

Estimation of Stochastic Volatility Models : An Approximation to the Nonlinear State Space Representation

Junji Shimada *
and
Yoshihiko Tsukuda **

March, 2004

Keywords : Stochastic volatility, Nonlinear state space representation,
Laplace approximation.

JEL Code : C13, C22

Acknowledgement: The first version of this paper was written while the second author was visiting the University of Western Ontario. The paper was presented at International Symposium on Financial Time Series sponsored by Tokyo Metropolitan University (February, 2004, Tokyo). We are grateful for helpful comments from J. Knight, R. A. Davis and H. K. van Dijk.

*Tohoku University, Graduate School of Economics, Kawauchi, Aoba-ku, Sendai, 980-8576, Japan. E-mail : jshimada@mail.cc.tohoku.ac.jp.

**Tohoku University, Graduate School of Economics, Kawauchi, Aoba-ku, Sendai, 980-8576, Japan. E-mail : tsukuda@econ.tohoku.ac.jp.

Abstract

The stochastic volatility (SV) model can be regarded as a nonlinear state space model. This paper proposes the Laplace approximation method to the nonlinear state space representation and applies it for estimating the SV models. We examine how the approximation works by simulations as well as various empirical studies. The Monte-Carlo experiments for the standard SV model indicate that our method is comparable to the Monte-Carlo Likelihood (MCL: Durbin and Koopman (1997)), Maximum Likelihood (Fridman and Harris (1998)) and MCMC methods in the sense of mean square error in finite sample. The empirical studies for stock markets reveal that our method provides very similar estimates of coefficients to those of the MCL. We show a relationship of our Laplace approximation method to importance sampling.

1 Introduction

The financial time series such as stock returns show heteroskedasticity. The squared returns process exhibits pronounced serial correlation whereas the returns process itself exhibits little or no serial correlation. The autoregressive conditional heteroskedasticity (ARCH) model is one way of describing the financial time series (Engel (1982), Bollerslev (1986) and Nelson (1991), among others). The ARCH type models specify the volatility of the current return as a deterministic function of the past observations and have been widely used in applied empirical research.

Alternatively, volatility may be modeled as an unobservable component following some latent stochastic process, such as an autoregressive model. Models of this kind are called as stochastic volatility (SV) models (Taylor (1994), Andersen (1994)). An appealing feature of the SV model is its close relationship to financial economic theories. The joint distribution of the security returns and trading volumes was incorporated into the SV model (Clark (1973), Tauchen and Pitt (1983)). The SV model was used to approximate the stochastic volatility diffusion process for evaluating the option prices (Hull and White (1987) and Melino and Turnbull (1990)).

Despite theoretical advantages, the SV models have not been popular as the ARCH models in practical applications. The main reason is that the likelihood of the SV models is not easy to evaluate unlike the ARCH models. The generalized method of moments (GMM) is less efficient but not dependent on the likelihood for estimating the SV models (Andersen and Sørensen (1996)).

Recent developments in Markov Chain Monte-Carlo (MCMC) methods have increased the popularity of Bayesian inference in many fields of research including the SV models. In their epoch making work Jacquire et al.(1994) applied a Bayesian analysis for estimating the SV model. They proposed a method which samples alternately parameters and unobservable volatilities. Shephard and Pitt (1997) improved sampling technique for volatilities by approximating a joint density of multiple volatilities by the second order Taylor expansion. Kim et al.(1998) extended the sampling technique of Shephard and Pitt (1997) and provided an excellent method of sampling the parameters and volatilities.

The classical analysis based on the likelihood for estimating the SV model has been extensively studied in the recent years. Danielson (1994) approximates

the marginal likelihood of the observable process by simulating the latent volatility conditional on the available information. Shephard and Pitt (1997) gave an idea of evaluating likelihood by exploiting sampled volatility. Durbin and Koopman(1997) explored the idea of Shephard and Pitt (1997) and evaluated the likelihood by Monte-Carlo integration. Sandmann and Koopman (1998) applied this method for the SV model. The method of Monte-Carlo maximum likelihood was reviewed by Durbin and Koopman (2000) from both classical and Bayesian perspectives. Fridman and Harris (1998) and Watanabe (1999) integrated out the latent volatilities by the numerical method of Kitagawa (1987) for evaluating the likelihood. While their numerical methods give the likelihood to any degree of accuracy depending on the computational costs, their algorithms are not easy to extend to multivariate models.

The purpose of this paper is to propose the Laplace approximation (LA) method to the nonlinear state space representation, and to show that the LA method is workable for estimating the SV models including the multivariate SV model and the dynamic bivariate mixture (DBM) model. The SV model can be regarded as a nonlinear state space model. The LA method approximates the logarithm of the joint density of current observation and volatility conditional on the past observations by the second order Taylor expansion around its mode, and then applies the nonlinear filtering algorithm. This idea of approximation is found in Shephard and Pitt (1997) and Durbin and Koopmann (1997). The Monte-Carlo Likelihood (MCL: Sandmann and Koopman (1998)) is now a standard classical method for estimating the SV models. It is based on importance sampling technique. Importance sampling is usually regarded as an exact method for maximum likelihood estimation. We show that the LA method of this paper approximates the weight function by unity in the context of importance sampling. We do not need to carry out the Monte Carlo integration for obtaining the likelihood since the approximate likelihood function can be analytically obtained. If one-step ahead prediction density of observation and volatility variables conditional on the past observations is sufficiently accurately approximated, the LA method is workable.

We examine how the LA method works by simulations as well as various empirical studies. In order to investigate the finite sample properties of the LA approach, we conduct Monte-Carlo experiments for the standard SV model. We

compare the LA approach with the MCL, Maximum Likelihood (Fridman and Harris (1998)) and MCMC in terms of the estimates of parameters and the smoothing estimates of volatilities. The Monte-Carlo experiments reveal that our method is comparable to the MCL, Maximum Likelihood and MCMC methods.

We apply this method to the univariate SV models with normal distribution or t-distribution, the bivariate SV model and the dynamic bivariate mixture model, and empirically illustrate how the LA method works for each of the extended models. The empirical results on the stock markets reveal that our method provides very similar estimates of coefficients to those of the MCL. As a result, this paper demonstrates that the LA method is workable through both simulation studies and empirical studies. Naturally, the workability is limited to the cases examined in this paper. But we believe the LA method is applicable to many SV models based on our study of this paper.

The paper is organized as follows. Section 2 discusses the algorithm of the LA approach to the nonlinear state space representation for the standard SV model. Although we state the algorithm for the standard SV model for the clarity of exposition, we emphasize that the algorithm is easily applicable to the more complicated SV models. In Section 3, we conduct the Monte-Carlo experiments to investigate the finite sample properties of the LA approach and compare the LA with the MCL and MCMC methods. In Section 4, we examine how the LA approach works when it is used for the actual stock market data. The LA approach is applied to the four types of SV models: the univariate SV model with either normal or t-distribution, the bivariate SV model and the DBM model and compared with the MCL. Section 5 states concluding remarks. An analytical relationship between the LA method and importance sampling technique is stated in Appendix B.

