Exact Maximum Likelihood and Bayesian Estimation of the Stochastic Volatility Model

Anderson C. O. Motta

Luiz K. Hotta

Abstract

This paper considers the classical and Bayesian approaches to the estimation of the stochastic volatility (SV) model. The estimation procedures rely heavily on the fact that SV model can be written in the State Space Form (SSF) with non-Gaussian disturbances. The first widely employed estimation procedure to use this model was the quasi-maximum likelihood estimator proposed by Harvey et al. The Bayesian approach was proposed by Jacquier et al. (1994). Lately, many papers have appeared in the literature dealing with non-Gaussian state space models which directly influenced the estimation of the SV model. Some of these methods proposed to estimate the SV model are compared using the São Paulo stock exchange index (IBOVESPA) and simulated series. The influence of outliers is also considered.

Resumo

Este artigo considera a estimação clássica e a abordagem Bayesiana do modelo de Volatilidade Estocástica. Os estimadores apresentados utilizam fortemente o fato de que este modelo pode ser colocado na forma de espaço de estados com perturbações não-Gaussianas. Harvey e colaboradores utilizaram esta possibilidade de poder representar o modelo na forma de espaço de estados para sugerir o estimador de máxima quase verossimilhança. Posteriormente Jacquier et al. (1994) propuseram uma estimação com abordagem Bayesiana. Recentemente, muitos artigos surgiram na literatura abordando modelos de espaço de estados não-Gaussianos que tiveram uma influência direta na estimação do modelo de volatilidade estocástica. Alguns dos métodos propostos são comparados utilizando séries simuladas e o Índice da Bolsa de Valores do Estado de São Paulo (IBOVESPA). Também é considerada a influência de outliers.

*Universidade Estadual de Campinas, Brasil
1. Introduction.

This paper considers the classical and Bayesian approaches to the estimation of the stochastic volatility (SV) model. The estimation procedures rely heavily on the fact that SV model can be written in the State Space Form (SSF) with non-Gaussian disturbances. Kalman (1960) proposed a recursive algorithm for SSF models, known as Kalman Filter (hereinafter referred to as KF), which gives the estimate of the filtered and smoothed densities. When the disturbances are Gaussian we have the Gaussian SSF (GSSF) model. In this case, we can use the prediction error decomposition to evaluate the likelihood function (see, for instance, Harvey (1989), pg.125).

Since the stochastic volatility model introduced by Taylor (1986), after a transformation, can be written in the SSF, Ruiz (1994) and Harvey et al. (1994) suggested to estimate the SV model by the quasi-maximum likelihood method.

On the other hand, Jacquier et al. (1994) proposed a Bayesian approach for the SV model which does not use the Gaussian approximation.

Lately there have been many developments in the literature dealing with nonlinear and non-Gaussian SSF (NGSSF) models that have been used in the classical maximum likelihood and Bayesian approaches. An important work using the classical approach is the paper written by Durbin and Koopman (1997a) (hereinafter called DK). On the Bayesian approach, Shephard and Pitt (1997) (hereinafter called SP) proposed an efficient algorithm writing the SV model in the SSF, which was later improved by Kim et al. (1998).

The main goal of this work is to present and compare the methods proposed in the classical and Bayesian approaches in a unified way. Since the SV model is mainly applied to financial time series, prone to outliers, we also analyze the effect of outliers on SV model estimators.

The structure of the paper is as follows. Initially we present the linear SSF models under the classical and Bayesian approaches in Section two and the SV model in Section three. Section four and five present, respectively, the classical and Bayesian treatment of the SV model. Section six applies the methodology to simulated and empirical series. The simulated series contain the two most common type of outliers, the additive outlier (AO) and the innovational outlier (IO). The empirical series is the IBOVESPA index series from Jan/02/1995 to Dec/27/2000 with a total of 1, 500 observations. Finally, Section 6 concludes.

2. Linear State Space Form Model.

A linear GSSD can be written as:

\[
\begin{align*}
\text{Measurement Equation:} & \quad y_t = Z_t \alpha_t + c_t + G_t \epsilon_t \\
\text{Transition Equation:} & \quad \alpha_t = T_t \alpha_{t-1} + d_t + R_t \eta_t,
\end{align*}
\]

where \(\{ \epsilon_t \}\) and \(\{ \eta_t \}\) are independent Gaussian white noise processes with zero mean and covariance matrices equal to \(H_t\) and \(Q_t\), respectively. We can incorporate the effect of covariables through the vectors \(c_t\) and \(d_t\). We also generally suppose that the initial values
$E(\alpha_0) = a_0$ and $Var(\alpha_0) = \Sigma_0$ are known and that the disturbances $\epsilon_t$ and $\eta_t$ are not correlated with $\alpha_0$ for $t = 1, \ldots, T$.

This formulation is especially important because the KF recursively yields the optimal estimate of the states $h_t$, based on the information up to time $t_1$, and the corresponding error covariance matrix (see, for instance, Harvey, 1989). Using the prediction error decomposition, the KF also gives the likelihood. Further, de Jong and Shephard (1995) developed the simulation smoother algorithm to sample the states conditionally to the observations.

2.1 Non-Gaussian State Space Form Model.

The non-Gaussian state space form (NGSSF) model can be written as:

\[ y_t \sim f(y_t|\theta_t), \quad \theta_t = Z_t \alpha_t + c_t \]  \hspace{1cm} (3)
\[ \alpha_t = T_t \alpha_{t-1} + d_t + H_t \eta_t, \quad \eta_t \sim NID(0, I) \]  \hspace{1cm} (4)

where $f(y_t|\theta_t)$ is a known density, $\theta_t = Z_t \alpha_t + c_t$ is the signal at time $t$, which is related to the state. In some works, as in DK and SP, it is assumed that $f(y_t|\theta_t)$ is log-concave in $\theta_t$. This is true for the SV model.

Some authors, such as Mariano et al., formulated the model as:

\[ y_t = f_t(\alpha_t, \epsilon_t) \]  \hspace{1cm} (5)
\[ \alpha_t = h_t(\alpha_{t-1}, \eta_t), \]  \hspace{1cm} (6)

where $\{\epsilon_t\}$ and $\{\eta_t\}$ are independent white noise processes, not necessarily Gaussian, with zero mean and covariance matrices equal to
Anderson C. O. Motta and Luiz K. Hotta

$H_t$ and $Q_t$, respectively. We assume that the functional form of $f_t$ and $h_t$ are known and they are invertible, i.e.,

\[
y_t = f_t(\alpha_t, \varepsilon_t) \Rightarrow q_t(y_t, \alpha_t) = \varepsilon_t,
\]
\[
\alpha_t = h_t(\alpha_{t-1}, \eta_t) \Rightarrow g_t(\alpha_t, \alpha_{t-1}) = \eta_t.
\]

2.1.1 Estimation, Filtering and Smoothing.

For the NGSSF models, the analytical form of the conditional densities are not known, but they can be evaluated as:

**Prediction Equation:**

Denoting $Y_t = (y_1, \ldots, y_t)$ we have

\[
p(\alpha_t|Y_{t-1}) = \int p(\alpha_t|\alpha_{t-1})p(\alpha_{t-1}|Y_{t-1}) d\alpha_{t-1}
\]

with the joint density of $\alpha_t, \ldots, \alpha_1, y_{t-1}, \ldots, y_1$ given by:

\[
p(\alpha_t, \ldots, \alpha_1, Y_{t-1}) = p(\alpha_t|\alpha_{t-1}) \cdots p(\alpha_1|\alpha_0)p(y_{t-1}|\alpha_{t-1}) \cdots p(y_1|\alpha_1),
\]

where

\[
p(\alpha_t, \alpha_{t-1}, Y_{t-1}) = \int p(\alpha_{t-1}, Y_{t-2})p(y_{t-1}|\alpha_{t-1})p(\alpha_t|\alpha_{t-1})dY_{t-2},
\]

and the joint density $p(\alpha_{t-1}, Y_{t-2})$ is given by:

\[
p(\alpha_{t-1}, Y_{t-2}) = \int \cdots \int p(\alpha_t|\alpha_{t-1}) \cdots p(\alpha_1|\alpha_0)p(y_{t-1}|\alpha_{t-1})
\]
\[
\cdots p(y_1|\alpha_1) \, d\alpha_{t-2}, \ldots, d\alpha_1.
\]
Since the marginal density of $Y_{t-1}$ is given by:

$$p(Y_{t-1}) = \int \int p(\alpha_t|\alpha_{t-1})p(y_{t-1}|\alpha_{t-1})p(\alpha_{t-1}, Y_{t-2}) \, d\alpha_t \, d\alpha_{t-1}$$

$$= \int p(y_{t-1}|\alpha_{t-1})p(\alpha_{t-1}, Y_{t-2}) \, d\alpha_{t-1},$$

the conditional density is given by:

$$p(\alpha_t, \alpha_{t-1}|Y_{t-1}) = p(\alpha_t|\alpha_{t-1})p(\alpha_{t-1}|Y_{t-1})$$

and the prediction equation is given by:

$$p(\alpha_t|Y_{t-1}) = \int p(\alpha_t|\alpha_{t-1})p(\alpha_{t-1}|Y_{t-1}) \, d\alpha_{t-1}. \quad (8)$$

