \Delta t \left( \frac{\alpha \rho}{\sigma} - 0.5 \right) - \frac{\rho}{\sigma}, \\
K_2 = \eta_2 \Delta t \left( \frac{\alpha \rho}{\sigma} - 0.5 \right) + \frac{\rho}{\sigma}, \quad K_3 = \eta_1 \Delta t (1 - \rho^2), \\
K_4 = \eta_2 \Delta t (1 - \rho^2)
\]

for certain constants \( \eta_1 \) and \( \eta_2 \). Here, we adopt a central discretization \( \eta_1 = \eta_2 = 0.5 \), also suggested in van Hastrecht and Pelsser (2008).

To summarize, the BK scheme for the variance process \( V(t) \) is combined with a discretization scheme for the log-prices \( \log(S(t)) \) in (10), as follows:

(i) For some \( u < t \), given \( V(u) \), generate a sample from the distribution of \( V(t) \) (sampling from a constant times a non-central chi-square random variable (see (7), (8) and (9))).

(ii) Draw a random sample \( Z \) from the standard normal distribution.

(iii) Given \( \log(S_{t-1}), V(u), V(t) \) and \( Z \), compute \( \log(S_t) \) as defined in (10).
3 Volatility of functional data: model and estimation

3.1 Preliminaries

This study is particularly oriented to modeling volatility of share prices from the Swedish limit order book (LOB) data. In general, any LOB represents a record of the unexecuted bid and ask limit orders. Basically, the data in an LOB consist of the bid and ask prices, the corresponding bid and ask quantities of shares, respectively, and the registered time points of each change in the LOB. The Swedish LOB is considered as highly transparent since all important information concerning the five best price levels and the corresponding quantities are available to the traders, see Elezović (2006, 2008) for a brief description of this market.

A convenient measure that summarizes the contents of the LOB at a given time point \( t \) is proposed by Olsson (2005) as an average price per share for a given quantity of shares \( q \), as follows

\[
p_t(q) = \frac{\sum_{l=1}^{k-1} (p_{l,t} Q_{l,t}) + p_{k,t} \left( q - \sum_{l=1}^{k-1} Q_{l,t} \right)}{q}, \quad \text{for } k = 1, 2, \ldots, K,
\]

where \( K \) is the number of the available price levels in the LOB (in the Swedish LOB \( K = 5 \)) and \( p_{k,t} \) represents the quoted price at time \( t \) and level \( k \) in the LOB. The level \( k \) is determined by its relation to the given quantity \( q \) (demanded or supplied), so that

\[
\sum_{l=1}^{k-1} Q_{l,t} < q \leq \sum_{l=1}^{k} Q_{l,t},
\]

which means that \( k \) represents the level in the LOB where the quantity of shares \( q \) would be executed. \( Q_{l,t} \) is the bid (ask) quantity of shares available at price \( p_{l,t} \).

In the financial literature, the price functions in (12) are sometimes called bid (ask) curves (see, e.g., Gourieroux and Jasiak, 2001). In
Elezović (2008) the bid (ask) curves are computed for given relative quantities \( wQ_t \), such that

\[
\tilde{p}_t (w) = \tilde{p}_t (wQ_t),
\]

where \( \tilde{p}_t (w) \) is relative bid (ask) curve for a given weights-percentage \( w \) and \( Q_t \) is total quantities available in the LOB at time \( t \), such that

\[
0 < w \leq 1 \quad \text{and} \quad Q_t = \sum_{l=1}^{5} Q_{l,t}.
\]

By using \( \tilde{p}_t (w) \) instead of the ordinary quoted prices \( p_t \), one may define the functional RQV over each \( i \)th interval (usually a day) denoted as \( [Y_M]_i (w) \), as follows

\[
[Y_M]_i (w) = \sum_{j=1}^{M} r_{j,i}^2 (w), \quad j = 1, 2, \ldots, M; \quad i = 1, 2, \ldots, T,
\]

where the functional returns \( r_{j,i} (w) \) are defined by extending the expression in (35), Appendix A, to obtain

\[
r_{j,i} (w) = \tilde{p}((i-1)h + \frac{j}{M}) (w) - \tilde{p}((i-1)h + \frac{j-1}{M}) (w), \quad j = 1, 2, \ldots, M.
\]

From now on, the obtained \([Y_M]_i (w)\) in (14) will be denoted as \( y^*_i (w) \) for simplicity.