2 The Laplace approximation to the nonlinear state space representation

Section 2 discusses the algorithm of the LA approach to the nonlinear state space representation for the univariate SV model with normal distribution. Although we state the algorithm for the standard univariate SV model for the clarity

of exposition, we emphasize that the algorithm is easily extended to the more complicated SV models.

2.1 The univariate stochastic volatility model

We consider the univariate SV model proposed by Taylor(1986),

$$y_t = \exp(h_t/2)\epsilon_t \quad , \quad \epsilon_t \sim NID(0, 1) \quad (1)$$

$$h_{t+1} = \alpha + \beta h_t + \eta_{t+1}, \quad \eta_{t+1} \sim NID(0, \sigma_\eta^2) \quad (2)$$

$$h_1 \sim N\left(\frac{\alpha}{1-\beta}, \frac{\sigma_\eta^2}{1-\beta^2}\right) \quad (3)$$

where $\sigma_t^2 = \exp(h_t)$ is the volatility of y_t . The log volatility h_t is specified by the AR(1) process with Gaussian innovation noise. The density functions of y_t given h_t and of h_t given h_{t-1} are respectively

$$f(y_t|h_t) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y_t^2}{2} \exp(-h_t) - \frac{h_t}{2}\right\} \quad (4)$$

$$f(h_t|h_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left\{-\frac{(h_t - \alpha - \beta h_{t-1})^2}{2\sigma_\eta^2}\right\}. \quad (5)$$

This model can be regarded as a nonlinear state space model. In order to evaluate the likelihood, we have to integrate out the latent log volatilities.

2.2 Filtering and evaluation of the likelihood

To evaluate the likelihood, we need to carry out the filtering algorithm for $t = 1, \dots, T$ given the initial distribution.

(i) one step ahead prediction of y_t :

$$\begin{aligned} f(y_t|Y_{t-1}) &= \int_{-\infty}^{\infty} f(y_t, h_t|Y_{t-1})dh_t \\ &= \int_{-\infty}^{\infty} f(y_t|h_t)f(h_t|Y_{t-1})dh_t \end{aligned} \quad (6)$$

where $Y_t = (y_t, y_{t-1}, \dots, y_1)$ for $t = 1, \dots, T$ and Y_0 is empty.

(ii) updating of h_t :

$$\begin{aligned} f(h_t|Y_t) &= f(h_t|y_t, Y_{t-1}) \\ &= \frac{f(y_t, h_t|Y_{t-1})}{f(y_t|Y_{t-1})} \\ &= \frac{f(y_t|h_t)f(h_t|Y_{t-1})}{f(y_t|Y_{t-1})}, \end{aligned} \quad (7)$$

(iii) one step ahead prediction of h_t :

$$\begin{aligned} f(h_{t+1}|Y_t) &= \int_{-\infty}^{\infty} f(h_{t+1}, h_t|Y_t)dh_t \\ &= \int_{-\infty}^{\infty} f(h_{t+1}|h_t)f(h_t|Y_t)dh_t. \end{aligned} \quad (8)$$

If we have $f(y_t|Y_{t-1}), t = 1, \dots, T$, we can calculate the log likelihood

$$L(\theta|Y_T) = \sum_{t=1}^T \log f(y_t|Y_{t-1}). \quad (9)$$

where $\theta = (\alpha, \beta, \sigma_\eta)$.

It is difficult to solve the integrations in the equations (6) and (8) analytically, because the SV model is not a linear Gaussian state space model. Kitagawa(1987) suggested a linear spline technique for approximating the nonlinear filter. Fridman and Harris(1998) and Watanabe(1999) applied this technique to the SV model.

We propose an alternative filtering algorithm which analytically evaluates the integrations in (6) and (8). The LA approach is constructed from the following two steps. First, we approximate $\log f(y_t|h_t)f(h_t|Y_{t-1})$ up to the second order Taylor expansion around the mode of $f(y_t|h_t)f(h_t|Y_{t-1})$, i.e.

$$\begin{aligned} \ell(h_t) &\equiv \log f(y_t|h_t)f(h_t|Y_{t-1}) \\ &\simeq \ell(h_t^*) + \frac{1}{2}\ell''(h_t^*)(h_t - h_t^*)^2, \end{aligned} \quad (10)$$

where

$$h_t^* = \arg \max_{h_t} f(y_t|h_t)f(h_t|Y_{t-1}), \quad t = 1, \dots, T. \quad (11)$$

The method of approximation in the equation (10) is the key idea of the LA approach. This idea of approximation is found in Shephard and Pitt(1997, p656) in the context of pseudo-dominating Metropolis sampler. From the equation (3), we write the initial normal density of h_1 as $f_N(h_1|\mu_{1|0}, s_{1|0}^2)$ where $\mu_{1|0} = \mu/(1-\beta)$ and $s_{1|0}^2 = \sigma_\eta^2/(1-\beta^2)$. Given $\mu_{t|t-1}$ and $s_{t|t-1}^2$, the second derivative in the right hand side of (10) is expressed as

$$\begin{aligned} \ell''(h_t^*) &= \frac{d^2 \log f(y_t|h_t^*)}{dh_t^2} - \frac{1}{s_{t|t-1}^2} \\ &= -\frac{y_t^2}{2} \exp(-h_t^*) - \frac{1}{s_{t|t-1}^2}. \end{aligned} \quad (12)$$

Second, instead of the algorithm of (6)-(8), we conduct the following filtering algorithm.

(i) one step ahead prediction of y_t :

$$\begin{aligned} f(y_t|Y_{t-1}) &= \int_{-\infty}^{\infty} f(y_t|h_t)f(h_t|Y_{t-1})dh_t & (13) \\ &\simeq \int_{-\infty}^{\infty} \exp\left\{\ell(h_t^*) - \frac{1}{2s_{t|t}^2}(h_t - \mu_{t|t})^2\right\} dh_t \\ &= \sqrt{2\pi s_{t|t}^2} \exp(\ell(h_t^*)), \end{aligned}$$

where

$$\mu_{t|t} = h_t^* \tag{14}$$

$$s_{t|t}^2 = -\ell''(h_t^*)^{-1} \tag{15}$$

(ii) updating of h_t :

$$\begin{aligned} f(h_t|Y_t) &= \frac{f(y_t|h_t)f(h_t|Y_{t-1})}{f(y_t|Y_{t-1})} & (16) \\ &\simeq \frac{1}{\sqrt{2\pi s_{t|t}^2}} \exp(-\ell(h_t^*)) \exp\left\{\ell(h_t^*) - \frac{1}{2s_{t|t}^2}(h_t - \mu_{t|t})^2\right\} \\ &= f_N(h_t|\mu_{t|t}, s_{t|t}^2), \end{aligned}$$

(iii) one step ahead prediction of h_t :

$$\begin{aligned} f(h_{t+1}|Y_t) &= \int_{-\infty}^{\infty} f(h_{t+1}|h_t)f(h_t|Y_t)dh_t & (17) \\ &= f_N(h_{t+1}|\mu_{t+1|t}, s_{t+1|t}^2), \end{aligned}$$

where

$$\mu_{t+1|t} = \alpha + \beta\mu_{t|t} \tag{18}$$

$$s_{t+1|t}^2 = \beta^2 s_{t|t}^2 + \sigma_\eta^2. \tag{19}$$

One step ahead prediction of h_t has the same expression as that of the standard Kalman filter.