**Updating Equation:**

$$p(\alpha_t|Y_t) = \frac{p(y_t|\alpha_t)p(\alpha_t|Y_{t-1})}{p(y_t|Y_{t-1})}, \quad (9)$$

$$p(y_t|Y_{t-1}) = \int p(y_t|\alpha_t)p(\alpha_t|Y_{t-1}) \, d\alpha_t. \quad (10)$$

In the prediction and the updating equations it is necessary to evaluate an integral which can be seen as an expectation and which can be evaluated by Monte Carlo simulation. This can be done by the particle filter, which was proposed by Carpenter et al. (1998), although the term had already been used before by Kitagawa (1996). Particle filter is a class of filter which approximates recursively by simulation the distribution of a random variable by
particles $\alpha_t^1, \ldots, \alpha_t^M$, with discrete probability density $\pi_t^1, \ldots, \pi_t^M$. The discrete observations are considered a random sample from the population and the continuous random variable is approximated by a discrete random variable with random support. Most of the researchers take $\pi_t^j$ as equal to $1/M$, with $M$ being sufficiently large.

The particle filter considers the discrete support generated by the particles as the true density. This allows us to generate an approximation for this density. For instance, for the prediction density we have:

$$\hat{p}(\alpha_t|Y_{t-1}) = \sum_{j=1}^{M} p(\alpha_t|\alpha_{t-1}^j)\pi_{t-1}^j,$$

an “empirical prediction density” as an estimate of the desired density. Similarly, the estimate of the filtering density of equation (9) is obtained as:

$$\hat{p}(\alpha_t|Y_t) \propto p(y_t|\alpha_t) \sum_{j=1}^{M} p(\alpha_t|\alpha_{t-1}^j)\pi_{t-1}^j$$

an “empirical filtering density”. Thus, the particle filter samples from this density in order to obtain new particles $\alpha_{t+1}^1, \ldots, \alpha_{t+1}^M$ with weights $\pi_{t+1}^1, \ldots, \pi_{t+1}^M$. One way of sampling from the filtered density is to consider the function $\sum_{j=1}^{M} f(\alpha_t|\alpha_{t-1}^j)\pi_{t-1}^j$ as the prior density (11) which, combined with the likelihood $\hat{p}(y_t|\alpha_t)$, gives the posterior (12). We can sample from (11) taking $\alpha_{t-1}^j$ with probability $\pi_{t-1}^j$ and sampling from $p(\alpha_t|\alpha_{t-1}^j)$. We can obtain $p(y_t|\alpha_t)$ except for a constant, and we can use traditional methods as Monte Carlo-Markov Chain (MCMC), importance sampling/resampling (S-R) in order to sample from this distribution.
Pitt and Shephard (1999) proposed to robustify the particle filter method using an auxiliary variable. The idea is to sample from the joint density $p(\alpha_t, k|Y_t)$ where $k$ is the mixture index in equation (11). Thus, they defined

$$p(\alpha_t, k|Y_{t-1}) = p(y_t|\alpha_t)p(\alpha_t|\alpha_{t-1}^{k})\pi_{t-1}^{k}, \quad k = 1, \ldots, M. \quad (13)$$

Sampling from this joint density and not considering the index we can obtain the sample from the empirical filtering density (12). The sample from the joint density can be done by traditional methods as before (see Pitt and Shephard (1999) for details). Kim et al. (1998) also used this technique in the analyses of the SV model.

The smoothing algorithm is given by:

$$p(\alpha_t|Y_T) = p(\alpha_t|Y_t) \int \frac{p(\alpha_{t+1}|Y_T)p(\alpha_{t+1}|\alpha_t)}{p(\alpha_{t+1}|Y_t)} d\alpha_{t+1}, \quad (14)$$

where $p(\alpha_t|Y_t)$ and $p(\alpha_{t+1}|Y_t)$ are obtained from the filtering algorithm given by equation (9), while $p(\alpha_{t+1}|\alpha_t)$ is obtained from the transition equation (6).

The likelihood can be evaluated as:

$$p(Y_T) = \prod_{t=1}^{T} p(y_t|Y_{t-1}) = \prod_{t=1}^{T} \left( \int p(y_t|\alpha_t)p(\alpha_t|Y_{t-1}) d\alpha_t \right).$$

A process $y_t$ follows an SV model when it is generated by the following equations (see, for instance, Taylor, 1994 and Ghysels et al., 1996):

$$y_t = e^{\frac{1}{2}h_t} \xi_t, \quad \xi_t \sim NID(0, 1) \quad (16)$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \quad \eta_t \sim NID(0, \sigma^2), \quad (17)$$

where $\xi_t$ and $\eta_t$ are independent. We consider that $h_t$ is stationary, i.e., $|\phi| < 1$.

Taking the logarithm of the square in equation (16) we have:

$$\log(y_t^2) = h_t + \log(\xi_t^2). \quad (18)$$

Since $\xi_t$ follows a standard Gaussian distribution it follows that $\log(\xi_t^2)$ has a log-chi-square distribution. It can be shown that $E[\log(\xi_t^2)] \approx -1.27$, and $V[\log(\xi_t^2)] = \frac{\pi^2}{2}$, and using the transformation $\xi_t = \log(\xi_t^2) - E[\log(\xi_t^2)] = \log(\xi_t^2) + 1.27$, equations (18) and (17) are rewritten as:

$$\log(y_t^2) = -1.27 + h_t + \xi_t, \quad \xi_t \sim ID(0, \frac{\pi^2}{2}), \quad (19)$$

$$h_t = \alpha + \phi h_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma^2)$$

with $\xi_t \perp \eta_t$. By definition of $\xi_t$ we have that $E[\xi_t] = 0$ and $Var[\xi_t] = \frac{\pi^2}{2}$.
3.1 State Space Form Formulation.

In this paper we also consider the effect of two types of outliers: the Additive Outlier (AO) and the Innovational Outlier (IO). The model with the presence of one AO in $t_{ao}$-th observation and one IO in the $t_{io}$-th in the SSF is given by:

\[
\log(y_t^2) = -1.27 + h_t + \xi_t + \Delta_t I_{ao} \tag{20}
\]

\[
h_t = \alpha + \phi h_{t-1} + \eta_t + \Lambda_t I_{io}, \tag{21}
\]

where $\Delta_t$ and $\Lambda_t$ are, respectively, the sizes of the AO and IO at the $t$-th observation. $I_{io}$ is an indicator function equal to one if $t = t_{io}$ and zero, otherwise $I_{ao}$ is defined similarly.


This section presents some numerical methods to evaluate the likelihood of the NGSSF models. The likelihood can be considered the basis for most of the techniques used in the classical inference approach. Section 4.1 presents the methodology developed by DK in a series of papers (Durbin and Koopman, 1992, 1997a, 1997b, 2000) and in a book (Durbin and Koopman, 2001). The method consists in estimating the likelihood through the correction of the quasi-likelihood. The quasi-likelihood and the corrections are evaluated using the KF. Control and antithetic variables can also be used in order to improve the simulation efficiency. Subsection 4.2 presents the method proposed by SP. Since we are interested in applications to the SV model we do not present a general method, for instance the methods developed by Tanizaki and Mariano (1994,1998) for NGSSF models.
4.1 Correction of the Quasi-Likelihood Using State Sampling.