### 3.2 Model and estimation procedure

model in contrast to estimating the coefficient functions of a functional linear model.

According to this procedure, a chosen time series model is fitted to each \( y^* (w) \) to obtain the raw estimates of the coefficient functions. Afterwards, the obtained raw estimates are smoothed in order to gain more efficiency by producing smoothed estimates of the coefficient functions. Here, the first step of the procedure involves estimating an autoregressive moving average (ARMA) model of order (1,1) of the following form:

\[
y^*_i (w) = \gamma_0 + \gamma_1 y^*_{i-1} (w) + \varepsilon_i (w) + \eta_1 \varepsilon_{i-1} (w).
\] (16)

For the purpose of this work, it suffices to assume that \( \varepsilon_i (w) \) is a random process, independent through time (days \( i, i = 1, 2, \ldots, T \)) but allowed to be correlated for different weights \( w_q \), where \( q = 1, 2, \ldots, \kappa \) and \( \kappa \) is given as a positive integer greater than 1. In other words

\[
\text{Cov} (\varepsilon_i (w), \varepsilon_{i+h} (w)) = 0, \quad \text{for } h \neq 0,
\]
\[
\text{Cov} (\varepsilon_i (w_j), \varepsilon_i (w_k)) = \psi (w_j, w_k), \quad \text{for } j \neq k,
\]
\[
= \sigma^2 \varepsilon, \quad \text{for } j = k.
\] (17)

After fitting an ARMA(1,1) model for each given \( w \), the obtained raw estimates of the autoregressive (of order 1) coefficient functions are obtained, written in vector notation as

\[
\hat{\gamma}_1 = (\hat{\gamma}_1 (w_1), \hat{\gamma}_1 (w_2), \ldots, \hat{\gamma}_1 (w_\kappa))^T,
\] (18)

where \( \kappa \) represents the total number of weight-percentages used. Furthermore, the raw estimates are smoothed by a smoothing technique applied to the data \( \{w_q, \hat{\gamma}_1 (w_q)\} \). This smoothing step represents the second step of this procedure resulting in the smooth coefficient function \( \hat{\gamma}_1 (w) \). As suggested in Fan and Zhang (2000), a convenient smoothing technique would be linear in the responses which is basically adopted here. Then, assuming that \( \gamma_1 (w) \) is \((d + 1)\) times continuously differentiable, a non-parametric linear estimator of the
\( g \)th derivative of \( \gamma_1^{(g)}(w) \), for some \( 0 \leq g < d + 1 \), is given by

\[
\dot{\gamma}_1^{(g)*}(w) = \sum_{q=1}^{\kappa} H(w_q, w) \dot{\gamma}_1(w_q),
\]

where the weights \( H(w_q, w) \) are constructed according to the relevant smoothing technique. Note that the same two-step procedure may also be applied to the coefficient curves of raw estimates of both the intercept coefficients and the moving average coefficients.

Within a general framework of the Constant Elasticity of Variance SV models to which the Heston model belongs, the actual variance \( \sigma_i^{[2]} \), as given in (40) in Appendix A, has an autocorrelation function of an ARMA(1,1) model (Barndorff-Nielsen and Shephard, 2002a). A basic assumption here is that the RQV has the same autocorrelation function and the same autoregressive roots as the corresponding actual variances. As noted in Appendix A, \( q_i^* \) is a consistent estimator of \( \sigma_i^{[2]} \) for all SV models. This implies that, in the limit, the RQV process and a corresponding actual variance process should have the same ARMA(1,1) representation (Meddahi, 2003).

A connection between the autoregressive parameter from the ARMA(1,1) model and the parameter \( \alpha \) from the Heston model in (1), may be written in an explicit manner as

\[
\gamma_1 = \exp(-\alpha \Delta t),
\]

where the interval \( \Delta t > 0 \) (Barndorff-Nielsen and Shephard, 2002a). This expression will be used to evaluate the performance of the two-step procedure outlined above.

