2008

An Efficient Method for Maximum Likelihood Estimation of a Stochastic Volatility Model

Junying, Shirley HUANG
Singapore Management University, shirleyhuang@smu.edu.sg

Jun YU
Singapore Management University, yujun@smu.edu.sg

Statistics and Its Interface Volume 1 (2008) 289–296

An efficient method for maximum likelihood estimation of a stochastic volatility model*

Shirley J. Huang† and Jun Yu‡

In this paper an efficient, simulation-based, maximum likelihood (ML) method is proposed for estimating Taylor's stochastic volatility (SV) model. The new method is based on the second order Taylor approximation to the integrand. The approximation enables us to transfer the numerical problem in the Laplace approximation and that in importance sampling into the problem of inverting two high dimensional symmetric tri-diagonal matrices. A result recently developed in the linear algebra literature shows that such an inversion has an analytic form, greatly facilitating the computations of the likelihood function of the SV model. In addition to provide parameter estimation, the new method offers an efficient way to filter, smooth, and forecast latent log-volatility. The new method is illustrated and compared with existing ML methods using simulated data. Results suggest that the proposed method greatly reduces the computational cost in estimation without sacrificing the statistical efficiency, at least for the parameter settings considered.

AMS 2000 subject classifications: Primary 62G05, 62G09; secondary 62P20.
Keywords and phrases: Maximum Likelihood, Stochastic Volatility Models, Laplace Approximation, Importance Sampler.

1. INTRODUCTION

In finance, volatility clustering refers to the observation that large changes tend to be followed by large changes. A quantitative manifestation of this fact is that, while returns themselves are uncorrelated, absolute returns or their squares display a positive, significant autocorrelation function. The best known models for capturing volatility clustering are of ARCH-type, where the conditional variance is explicitly modeled as a deterministic function of past returns. Such a model facilitates statistical analysis because the likelihood function is readily available in closed form.

A popular alternative to the ARCH-type models for capturing volatility clustering is the stochastic volatility (SV) model (Taylor, 1982), where volatility is latent, following a separate stochastic process. A salient feature of this model is that the likelihood function is expressed by a high dimensional integral which cannot be solved analytically due to the presence of a latent process. As a result, the maximum likelihood (ML) estimation is not trivial as in the ARCH models. In recent years, several simulation-based ML methods has been proposed in the econometrics literature for estimating the SV model. A partial list of references is Danielsson and Richard (1993), Danielsson (1994), Shephard and Pitt (1997), Durbin and Koopman (1997), Kim, Shephard and Chib (1998), Sandmann and Koopman (1998), Liesenfeld and Richard (2003, 2006), Richard and Zhang (2007), Durham (2006, 2007), and Skaug and Yu (2007). The idea in these papers is to evaluate the likelihood function numerically by integrating out the latent volatility process via importance sampling techniques, followed by numerical maximization of the approximate likelihood function.

In particular, the work by Shephard and Pitt (1997), Durbin and Koopman (1997), Sandmann and Koopman (1998), seems to have received much attention. In these papers, the likelihood function is approximated 1) by approximating the distribution of volatilities, conditional on returns, with a multivariate normal distribution (i.e. Laplace approximations); 2) by making independent draws from the multivariate normal distributions for Monte Carlo integrations (i.e. importance sampling). Both in the Laplace approximation stage and in the importance sampling stage, one has to calculate, numerically and repeatedly, the inverses of high dimensional matrices. These numerical inversions are the main contributors to the computational cost for evaluating the likelihood function. Another popular ML method was proposed by Richard and Zhang (2007).

The purpose of this paper is to develop a new simulation-based ML method whose computational cost is much reduced relative to that of the existing ML methods. Our method is built upon the work of Shephard and Pitt (1997), Durbin and Koopman (1997), Sandmann and Koopman (1998), Durham (2006, 2007), and Skaug and Yu (2007). In particular, the proposed method approximates the integrand in the likelihood function using the second order Taylor expansions and hence translates the numerical problem in the

*We would like to thank the editor and a referee for helpful comments.
†Huang gratefully acknowledges financial support from the Office of Research at Singapore Management University under Grant No. 07-C207-SMU-028.
‡Yu gratefully acknowledges financial support from the Ministry of Education AcRF Tier 2 fund under Grant No. T206B4301-RS.

Laplace approximation into the problem of inverting symmetric tri-diagonal matrices. In a recent contribution in the linear algebra literature, Yamani and Abdekmonen (1997) showed that the inverse of symmetric tri-diagonal matrices has an analytic form. The use of the explicit inverse greatly decreases the computational cost for calculating the likelihood function. Furthermore, the new method provides a computationally efficient way to filter, smooth, and forecast latent log-volatility.

A drawback of the proposed method is that it may not be easily generalized to more flexible stochastic volatility models. For example, a highly important stylized fact in stock returns is the leverage effect (Black, 1976, Harvey and Shephard, 1996, Yu, 2005). However, it is well known that for some financial time series, such as the exchange rate returns and interest rates, the leverage effect is insignificant (Meyer and Yu, 2000) and our method should be useful for these data.

The rest of the paper is organized as follows. Section 2 reviews the ML method based on the Laplace approximation and the importance sampler and introduces the new method. Section 3 explains how the proposed method facilitates smoothing, predicting and filtering the latent variable. In Section 4 we examine the relative performance of the proposed method using simulated data. Section 5 concludes.