The distribution of h_t and h_{t+1} conditional on Y_t are normal for all $t = 1, \dots, T$, and we can calculate the approximated likelihood. We have the estimates of parameters of the SV models by maximizing the likelihood. The relationship of the Laplace approximation of this paper to importance sampling is stated in Appendix B.

2.3 Smoothing of the volatility process

The approximation of smoothing density of h_t can be also easily calculated. The smoothing density of h_t is expressed as

$$\begin{aligned} f(h_t|Y_T) &= \int_{-\infty}^{\infty} f(h_t, h_{t+1}|Y_T) dh_{t+1} \\ &= f(h_t|Y_t) \int_{-\infty}^{\infty} \frac{f(h_{t+1}|Y_T) f(h_{t+1}|h_t, Y_t)}{f(h_{t+1}|Y_t)} dh_{t+1}. \end{aligned} \quad (20)$$

Since $f(h_{t+1}|Y_T)$ and $f(h_{t+1}|Y_t)$ are approximated by normal density functions, the integration of (20) can be analytically solved. The smoothing density of h_t is

$$f(h_t|Y_T) \simeq f_N(h_t|\mu_{t|T}, s_{t|T}^2) \quad (21)$$

where

$$\mu_{t|T} = \mu_{t|t} + J_t(\mu_{t+1|T} - \mu_{t+1|t}) \quad (22)$$

$$s_{t|T}^2 = s_{t|t}^2 + J_t^2(s_{t+1|T}^2 - s_{t+1|t}^2) \quad (23)$$

$$J_t = \beta s_{t|t}^2 / s_{t+1|t}^2. \quad (24)$$

Hence, the smoothing estimates of volatility are obtained by

$$\begin{aligned} \sigma_{t|T}^2 &\equiv \text{Var}(y_t|Y_T) \\ &= E(\exp(h_t)|Y_T) \\ &= \exp(\mu_{t|T} + s_{t|T}^2/2) \end{aligned} \quad (25)$$

with variance

$$\begin{aligned} \text{Var}(\sigma_{t|T}^2) &= E(\exp(2h_t)|Y_T) - \{E(\exp(h_t)|Y_T)\}^2 \\ &= \exp(2\mu_{t|T} + 2s_{t|T}^2) \{1 - \exp(-s_{t|T}^2)\}. \end{aligned} \quad (26)$$

The smoothing estimates of the square root of volatility are also calculated by

$$\begin{aligned} \sigma_{t|T}^* &\equiv E(\exp(h_t/2)|Y_T) \\ &= \exp(\mu_{t|T}/2 + s_{t|T}^2/8) \end{aligned} \quad (27)$$

with variance

$$\text{Var}(\sigma_{t|T}^*) = \exp(\mu_{t|T} + s_{t|T}^2/2) \{1 - \exp(-s_{t|T}^2/4)\}. \quad (28)$$

3 Simulation experiments

We conduct simulation experiments for the univariate SV model with normal distribution to investigate the finite sample properties of the LA and to compare the LA with Monte-Carlo likelihood (MCL), the maximum likelihood of Fridman and Harris (1998) (F&H's ML), and MCMC in terms of the estimation of parameters as well as the estimation of volatilities.

3.1 Monte-Carlo set-up

Following the design of Jaquire, et al.(1994) and Sandmann and Koopman(1998), the nine sets of parameters are selected, thus facilitating direct comparison with the MCL, F&H's ML and MCMC methods. The values of the autoregressive parameter β are set to 0.90, 0.95, and 0.98. For each value of β , the value of σ_η are selected so that the coefficient of variation (CV) of h_t takes the values 10.0, 1.0, and 0.1. The values of the location parameter α are chosen so that the expected value of h_t is $E[h_t] = 0.009$.

We generate $\{h_t\}_{t=1}^T$ and $\{y_t\}_{t=1}^T$ that follow the SV model in the equations (1)-(3), and estimate the parameters and calculate the smoothing estimates of volatilities by using the LA approach. The sample size is $T = 500$ or 2000 . We maximize the likelihood function by using the simplex method at the first stage and the Newton-Rapson method at the second stage. Numerical derivatives are used throughout. Fortran 90 is used for programming. The computing time for the number of simulated realizations of the process $K = 500$ and the length of sample $T = 500$ takes about four minutes on Pentium 800MHz PCs for each set of parameters.

3.2 Parameter Estimates

Results from the sampling experiments for $T = 500$ are presented in Table.1 which is divided into three panels in accordance with the CV. For each value of CV the mean estimates for the LA, MCL, F&H's ML, and MCMC estimators are reported. The results of the MCMC, MCL, F&H's ML estimators are respectively taken from Table.7 of Jaquire et al.(1994), Table.2 of Sandmann and Koopman(1998), and Table.1 of Fridman and Harris(1998). The entries denotes

the mean estimate for each estimator, and the values in the parentheses denote the root mean squared errors (RMSE).

We observe from Table.1 the following facts : (i) All four estimators exhibit similar efficiency across most parameter values except for the MCL estimator in the case of $CV = 0.1$. (ii) All four estimators deteriorate as CV decreases. For $CV = 0.1$, The LA, F&H's ML and MCMC estimators exhibit similar efficiency. The MCL estimates of α are most efficient, but the MCL estimates of β are least efficient in this region. (iii) In terms of bias, the LA and F&H's ML have a common property. In all nine cases, both the LA and F&H's ML exhibit small downward bias for estimating β , but upward bias for estimating α and σ_η . On the other hand, in terms of the RMSE the F&H's ML estimator of β are smaller than the LA estimator for $CV = 10$ and 1 , but the magnitudes of RMSE are reversed for $CV = 0.1$.

Next, we examine the effects of increase in the sample size. Table.2 presents the performance of the estimators in the case of $T = 2000$ and $CV = 1.0$. We compare the LA with the MCL, MCMC, and NFML of Watanabe(1999). The NFML is similar in idea to that of Fridman and Harris(1998). Fridman and Harris(1998) does not carry out experiments for the present case. Table.2 shows the means and the RMSE of the LA, MCL, NFML and MCMC. The N of the NFML stands for the number of segments in numerical integration. Watanabe(1999) uses the two numbers $N = 25$ and 50 . The results of the MCL, NFML and MCMC estimators are taken from Tables 3 of Sandmann and Koopman(1998), Table.1 of Watanabe(1999) and Table.9 of Jacquire et al.(1994).

We observe from Table.2 the following facts : (i) The bias and the RMSE of the LA are smaller than those in Table.1. This implies that the parameters are more accurately estimated when the sample size increases. (ii) In terms of the mean, the LA estimator is comparable to the NFML ($N = 25$), but the LA is dominated by the NFML ($N = 50$) and MCL. However, in terms of the RMSE, the LA dominates NFML ($N = 25$), and it is as good as the NFML ($N = 50$).