In addition, it is worth noting that the moving average root from the ARMA(1,1) model is also determined by the expression in (20) but has to be found numerically (Barndorff-Nielsen and Shephard, 2002a).
4 Simulation study

4.1 Purpose

In the simulations presented here, the performance of the two-step estimation procedure for the functional time series model (16) is studied. First of all, the functional stock prices are generated from the Heston model with pre-specified parameters. From these prices the corresponding $y_i^*(w)$ are computed and used in the two-step procedure, as the observed time series. Then, the performance of fit is evaluated by inspecting the MSE and related risk measures for the smooth autoregressive coefficient functions in comparison with the corresponding measures for the raw autoregressive coefficient functions.

4.2 Design

The three simulation methods for the Heston SV model, explained in the previous section, are used here: the EM method with the reflection fix by Higham and Mao (2005), the IJK-IMM procedure with the full-truncation fix and the BK scheme. The crucial parameter $\alpha$ is assumed to be a function of $w$ while the remaining parameters are kept constant. For the purpose of this study, $\alpha(w)$ is computed for a set of 19 relative quantities $w = 0.05, 0.10, 0.15, \ldots, 0.95$, choosing the functional form

$$\alpha(w) = \beta_0 + \beta_1 w + \beta_2 w^2.$$  (21)

The motivation for the choice of the functional form for $\alpha(w)$ and the remaining parameters comes from real LOB data as described in Appendix B. The values of all parameters are shown in Table 1.

Simulations are generally performed over a fixed sample period $[0, T]$, which is divided into $N$ intervals, each having length $\Delta t = \frac{T}{N}$, where $T$ corresponds to the number of years and $N$ corresponds to the number of days in the sample. As the major goal with this scheme is to simulate $y_i^*(w)$ from the intra-day data, the interval $\Delta t$ is further divided by $M$, the number of intra-daily observations, which results in the discrete time increments $\Delta^* = \Delta t/M$. In this way, both the
Table 1: Parameter values for simulations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>0</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.359</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.01</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.6534</td>
</tr>
<tr>
<td>$\rho^*_{t_0}$</td>
<td>0.0135</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.0038</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>(-0.00445)</td>
</tr>
<tr>
<td>$V_{t_0}$</td>
<td>1.58e-05</td>
</tr>
<tr>
<td>$\rho$</td>
<td>(-0.5)</td>
</tr>
</tbody>
</table>

Squared volatilities and the log-prices in (3b) are assumed to vary from one time increment to another.

As a proxy for actual variance $\sigma_i^2$ over the $i$th day, as defined in (39), the following expression is used:

$$\tilde{\sigma}_i^2 = \sum_{j=1}^{M} V_{i,t_j} \left( \Delta t / M \right). \tag{22}$$

In these simulations $T = 4$ and $N = 1028$, which leads to the sampling time interval $\Delta t = 1/257$. The total number of replications is set to $K = 1000$. $N$ is first set to 2056, but the initial 1028 observations are discarded to avoid possible errors due to transient effects.

For the EM procedure, two $M$ values are used: $M = 12$ which corresponds to an interval of 40 minutes in which case $\Delta^* = \Delta t / M = \frac{4}{1028 \times 12} \approx 0.000324$ and $M = 96$ corresponding to an interval of 5 minutes with $\Delta^* = \Delta t / M = \frac{4}{1028 \times 96} \approx 0.0000405$. The IJK-IMM discretization and the BK approximation are only simulated over $M = 12$, because of computational intensity of the procedures.

### 4.3 Performance of fit: comparison between smoothed and raw coefficient functions

As mentioned before, the main objective here is to compare the raw estimates and the smoothed estimates using the MSE (and some related measures) of the corresponding coefficient curves. The population MSE (of, e.g., raw estimates) may be defined as
\[ E \left\{ [\hat{\gamma}_1 (w_q) - \gamma_1 (w_q)]^2 \right\} = E \left\{ [\hat{\gamma}_1 (w_q) - E (\hat{\gamma}_1 (w_q))]^2 \right\} \]
\[ + \{ E [\hat{\gamma}_1 (w_q) - \gamma_1 (w_q)] \}^2 \]
\[ = \text{Var} \{ \hat{\gamma}_1 (w_q) \} + \{ \text{Bias} [\hat{\gamma}_1 (w_q)] \}^2. \] (23)

Here, the MSE is computed at each single point \( w_q \) as the average MSE over all \( K \) Monte Carlo replications, as follows
\[ \text{MSE} [\hat{\gamma}_1 (w_q)] = \frac{1}{K} \sum_{k=1}^{K} [\hat{\gamma}_{1,k} (w_q) - \gamma_1 (w_q)]^2, \] (24)
where the “true” parameters \( \gamma_1 (w_q) \) are obtained by transforming \( \alpha (w) \) from the polynomial model in (21) using the expression in (20).