2. SIMULATION-BASED MAXIMUM LIKELIHOOD METHOD

2.1 A conventional ML method

In this paper, we focus on the so-called basic SV model of Taylor (1982) which is defined by

$$\begin{cases} X_t &= \sigma e^{h_t/2}\epsilon_t, \quad t = 1, \ldots, T, \\ h_{t+1} &= \phi h_t + \gamma\eta_t, \quad t = 1, \ldots, T-1 \end{cases} \tag{1}$$

where X_t is the return of an asset, $\epsilon_t \overset{iid}{\sim} N(0,1)$, $\eta_t \overset{iid}{\sim} N(0,1)$, and $h_1 \sim N(\mu, \gamma^2/(1-\phi^2))$. As pointed out in Yu (2005), a negative value for $corr(\epsilon_t, \eta_t)$ produces the leverage effect. In this paper, however, we assume $corr(\epsilon_t, \eta_t) = 0$.

Let $\theta = (\sigma, \phi, \gamma)'$ be the parameters of interest, $\mathbf{X} = (X_1, \ldots, X_T)'$ and $\mathbf{h} = (h_1, \ldots, h_T)'$. The likelihood function of the basic SV model is given by

$$\ell(\theta) \equiv p(\mathbf{X}; \theta) = \int p(\mathbf{X}, \mathbf{h}; \theta) d\mathbf{h}, \tag{2}$$

where $p(\cdot; \theta)$ represents the density, given θ. In general (2) is high-dimensional integral which does not have closed form expression due to the non-linear dependence of $\mathbf{X}$ on $\mathbf{h}$.

To perform ML estimation to the SV model, one must approximate the high-dimensional integral (2) numerically. Since a typical financial time series has at least several hundreds observations, using traditional numerical integration methods, such as quadratures, to approximate the high-dimensional integral (2) is numerically infeasible. This is the motivation of the use of Monte Carlo integration methods in much of the SV literature. One of the best received Monte Carlo methods is to first match the integrand with a multivariate normal distribution and second draw a sequence of independent variables from the multivariate normal distribution. The technique in Stage 1 is known as the Laplace approximation while the technique in Stage 2 is known as the importance sampler. In this paper we call the method LA-IS.

To fix the idea in details, in Stage 1, we match $p(\mathbf{X}, \mathbf{h}; \theta)$ and a multivariate normal distribution for $\mathbf{h}$ as closely as possible (up to a constant proportion). More precisely, the mean and co-variance in the multivariate normal are taken to be $\mathbf{h}^*$ and $-\Omega^{-1}$, respectively, where

$$\mathbf{h}^* = \arg\max_{\mathbf{h}} \ln p(\mathbf{X}, \mathbf{h}; \theta) \tag{3}$$

and

$$\Omega = \frac{\partial^2 \ln p(\mathbf{X}, \mathbf{h}^*; \theta)}{\partial \mathbf{h} \partial \mathbf{h}'}. \tag{4}$$

Then the Laplace approximation to the integrand (2) is

$$p(\mathbf{X}, \mathbf{h}; \theta) \approx N(\mathbf{h}; \mathbf{h}^*, -\Omega^{-1}), \tag{5}$$

where $N(\cdot; \mu, \Sigma)$ represents the density of a (multivariate) normal distribution with mean of μ and co-variance of Σ.

The Laplace approximation is exact when $p(\mathbf{X}, \mathbf{h}; \theta)$ is Gaussian in $\mathbf{h}$. This is of course not the case for the SV model. Moreover, for the SV model $\mathbf{h}^*$ does not have the analytical expression. To find $\mathbf{h}^*$, Newton's method has been proposed in Shephard and Pitt (1997) and in Durham (2006) which involves recursive calculations of $\mathbf{h} = \mathbf{h}_- - \Omega^{-1}\mathbf{h}_-$, based on a certain initial vector of log-volatilities, $\mathbf{h}_0$.[1] Obviously, one has to invert Ω, either analytically or numerically.

Based on the Laplace approximation, the likelihood function can be written as

$$\begin{aligned} p(\mathbf{X}; \theta) &= \int p(\mathbf{X}, \mathbf{h}; \theta) d\mathbf{h} \\ &= \int \frac{p(\mathbf{X}, \mathbf{h}; \theta)}{N(\mathbf{h}; \mathbf{h}^*, -\Omega^{-1})} N(\mathbf{h}; \mathbf{h}^*, -\Omega^{-1}) d\mathbf{h}. \end{aligned} \tag{6}$$

The idea of importance sampling is to draw samples $\mathbf{h}^{(1)}, \ldots, \mathbf{h}^{(S)}$ from $N(\cdot; \mathbf{h}^*, -\Omega^{-1})$ so that we can approximate $p(\mathbf{X}; \theta)$ by

$$\frac{1}{S} \sum_{s=1}^{S} \frac{p(\mathbf{X}, \mathbf{h}^{(s)}; \theta)}{N(\mathbf{h}^{(s)}; \mathbf{h}^*, -\Omega^{-1})}. \tag{7}$$

[1] Since the unconditional mean of $\mathbf{h}$ is 0, a natural choice of the initial vector is 0.

As in Newton's method, one has to invert Ω, either analytically or numerically, during the importance sampling. After the likelihood function is obtained, a numerical optimization procedure, such as the quasi Newton method, can be applied to obtain the ML estimator.[2]

The convergence of (7) to the likelihood function $p(\mathbf{X};\theta)$ with $S \to \infty$ is ensured by Komogorov's strong law of large numbers. The square root rate of convergence is achieved if and only if the following condition holds

$$Var\left(\frac{p(\mathbf{X},\mathbf{h}^{(s)};\theta)}{N(\mathbf{h}^{(s)};\mathbf{h}^*,-\Omega^{-1})}\right) < \infty.$$

See Koopman and Shephard (2004) for further discussions on the conditions and a test for the convergence.