Consider the model in the SSF given by equations (3) and (4). Denote by $p(y), p(\theta), p(y, \theta)$ and $p(y|\theta)$, respectively, the marginal densities of $y = (y_1, \cdots, y_T)$ and $\theta$, the joint densities of $(y, \theta)$ and the conditional distribution of $(y$ on $\theta)$.

Let $\Psi$ be the set of all parameters. Since the likelihood is given by:

$$L(\Psi) = p(y) = \int p(y|\theta)p(\theta) d\theta$$

it can be estimated by simulation, generating $N$ independent samples $\theta^{(1)}, \ldots, \theta^{(N)}$ from the density $p(\theta)$ and taking the mean value of the values $p(y|\theta^{(i)})$ for $i = 1, \ldots, N$.

DK proposed the importance sampling method to estimate this expectation. Denoting the importance density by $\tilde{p}(\theta|y)$, the likelihood is equal to:

$$L(\Psi) = \tilde{E} \left\{ p(y|\theta) \frac{p(\theta)}{\tilde{p}(\theta|y)} \right\}$$

and taking $N$ samples for $\theta^{(i)}$ from density $\tilde{p}(\theta|y)$ we can estimate $L(\Psi)$ by:

$$\hat{L}_1(\Psi) = \frac{1}{N} \sum_{i=1}^{N} p(y|\theta^{(i)}) \frac{p(\theta^{(i)})}{\tilde{p}(\theta^{(i)}|y)}.$$  \hspace{1cm} (22)

DK suggested to take $\tilde{p}(\theta|y)$ from a GSSF model. Denoting the likelihood associated to this model by $L_g(\Psi)$ and the true likelihood by $L(\Psi)$, and the respective densities by $p(\cdot)$ and $g(\cdot)$, we have:
Exact Maximum Likelihood and Bayesian Estimation of the Stochastic Volatility

\[
L(\Psi) = \int p(y, \theta) \, d\theta = \int p(y|\theta) p(\theta) \, d\theta \quad (23)
\]

\[
L_g(\Psi) = \frac{g(y, \theta)}{g(\theta|y)} = \frac{g(y|\theta)p(\theta)}{g(\theta|y)}. \quad (24)
\]

Substituting \( p(\theta) = L_g(\Psi) \frac{g(\theta|y)}{g(y|\theta)} \) into the first equation we have:

\[
L(\Psi) = \int p(y|\theta) \frac{L_g(\Psi) g(\theta|y)}{g(y|\theta)} \, d\theta = L_g(\Psi) \int \frac{p(y|\theta)}{g(y|\theta)} g(\theta|y) \, d\theta
\]

\[
= L_g(\Psi) E_g \left\{ \frac{p(y|\theta)}{g(y|\theta)} \right\} \quad (25)
\]

showing that we only have to be concerned with the correction since the approximate likelihood can be obtained using the KF. DK and SP suggested the use of \( g(\theta|y) \) as the importance density in the simulation. The sampling can be done using the de Jong and Shephard (1995) smoother algorithm.

4.1.1 Selection of the Approximating Model.

The choice of the approximating model is analogous to the choice of the mean \( \mu_t \) and the scale factor \( A_t \) of the disturbances \( \epsilon_t \). DK proposed to select a density which matched the mode and the curvature of the true density as close as possible. They took \( \hat{\theta} = E_g(\theta) \), where \( \theta \) is evaluated by the de Jong and Shephard (1995) algorithm considering the Gaussian model. Since the empirical distribution of \( \theta_t \), generated by the smoother algorithm is normally distributed around \( \hat{\theta}_t \) they selected \( \mu_t \) and \( A_t \) such that the density was close to \( \hat{\theta} \). DK suggested to take the first and second derivatives of the
logarithm of the ratio between the true and approximating model equal to zero. Since the log-ratio is equal to:

\[ l(\theta) = \frac{p(\theta|y)}{p(y|\theta)p(y)} \]

and \( g(y_t|\theta_t) \sim N(\mu_t + \theta_t, A_t) \) we have

\[ \frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t} \bigg|_{\theta_t=\hat{\theta}_t} - A_t^{-1}(y_t - \hat{\theta}_t - \mu_t) = 0 \] (27)

\[ \frac{\partial^2 \log p(y_t|\theta_t)}{\partial \theta \partial \theta'} \bigg|_{\theta_t=\hat{\theta}_t} + A_t^{-1} = 0. \] (28)

These equations are solved iteratively starting with \( \mu_t = 0 \) and \( A_t \) an arbitrary positive definite matrix. The system has a solution whenever \( A_t \) is a positive definite matrix, which happens in the SV model.

Since the density of the SV model is given by:

\[ \log f(y_t|\theta_t) = \frac{1}{2} \left( \frac{1}{2} \right)^2 \frac{1}{2} \frac{y_t^2}{2\sigma} \] (29)

we have \( A_t = 2\sigma^2 y_t^- \) positive definite. The solution can be obtained iteratively. At each iteration \( \theta \) is the smoothed solution of the GSSF model:

\[ \tilde{y}_t = \theta_t + \epsilon_t, \quad \epsilon_t \sim N(0, A_t), \quad \tilde{y}_t = y_t \] (30)

\[ \alpha_t = T_t\alpha_{t-1} + R_t\eta_t, \quad \eta_t \sim N(0, Q_t), \quad \theta_t = Z_t\alpha_t. \] (31)
4.1.2 Log-likelihood Estimation and Correction of the Bias.

Consider a sample $\theta^{(i)}$ obtained by the smoother algorithm and let $w_i = w(\theta^{(i)})$, for $i = 1, \ldots, N$, where $w(\theta) = p(y|\theta)[g(y|\theta)]^{-1}$.

The Monte Carlo estimation of $L(\Psi)$ based on the importance sampling is equal to:

$$\hat{L}_2(\Psi) = L_g(\Psi)\bar{w}, \quad (32)$$

where $\bar{w} = N^{-1} \sum_{i=1}^{N} w_i$. In general, we are interested in estimating the log-likelihood $L(\Psi)$. In order to correct the bias of log $\hat{L}_2(\Psi)$ we take:

$$\log \bar{w} - \log \mu_w = \log \left( 1 + \frac{\bar{w} - \mu_w}{\mu_w} \right) = \frac{\bar{w} - \mu_w}{\mu_w} - \frac{1}{2} \left( \frac{\bar{w} - \mu_w}{\mu_w} \right)^2$$

$$+ O_p(N^{-\frac{3}{2}}).$$

Therefore,

$$E_g(\log \bar{w}) = \log \mu_w + \frac{\sigma^2_w}{2N\mu^2_w} + O(N^{-\frac{3}{2}}), \quad (33)$$

where $\sigma^2_w = \text{var}(w_i)$. Substituting $\mu_w$ for $\bar{w}$ and $\sigma^2_w$ for $s^2_w$, where $s^2_w = (N-1)^{-1} \sum_{i=1}^{N} (w_i - \bar{w})^2$ we have that an approximate unbiased estimator for log $L(\Psi)$ is given by:

$$\log \hat{L}_3(\Psi) = \log L_g(\Psi) + \log \bar{w} + \frac{s^2_w}{2N\bar{w}^2}. \quad (34)$$
Antithetic Variables

Let \( \theta \) be a sample from the density \( g(\theta|y) \) using the smoother algorithm with \( \mu_t \) and \( A_t \) selected as before. The first antithetic variable, balanced by location, is given by:

\[
\tilde{\theta}_t = 2\hat{\theta}_t - \theta_t, \quad t = 1, \ldots, n
\]

which is negatively correlated with \( \theta_t \).