The corresponding decomposition of the MSE in (24) into variance and the squared bias may be written as
\[ \text{MSE} [\hat{\gamma}_1 (w_q)] = \frac{1}{K} \sum_{k=1}^{K} [\hat{\gamma}_{1,k} (w_q) - \bar{\gamma}_1 (w_q)]^2 \]
\[ + \left\{ \frac{1}{K} \sum_{k=1}^{K} [\hat{\gamma}_{1,k} (w_q) - \gamma_1 (w_q)] \right\}^2. \] (25)

For comparison purposes, the un-weighted average squared error (UASE) and mean absolute deviation error (MADE) are also computed, for the whole autoregressive coefficient curve at each replication \( k \) following Fan and Zhang (2000), as follows
\[ \text{UASE}_k (\gamma_1 (w)) = \frac{1}{K} \sum_{q=1}^{\kappa} [\gamma_{1,k} (w_q) - \gamma_{1,k} (w_q)]^2, \] (26a)
\[ \text{MADE}_k (\hat{\gamma}_1 (w)) = \frac{1}{K} \sum_{q=1}^{\kappa} |\hat{\gamma}_{1,k} (w_q) - \gamma_{1,k} (w_q)|. \] (26b)
By replacing $\hat{\gamma}$ with $\hat{\gamma}^*$ in (24, 25, 26a and 26b), the same measures may be computed for the smooth coefficient functions.

In addition, the difference expressed as

$$d_{1,k} = \text{UASE}_k \{ \hat{\gamma}^*_1 (w) \} - \text{UASE}_k \left\{ \hat{\gamma}_1^* \left( \hat{\sigma}_{i^*}^{[2]} (w) \right) \right\},$$

is compared to the following difference

$$d_{2,k} = \text{UASE}_k \{ \hat{\gamma}_1 (w) \} - \text{UASE}_k \left\{ \hat{\gamma}_1 \left( \hat{\sigma}_i^{[2]} (w) \right) \right\},$$

for each $k$, where the expressions $\hat{\gamma}_1 \left( \hat{\sigma}_i^{[2]} (w) \right)$ and $\hat{\gamma}^*_1 \left( \hat{\sigma}_{i^*}^{[2]} (w) \right)$ refer to the raw estimate and the smooth estimate, respectively, of the autoregressive coefficient functions obtained from the two-step estimation procedure applied to the time series of proxies for the actual variances $\hat{\sigma}_i^{[2]} (w)$ in (22). As earlier, $\hat{\gamma}_1 (w)$ and $\hat{\gamma}^*_1 (w)$ represent the raw respective smooth estimate from the same procedure applied to the time series $y_i^* (w)$. Moreover, define

$$d_1 = \frac{1}{K} \sum_{k=1}^{K} d_{1,k} \quad \text{and} \quad d_2 = \frac{1}{K} \sum_{k=1}^{K} d_{2,k}.$$  

(29)

As before, UASE in (27) and (28) may be replaced with MADE. The comparison of $d_1$ and $d_2$ from (29) is interesting to study because of the convergence of RQV to the actual variance, as explained in Appendix A. If $d_1 - d_2 < 0$, the smooth estimates of the autoregressive coefficient functions may be considered as having better performance than the corresponding raw estimates, which would in turn justify the use of the two-step estimation procedure.