For Taylor's SV model, the integrand in (2) can be written as

(8)
$$p(\mathbf{X},\mathbf{h};\theta) = N\left(h_1, 0, \frac{\gamma^2}{1-\phi^2}\right) \times \prod_{t=2}^{T} N\left(h_t, \phi h_{t-1}, \gamma^2\right) \prod_{t=1}^{T} N\left(X_t, 0, \sigma^2 e^{h_t}\right),$$

and hence

(9)
$$\ln p(\mathbf{X},\mathbf{h};\theta) = \ln N\left(h_1, 0, \frac{\gamma^2}{1-\phi^2}\right) + \sum_{t=2}^{T} \ln N\left(h_t, \phi h_{t-1}, \gamma^2\right) + \sum_{t=1}^{T} \ln N\left(X_t, 0, \sigma^2 e^{h_t}\right).$$

It is easy to show that

$$\frac{\partial N(x;\mu,\sigma^2)/\partial x}{N(x;\mu,\sigma^2)} = -\frac{x-\mu}{\sigma^2},$$

$$\frac{\partial N(x;\mu,\sigma^2)/\partial \mu}{N(x;\mu,\sigma^2)} = -\frac{\mu-x}{\sigma^2},$$

$$\frac{\partial N(x;\mu,\sigma^2)/\partial \sigma^2}{N(x;\mu,\sigma^2)} = -\frac{1}{\sigma^2}\left(1-\frac{(x-\mu)^2}{\sigma^2}\right).$$

[2]It is important to emphasize the Common Random Numbers (CRNs) technique during the numerical optimizations to enforce a smooth surface of the objective function. That is, the S simulated paths are always obtained from a fixed set of canonical random numbers, in this case, the standardized normals. Moreover, the numerical standard errors can be obtained by conducting the simulated ML under different set of CRNs. As argued in Durham (2006), the numerical standard errors are important for assessing the stability of the method.

Using these results, we obtain the gradient of the log-integrand, denoted as G,

(10)
$$G = \begin{pmatrix} \frac{\partial \ln p(\mathbf{X},\mathbf{h};\theta)}{\partial h_1} \\ \frac{\partial \ln p(\mathbf{X},\mathbf{h};\theta)}{\partial h_2} \\ \vdots \\ \frac{\partial \ln p(\mathbf{X},\mathbf{h};\theta)}{\partial h_{T-1}} \\ \frac{\partial \ln p(\mathbf{X},\mathbf{h};\theta)}{\partial h_T} \end{pmatrix} = \begin{pmatrix} \frac{\phi h_2 - h_1}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_1^2 \\ \frac{\phi h_3 - \phi^2 h_2 + \phi h_1}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_2^2 \\ \vdots \\ \frac{\phi h_T - \phi^2 h_{T-1} + \phi h_{T-2}}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_{T-1}^2 \\ \frac{h_T - \phi h_{T-1}}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_T^2 \end{pmatrix}$$

Furthermore, the Hessian matrix of the log-integrand, denoted as Ω, is

(11)
$$\Omega = \begin{pmatrix} -\frac{1}{\gamma^2} - \frac{1}{2}\epsilon_1^2 & \frac{\phi}{\gamma^2} & \cdots & 0 & 0 \\ \frac{\phi}{\gamma^2} & -\frac{1+\phi^2}{\gamma^2} - \frac{1}{2}\epsilon_2^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -\frac{1+\phi^2}{\gamma^2} - \frac{1}{2}\epsilon_{T-1}^2 & \frac{\phi}{\gamma^2} \\ 0 & 0 & \cdots & \frac{\phi}{\gamma^2} & -\frac{1}{\gamma^2} - \frac{1}{2}\epsilon_T^2 \end{pmatrix}$$

where $\epsilon_t = X_t \exp(-0.5h_t)/\sigma$.

From (10), we see that G contains a system of nonlinear equations in h_t and Ω is a symmetric tri-diagonal matrix. Finding the mode of $\ln p(\mathbf{X},\mathbf{h};\theta)$ is equivalent to solving the nonlinear system $G(\mathbf{h}) = 0$. Since there is no analytical expression for the root to this system, numerical methods are called for. Shephard and Pitt (1997) suggested using Newton's method, which recursively calculates $\mathbf{h} = \mathbf{h}_- - \Omega^{-1}\mathbf{h}_-$. The recursion starts with a certain initial vector for log-volatilities. A nature choice is $\mathbf{h}_0 = 0$. As argued in Shephard and Pitt, the algorithm can be slow if one has to iterate the procedure until full convergence. Instead they suggested using a fixed number of iterations to get $\mathbf{h}^*$.

2.2 The new ML method

The main component of the computation cost in the ML method reviewed in Section 2.1 is in the iterations of Newton's method and the calculations of Ω^{-1} in Newton's iterations and importance sampling. To reduce the computational cost, we propose to approximate the log-integrand (8) by the second order Taylor expansion:

(12)

$$\begin{aligned}\ln p(\mathbf{X},\mathbf{h};\theta) = & -T\ln(2\pi) - \frac{T}{2}\ln(\sigma^2\gamma^2) - \frac{1}{2}\sum_{t=1}^{T} h_t \\ & - \frac{1}{2\sigma^2}\sum_{t=1}^{T} X_t^2 e^{-h_t} + \frac{1}{2}\ln(1-\phi^2) \\ & - \frac{h_1^2}{2\gamma^2/(1-\phi^2)} - \sum_{t=2}^{T}\frac{(h_t-\phi h_{t-1})^2}{2\gamma^2} \\ \approx & -T\ln(2\pi) - \frac{T}{2}\ln(\sigma^2\gamma^2) - \frac{1}{2}\sum_{t=1}^{T} h_t \\ & + \frac{1}{2}\ln(1-\phi^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} X_t^2(1-h_t+h_t^2/2) \\ & - \frac{h_1^2}{2\gamma^2/(1-\phi^2)} - \sum_{t=2}^{T}\frac{(h_t-\phi h_{t-1})^2}{2\gamma^2}.\end{aligned}$$

Consequently, we obtain the approximation to the gradient of the log-integrand, denoted by $\tilde{G}$,

(13)

$$\tilde{G} = \begin{pmatrix} -\frac{1}{2} + \frac{X_1^2}{2\sigma^2} - \left(\frac{X_1^2}{2\sigma^2} + \frac{1}{\gamma^2}\right)h_1 + \frac{\phi}{\gamma^2}h_2 \\ -\frac{1}{2} + \frac{X_2^2}{2\sigma^2} + \frac{\phi}{\gamma^2}h_1 - \left(\frac{X_2^2}{2\sigma^2} + \frac{1+\phi^2}{\gamma^2}\right)h_2 + \frac{\phi}{\gamma^2}h_3 \\ \vdots \\ -\frac{1}{2} + \frac{X_{T-1}^2}{2\sigma^2} + \frac{\phi}{\gamma^2}h_{T-2} - \left(\frac{X_{T-1}^2}{2\sigma^2} + \frac{1+\phi^2}{\gamma^2}\right)h_{T-1} + \frac{\phi}{\gamma^2}h_T \\ -\frac{1}{2} + \frac{X_T^2}{2\sigma^2} + \frac{\phi}{\gamma^2}h_{T-1} - \left(\frac{X_T^2}{2\sigma^2} + \frac{1}{\gamma^2}\right)h_T \end{pmatrix}$$

Solving $\tilde{G} = 0$ for $\mathbf{h}$ is equivalent to calculating $\mathbf{h}^* = A^{-1}B$ where

(14)

$$A = \begin{pmatrix} -\left(\frac{X_1^2}{2\sigma^2} + \frac{1}{\gamma^2}\right) & \frac{\phi}{\gamma^2} & \cdots & 0 & 0 \\ \frac{\phi}{\gamma^2} & -\left(\frac{X_2^2}{2\sigma^2} + \frac{1+\phi^2}{\gamma^2}\right) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -\left(\frac{X_{T-1}^2}{2\sigma^2} + \frac{1+\phi^2}{\gamma^2}\right) & \frac{\phi}{\gamma^2} \\ 0 & 0 & \cdots & \frac{\phi}{\gamma^2} & -\left(\frac{X_T^2}{2\sigma^2} + \frac{1}{\gamma^2}\right) \end{pmatrix}$$

is a symmetric tri-diagonal matrix and

(15)
$$B = \begin{pmatrix} \frac{1}{2} - \frac{X_1^2}{2\sigma^2} \\ \vdots \\ \frac{1}{2} - \frac{X_T^2}{2\sigma^2} \end{pmatrix}.$$

Therefore, the problem in Newton's iterations is transformed into the inverse of a symmetric tri-diagonal matrix, A. This transformed problem is the same as that in importance sampling where one has to invert Ω which is also a symmetric tri-diagonal matrix.

Inverses of tri-diagonal matrices has been extensively studied both in the linear algebra literature and in the numerical analysis literature; see Meurant (1992) for an excellent survey of the literature. Many practical applications require the inverse of tri-diagonal matrices. For example, when solving elliptic or parabolic partial differential equations (pde's) with finite element methods, one has to consider tri-diagonal matrices for one-dimensional problems. In general, no explicit formulas exist for elements of the inverse and numerical iteration methods have to be used. However, for some special tri-diagonal matrices, there are explicit formulas for elements of the inverse.

In Kershaw (1969) and Hu and O'Connell (1996), the closed form inverse of a tri-diagonal matrix, (C_{ij}), is given, where $C_{ij} = C$ if $i = j$, $C_{ij} = 1$ if $j = i+1$ and zero otherwise. In Schlegel (1970) a slightly more general matrix was considered where $C_{ij} = C_i$ if $i = j$ and $C_i = C_{T-i+1}$. Obviously, neither results can be applied to the tri-diagonal matrices, A and Ω, encountered in our problem.

In this paper, we make the use of a recent result developed in Yamani and Abdelmonem (1997) to obtain the closed-form expression for elements in A^{-1} and Ω^{-1}. There are two advantages of using the analytical result relative to the numerical iteration methods. First, it reduces computational cost. Second, the error in analytical inverse is up to the machine precision, ensuring a smoother likelihood function.

The analytical formulae in Yamani and Abdelmonem (1997) apply to general symmetric tri-diagonal matrices. Let $C_{ij} = a_i$ if $i = j$, $C_{ij} = b_i$ if $j = i+1$ and zero otherwise. Let $p_1 = 1$, $p_2 = -\frac{a_1}{b_1}$, $q_1 = 0$, $p_2 = -\frac{1}{b_1}$, and

$$p_{i+1} = -p_{i-1} - \frac{a_i}{b_i}p_i,$$

$$q_{i+1} = -q_{i-1} - \frac{a_i}{b_i}q_i,$$

for $i = 3, \ldots, T$. Then

(16)
$$(C^{-1})_{ij} = \left(q_j - \frac{q_{T+1}}{p_{T+1}}p_j\right)p_i,$$

for $i, j = 1, \ldots, T$.