For the second antithetic variable assume that \( y_t \) univariate random variables. The sampled value \( \theta \) is given by a nonlinear nonsingular transformation of \( u \), with \( \theta = \hat{\theta} + Bu \), where the vector \( u \) comes from a standard \( N(0, I_n) \) and \( \text{var}(\theta) = V = BB' \). Define

\[
c = (\theta - \hat{\theta})'V^{-1}(\theta - \hat{\theta}) = u'u \sim \chi^2_n. \tag{35}
\]

For a given \( c \), let \( q = Pr(\chi^2_n < c) = F(c) \), and \( \bar{c} = F^{-1}(1 - q) \). Therefore \( \bar{c} \sim \chi^2_{(n)} \), and defining \( \bar{\theta} = \hat{\theta} + (\bar{c}/c)^{1/2} (\theta - \hat{\theta}) \) then \( \bar{\theta} \) has the same distribution as \( \theta \). Besides, when \( |\theta_t - \hat{\theta}_t| \) is large, \( |\bar{\theta}_t - \hat{\theta}_t| \) is small and vice versa. The sample pair \( \theta \) and \( \bar{\theta} \) with these properties are called antithetic variables balanced by scale.

DK noticed that the decrease in the variance of the estimator of \( \log L(y) \) with the use of these antithetic variables is always larger than with the use of antithetic variables balanced by location. Using the four antithetic variables we have four values for \( w(\theta) \) for each sample value generated by the smoother algorithm. Let \( \bar{\theta}_{(i)} \) be a variable balanced by scale which is antithetic for \( \bar{\theta}^{(i)} \) and let \( \bar{w}_{(i)} = w(\bar{\theta}^{(i)}) \). Then,

\[
w^* = \frac{1}{4}(w_i + \bar{w}_i + \bar{w}_i + \bar{w}_i) \tag{36}
\]
Exact Maximum Likelihood and Bayesian Estimation of the Stochastic Volatility

and we can estimate \( \log L(\Psi) \) by \( \log L_4(\Psi) \) which is given by:

\[
\log \hat{L}(\psi) = \log L_9(\psi) + \log \bar{w} + \frac{s_w^2}{2N(\bar{w})^2}. \tag{37}
\]

Control Variables

DK used the Taylor approximation of \( w(\theta) \) for \( \theta \) close to \( \hat{\theta} \). Using the notation:

\[
\begin{align*}
    l &= l(\theta), \quad l'_t = \frac{\partial l(\theta)}{\partial \theta_t}, \quad l''_t = \frac{\partial^2 l(\theta)}{\partial \theta_t^2}, \ldots \\
    w &= w(\theta), \quad w'_t = \frac{\partial w(\theta)}{\partial \theta_t}, \quad w''_t = \frac{\partial^2 w(\theta)}{\partial \theta_t^2}, \ldots
\end{align*}
\]

and \( w|_{\hat{\theta}} = \hat{w}, \quad l''_t|_{\hat{\theta}} = \hat{l''_t} \) and \( l''''_t|_{\hat{\theta}} = \hat{l''''_t} \), and considering that \( \mu_t \) and \( A_t \) are such that the first and second derivatives are zero, we have the following Taylor expansion:

\[
w(\theta) = \hat{w} \left\{ 1 + \frac{1}{6} \sum_{t=1}^{n} \hat{l''_t}(\theta_t - \hat{\theta}_t)^3 + \frac{1}{24} \sum_{t=1}^{n} \hat{l''''_t}(\theta_t - \hat{\theta}_t)^4 + \ldots \right\}. \tag{38}
\]

DK considered \( \hat{w}(\bar{\alpha}) \) as the control variable to estimate \( \log L(\Psi) \), where \( \bar{\alpha} = N^{-1} \sum_{i=1}^{N} a_i \) and \( a_i \) is given by:

\[
a_i = \frac{1}{6} \sum_{t=1}^{n} \hat{l''_t}((\theta_t^{(i)} - \hat{\theta}_t)^3 + \frac{1}{24} \sum_{t=1}^{n} \hat{l''''_t}((\theta_t^{(i)} - \hat{\theta}_t)^4. \tag{39}
\]

The first term vanishes as a result of the antithetic variables being balanced by location. Let \( \nu_t = \text{var}(\varepsilon_t|y) \).
which can be obtained by the smoother algorithm. Since $\theta_t = y_t$
$\text{var}(\theta_t | y_t)$

$$E_g(a_i) = \frac{1}{8} \sum_{t=1}^{n} \hat{l}'''' t^2,$$

where $a_i$ is obtained taking the mean of the four values of

$$\frac{1}{24} \sum_{t=1}^{n} \hat{l}''''(\theta_t^{(i)} - \hat{\theta})$$

with $\theta_t^{(i)} = \theta_t^{(1)}, \hat{\theta}_t^{(1)}, \theta_t^{(2)}, \hat{\theta}_t^{(2)}$

$$w_t^i = w_i - \hat{w} \left( a_i - \frac{1}{8} \sum_{t=1}^{n} \hat{l}'''' v_t^2 \right). \quad (40)$$

Thus, the estimate of $\log L(\Psi)$ using the control variable $a_i$ is

$log \hat{L}_5(\Psi)$ given by (34), with $w_t^i$ in place of $w_i$. From (38) and (39)
we have

$$E_g\{w(\theta)\} \approx \hat{w}\{1 + E_g(a_i)\} \quad (41)$$

Therefore the approximated log-likelihood is given by:

$$\log \hat{L}_0 = \log L_g + \log \hat{w} + \log \left( 1 + \frac{1}{8} \sum_{t=1}^{n} \hat{l}'''' v_t^2 \right). \quad (42)$$

4.1.3 Algorithm to Estimate the Likelihood.

The estimation algorithm using the antithetic and control variables can be implemented by the following steps:

1. Identify the approximated Gaussian density $g(y|\theta)$ for $p(y|\theta)$, based on the fact that $y_t$
solutions of equations (27) and (28);
2. Evaluate $\log L_g(\Psi)$ using the Kalman filter;

3. Take $g(y|\theta)$ as the importance density and sample $\theta^{(i)}$ from $g(\theta|y)$ using the de Jong and Shephard (1995) smoother algorithm for $i = 1, \ldots, N$;

4. For each $i$ evaluate $w_i^\dagger$ by (40) and evaluate $\log \hat{L}_g(\Psi)$ by equation (34) substituting $w_i$ for $w_i^\dagger$.

5. Bayesian Analysis of the Stochastic Volatility Model.

The Bayesian approach to the SV model used in this work is based on the pioneering paper by Jacquier et al. (1994) and on the papers by SP and Kim et al. (1998).

Consider the SV model given by:

\begin{align*}
y_t &= h_t^{1/2} \varepsilon_t, \quad t \geq 1, \tag{43} \\
\log h_t &= \alpha + \phi \log h_{t-1} + \sigma_\eta \eta_t, \tag{44}
\end{align*}

with prior distributions of the parameters $(\alpha, \phi, \sigma_\eta^2)$. Let $h = (h_1, \ldots, h_T)$, be the $T$-dimensional vector of the volatility and $\theta = (\alpha, \phi, \sigma_\eta^2)$, the volatility parameters. Observe that now $h$ denotes the conditional variance, while before, it denoted the logarithm of the conditional variance.

In the MCMC algorithm we alternatively sample the latent vector $h_t$ and the vector of coefficients $\theta$. Subsection 5.1 presents two algorithms to sample the latent vector. The first algorithm presented in Subsection 5.1.1, called the simple sampling, samples each component of the latent vector consecutively while the second algorithm presented in subsection 5.1.2 samples the latent vector jointly using Kalman filter smoother algorithm. In order to apply this algorithm
the non-Gaussian distribution is approximated by a mixture of Gaussian densities. The first algorithm was proposed by Jacquier et al. (1994) while the second one was proposed by Kim et al. (1998). The sampling and the priors of the coefficients are given in Subsection 5.2

5.1 Linear State Space Form Model.

In this subsection we show how to estimate the conditional densities which will be used later. The joint density of $h, \theta$ conditioned to $y$, the marginal densities of $h$ and $\theta$ conditioned on $y$ are given by:

$$p(h, \theta | y) = \frac{p(y|h, \theta)p(h|\theta)p(\theta)}{p(y)}$$  \hspace{1cm} (45)

$$p(y) = \int \int p(y|h, \theta)p(h|\theta)p(\theta) \, d\theta \, dh$$  \hspace{1cm} (46)

$$p(\theta | y) = \int p(h, \theta | y) \, dh.$$  \hspace{1cm} (47)

Since there is not an analytical function for $p(h, \theta | y)$, Jacquier et al. (1994) augmented the observed values $y$ by the latent variables $h$, as proposed formerly by Carlin et al. (1992) for the NGSSF model. The former densities are now given by:

$$p(h|y) = \int p(h, \theta | y) \, d\theta = \int p(h|\theta, y)p(\theta|y) \, d\theta$$  \hspace{1cm} (48)

$$p(h, \theta | y) = p(\theta|h, y)p(h|y)$$  \hspace{1cm} (49)

$$p(\theta | y) = \int p(h, \theta | y) \, dh = \int p(\theta|h, y)p(h|y) \, dh$$  \hspace{1cm} (50)

$$p(h, \theta | y) = p(h|\theta, y)p(\theta|y).$$  \hspace{1cm} (51)
Therefore, we end up with a recursive problem; it is necessary to have the density of $h|y$ to estimate $p(\theta|y)$ and vice versa.