4.4 Results

All simulations are done in the statistical software $R$ (R Development Core Team, 2008) using our own code and some available packages. The first step of the analysis is concerned with obtaining the raw estimates of the coefficient functions, which is performed using the function $arma$ from the package $tseries$. The second step involves
smoothing performed by two methods for comparison purposes: the local polynomial fitting using Gaussian kernel, which is done with the function `locfit` from the package `locfit`, and the super smoother method with the function `supsmu` from the package `stats`. Both methods are essentially versions of a local linear regression (or local polynomial fitting of order one), see Loader (1999, Ch. 2) for more details about this topic.

The bandwidths in the `locfit` procedure are selected by the “rule of thumb” method described in Fan and Gijbels (1996, Ch. 4.2). This method is implemented in the function `thumbBw` in the package `locpol`, also available in R. The super smoother procedure is also a version of local polynomial fitting but with a variable bandwidth. Choice of the bandwidth is determined by a local cross validation. The super smoother method is particularly convenient because of its computational speed.

For comparison purposes some other methods are also used, such as local polynomial regression of various orders, splines and different kernel regression. It seems that the local polynomial regression of order one works best, which is why only the previously mentioned two methods are presented in the paper.

Figure 1 shows the MSE, biases and variances of the raw and the smooth estimates of the coefficient functions, computed according to (24) and (25), when the EM discretization is used. Note that the MSE of the smooth estimates is generally smaller than the MSE of the raw estimates, except for very small and very large $w$. The graphs also show that the MSE:s are noticeably smaller for $M = 96$ than for $M = 12$, in most of the cases.

Figure 2 shows the corresponding measures, as in Figure 1, but for the IJK-IMM and BK approximations and for $M = 12$ only. The differences between the MSE:s of the smooth and raw estimates are even more obvious when the BK approximation is used than with the other two procedures. Behaviors of MSE from the IJK-IMM procedure and MSE from the Euler method are similar here.

Comparison of the performance of fit of the smooth estimates and the corresponding raw estimates is done by testing the null-hypothesis
Figure 1: EM scheme: MSE, bias and variance over 1000 replications—raw estimates (Raw Est), local regression smoother (Local Reg.), super smoother (Sup. Smoother); M=12 and M=96.

of equal means between the two risk measures, UASE and MADE, for the respective coefficient functions. According to both the Student t-tests and Wilcoxon rank tests, the null-hypothesis of equal means are strongly rejected for both UASE and MADE, in favor of the alternative hypothesis that the means of the risk measures for the smooth
Figure 2: BK (left panels) and KJ (right panels) which is short for IJK-IMM: MSE, bias and variance over 1000 replications - raw estimates (RE), local regression smoother (LR), super smoother (SSm); M=12. Coefficient functions are smaller than the corresponding risk measures for the raw estimates coefficient functions (p-values close to zero).

Moreover, equality in mean of $d_{1,k}$ in (27) and $d_{2,k}$ (28) is also tested. The tests reject the null-hypothesis of equal means of $d_{1,k}$ and $d_{2,k}$ in a vast majority of cases, indicating superiority of the
smooth estimates over the raw estimates. The only exceptions are non-significant p-values for the t-tests from the EM procedure applied to MADE, for $M = 96$. On the contrary, the Wilcoxon tests reject the null-hypothesis even in these cases.

As a general conclusion, it may be considered that the graphical inspection and the related testing procedures give evidence of the significant difference in means between all three risk measures in favor of the smooth estimates.

5 Summary

In this paper a two-step volatility estimation procedure, presented in Elezović (2008), is evaluated through a simulation study. The Heston stochastic volatility model is used as a parametric model to simulate spot squared volatilities and the corresponding functional realized quadratic variations. Then, a theoretical relationship between the autoregressive parameters from ARMA(1,1) representation of actual variances and the corresponding mean-reverting parameters from the Heston model, is utilized to evaluate the performance of fit of the mentioned procedure. This is done by fitting an ARMA(1,1) model to each series of the functional realized quadratic variations to obtain the so-called raw estimates of coefficient functions. Furthermore, in the second step of the procedure, a smoothing technique is applied to the raw estimates to obtain the smooth coefficient functions.

Comparisons between the smooth coefficient functions and the corresponding raw estimate functions are done with MSE and some related risk measures. The graphical inspections show that the risk measures for the smooth autoregressive coefficient functions are generally smaller than the corresponding measures for the raw estimates of the autoregressive coefficients. Both the Student t-tests and the Wilcoxon tests overwhelmingly confirm the results from the graphical presentation indicating superiority of the smooth estimates.