Since in our matrices, A and Ω, the elements in the off-diagonal are the same, the recursive formulas in p and q can be simplified to

(17)
$$p_{n+1} = -p_{i-1} - \frac{a_i}{b}p_i,$$

(18)
$$q_{n+1} = -q_{i-1} - \frac{a_i}{b}q_i.$$

In sum, our procedure on calculating the log-likelihood involves the following steps

1. Obtain $\mathbf{h}^* = A^{-1}B$ and $\Omega^{-1}(\mathbf{h}^*)$ using the closed form expression for the inverse based on (16), (17) and (18);
2. Obtain S independent draws from $N(\cdot;\mathbf{h}^*, -\Omega^{-1})$, denoted by $\mathbf{h}^{(s)}$, $s = 1, \ldots, S$;

3. Calculate $\frac{1}{S}\sum_{s=1}^{S}\frac{p(\mathbf{X},\mathbf{h}^{(s)};\theta)}{N(\mathbf{h}^{(s)};\mathbf{h}^*,-\Omega^{-1})}$.

The convergence of the above quantity to the likelihood function $p(\mathbf{X};\theta)$ with $S\to\infty$ is ensured by Komogorov's strong law of large numbers, as in the conventional ML method. The square root rate of convergence is achieved if and only if the following condition holds

$$Var\left(\frac{p(\mathbf{X},\mathbf{h}^{(s)};\theta)}{N(\mathbf{h}^{(s)};\mathbf{h}^*,-\Omega^{-1})}\right)<\infty.$$

3. SMOOTHING, PREDICTING AND FILTERING LOG-VOLATILITY

Following the suggestion of Skaug and Yu (2007), we now show how to smooth, filter and forecast log-volatility in a numerically efficient way. Our method can be regarded as an alternative to the "reprojection" procedure developed by Gallant and Tauchen (1998) and the particle filter procedures developed by Kitagawa (1996) or Pitt and Shephard (1999). A major advantage of the proposed algorithms is the ease of implementation and low computational cost. This is in the sharp contrast to the filtration methods available in the literature.

The smoother is obtained as a by-product of the Laplace approximation, i.e., when we calculate $\mathbf{h}^*=A^{-1}B$, evaluated at the ML estimates. Obviously, $\mathbf{h}^*$ depends on the entire return series $X_1,\ldots,X_T$.

To obtain the prediction of future values of the log-volatility, we extend the $\mathbf{h}$ vector with as many time periods as needed. If we want a K-step prediction we take $\mathbf{h}=(h_1,\ldots,h_T,h_{T+1},\ldots,h_{T+K})'$. Without loss of generality, assume $K=1$ and hence $\mathbf{h}=(h_1,\ldots,h_T,h_{T+1})'$. With the new $\mathbf{h}$ Equation (9) can be written as

$$\begin{aligned}\ln p(\mathbf{X},\mathbf{h};\theta)=&\ \ln N\left(h_1,0,\frac{\gamma^2}{1-\phi^2}\right)\\&+\sum_{t=2}^{T+1}\ln N\left(h_t,\phi h_{t-1},\gamma^2\right)\\&+\sum_{t=1}^{T}\ln N\left(X_t,0,\sigma^2e^{h_t}\right),\end{aligned}\tag{19}$$

and A, B have the form of

$$A=\begin{pmatrix}-\left(\frac{X_1^2}{2\sigma^2}+\frac{1}{\gamma^2}\right) & \frac{\phi}{\gamma^2} & \cdots & 0 & 0\\ \frac{\phi}{\gamma^2} & -\left(\frac{X_2^2}{2\sigma^2}+\frac{1+\phi^2}{\gamma^2}\right) & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & -\left(\frac{X_T^2}{2\sigma^2}+\frac{1+\phi^2}{\gamma^2}\right) & \frac{\phi}{\gamma^2}\\ 0 & 0 & \cdots & \frac{\phi}{\gamma^2} & -\frac{1}{\gamma^2}\end{pmatrix},\tag{20}$$

is a symmetric tri-diagonal matrix and

$$B=\begin{pmatrix}\frac{1}{2}-\frac{X_1^2}{2\sigma^2}\\ \vdots\\ \frac{1}{2}-\frac{X_T^2}{2\sigma^2}\\ 0\end{pmatrix}.\tag{21}$$

Since A maintain the symmetric tri-diagonal structure, the inverse can be obtained in the same manner as before. The last element in $\mathbf{h}^*$, calculated by $A^{-1}B$ and evaluated at the ML estimates, is the one-step-ahead forecast of log-volatility.

The same idea is applied to obtain the filtered latent variable (i.e. the estimate of h_t conditional on $X_1,\ldots,X_t$). For example, to obtain the filter h_t, conditional on $X_1,\ldots,X_t$, we define $\mathbf{h}=(h_1,\ldots,h_t)'$ and $\mathbf{X}=(X_1,\ldots,X_t)'$. With the new $\mathbf{h}$ and $\mathbf{X}$, Equation (9) can be written as

$$\begin{aligned}\ln p(\mathbf{X},\mathbf{h};\theta)=&\ \ln N\left(h_1,0,\frac{\gamma^2}{1-\phi^2}\right)\\&+\sum_{i=2}^{t}\ln N\left(h_i,\phi h_{i-1},\gamma^2\right)\\&+\sum_{i=1}^{t}\ln N\left(X_i,0,\sigma^2e^{h_i}\right),\end{aligned}\tag{22}$$