From the simulation experiments, we may conclude that the small sample performance of the LA estimator is comparable to the MCL, F&H's ML, NFML, and the Bayesian MCMC methods for estimating the parameters.

Table.1 Mean and root mean square error of the estimators :

T = 500

CV = 10	α	β	σ_η	α	β	σ_η	α	β	σ_η
TRUE	-0.821	0.9	0.675	-0.411	0.95	0.484	-0.164	0.98	0.308
LA	-0.905 (0.278)	0.880 (0.037)	0.727 (0.097)	-0.510 (0.226)	0.931 (0.031)	0.534 (0.089)	-0.259 (0.172)	0.965 (0.023)	0.343 (0.066)
F&H's ML	-0.896 (0.280)	0.890 (0.034)	0.685 (0.080)	-0.505 (0.180)	0.940 (0.020)	0.495 (0.070)	-0.100 (0.080)	0.986 (0.010)	0.320 (0.050)
MCL	-0.837 (0.034)	0.915 (0.025)	0.579 (0.119)	-0.417 (0.021)	0.953 (0.020)	0.436 (0.077)	-0.166 (0.010)	0.977 (0.020)	0.290 (0.053)
MCMC	-0.679 (0.220)	0.916 (0.026)	0.562 (0.120)	-0.464 (0.160)	0.940 (0.020)	0.460 (0.055)	-0.190 (0.080)	0.980 (0.010)	0.350 (0.060)
CV = 1.0	α	β	σ_η	α	β	σ_η	α	β	σ_η
TRUE	-0.736	0.9	0.363	-0.368	0.95	0.26	-0.147	0.98	0.166
LA	-0.926 (0.424)	0.872 (0.059)	0.422 (0.108)	-0.526 (0.390)	0.927 (0.053)	0.303 (0.089)	-0.278 (0.246)	0.961 (0.034)	0.200 (0.067)
F&H's ML	-0.870 (0.430)	0.880 (0.050)	0.370 (0.080)	-0.510 (0.306)	0.930 (0.040)	0.280 (0.070)	-0.090 (0.060)	0.987 (0.015)	0.180 (0.040)
MCL	-0.745 (0.022)	0.897 (0.100)	0.325 (0.080)	-0.372 (0.011)	0.93 (0.102)	0.233 (0.075)	-0.148 (0.010)	0.97 (0.071)	0.161 (0.050)
MCMC	-0.870 (0.340)	0.880 (0.046)	0.350 (0.067)	-0.560 (0.340)	0.920 (0.046)	0.280 (0.065)	-0.220 (0.140)	0.970 (0.020)	0.230 (0.080)
CV = 0.1	α	β	σ_η	α	β	σ_η	α	β	σ_η
TRUE	-0.706	0.9	0.135	-0.353	0.95	0.096	-0.141	0.98	0.061
LA	-1.227 (1.552)	0.827 (0.217)	0.178 (0.137)	-0.763 (1.161)	0.892 (0.163)	0.133 (0.115)	-0.489 (0.976)	0.931 (0.136)	0.099 (0.107)
F&H's ML	-1.360 (1.720)	0.810 (0.240)	0.160 (0.120)	-0.810 (1.150)	0.886 (0.160)	0.120 (0.090)	-0.537 (1.130)	0.924 (0.160)	0.088 (0.090)
MCL	-0.709 (0.010)	0.443 (0.770)	0.156 (0.112)	-0.355 (0.010)	0.526 (0.735)	0.136 (0.108)	-0.142 (0.001)	0.572 (0.726)	0.113 (0.113)
MCMC	-1.540 (1.350)	0.780 (0.190)	0.150 (0.082)	-1.120 (1.150)	0.840 (0.160)	0.120 (0.074)	-0.660 (0.830)	0.910 (0.120)	0.140 (0.099)

Note : The table shows the mean and the RMSE (in parentheses). These entries are calculated from the K = 500 simulated samples with the T = 500 length of samples. MCL, F&H's ML and MCMC are respectively obtained from Table.2 of Sandmann and Koopman(1998), Table.1 of Fridman and Harris(1998) and Table.7 of Jacquire et al.(1994) respectively. The RMSE of MCL is calculated from the bias and the standard deviation in Table.2 of Sandmann and Koopman(1998).

Table.2 Mean and root mean square error of the estimators :

T = 2000

Method	α	β	σ_η	α	β	σ_η	α	β	σ_η
	-0.736	0.9	0.363	-0.368	0.95	0.26	-0.147	0.98	0.166
LA	-0.819 (0.161)	0.886 (0.022)	0.411 (0.045)	-0.427 (0.107)	0.940 (0.015)	0.293 (0.035)	-0.179 (0.065)	0.975 (0.009)	0.183 (0.026)
MCL	-0.745 (0.013)	0.913 (0.024)	0.317 (0.055)	-0.372 (0.011)	0.954 (0.011)	0.239 (0.037)	-0.148 (0.001)	0.980 (0.010)	0.1584 (0.022)
NFML									
N=25	-0.812 (0.199)	0.890 (0.027)	0.406 (0.068)	-0.426 (0.124)	0.942 (0.017)	0.294 (0.052)	-0.194 (0.083)	0.974 (0.011)	0.197 (0.043)
N=50	-0.766 (0.168)	0.895 (0.023)	0.368 (0.041)	-0.406 (0.106)	0.945 (0.014)	0.264 (0.032)	-0.178 (0.067)	0.976 (0.009)	0.169 (0.024)
MCMC	-0.762 (0.150)	0.896 (0.020)	0.359 (0.034)						

Note : The table shows the mean and the RMSE (in parentheses). These entries are calculated from the K = 500 simulated samples with the T = 2000 length of samples for LA, MCL and MCMC and from the K = 1000 with T = 2000 for NFML. The MCL, NFML and MCMC are respectively obtained from Table.3 of Sandmann and Koopman(1998), Table.1 of Watanabe(1999) and Table.9 of Jacquire et al.(1994). The N of NFML stands for the number of segments in numerical integration.

3.3 Volatility Estimates

Next, the finite sample performance of the LA estimators of volatilities is compared with that of the F&H's ML and MCMC methods. Sandmann and Koopman(1998) did not report the volatility estimates. Following Jacquier et al.(1994), the criterion for evaluating the performance is the grand root mean squared error(GRMSE),

$$\text{GRMSE} = \sqrt{\frac{1}{K(T-199)} \sum_{i=1}^K \sum_{t=100}^{T-100} (\sigma_{i,t}^2 - \hat{\sigma}_{i,t}^2)^2}, \quad (29)$$

where $\sigma_{i,t}^2$ is the true volatility simulated at the period t in the *i*th simulation and $\hat{\sigma}_{i,t}^2$ denotes the smoothing estimate of volatility given by the equation (25).

We observe from Table.3 the following facts : (i) The GRMSE of all the three estimators decreases as CV decreases, and as the true value of β increases to 1.0.