### 5.1.1 Simple Sampling from the Latent Variable.

We cannot use MCMC directly in the SV model because we do not know the analytical form of the density $p(y|\theta) = \int p(y|h, \theta)p(h|\theta) dh$. We can overcome this problem by working with the density $p(\theta, h|y)$ and using the traditional MCMC sample from the joint posterior density and evaluate the quantities of interest without the likelihood. The algorithm can be implemented by the following recursive steps:

1. Given the initial vectors $h^{(i)}$ and $\theta^{(i)}$, sample $h^{(i+1)}$ from the density $p(h|\theta^{(i)}, y)$;
2. Given $h^{(i+1)}$, sample $\theta^{(i+1)}$ from $p(\theta|h^{(i+1)}, y)$

We need the analytical form of $p(h|\theta^{(i)}, y)$ and $p(\theta|h^{(i+1)}, y)$ in order to sample from these densities. Consider $p(\theta|h^{(i+1)}, y)$ as:

$$p(\theta|h, y) = \frac{p(\theta, h, y)}{\int p(\theta, h, y) d\theta} = p(y|\theta, h)p(h|\theta)p(\theta).$$ \hspace{1cm} (52)

Since $\epsilon_t \sim N(0, 1)$ and $\eta_t \sim N(0, \sigma^2_\eta)$ it follows that $y_t | E(\log h_t|\theta, \Psi_t) = \alpha + \phi \log h_{t-1}$ and $V(\log h_t|\theta, \Psi_t) = \sigma^2_\eta$. Also

$$p(y|h) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi h_t}} \exp \left(-\frac{y_t^2}{2h_t}\right), \text{ and}$$ \hspace{1cm} (53)

$$p(\log h_t|\theta, \Psi_t) = \frac{1}{\sqrt{2\pi\sigma^2_\eta}} \exp\left(-\frac{(\log h_t - \alpha - \phi \log h_{t-1})^2}{2\sigma^2_\eta}\right),$$ \hspace{1cm} (54)

where $\Psi_t = (h_t, h_{t-1}, \ldots)$. Therefore the density $p(h_t|\theta, \Psi_t)$ is given by:
Thus, \( h_t \) has log-normal density and the joint density is given by:

\[
p(h_t|\theta, \Psi_t) = \left( \frac{\partial \log h_t}{\partial h_t} \right) p(\log h_t|\theta, \Psi_t) = \frac{1}{h_t} p(\log h_t|\theta, \Psi_t). \quad (55)
\]

Since \( h_0 \) is not known and \( \log h_1|\theta \sim N\left(\frac{\alpha}{1-\phi}, \frac{\sigma_\eta^2}{1-\phi^2}\right) \), the density of \( h_1|\theta \) is the nonconditional distribution given by:

\[
p(h_1|\theta) = p(h_1|\theta) \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma_\eta}} \exp\left( -\frac{(\log h_t - \alpha - \phi \log h_{t-1})^2}{2\sigma_\eta^2} \right). \quad (56)
\]

Given \( \theta^{(i)} \) the sampling from the multivariate \( p(h^{(i+1)}|\theta, y) \) distribution can be done by sampling \( h^{(i+1)}_t \) from \( h_t|h^{(i)} \), \( y, \theta^{(i)} \) for \( t \) from 1 to \( T \), where \( h_{<t} = (h_1, \ldots, h_{t-1}, h_{t+1}, \ldots, h_T) \).

The functional form of the density \( p(h_t|h_{<t}, \theta, y) \) is evaluated through:
Exact Maximum Likelihood and Bayesian Estimation of the Stochastic Volatility

\[ p(h_t|\mathbf{h}_{-t}, \theta, y) = \frac{p(h, \theta, y)}{\int p(h, \theta, y) \, dh_t} = \frac{p(y|h)p(h|\theta)p(\theta)}{\int p(y|h)p(h|\theta)p(\theta) \, dh_t} \]

\[ \propto p(y|h)p(h|\theta) \propto p(y_t|h_t)p(h_t|h_{t-1}, \theta)p(h_{t+1}|h_t, \theta) \]

\[ \propto \frac{1}{\sqrt{h_t}} \exp\left(-\frac{y_t^2}{2h_t}\right) \frac{1}{h_t} \exp\left(-\left(\log h_t - \alpha - \phi \log h_{t-1}\right)^2 - \left(\log h_{t+1} - \alpha - \phi \log h_t\right)^2\right) \]

\[ \propto \frac{1}{h_t^{1.5}} \exp\left(-\frac{y_t^2}{2h_t}\right) \exp\left(-\frac{1}{2\sigma^2_{\eta}} \left(\log h_t - \alpha - \phi \log h_{t-1}\right)^2 + \left(\log h_{t+1} - \alpha - \phi \log h_t\right)^2\right), \]

where

\[ * = -\frac{1 + \phi^2}{2\sigma^2_{\eta}} \left((\log h_t - \gamma_t)^2 + cte - \gamma_t^2\right) \propto -\frac{1}{2\sigma^2} (\log h_t - \gamma_t)^2, \]

\[ \gamma_t = \frac{\alpha(1 - \phi) + \phi(\log h_{t+1} + \log h_{t-1})}{1 + \phi^2} \quad \text{and} \quad \sigma^2 = \frac{\sigma^2_{\eta}}{1 + \phi^2}. \]

Therefore the analytical form of \( p(h_t|\mathbf{h}_{-t}, \theta, y) \) is given by:

\[ p(h_t|\mathbf{h}_{-t}, \theta, y) \propto \frac{1}{h_t^{1.5}} \exp\left(-\frac{y_t^2}{2h_t}\right) \exp\left(-\frac{(\log h_t - \gamma_t)^2}{2\sigma^2}\right). \quad (57) \]
Jacquier et al. (1994) suggested using Metropolis-Hastings’ (hereinafter called M-H) algorithm to get the desired sample. Let \( X_i = h_t(i) \), where \( t = 1, \ldots, T \). The M-H algorithm proposed by Jacquier et al. (1994) to sample from \( p(\theta|y) \) and \( p(h|y) \) is given as follows:

1. Given \( h(i), \theta(i) \), generate candidates for \( h_t(i+1) \) sampling from \( q(h_t(i+1)|h_{<t}, \theta, y) \);

2. (i) The candidate \( X_{i+1} = h_t(i+1) \) is accepted with probability

\[
\min \left\{ \frac{p(h_t(i+1)|h_{<t}, \theta(i), y)}{cq(h_t(i+1)|h_{<t}, \theta(i), y)}, 1 \right\}
\]

and go to (3);

(ii) Otherwise reject \( h_t(i+1) \), put \( X_{i+1} = h_t(i) \) and stop.

3. (i) Accept the candidate \( X_{i+1} = h_t(i+1) \) with probability

\[
\alpha(h_t(i), h_t(i+1)) = \min \left\{ \frac{\pi(h_t(i+1))\min\{\pi(h_t(i), 1)\}}{\pi(h_t(i))\min\{\pi(h_t(i+1), 1)\}}, 1 \right\}
\]

and stop; (ii) Otherwise reject \( h_t(i+1) \), put \( X_{i+1} = h_t(i) \) and stop.