and A, B have the form of

$$A=\begin{pmatrix}-\left(\frac{X_1^2}{2\sigma^2}+\frac{1}{\gamma^2}\right) & \frac{\phi}{\gamma^2} & \cdots & 0 & 0\\ \frac{\phi}{\gamma^2} & -\left(\frac{X_2^2}{2\sigma^2}+\frac{1+\phi^2}{\gamma^2}\right) & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & -\left(\frac{X_{t-1}^2}{2\sigma^2}+\frac{1+\phi^2}{\gamma^2}\right) & \frac{\phi}{\gamma^2}\\ 0 & 0 & \cdots & \frac{\phi}{\gamma^2} & -\left(\frac{X_t^2}{2\sigma^2}+\frac{1}{\gamma^2}\right)\end{pmatrix},\tag{23}$$

is a symmetric tri-diagonal matrix and

$$B=\begin{pmatrix}\frac{1}{2}-\frac{X_1^2}{2\sigma^2}\\ \vdots\\ \frac{1}{2}-\frac{X_t^2}{2\sigma^2}\end{pmatrix}.\tag{24}$$

Once again, A maintains the symmetric tri-diagonal structure and the inverse is obtained as before. The last element in $\mathbf{h}^*$, calculated by $A^{-1}B$ and evaluated at the ML estimates, is the filtered log-volatility.

4. PERFORMANCE OF THE PROPOSED METHOD

4.1 Another existing ML method

Richard and Zhang (2007) developed an alternative but related ML method. It is based on importance sampling and termed as Efficient Importance Sampling (EIS) procedure. Relative to the ML method reviewed in Section 2.1, EIS

Table 1. Comparison of the two existing ML methods with the proposed ML for estimating Taylor's SV model with $T = 2000$. LA-IS is the ML method of Shephard and Pitt (1997), Durbin and Koopman (1997). EIS is the ML method of Richard and Zhang (2007)

	IS-ML			EIS			Proposed ML		
	σ	ϕ	γ	σ	ϕ	γ	σ	ϕ	γ
True Value	1	0.9	0.1	1	0.9	0.1	1	0.9	0.1
Mean	.9999	.8779	.1030	.9952	.8779	.1041	.9994	.8777	.1046
Std Err	.0196	.0823	.0486	.0211	.856	.0478	.0196	.0824	.0494
RMSE	.0196	.0852	.0487	.0216	.856	.0480	.0196	.0854	.0496
CPU	222,715s			134,655s			28,467s		

Table 2. Comparison of the two existing ML methods with the proposed ML for estimating Taylor's SV model with $T = 1000$. LA-IS is the ML method of Shephard and Pitt (1997), Durbin and Koopman (1997). EIS is the ML method of Richard and Zhang (2007)

	IS-ML			EIS			Proposed ML		
	σ	ϕ	γ	σ	ϕ	γ	σ	ϕ	γ
True Value	1	0.9	0.1	1	0.9	0.1	1	0.9	0.1
Mean	1.007	.8442	.0936	1.006	.8449	.0939	1.008	.8451	.0949
Std Err	.0539	.1900	.0653	.0530	.1890	.0654	.0548	.1950	.0691
RMSE	.0544	.1980	.0656	.0530	.1969	.0657	.0554	.2026	.0693
CPU	143,690s			86,881s			18,367s		

minimizes the Monte Carlo sampling variance of the approximation to the integrand by factorizing the importance density. To fix the idea, assume $g(\mathbf{h}|\mathbf{X})$ is the importance density which can be constructed as

(25)
$$g(\mathbf{h}|\mathbf{X}) = \prod_{t=1}^{T} g(h_t|h_{t-1}, \mathbf{X}) = \prod_{t=1}^{T} \left\{ C_t e^{c_t h_t + d_t h_t^2} p(h_t|h_{t-1}) \right\},$$

where c_t, C_t and d_t depend on $\mathbf{X}$ and h_{t-1} with $\{C_t\}$ be a normalization sequence so that g is a normal distribution. The sequences $\{c_t\}$ and $\{d_t\}$ should be chosen to match $p(\mathbf{X}, \mathbf{h}; \theta)$ and $g(\mathbf{h}|\mathbf{X})$ which, as we shown in the last section, requires a high-dimensional non-linear regression. The caveat of EIS is to match each component in $g(\mathbf{h}|\mathbf{X})$ (i.e. $C_t e^{c_t h_t + d_t h_t^2} p(h_t|h_{t-1})$), to the corresponding element in the integrand $p(\mathbf{X}; \mathbf{h})$ (i.e. $p(X_t|h_t)p(h_t|h_{t-1})$) in a backward manner, with $t = T, T-1, \ldots, 1$. It is easy to show that C_t depends only on h_{t-1} but not on h_t. As a result, the recursive matching problem is equivalent to running the following linear regression backward:

(26)
$$\ln p(X_t|h_t^{(s)}) - \ln C_{t+1} = a + c_t h_t^{(s)} + d_t (h_t^{(s)})^2, \ s = 1, \ldots, S,$$

where $h_t^{(1)}, \ldots, h_t^{(S)}$ are drawn from the importance density and $h_t^{(i)}$ and $h_t^{(i)}$ are treated as the explanatory variables in the regression model with $C_{T+1} = 1$.