Table.3 GRMSE of volatility estimates :

	β		
	0.90	0.95	0.98
CV = 10			
LA	18.39	14.65	10.95
F&H's ML	21.10	17.00	12.20
MCMC	22.10	18.70	12.50
CV = 1.0			
LA	6.21	5.36	4.44
F&H's ML	5.90	5.30	5.04
MCMC	6.00	5.30	5.10
CV = 0.1			
LA	2.65	2.41	2.04
F&H's ML	2.60	2.40	2.20
MCMC	2.60	2.46	2.27

Note : GRMSE \times 10000 is displayed. These entries are calculated from the $K = 500$ simulated samples with the $T = 500$ length of samples. The F&H's ML and the MCMC are respectively obtained from Table.3 of Fridman and Harris(1998) and Table.10 of Jacquire et al.(1994).

(ii) The GRMSE of the LA estimator are smallest among the three estimators for the case of $\beta = 0.98$, but the GRMSE of the LA are largest for the case of $\beta = 0.90$. (iii) The GRMSE of the LA are smaller than those of the F&H's ML for 5 cases out of the 9 cases, while the GRMSE of the LA are smaller than those of the MCMC for 6 out of 9.

We calculated the GRMSE of volatility estimates for $T = 2000$, and compared the results with NFML. The results is similar to the case of $T = 500$, although we do not report it here.

Simulation experiments in Sections 3.2 and 3.3 reveals that the LA method is comparable to the MCL, F&H's ML and MCMC methods. The LA approach is flexible and easily extended to the more complicated SV models as is shown in Section 4.

4 Comparison of the LA with MCL via empirical studies on the stock markets

This section empirically illustrates how the LA approach works when it is applied to the daily returns on the stock markets and compares the LA with the MCL.

4.1 Univariate SV models

(i) The data set

The continuously compounded returns are calculated from the daily closing prices for the Tokyo Stock Price Index (TOPIX) from January 4, 1995 to December 30, 2001. The sample size is 1578. Though y_t is assumed to follow a stationary process with zero mean and no autocorrelation in the simple SV model, the returns on the stock prices often have weak autocorrelations. We remove the mean and autocorrelations from the return series by using the first-order autoregression:

$$R_t = a + bR_{t-1} + \zeta_t, \quad \zeta_t \sim NID(0, \sigma_\zeta^2) \quad (30)$$

where R_t denotes the daily returns¹⁾.

Table.4 Estimation results of preliminary regression

Parameter	a	b	σ_ζ	LB(12)
Estimate	-0.009	0.074	1.565	19.304
Standard Error	(0.032)	(0.020)	(0.037)	

Note: White(1980)'s heteroskedasticity corrected standard errors are in the parentheses. The last column denotes the heteroskedasticity corrected Ljung-Box statistic for twelve lags of the residual autocorrelations which is calculated from the method of Diebold(1988). Its p-values is 0.081.

Table.4 shows the estimated coefficients and their heteroskedasticity consistent standard errors (White(1980)). Since the TOPIX returns have the significant first order autocorrelation, we define y_t as the residuals from the regression (30) in the following analysis.

¹⁾ We estimated the AR models with lag lengths 1 through 4. The SBIC was maximized at the lag length of 1.

(ii) The SV model with normal distribution

We estimate the SV model in the equations (1)-(3) by maximizing the likelihood function (9). We assume the consistency and the asymptotic normality of the LA estimator even if we are not able to prove them.

Table.5 Estimates of the SV model

Parameter	α	β	σ_η
LA	0.020 (0.008)	0.948 (0.018)	0.224 (0.043)
MCL	0.007 (0.005)	0.962 (0.014)	0.177 (0.032)
NFML	0.009 (0.006)	0.957 (0.015)	0.193 (0.035)
MCMC	0.195	0.952	0.202

Note: The standard errors of estimators are in the parentheses. The number of draws in the MCL is $M = 5$. The number of segments in the NFML is $N = 100$. In the MCMC, we discard the first 1500 sample draws, and we use the after 2500 sample draws to estimate parameters.

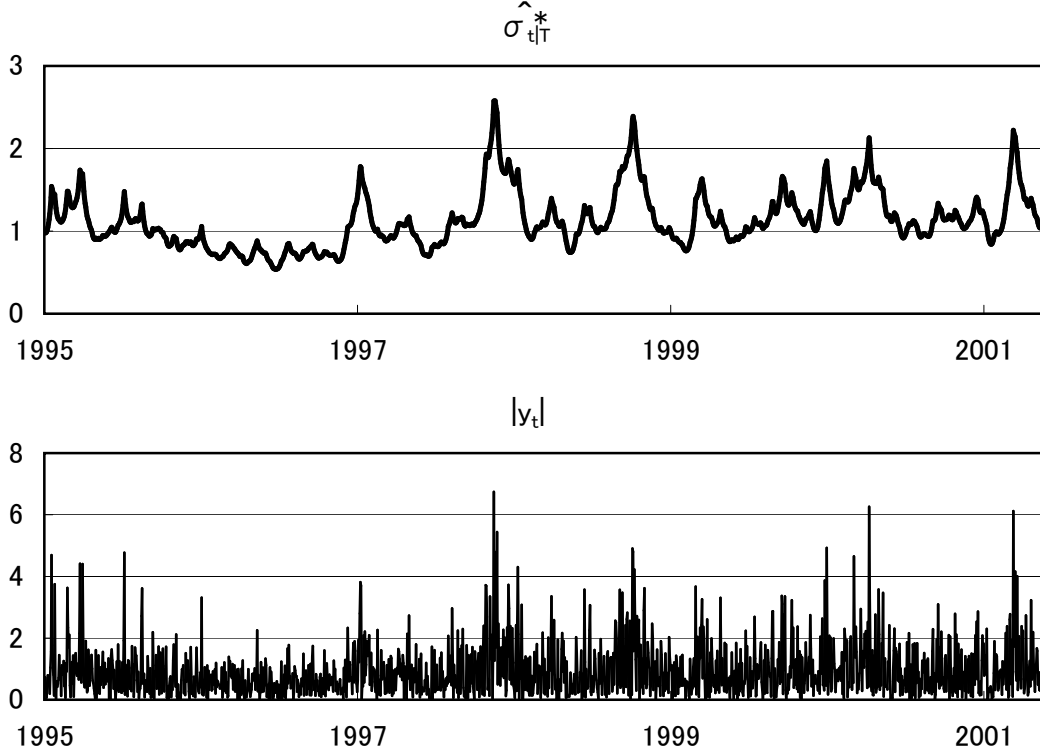
For the purpose of comparison, we calculate the estimates of the MCL, NFML and MCMC in addition to the estimates of the LA. Table.5 shows the results²⁾. The estimators for all methods have virtually identical values excepts for the estimate of α by the MCMC. The all estimates of β are significant and indicates strong persistency of volatility.

We calculate the smoothing estimates of the standard deviations using the formula in the equation (27). Figure.1 plots the smoothing estimates of the square root of volatility by using the LA and the absolute values of y_t . Although the volatilities are not observable, one may think that the absolute returns reflect the fluctuation of the volatilities. We can see that the smoothed estimates of squared root of the volatilities ($\hat{\sigma}_{t|T}^*$) move in correlation to the absolute returns. We also calculated the smoothing estimates by using other methods. Since the graphs of alternative methods are not distinguishable from the graph of the LA, we do not

²⁾ The algorithms for the MCL, NFML and MCMC were written by using Fortran 90. The entries in the parentheses stand for the standard errors.

present them.