4. Given \( p(h|y) \), sample \( \theta \) from \( \theta|h, y \);

5. Return to (2).

It is necessary to know the functional forms of the densities \( q(h_t|h_{<t}, \theta, y) \) and \( p(\theta|h, y) \). The density \( p(\theta|h, y) \) was already discussed. Consider the functional form of \( p(h_t|h_{<t}, \theta, y) \) given by equation (57). The second term is a log-normal density which was approximated by Jacquier et al. (1994) by an inverse Gamma distribution equating the first two moments of the log-normal and the
Exact Maximum Likelihood and Bayesian Estimation of the Stochastic Volatility

inverse gamma distributions. The parameters of the inverse gamma distribution are given by:

\[ \lambda = (\nu - 1) \exp(\gamma_t + \frac{1}{2}\sigma^2) \quad \text{and} \quad \nu = \frac{1 - 2\exp(\sigma^2)}{1 - \exp(\sigma^2)}. \]

Therefore we approximate \( q(h_t|h_{<t}, \theta, y) \) by the density:

\[ \frac{1}{\sqrt{h_t}} \exp\left(\frac{y_t^2}{2h_t}\right) \frac{1}{h_t^{(\nu+1)}} \exp\left(-\frac{\lambda}{h_t}\right) \]

which gives the parameters of the inverse gamma posterior distribution:

\[ \nu_1 = \frac{1 - 2\exp(\sigma^2)}{1 - \exp(\sigma^2)} + \frac{3}{2} \quad \text{and} \quad \lambda_1 = (\nu - 1) \exp(\gamma_t + \frac{1}{2}\sigma^2) + \frac{1}{2}y_t^2, \]

where \( \gamma_t \) and \( \sigma^2 \) are given by equation (57).

5.1.2 Joint Sampling of the Latent Variable.

This subsection presents some aspects of the joint sampling. The SV model is represented by equations (16) and (17).

Approximation by mixture

Kim et al. (1998) used a mixture of normal distributions in order to approximate the exact likelihood. With this approximation we can implement the adapted Gibbs sampling, which allows us to sample all the volatilities at once. Since the disturbance in the measurement equation is non-Gaussian, it is approximated by a mixture of normal distributions given by
\[ f(z_t) = \sum_{i=1}^{K} q_i f_N(z_t|m_i - 1.2704, v_i^2), \] (59)

where \( f_N(z_t|m_i - 1.2704, v_i^2) \) denotes the density of a normal distribution of mean \( m_i - 1.2704 \) and variance \( v_i^2 \) at \( z_t \).

Using an indicator random variable \( s_t \) in order to represent the mixture we have:

\[ z_t|s_t = i \sim N(m_i - 1.2704, v_i^2), \quad \text{Pr}(s_t = i) = q_i. \] (60)

Kim et al. (1998) following a suggestion by Titterington et al. (1985, p. 133) found that with \( K = 7 \) the first four moments of a \( \log \chi^2 \) distribution could be approximated closely.

The mixture model can be better sampled in the MCMC approach using the representation (60). The posterior density of interest is \( \pi(s, h, \phi, \sigma^2, \mu|y^*) \), where \( s = (s_1, \ldots, s_n) \). In this case \( h \) and \( s \) can be sampled separately in one block.

1. Initialize \( s, \phi, \sigma^2 \) and \( \mu \);
2. Sample \( h \) from \( h|y^*, s, \phi, \sigma^2, \mu \);
3. Sample \( s \) from \( s|y^*, h \);
4. Update \( \phi, \sigma^2, \mu \) as will be shown in section 5.2 of sampling of parameters;
5. Return to (2).

Now we can sample from the multivariate Gaussian distribution \( h|y^*, s, \phi, \sigma_\eta, \mu \), because \( y^*|s, \phi, \sigma_\eta, \mu \) can be written as a GSSF, and we can use the KF to sample from \( h|y^*, s, \sigma_\eta, \mu \) using the smoother algorithm. The sample of \( s \) from \( s|y^*, h \) is done by sampling independently each \( s_t \) from the following probability function:
Exact Maximum Likelihood and Bayesian Estimation of the Stochastic Volatility

\[ f(s) \propto q_i f_N(y_t^*) \]

Since \( y \)

then \( y \)

**Integrating the log-volatilities**

The previous algorithm significantly improves the behavior of the correlations in the simulation. Kim *et al.* (1998) suggested sampling from the joint density \( \pi(\phi) \)

from \( \pi(\phi) \)

from \( \pi(h, N(\tilde{\mu}, \sigma^2_\eta)) \)

\[ \pi(\phi, \)

\[ \frac{f(y^*)}{\pi(\mu)} \]

\[ \propto \pi(\phi) \left( \prod_{t=1}^{n} -1/2 \right) \sum_{t=1}^{n} \frac{v^2_t}{F_t} \exp \left( \frac{1}{2\sigma^2} -2 \tilde{\mu}^2 \right) \]

\[ (61) \]

where \( v_t \) is the one-step-ahead prediction error and \( F_t \) its mean square error. The values \( v_t, F_t, \tilde{\mu}, \sigma^2_\mu \)

mentioned on \( s \).

This implies that given \( \{ \phi' \)

\( \phi, \)

sity \( g(\phi), \)

\[ \min \left\{ 1, \frac{\pi(\phi^{(i)}, \sigma^2_\eta^{(i)})}{\pi(\phi^{(i-1)}, \sigma^2_\eta^{(i-1)})} \right\} \]

\[ (62) \]
The integration is given by:

1. Initialize \((s, \phi, \sigma_\eta, \mu)\);
2. Sample \((\phi, \sigma_\eta)\) from \(\pi(\phi; \sigma_\eta^2|y^*, s)\) using the candidate from M-H based on \(g(\sigma_\eta^2, \phi)\), and accepting probability (62);
3. Sample \(h, \mu|y^*, s, \phi, \sigma_\eta^2\) using the augmented smoother algorithm;
4. Sample \(s|y^*, h\) as in the previous algorithm;
5. Return to 2.

One advantage of this sampler is that we can sample from \((h, \mu)\) from the smoother algorithm. The sampling from \(\pi(\phi, \sigma_\eta^2|y)\) is done by using M-H algorithm from Gilks et al. (1995).

**Reweighting** After getting a sample from the posterior distributions of the parameters Kim et al. (1998) proposed to reweight the sample in order to improve the estimates. Assuming that the samples are taken from \(k(\theta, h|y^*)\), where \(\theta = (\phi, \sigma_\eta, \mu)\), define

\[
\begin{align*}
    w(\theta, h) &= \log \frac{f(\theta, h|y)}{k(\theta, h|y)} = \text{cte} + \log f(y|h) - \log k(y^*|h), \\
    f(y|h) &= \prod_{t=1}^{n} f_N(y_t | e^{ht}) \\
    k(y^*|h) &= \prod_{t=1}^{n} \sum_{i=1}^{k} q_i f_N(y_t|h_t)
\end{align*}
\]

Thus, for any function of \(\theta\) we have

\[
E(g(\theta)|y) = \int g(\theta)f(\theta|y)\,d\theta = \frac{\int g(\theta)e^{w(\theta,h)}k(\theta, h|y^*)\,d\theta\,dh}{\int e^{w(\theta,h)}k(\theta, h|y^*)\,d\theta\,dh} \quad (63)
\]

Brazilian Review of Econometrics 23 (2) November 2003 209
and the estimate by reweighting is given by:

\[ E(g(\hat{\theta})|y) = \sum_j g(\theta^j) \frac{e^{w(\theta^j, h^j)}}{\sum_i e^{w(\theta^i, h^i)}}. \] (64)

5.2 Sampling the Coefficients of the Model.

Jacquier et al. (1994), SP and Kim et al. (1998) suggested conjugate priors for all the coefficients of the model. Jacquier et al. (1994) considered a vector \( \Theta \)

\[ \Theta = (\alpha, \beta) \]

then

\[ p(\Theta) = p(\Delta, \sigma_\eta) \Delta|\sigma \]

Thus, we have:

\[ p(\Delta|\sigma_\eta) = \frac{|A|^{\frac{1}{2}}}{2\pi \sigma_\eta^2} \exp\left(-\frac{(\Delta - \overline{\beta})^T A (\Delta - \overline{\beta})}{2}\right) \]

\[ \sigma_\eta \]

\[ p(\sigma_\eta) = \frac{\lambda^\nu}{\Gamma(\nu)\sigma_\eta^{(\nu+1)}} \exp\left(-\frac{\lambda}{\sigma_\eta}\right), \]

where \( \alpha, \beta, \sigma_\eta, A, \nu, \lambda \) are prior parameters. With the conjugate priors selected by Jacquier et al. (1994) the posterior is a product of a multivariate normal with an inverse gamma.