The method to approximate the likelihood involves the following procedures:

1. Draw initial $\mathbf{h}^{(s)}$ from Equation (1) with $s = 1, \ldots, S$.
2. Estimate c_t and d_t from (26) and do it backward with $C_{T+1} = 1$.
3. Draw $\mathbf{h}^{(s)}$ from importance density $g(\mathbf{h}|\mathbf{X})$ based on c_t and d_t.
4. Repeat Steps 2-3 until convergence. Denote the resulting sampler by $\mathbf{h}^{(s)}$.
5. Approximate the likelihood by

$$\frac{1}{S} \sum_{s=1}^{S} \left\{ \prod_{t=1}^{T} \frac{p(X_t|h_t^{(s)})}{C_t \exp\left(c_t h_t^{(s)} + d_t (h_t^{(s)})^2\right)} \right\}.$$

4.2 Performance of the proposed method

To check the performance of the proposed ML method relative to the LA-IS method reviewed in Section (2.1) and the EIS method reviewed in Section (4.1), we fit the Taylor's SV model to simulated return series using the three ML methods. In total we design four experiments. In the first experiment we simulate 2,000 observations simulated from Model (1) with $\sigma = 1$, $\phi = 0.9$, and $\gamma = 0.1$. In the second experiment we simulate 1,000 observations simulated from Model (1) with $\sigma = 1$, $\phi = 0.9$, and $\gamma = 0.1$. In the third experiment we simulate 2,000 observations simulated from Model (1) with $\sigma = 1$, $\phi = 0.95$, and $\gamma = 0.05$. In the last experiment we simulate 1,000 observations simulated from Model (1) with $\sigma = 1$, $\phi = 0.95$, and $\gamma = 0.05$.

In all three methods, we select $S = 64$ and use the same common random numbers. For the conventional ML method, we iterate Newton's method for six times. For the EIS method, we iterate the recursive regression also for six

Table 3. Comparison of the two existing ML methods with the proposed ML for estimating Taylor's SV model with $T = 2000$. LA-IS is the ML method of Shephard and Pitt (1997), Durbin and Koopman (1997). EIS is the ML method of Richard and Zhang (2007)

	IS-ML			EIS			Proposed ML		
	σ	ϕ	γ	σ	ϕ	γ	σ	ϕ	γ
True Value	1	0.95	0.05	1	0.95	0.05	1	0.95	0.05
Mean	1.002	.9340	.0507	1.001	.9341	.0505	.9994	.9330	.0496
Std Err	.0196	.0567	.0235	.0196	.0568	.0234	.0201	.0575	.0242
RMSE	.0197	.0589	.0235	.0196	.0590	.0234	.0201	.0600	.0242
CPU	616,815s			373,863s			78,968s		

Table 4. Comparison of the two existing ML methods with the proposed ML for estimating Taylor's SV model with $T = 1000$. LA-IS is the ML method of Shephard and Pitt (1997), Durbin and Koopman (1997). EIS is the ML method of Richard and Zhang (2007)

	IS-ML			EIS			Proposed ML		
	σ	ϕ	γ	σ	ϕ	γ	σ	ϕ	γ
True Value	1	0.95	0.05	1	0.95	0.05	1	0.95	0.05
Mean	1.001	.9128	.049	1.001	.9127	.049	.9990	.9106	.049
Std Err	.052	.1310	.0252	.052	.1311	.0250	.055	.1333	.0254
RMSE	.052	.1362	.02522	.052	.1363	.0250	.055	.1390	.0254
CPU	277,918s			198,021s			35,529s		

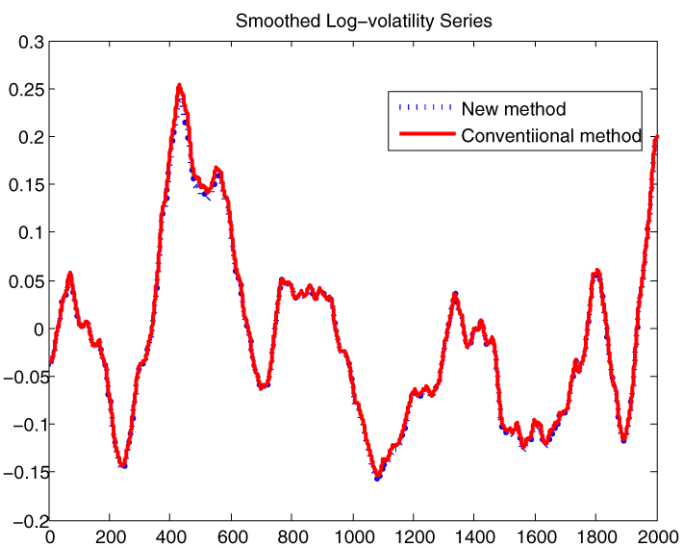

Figure 1. Smoothed log-volatilities obtained from the LA-IS method (solid line) and from the proposed method (dotted line) based on Taylor's SV model using simulated data.

times. The experiment is repeated 1,000 times to obtain the mean, the standard error and the root mean square error (RMSE) for each estimate.

Tables 1–4 report the mean, the standard error and the RMSE of the three parameters across 1,000 replications in all four experiments, respectively. We also report the CPU time of 1,000 replications for the three ML methods, using Matlab7.5 in a Pentium IV 3.2 GHz PC running on WINDOWS XP. First, all three methods work well on estimating σ. The estimates have little bias and small standard errors. Also, all three methods work well on estimating γ. However, all methods estimate ϕ with noticeable biases. The estimation bias is of Hurwicz type (Hurwicz, 1950) and has been documented in Phillips and Yu (2008) in the context of SV models. Second, the three sets of estimates are very close to each other, suggesting little approximation error in the proposed ML method. This result reinforces what has been documented in the literature about the relative performance of the existing ML method. For example, it was found in Lee and Koopman (2004) that LA-IS and EIS estimators do not differ much. Third, the proposed ML method is much faster to compute, reducing the computational cost by nearly 90% and 80%, respectively, over the two existing ML methods. This reduction in computational cost is important when simulation is needed in applications. For example, it would be much more efficient numerically to obtain the binding function of $\hat{\phi}$ in indirect inference (Phillips and Yu, 2008) based on the proposed ML method.