Figure.1 Smoothing estimates of square root of the volatility and the absolute values of y_t



(iii) The SV model with t-distribution

It is widely known that the densities of many financial time series exhibit larger kurtosis than that can be explained by the standard SV models with normal error distribution. The SV model with a fat tail distribution have been proposed to deal with this problem.

We examine the SV model with t-distribution. To estimate this type of SV model, we have only to replace the equation (4) with

$$f(y_t|h_t) = \frac{1}{\sqrt{\pi(\nu-2)}} \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]} e^{-h_t/2} \left(1 + \frac{y_t^2 e^{-h_t}}{\nu-2}\right)^{-\frac{\nu+1}{2}} \quad (31)$$

where ν represents a parameter of degree of freedom and $\Gamma[\cdot]$ stands for the Gamma function, and to replace the equation (12) with

$$\ell''(h_t^*) = -\frac{\nu+1}{2} \left\{ \frac{y_t^2}{\nu-2} \exp(-h_t^*) \right\} \left\{ 1 + \frac{y_t^2}{\nu-2} \exp(-h_t^*) \right\}^{-2} - \frac{1}{s_{t|t-1}^2}. \quad (32)$$

Smoothing estimates of volatilities can be calculated in a similar manner to the case of normal distribution. We use the data set in section 4.1 (i).

Table.6 Estimation results of the SV model with t-distribution

Parameter	α	β	σ_η	$1/\nu$
LA	0.010 (0.005)	0.972 (0.011)	0.141 (0.029)	0.081 (0.013)
MCL	0.005 (0.004)	0.978 (0.009)	0.123 (0.025)	0.088 (0.011)
NFML	0.006 (0.004)	0.975 (0.010)	0.133 (0.027)	0.088 (0.016)

Note: The standard errors are in the parentheses. The number of draws in the MCL is $M = 5$. The number of segments in the NFML is $N = 100$.

Table.6 shows the estimates of parameters for the SV model with t-distribution. This result is comparable to Sandmann and Koopman (1998). They estimated the SV models with either normal or t-distribution. The estimate of $1/\nu$ and its standard error suggest that the error distribution of the process follows a fat tail distribution. Comparing the estimates of this section with those in Table.5, it turns out that the estimate of β with t-distribution is higher than that with standard normal distribution. Sandmann and Koopman (1998) observed the same result by using the data for the S&P500 stock index.

4.2 A multivariate SV model

We consider the multivariate SV model proposed by Harvey, Ruiz and Shephard(1994). Let a $p \times 1$ vector \mathbf{y}_t follow the stochastic process

$$\mathbf{y}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\epsilon}_t \quad , \quad \boldsymbol{\epsilon}_t \sim NID(\mathbf{0}, \mathbf{I}_p) \quad (33)$$

where $\mathbf{0}$ is a $p \times 1$ vector of zeros, \mathbf{I}_p is a $p \times p$ identity matrix and $\boldsymbol{\Sigma}_t$ is a $p \times p$ time-varying volatility matrix. The matrix $\boldsymbol{\Sigma}_t$ consists of a $p \times p$ time-invariant correlation matrix \mathbf{R} and a $p \times p$ time-varying scale matrix $\mathbf{H}_t = \text{diag}(\exp(h_{1t}), \dots, \exp(h_{pt}))$ as,

$$\boldsymbol{\Sigma}_t = \mathbf{H}_t^{1/2} \mathbf{R} \mathbf{H}_t^{1/2}. \quad (34)$$

The $p \times 1$ vector $\mathbf{h}_t = (h_{1t}, \dots, h_{pt})'$ is specified by the stationary VAR(1) process with a Gaussian noise

$$\mathbf{h}_{t+1} = \mathbf{a} + \mathbf{B}\mathbf{h}_t + \boldsymbol{\eta}_{t+1} \quad , \quad \boldsymbol{\eta}_{t+1} \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \quad (35)$$

$$\mathbf{h}_1 \sim N((\mathbf{I} - \mathbf{B})^{-1}\mathbf{a}, \mathbf{W}) \quad (36)$$

where \mathbf{a} and \mathbf{B} are respectively a $p \times 1$ vector and a $p \times p$ matrix of coefficients and $\boldsymbol{\Sigma}_\eta$ is a $p \times p$ covariance matrix of $\boldsymbol{\eta}_t$. The covariance matrix of $\boldsymbol{\eta}_1$ satisfies that

$$\mathbf{W} = \mathbf{B}\mathbf{W}\mathbf{B}' + \boldsymbol{\Sigma}_\eta \quad (37)$$

The multivariate SV model defined by the equations (33)-(37) is a natural extension of the univariate SV model defined by the equations (1) - (3). The density function of \mathbf{y}_t given \mathbf{h}_t and that of \mathbf{h}_{t+1} given \mathbf{h}_t are respectively

$$f(\mathbf{y}_t|\mathbf{h}_t) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}_t|^{-1/2} \exp\{-0.5 \mathbf{y}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{y}_t\} \quad (38)$$

$$f(\mathbf{h}_{t+1}|\mathbf{h}_t) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}_\eta|^{-1/2} \exp\{-0.5 (\mathbf{h}_{t+1} - \mathbf{a} - \mathbf{B}\mathbf{h}_t)'\boldsymbol{\Sigma}_\eta^{-1}(\mathbf{h}_{t+1} - \mathbf{a} - \mathbf{B}\mathbf{h}_t)\}.$$

Replacing the equations (4) and (5) with the equations (38) and (39) respectively, we can apply the LA approach to the multivariate SV model. The details of the filtering and smoothing algorithms are stated in the Appendix.

We apply the above model to the daily returns of the NYSE Composite Index and TOPIX for the purpose of numerical illustrations. The continuously compounded returns are calculated from the daily closing prices for the TOPIX and the NYSE Composite Index from January 4, 1995 to December 30, 2001. The sample size is 1533³⁾. The returns on each index are respectively adjusted for the AR(1) model by the same fashion as explained in Section 4, and the residuals are used for analyzing the model⁴⁾. We define y_{1t} as the returns on the US market and y_{2t} on the Japanese market.

Table.7 shows the estimates of the multivariate SV model. The contemporary correlation of the returns between the US and Japanese markets is very small

³⁾If Japanese or US market closed on a day, we assume both two markets are closed on that day.

⁴⁾We omit the estimation results of AR(1) model for the NYSE index to save the space.

Table.7 Estimation results of the multivariate SV model

	LA	MCL
a_1	0.031 (0.014)	0.011 (0.011)
a_2	0.042 (0.018)	0.019 (0.012)
\mathbf{B}_{11}	0.998 (0.016)	0.997 (0.013)
\mathbf{B}_{12}	-0.075 (0.028)	-0.053 (0.021)
\mathbf{B}_{21}	0.012 (0.019)	0.010 (0.014)
\mathbf{B}_{22}	0.909 (0.037)	0.938 (0.025)
$\Sigma_{\eta 11}$	0.065 (0.019)	0.043 (0.012)
$\Sigma_{\eta 12}$	0.066 (0.022)	0.040 (0.013)
$\Sigma_{\eta 22}$	0.082 (0.031)	0.049 (0.016)
\mathbf{R}_{12}	0.105 (0.027)	0.091 (0.024)

Note: The standard errors are in the parentheses. The number of draws $M = 5$ in the MCL.