Sampling \( \sigma_\eta^2 \),

Kim et al. (1998) sampled \( \sigma_\eta^2 \) as SP, a conjugate prior \( \sigma_\eta^2|\phi \),

from the inverse Gamma:
\[
\sigma^2_\eta | y, h, \phi, \mu \sim IG\left\{ \frac{n + \sigma^\prime}{2}, \right\}
\]
\[
S_{\sigma} + \frac{(h_1 - \mu)^2(1 - \phi^2) + \sum_{t=1}^{n-1}((h_{t+1} - \mu) - \phi(h_t - \mu))^2}{2}
\]

Kim et al. (1998) fixed \( \sigma^\prime = 5 \) and \( S_{\sigma} = 0.01 \times \sigma^\prime \). SP using the relationship between gamma and \( \chi^2 \) distributions and a prior \( \sigma^2_\eta | \phi \sim \chi_p^{-2}S_0 \) found the posterior

\[
\sigma^2_\eta | y, \alpha, \phi \sim \chi_{n+p}^{-2} \left\{ S_0 + \alpha_1^2(1 - \phi^2) + \sum_{t=1}^{n-1}(\alpha_{t+1} - \phi \alpha_t)^2 \right\}, \quad (65)
\]

where SP used a parametrization \( \alpha_t = h_t - \mu \) in comparison to the model used by Kim et al. (1998).

The parameter \( \phi \) is sampled from the complete conditional density using the A-R algorithm. Thus, considering the re-parametrization \( \phi = 2\phi^* - 1 \), where the distribution of \( \phi \) is a beta distribution with parameters \( (\phi^{(1)}, \phi^{(2)}) \), the prior of \( \phi \) is given by:

\[
\pi(\phi) \propto \left\{ \frac{(1 + \phi)}{2} \right\}^{\phi^{(1)} - 1} \left\{ \frac{(1 - \phi)}{2} \right\}^{\phi^{(2)} - 1}, \quad \phi^{(1)}, \phi^{(2)} > \frac{1}{2}.
\]

Kim et al. (1998) used this re-parametrization with \( \phi^{(1)} = 20 \) and \( \phi^{(2)} = 1.5 \) which means a prior mean equal to 0.86.

Given the prior (66), the complete conditional density of \( \phi \) is proportional to \( \pi(\phi)f(h|\mu, \phi, \sigma^2_\eta) \), where
Exact Maximum Likelihood and Bayesian Estimation of the Stochastic Volatility

\[
\log f(h|\mu, \phi, \sigma_\eta^2) \propto -\frac{(h_t - \mu)^2(1 - \phi)}{2\sigma_\eta^2} + \frac{1}{2} \log(1 - \phi^2) - \frac{\sum_{t=1}^{n-1} \{(h_{t+1} - \mu) - \phi(h_t - \mu)\}^2}{2\sigma_\eta^2}.
\]  (67)

This function is concave in \( \phi \) for any values of \( \phi^{(1)} \) and \( \phi^{(2)} \). Thus, we can sample \( \phi \) using an A-R algorithm. Kim et al. (1998) took a Taylor expansion around the value

\[
\hat{\phi} = \frac{\sum_{t=1}^{n-1} (h_{t+1} - \mu)(h_t - \mu)}{\sum_{t=1}^{n-1} (h_t - \mu)^2}.
\]  (68)

Therefore, given a proposed value \( \phi^* \) from \( N(\hat{\phi}, \sigma_\phi^2) \), where \( \sigma_\phi^2 = \sigma^2_\eta \sum_{t=1}^{n-1} (h_t - \mu)^2 \)^{-1} this value is accepted with probability

\[
\exp\{g(\phi^*) - g(\phi^{(i-1)})\},
\]  where

\[
g(\phi) = \log \pi(\phi) - \frac{(h_t - \mu)^2(1 - \phi^2)}{2\sigma_\eta^2} + \frac{1}{2} \log(1 - \phi^2). \]  (69)

If the proposed value is rejected, we take \( \phi^{(i)} = \phi^{(i-1)} \).

**Sampling \( \mu \).**

Supposing a diffuse prior for \( \mu \), the sample is taken from a complete conditional distribution given by \( \mu|h, \phi, \sigma_\eta^2 \sim N(\hat{\mu}, \sigma_\mu^2) \)

\[
\hat{\mu} = \sigma_\mu^2 \left\{ \frac{(1 - \phi^2)}{\sigma_\eta^2} h_1 + \frac{(1 - \phi)}{\sigma_\eta^2} \sum_{t=1}^{n-1} (h_{t+1} - \phi h_t) \right\}, \text{ and } \]  (70)
\[ \sigma^2_{\mu} = \sigma^2_{\eta} \left\{ (n - 1)(1 - \phi^2) + (1 - \phi^2) \right\}^{-1}. \quad (71) \]

In this paper we sampled \( \mu \) and used the inverse transformation \( \beta = \exp(\mu/2) \). SP sampled \( \beta^2 | y, \alpha \quad \chi^{-2}_n \sum y_t^2 \exp(-\alpha_t) \).


This section applies the methodologies presented to empirical, IBOVESPA series, and simulated series. We used the simulated series to study the effects of AO and IO on the likelihood and on the estimates of the parameters and the volatility. We used a Pentium III, 550 MHz, 64Mb computer.

6.1 Simulated Series.

We ran the simulations with four set of parameters. The parameter values were taken from SV models fitted in the literature and are presented in Table 1. All the analyses were done for the four cases but we will only present the results for cases I and IV, although the comments are also valid for cases II and III.

<table>
<thead>
<tr>
<th>Cases</th>
<th>( \phi )</th>
<th>( \sigma )</th>
<th>( \beta )</th>
<th>( \Delta_{250} )</th>
<th>( \Lambda_{600} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.98</td>
<td>0.12</td>
<td>1.0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>II</td>
<td>0.95</td>
<td>0.10</td>
<td>1.0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>III</td>
<td>0.90</td>
<td>0.50</td>
<td>1.0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>IV</td>
<td>0.98</td>
<td>0.12</td>
<td>1.0</td>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 1: Models used in the simulation.

\( \Delta = \text{size of the AO and } \Lambda = \text{size of the IO} \)

For each model we simulated series of size 1000. The basic series was generated without outlier and each basic series was modified in three different ways. In each modification we used the same set
of disturbances used in the basic series. In the first modification we introduced an AO at the 250-th observation and an IO at the 600-th observation. This means that both series are equal until the 599-th observation, except for the 250-th observation. In the second modification a single AO was introduced at the 250-th observation while in the last modification we introduced a single IO at the 600-th observation. The AO size was equal to \( \Delta \) sample standard deviation of the basic series while the size of the IO was equal to \( \Lambda \) times the standard deviation of the autoregressive process, i.e., equal to \( \Lambda[\sigma_\eta^2(1 - \phi^2)^{-1}]^{0.5} \).

Tables 2 presents some descriptive statistics for cases I and IV. As expected the presence of an AO in the series has larger influence on the mean of the returns, while the presence of an IO increases the variance of the series.