Moreover, the differences between the risk measures for RQV and the corresponding risk measures for actual variance are shown to be significantly smaller for the smooth coefficient functions than for the
corresponding raw estimates, in most of the cases. Accordingly, we find that the two-step estimation procedure, under the framework presented here, may be considered as an improvement in volatility estimation.

Acknowledgements

The author is grateful to Professor Xavier de Luna for his guidance during the preparation of this work. The author is also grateful to Anders Muszta for valuable comments and suggestions that lead to an improved presentation of the paper.
References


Appendices

A Notes on the theory of realized quadratic variation as a measure of volatility

Assume that the log-price of an asset, \( p^*(t) = \log(S(t)) \), follows the solution to the stochastic differential equation (SDE) of the following form (Barndorff-Nielsen and Shephard, 2002a)

\[
dp^*(t) = \{ \mu + \beta \sigma^2(t) \} dt + \sigma(t) dW(t),
\]

where the mean process \( \{ \mu + \beta \sigma^2(t) \} \) and the instantaneous (spot) volatility \( \sigma(t) \) are assumed to be independent of the standard Brownian motion \( W(t) \). Here, \( \sigma^2(t) \) is called the instantaneous (spot) variance, \( \mu \) and \( \beta \) are given parameters. The formula in (30) and its solution

\[
p^*(t) = \mu^*(t) + \int_0^t \sigma(s) dW(s), \quad t \geq 0,
\]

build a general and very flexible stochastic volatility (SV) model, where

\[
\mu^*(t) = \mu t + \beta \sigma^{2*}(t)
\]

\[
= \mu t + \beta \int_0^t \sigma^2(s) ds.
\]

The term \( \sigma^{2*}(t) \), called the integrated variance (IV), is an object of interest which econometricians usually want to estimate. Within this framework, \( \sigma(\cdot) \) and \( \mu^*(\cdot) \) should obey some weak regularity conditions, as outlined in Barndorff-Nielsen and Shephard (2002b). Moreover, the spot volatility \( \sigma(\cdot) \) is assumed to be random and serially
dependent, which is particularly important for capturing properties of financial returns (e.g. Nielsen and Frederiksen, 2007). An important characteristic of the model defined by the two equations (30) and (31) is that

\[ p^* (t) | \mu^* (t), \sigma^{2*} (t) \sim N \{ \mu^* (t), \sigma^{2*} (t) \} . \]  

(33)

Consider now a financial return process \( (r_i) \) defined over an \( i \)th fixed interval of time having length \( h > 0 \), as follows

\[ r_i = p^* (ih) - p^* ((i - 1)h), i = 1, 2, \ldots, \]  

(34)

where \( p^* (ih) \) is usually treated as the observed (log) price at the end of the \( i \)th such period. Within this interval there might exist a collection of intra-\( h \) returns, defined for the \( i \)th period as

\[ r_{j,i} = p^* \left( (i - 1)h + \frac{hj}{M} \right) - p^* \left( (i - 1)h + \frac{h(j - 1)}{M} \right), j = 1, 2, \ldots, M, \]  

(35)

where \( M \) is the number of such intra-\( h \) returns over the \( i \)th interval. Variability of the financial \( i \)th return over this interval may be measured by the realized quadratic variation (RQV) defined as the sum of squared returns

\[ [Y_M]_i = \sum_{j=1}^{M} r_{j,i}^2, \quad i = 1, 2, \ldots, \quad j = 1, 2, \ldots, M, \]  

(36)

It is well known (see, e.g., Barndorff-Nielsen and Shephard, 2002b) that, as \( M \to \infty \), \([Y_M]_i\) is a consistent estimator of the corresponding unknown quadratic variation (QV) for all semimartingales (a class of stochastic processes that the SV models also belong to). This continuous QV, denoted as \([p^*, p^*] (t)\), for the process \( p^* (t) \), is defined as the limit in probability of the following form

\[ [p^*, p^*] (t) = \plim_{M \to \infty} \sum_{j=0}^{M-1} (p^*(t_{j+1}) - p^*(t_j))^2, \]  