($\mathbf{R}_{12} = 0.105$ for LA and $= 0.091$ for MCL). The eigenvalues of the matrix \mathbf{B} are 0.986, 0.921 for LA and 0.986, 0.946 for MCL, which observation implies that the volatility process of the both markets are stationary but highly auto-correlated. The correlation coefficient of the shocks to the volatility processes is $\Sigma_{\eta_{12}}(\Sigma_{\eta_{11}}\Sigma_{\eta_{22}})^{-1/2} = 0.908$ for LA and $= 0.878$ for MCL. High correlation coefficient may indicate that the shock caused by an event simultaneously affects the volatilities of the US and Japanese markets. Hence the volatility process of the two markets are contemporaneously and strongly correlated. The both estimates of LA and MCL give similar values for most of the parameters.

4.3 The dynamic bivariate mixture model

Tauchen and Pitt (1983) observed that the large fluctuations of returns have a tendency to coincide with the large trading volumes. Andersen (1996), Lisenfeld(1998) and Watanabe(2000) combined the model of Tauchen and Pitt (1983) with the information flow to the market. In their models, the returns on the stock and the trading volumes follow the system

$$y_t|h_t \sim N(0, \exp(h_t)) \quad (40)$$

$$V_t|h_t \sim N(\mu_V \exp(h_t), \sigma_V^2 \exp(h_t)) \quad (41)$$

where V_t is the trading volume and $\exp(h_t)$ is interpreted as the information flow to the market. This model is called the dynamic bivariate mixture (DBM) model. The DBM model is expressed by the equations (1)-(3) and

$$V_t = \mu_V \exp(h_t) + \sigma_V \exp(h_t/2)\epsilon_{Vt} \quad , \quad \epsilon_{Vt} \sim NID(0, 1). \quad (42)$$

In order to estimate the parameters of the model, we have only to replace the equation (4) with

$$f(y_t, V_t|h_t) = \frac{1}{2\pi\sigma_V} \exp\left\{-\frac{1}{2}\left(y_t^2 + \frac{V_t^2}{\sigma_V^2}\right)\exp(-h_t) + \frac{\mu_V V_t}{\sigma_V^2} - \frac{\mu_V^2}{2\sigma_V^2}\exp(h_t) - h_t\right\} \quad (43)$$

and the equation (12) with

$$\ell''(h_t^*) = -\frac{1}{2}\left(y_t^2 + \frac{V_t^2}{\sigma_V^2}\right)\exp(-h_t^*) - \frac{\mu_V^2}{2\sigma_V^2}\exp(h_t^*) - \frac{1}{s_{t|t-1}^2}, \quad (44)$$

and to apply the algorithm in Section 2.2. The smoothing estimates of the volatilities are obtained by the same algorithm as the one in Section 2.3.

We apply the DBM model to the Japanese market, and investigate the relationship of the returns on the TOPIX and the trading volumes. The trading volume is measured in terms of one-billion shares traded during the day. The sample periods are the same as those used in Section 4.1. The residual returns from the AR(1) process are analyzed by the same reason as explained in the previous section.

Table.8 Estimation results of the DBM model

Parameter	α	β	σ_η	μ_v	σ_v
LA	0.046 (0.009)	0.865 (0.016)	0.203 (0.008)	0.338 (0.012)	0.042 (0.004)
MCL	0.049 (0.009)	0.861 (0.016)	0.204 (0.008)	0.332 (0.013)	0.042 (0.004)

Note: Standard errors are in the parentheses.

Table.8 shows the estimates of parameters of the DBM model by the LA and the MCL methods. The volatility persistence parameter is highly significant ($\beta = 0.865(0.861)$ for LA(MCL)). However, the estimated values are lower than those of the standard SV model reported in Table.5. The empirical result is conformable to the studies of Andersen (1996) and Lisenfeld (1998) while we employed the different method of estimation for the different data set from their studies. The LA and MCL give almost identical estimates.

5 Concluding remarks

This paper proposed the Laplace approximation method to the nonlinear state space representation and applied it for estimating the SV models.

This method approximates the logarithm of the joint density of current observation and volatility variables conditional on the past observations by the second order Taylor expansion around its mode, and then applies the nonlinear filtering algorithm. The MCL (Sandmann and Koopman(1998)) is now a standard clas-

sical method for estimating the SV models. It is based on importance sampling technique. Importance sampling is regarded as an exact method for maximum likelihood estimation. We showed that the LA method of this paper approximates the weight function by unity in the context of importance sampling. We do not need to carry out the Monte Carlo integration for obtaining the likelihood since the approximate likelihood function can be analytically obtained. If one-step ahead prediction density of observation and volatility variables conditional on the past observations is sufficiently accurately approximated, the LA method is workable.

We examined how the approximation works by simulations as well as various empirical studies. We conducted the Monte-Carlo simulations for the univariate SV model for examining the small sample properties and compared them with other methods. Simulation experiments revealed that our method is comparable to the MCL, Maximum Likelihood (Fridman and Harris (1998)) and MCMC methods. We applied the LA method to the univariate SV models with normal distribution or t-distribution, the bivariate SV model and the dynamic bivariate mixture model, and empirically illustrated how it works for each of the extended models. The empirical results on the stock markets revealed that our method provides very similar estimates of coefficients to those of the MCL.

The interest of this paper is whether the LA method is workable for estimating SV models in practice. We showed workability in two ways by comparing the approximation with the MCL; first the simulation studies, second the empirical studies. Naturally, the workability is limited to the cases we have examined. But we believe the LA method is applicable to many SV models based on our study of this paper.

Appendix A : The LA algorithm for the multivariate SV model

We explain the algorithm of the LA approach for the multivariate SV model. The algorithm follows the same lines as those in Sections 2.2 and 2.3.

A.1 Filtering, prediction and evaluation of the likelihood

Let the initial normal density in the equation (36) be

$$f(\mathbf{h}_1) = f_N(\mathbf{h}_1 | \boldsymbol{\mu}_{1|0}, \mathbf{S}_{1|0}). \quad (\text{A.1})$$

where $\boldsymbol{\mu}_{1|0} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{a}$ and $\mathbf{S}_{1|0} = \mathbf{W}$. The joint density of $(\mathbf{y}_t, \mathbf{h}_t)$ conditional on the past observations is approximated by

$$\begin{aligned} \ell(\mathbf{h}_t) &\equiv \log f(\mathbf{y}_t | \mathbf{h}_t) f(\mathbf{h}_t | \mathbf{Y}_{t-1}) \\ &\simeq \ell(\mathbf{h}_t^*) + \frac{1}{2}(\mathbf{h}_t - \mathbf{h}_t^*)' \ell''(\mathbf{h}_t^*)(\mathbf{h}_t - \mathbf{h}_t^*). \end{aligned} \quad (\text{A.2})$$