<table>
<thead>
<tr>
<th>Case</th>
<th>Mean</th>
<th>S.D.</th>
<th>Skewness</th>
<th>Exc.Kurt.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>without outlier</td>
<td>0.0023</td>
<td>1.053</td>
<td>0.0250</td>
<td>1.124</td>
<td>-4.253</td>
<td>4.068</td>
</tr>
<tr>
<td>AO and IO</td>
<td>0.0063</td>
<td>1.204</td>
<td>0.0277</td>
<td>1.858</td>
<td>-5.847</td>
<td>5.017</td>
</tr>
<tr>
<td>AO</td>
<td>0.0063</td>
<td>1.063</td>
<td>0.0982</td>
<td>1.344</td>
<td>-4.253</td>
<td>4.685</td>
</tr>
<tr>
<td>IO</td>
<td>0.0023</td>
<td>1.195</td>
<td>-0.0215</td>
<td>1.770</td>
<td>-5.847</td>
<td>5.017</td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>without outlier</td>
<td>0.0023</td>
<td>1.053</td>
<td>0.0250</td>
<td>1.124</td>
<td>-4.253</td>
<td>4.068</td>
</tr>
<tr>
<td>AO and IO</td>
<td>0.0080</td>
<td>1.445</td>
<td>-0.0249</td>
<td>10.40</td>
<td>-11.60</td>
<td>8.599</td>
</tr>
<tr>
<td>AO</td>
<td>0.0089</td>
<td>1.082</td>
<td>0.0395</td>
<td>3.540</td>
<td>-4.253</td>
<td>7.908</td>
</tr>
<tr>
<td>IO</td>
<td>0.0074</td>
<td>1.424</td>
<td>-0.417</td>
<td>10.27</td>
<td>-11.60</td>
<td>8.599</td>
</tr>
</tbody>
</table>

Table 2: Descriptive statistics of simulated series for cases I and IV.

The simulated series of cases I and IV are presented in figure 1.
6.1.1 Convergence in the Bayesian Estimation.

The MCMC algorithm was initialized with $h_t = 0.02$ and $\mu = 0$, following Kim et al. (1998). They also checked that the initial values are not important.

We ran 1,000 iterations of the volatility estimation algorithm followed by 50,000 sampling iterations of parameters and volatility. After the burn-in we took 150,000 samples from the posterior density. These numbers of iteration in the burn-in stage were suggested by Kim et al. (1998).

The study of the convergence was done graphically. The analysis of the results confirmed the decrease of the correlation with the use of sampling by block.

Figure 1: Simulated series for cases I and IV with two outliers.
6.1.2 Likelihood and the Estimation of the Parameters.

The values of the likelihood using the true parameter values, but without considering the outliers in the fitted model, are shown in Table 3.

<table>
<thead>
<tr>
<th>Type</th>
<th>case I</th>
<th>case IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DK</td>
<td>SP</td>
</tr>
<tr>
<td>Without outlier</td>
<td>-1.410778597</td>
<td>-1.410696019</td>
</tr>
<tr>
<td>AO and IO</td>
<td>-1.533930887</td>
<td>-1.532933350</td>
</tr>
<tr>
<td>AO</td>
<td>-1.420669052</td>
<td>-1.420407575</td>
</tr>
<tr>
<td>IO</td>
<td>-1.525432642</td>
<td>-1.524605183</td>
</tr>
</tbody>
</table>

Table 3: Value of the Log-likelihood corrected for cases I and IV.

The analysis of the results in Table 3 shows that the proposed methods produce very similar likelihood values and that the likelihood evaluated by DK method is in general smaller than the likelihood evaluated by SP method. The difference between the methods increases with the presence of outliers. In general, the effect of the IO on the likelihood is larger than the effect of an AO. The true comparison, in fact, will be given by the comparison of the maximum likelihood estimates.

The maximization of the likelihood was done using the BFGS algorithm (Broyden-Fletcher-Goldfarb-Shannon), as implemented in the OX computer language. The standard deviation was estimated using the second derivative of the log-likelihood. The results are shown in Tables 4 and 5. The tables also show the Bayesian estimates given by the mean of the posterior.
Table 4: Estimates of parameters of case I: $\phi = 0.98$, $\sigma_\eta = 0.12$ and $\beta = 1.0$. The values in brackets is the estimated standard deviation.
All the point estimates of the parameter are very close, except for the Bayesian estimate by mixture. The estimated standard error in the Bayesian approach, in general, is larger than in the classical approach. The estimates of the persistence parameter decrease in the presence of outliers of type AO and IO. The effect of the IO is larger than the effect of AO in the classical estimate, but in the Bayesian approach the inverse occurs. The presence of outliers as expected increases the estimates of the variance $\sigma^2$. The AO has a smaller effect on the estimation of $\beta$ than the IO.

6.1.3 Estimates of the Volatility.

The smoothed estimates are obtained using de Jong and Shephard (1995) algorithm which is implemented in the SsfPack 2.2 program of Koopman et al. (1999), written in OX. The filtered estimates were evaluated using the particle filter used by Kim et al. (1998).

![Figure 2: Estimate of the volatility of the case IV series with two outliers.](image)
Table 5: Estimates of parameters of case IV: $\phi = 0.98$, $\sigma_\eta = 0.12$ and $\beta = 1.0$. The values in brackets is the estimated standard deviation.
The estimate of the volatility was almost the same using any method except for observations near the outlier. Figure 2 presents the true volatility and the filtered and smoothed estimates for case IV with the largest AO and IO. The difference in both estimates is large only near the outliers.

Figure 3 presents a closer look at the estimates near the outlier using different parameter estimates.

Figure 3: Estimate of the volatility of the case IV series near the two outliers.

6.2 IBOVESPA Series.

This subsection analyzes the IBOVESPA series from January 02, 1995 to December 27, 2000 with a total of 1,500 observations. This period includes several crises: February and March (Mexican crisis), October of 1997 (Asian crisis), August of 1998 (Russian crisis), January of 1999 (Brazilian real devaluation) and April of 2000 (Nasdaq’s crisis). We use the modified returns as suggested by Kim et al.(1998), which is defined as:
Anderson C. O. Motta and Luiz K. Hotta

\[ y_t = 100 \times \{(\log r_t - \log r_{t-1}) - n^{-1} \sum_{i=1}^{n} (\log r_i - \log r_{i-1})\}, \]

where \( n \) is the number of observations. The modified series has zero mean, asymmetry equal to 0.743, excess kurtosis equal to 12.94, standard deviation equal to 2.859, minimum equal to -17.94 and maximum equal to 28.73.

The parameters estimates are given in Table 6. The estimates with all the methods are very similar with persistence around 0.975. Since the estimates of the volatility are very similar we only present the filtered and smoothed estimates according to Kim et al. (1998) method in Figure 4.

Similarly to the simulated series, it was necessary to have a larger number of iterations in simple sample method in order to get good estimates. On the other hand, the integration algorithm yielded good results and low iteration correlation with only 2,000 iterations.

<table>
<thead>
<tr>
<th>Classical</th>
<th>Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durbin Koopman</td>
<td>Shepard Pitt</td>
</tr>
<tr>
<td>Simple Sampling</td>
<td>Mixture</td>
</tr>
<tr>
<td>( \hat{\phi} = 0.97329 )</td>
<td>0.97357</td>
</tr>
<tr>
<td>(0.00754)</td>
<td>(0.00756)</td>
</tr>
<tr>
<td>( \sigma_n = 0.052320 )</td>
<td>0.05179</td>
</tr>
<tr>
<td>(0.02631)</td>
<td>(0.02624)</td>
</tr>
<tr>
<td>( \beta = 2.12143 )</td>
<td>2.12132</td>
</tr>
<tr>
<td>(0.23446)</td>
<td>(0.23587)</td>
</tr>
</tbody>
</table>

Table 6: Estimates of the parameters of the IBOVESPA series with standard error in brackets.
Figure 4: IBOVESPA series: Absolute value of the returns, filtered and smoothed estimates according to Kim et al. (1998) method.

7. Concluding Remarks.

This paper presented an analysis of the SV model under the Bayesian and classical approaches. The main conclusion is that the Bayesian and classical approaches give almost the same estimates for the model parameters and for the volatility, but the classical approach is much quicker than the Bayesian approaches. The maximum likelihood method of DK took around 4 minutes while the simple, integration and mixture sampling Bayesian methods took, respectively, around 7, 36 and 39 minutes. The difference of computation time between the maximum likelihood and the simple integration method is not very large, but this method is not reliable, especially in the presence of outliers. Kim et al. (1998) already called attention to this point and this was confirmed in the simulation when we found a large difference among the estimates using this sampling compared to other sampling methods. We also found that the modifications
proposed by Kim et al. (1998) considerably improved the algorithm in terms of convergence and correlations between the iterations.

Acknowledgments
The first author thanks FAPESP and the second author thanks CNPq and PRONEX for the grants received. Also thanks to two anonymous reviewers for their helpful comments.


References


Exact Maximum Likelihood and Bayesian Estimation of the Stochastic Volatility