As an alternative way to check the relative performance of the proposed method, we obtain smoothed time series values of $\mathbf{h}$ using the LA-IS method and the proposed method, both based on an arbitrarily chosen simulated sequence. Figure 1 plots the two series. It can be seen the two series are almost identical to each other, suggesting that the proposed method works well for smoothing.

5. CONCLUSION

This paper introduces a new ML method to estimate the Taylor's SV model. Evaluation of the likelihood function

of this model involves high-dimensional numerical integrations. Conventional ML methods combine the importance sampling technique with the Laplace approximation. To locate the mode of the integrand, the conventional ML method relies on Newton's iterations. In the proposed ML method, we first approximate the integrand by the second order Taylor series and show that such an expansion avoids the need of Newton's iterations. Using the recently developed explicit expressions for the inverse of symmetric tri-diagonal matrices, we show that the calculation of the log-likelihood function is greatly simplified. We also discuss how to apply the new method to smooth, filter and forecast log-volatilities. Simulation studies show that our method substantially reduces the computational cost without sacrificing the statistical efficiency, at least for the parameter setting considered.

Received 19 October 2007

REFERENCES

[1] Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section* 177–181.

[2] Danielsson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *Journal of Econometrics* **64** 375–400.

[3] Danielsson, J. and Richard, J. F. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics* **8** s153–s173.

[4] Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society Series B* **62** 3–56. MR1745604

[5] Durham, G. S. (2006). Monte Carlo methods for estimating, smoothing, and filtering one and two-factor stochastic volatility models. *Journal of Econometrics* **133** 273–305. MR2250181

[6] Durham, G. S. (2007). SV mixture models with application to S&P 500 Index Returns. *Journal of Financial Economics*, forthcoming.

[7] Gallant, A. R. and Tauchen, G. (1998). Reprojecting partially observed systems with application to interest rate diffusions. *Journal of the American Statistical Association* **93** 10–24.

[8] Harvey, A. C., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies* **61** 247–264.

[9] Harvey, A. C. and Shephard, N. (1996). The estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* **14** 429–434.

[10] Hu, G. Y. and O'Connell, R. F. (1996). Analytical inversion of symmetric tridiagonal matrices. *Journal of Physics A: Mathematical and Theoretical* **29** 1511–1514. MR1395519

[11] Hurvicz, L. 1950. Least square bias in time series, in T. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, New York, Wiley, 365–383. MR0038640

[12] Kershaw, D. (1969). The explicit inverses of two commonly occurring matrices, *Mathematics of Computation* **23** 189–91. MR0238478

[13] Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65** 361–393.

[14] Kitagawa, G. (1996). Monte Carlo filter and smoother for Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics* **5** 1–25. MR1380850

[15] Koopman, S. and Shephard, N. (2004). Testing the assumptions behind the use of importance sampling, Working Paper, Nuffield College, University of Oxford. MR2077629

[16] Lee, K. M. and Koopman, S. J. (2004). Estimating stochastic volatility models: a comparison of two importance samplers. *Studies in Nonlinear Dynamics and Econometrics* **8** 1–15.

[17] Liesenfeld, R. and Richard, J. F. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance* **10** 505–531.

[18] Liesenfeld, R. and Richard, J. F. (2006). Classical and Bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews* **25** 335–360. MR2281691

[19] Meurant, G. (1992). A review of the inverse of symmetric tridiagonal and block tridiagonal matrices, *SIAM Journal on Matrix Analysis and Applications* **12** 707–728. MR1168018

[20] Phillips, P. C. B. and Yu, J. (2008). Simulation-based Estimation of Contingent-claims Prices, Review of Financial Studies, forthcoming.

[21] Pitt, M. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filter, *The Journal of the American Statistical Association* **94** 590–599. MR1702328

[22] Richard, J. F. and Zhang, W. (2007) Efficient high-dimensional importance, *Journal of Econometrics* **141** 1385–1411.

[23] Sandmann, G. and Koopman, S. J. (1998). Maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* **63** 289–306.

[24] Schlegel, D. (1970). The explicit inverses of a tridiagonal matrix, *Mathematics of Computation* **24** 665–665. MR0273798

[25] Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84** 653–667. MR1603940

[26] Skaug, H. J. and Yu, J. (2007). Automated Likelihood Based Inference for Stochastic Volatility Models. Working Paper, Singapore Management University.

[27] Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes — a study of the daily sugar prices 1961–75. In Anderson, O. D., Editor, *Time Series Analysis: Theory and Practice* **1** 203–226. North-Holland, Amsterdam.

[28] Yamani, H. A. and Abdelmonem, M. S. (1997). The analytical inversion of any finite symmetric tridiagonal matrix. *Journal of Physics A: Mathematical and Theoretical* **30** 2889–2893. MR1450354

[29] Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics* **127** 165–178. MR2166061

Shirley J. Huang
Lee Kong Chian School of Business
Singapore Management University
50 Stamford Road
Singapore 178899
E-mail address: shirleyhuang@smu.edu.sg
url: http://www.mysmu.edu/faculty/shirleyhuang

Jun Yu
School of Economics
Singapore Management University
90 Stamford Road
Singapore 178903
E-mail address: yujun@smu.edu.sg
url: http://www.mysmu.edu/faculty/yujun