(37)
where \( t_0 = 0 < t_1 < t_2 < \ldots < t_M = t \) and \( \sup_{1 \leq j \leq M} \{ t_j - t_{j-1} \} \rightarrow 0. \)

Furthermore, it has been shown (see, e.g., Barndorff-Nielsen and Shephard, 2002a) that \( [p^*, p^*](t) \) is measurable for all SV models. For these models \( [p^*, p^*](t) \) of the process \( p^*(t) \) equals the integrated variance:

\[
[p^*, p^*](t) = \sigma^{2*}(t) = \int_0^t \sigma^2(s) \, ds.
\] (38)

Then the actual variance \( \sigma_i^{[2]} \), which is crucial for SV models, defined over \( i \)th interval of length \( h \) is

\[
\sigma_i^{[2]} = \sigma^{2*}(ih) - \sigma^{2*}((i-1)h),
\] (39)

implying that

\[
\sigma_i^{[2]} = [p^*, p^*](ih) - [p^*, p^*]((i-1)h) = [p^*, p^*]_i.
\] (40)

As a consequence, the actual (e.g. daily) variance for the SV models is equal to the (daily) increments of the QV. Hence, the daily RQV consistently estimates the actual variance, as shown here

\[
[Y_M]_i \xrightarrow{P} \sigma_i^{[2]}, \quad \text{as } M \to \infty.
\] (41)

### B Parameter set up for simulation study

The functional form of \( \alpha(w) \) is chosen here by inspecting the time series of \( y_i^*(w) \) from the Swedish limit order book data for Ericsson B stock (see Elezović, 2008). The estimates of the autoregressive coefficients \( \hat{\gamma}_1 \) are obtained after fitting an ARMA(1,1) model to each \( y_i^*(w) \). These estimates are then converted to \( \alpha(w) \)-parameters using (20), and the polynomial model in (21) is fitted by the ordinary least squares procedure, obtaining the estimated parameter values for the coefficients \( \beta_0, \beta_1 \) and \( \beta_2 \), given in Table 1.

Figure 3 shows both the \( \alpha(w) \)-coefficients from the real data and the fitted \( \alpha(w) \)-coefficient curve. Note that the choice of functional
form for $\alpha(w)$ is not of particular importance here, hence any other plausible functional form might be used instead of (21).

The starting values of the remaining parameters in Table 1 are also chosen in such a way that resemblance with the real data is preserved as much as possible. For instance, the starting value for the stock price process $S_{t_0}$ in (3a) is chosen as the first observation of the observed time series of functional spreads (differences between the bid and ask prices) for $w = 50\%$, and correspondingly the starting value $V_{t_0}$ is chosen as the first observation of the $y^*(w = 50\%)$ series.

Values of the parameters $\theta$ and $\sigma$ are also computed using the
same observed time series and utilizing the expressions for the two first marginal moments for the spot variance (1b), when the limiting case \( t = +\infty \) is considered (Gourieroux and Jasiak, 2001), as shown here

\[
E(V_t) \approx \theta \tag{42a}
\]
\[
\text{Var}(V_t) \approx \frac{\theta \sigma^2}{2\alpha}. \tag{42b}
\]

The correlation coefficient \( \rho \) is arbitrarily set to \(-0.5\), following the recommendations in the empirical studies that usually assume a negative correlation to account for the opposite movements of asset returns and most measures of volatilities.

It is worth mentioning that, in some works, different ways of extracting necessary parameters for the Heston model from the real data, (see, e.g. Drăgulescu and Yakovenko, 2002; Daniel et al., 2005; Fioren- tini et al., 2002; Mariani et al., 2007; Aït-Sahalia and Kimmel, 2007). Most of the proposed methods in the literature utilize the closed form solution for the option prices given in Heston (1993) by some kind of calibration. This calibration usually implies optimization which minimizes the sum squared errors between the option prices from the traded options and the option prices obtained from the Heston model with different parameter sets.

Since the focus here is put on applications related to the limit order book data without any records of option prices, the calibration approaches are neglected here. In most related studies the unknown parameters are usually specified to make the simulated processes mimic the real data in some way (e.g. Andersen, 2007; Lord et al., 2006), which is roughly the approach considered here.