and

$$\mathbf{h}_t^* = \arg \max_{\mathbf{h}_t^*} f(\mathbf{y}_t | \mathbf{h}_t) f(\mathbf{h}_t | \mathbf{Y}_{t-1}) \quad (\text{A.3})$$

for $t = 1, \dots, T$. Given $\boldsymbol{\mu}_{t|t-1}$ and $\mathbf{S}_{t|t-1}$, the second derivative in (A.2) is expressed as

$$\ell''(\mathbf{h}_t) = -\frac{1}{4} \left\{ \tilde{\mathbf{Y}}_t \Sigma_t(\mathbf{h}_t^*)^{-1} \tilde{\mathbf{Y}}_t + \text{diag} \left(\left(\mathbf{y}'_t \Sigma_t(\mathbf{h}_t^*)^{-1} \tilde{\mathbf{Y}}_t \right)_i \right) \right\} - \mathbf{S}_{t|t-1}^{-1} \quad (\text{A.4})$$

where $\tilde{\mathbf{Y}}_t = \text{diag}(y_{1t}, \dots, y_{pt})$, $\mathbf{H}_t^* = \text{diag}(\exp(h_{1t}^*), \dots, \exp(h_{pt}^*))$, $\Sigma_t(\mathbf{h}_t^*) = \mathbf{H}_t^{*1/2} \mathbf{R} \mathbf{H}_t^{*1/2}$, and $\left(\mathbf{y}'_t \Sigma_t(\mathbf{h}_t^*)^{-1} \tilde{\mathbf{Y}}_t \right)_i$ is the i -th element of $\mathbf{y}'_t \Sigma_t(\mathbf{h}_t^*)^{-1} \tilde{\mathbf{Y}}_t$.

Then, we have the following algorithm.

(i) **one step ahead prediction of \mathbf{y}_t :**

$$f(\mathbf{y}_t | \mathbf{Y}_{t-1}) = (2\pi)^{p/2} |\mathbf{S}_{t|t}|^{1/2} \exp \{ \ell(\mathbf{h}_t^*) \} \quad (\text{A.5})$$

where

$$\boldsymbol{\mu}_{t|t} = \mathbf{h}_t^* \quad (\text{A.6})$$

$$\mathbf{S}_{t|t} = -\ell''(\mathbf{h}_t^*)^{-1}. \quad (\text{A.7})$$

(ii) **updating of \mathbf{h}_t :**

$$f(\mathbf{h}_t | \mathbf{Y}_t) = f_N(\mathbf{h}_t | \boldsymbol{\mu}_{t|t}, \mathbf{S}_{t|t}). \quad (\text{A.8})$$

(iii) one step ahead prediction of \mathbf{h}_t :

$$f(\mathbf{h}_{t+1}|\mathbf{Y}_t) = f_N(\mathbf{h}_t|\boldsymbol{\mu}_{t+1|t}, \mathbf{S}_{t+1|t}) \quad (\text{A.9})$$

where

$$\boldsymbol{\mu}_{t+1|t} = \mathbf{a} + \mathbf{B}\boldsymbol{\mu}_{t|t} \quad (\text{A.10})$$

$$\mathbf{S}_{t+1|t} = \mathbf{B}\mathbf{S}_{t|t}\mathbf{B}' + \boldsymbol{\Sigma}_\eta. \quad (\text{A.11})$$

The equations (A.5) - (A.11) exactly correspond to the equations (13)-(19) in Section 2.2.

A.2 Smoothing of the volatility process

The approximated smoothing density of \mathbf{h}_t is expressed as

$$f(\mathbf{h}_t|\mathbf{Y}_T) = f_N(\mathbf{h}_t|\boldsymbol{\mu}_{t|T}, \mathbf{S}_{t|T}) \quad (\text{A.12})$$

where

$$\boldsymbol{\mu}_{t|T} = \boldsymbol{\mu}_{t|t} + \mathbf{J}_t(\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{t+1|t}) \quad (\text{A.13})$$

$$\mathbf{S}_{t|T} = \mathbf{S}_{t|t} + \mathbf{J}_t(\mathbf{S}_{t+1|T} - \mathbf{S}_{t+1|t})\mathbf{J}'_t \quad (\text{A.14})$$

$$\mathbf{J}_t = \mathbf{S}_{t|t}\mathbf{B}'\mathbf{S}_{t+1|t}^{-1}. \quad (\text{A.15})$$

Then, smoothing estimates of volatility process is obtained by

$$\sigma_{it|T}^2 \equiv \text{Var}(y_{it}|Y_T) = \exp(\mu_{it|T} + s_{it|T}^2/2). \quad (\text{A.16})$$

with variance

$$\text{Var}(\sigma_{it|T}^2) = \exp(2\mu_{it|T} + 2s_{it|T}^2)\{1 - \exp(-s_{it|T}^2)\}. \quad (\text{A.17})$$

where $\mu_{it|T}$ is the i -th element of $\boldsymbol{\mu}_{t|T}$ and $s_{it|T}^2$ is the (i,i) -th element of $\mathbf{S}_{t|T}$. The equations (A.12) - (A.17) correspond to the equations (21)-(26) in Section 2.3.

Appendix B : Relationship between the LA and importance sampling

We state the relationship between the Laplace approximation and importance sampling. Let us define $y = (y_1, \dots, y_T)'$, $h = (h_1, \dots, h_T)'$. The marginal density of y is given as

$$f(y) = \int \dots \int \prod_{t=1}^T f(y_t|h_t) f_N(h_t|h_{t-1}) dh_1 \dots dh_T. \quad (\text{B.1})$$

Recall that $f(y_t, h_t|Y_{t-1})$ (one step ahead prediction of y_t and h_t conditional on Y_{t-1}) is approximated by

$$g(y_t, h_t|Y_{t-1}) = g(y_t|Y_{t-1}) f_N(h_t|Y_t), \quad (\text{B.2})$$

where $g(y_t|Y_{t-1}) = \sqrt{2\pi s_{t|t}^2} \exp(\ell(\mu_{t|t}))$ and $f_N(h_t|Y_t) \sim N(h_t|\mu_{t|t}, s_{t|t}^2)$. See equations (10), (14) and (15). Then, we have

$$f(y) = \int \dots \int \prod_{t=1}^T \left\{ g(y_t|Y_{t-1}) \frac{f(y_t, h_t|Y_{t-1})}{g(y_t, h_t|Y_{t-1})} f_N(h_t|h_{t+1}, Y_t) \right\} dh_1 \dots dh_T. \quad (\text{B.3})$$

where $f_N(h_t|h_{t+1}, Y_t) \sim N(\mu_{t|t+1}^*, s_{t|t+1}^{*2})$, $\mu_{t|t+1}^* = \mu_{t|t} + \beta s_{t|t}^2 s_{t|t+1}^{-2} (h_{t+1} - \mu_{t+1|t})$ and $s_{t|t+1}^{*2} = s_{t|t}^2 (1 - \beta^2 s_{t|t}^2 s_{t+1|t}^{-2})$. Here, we define $f_N(h_T|h_{T+1}, Y_T) = f_N(h_T|Y_T)$.

Finally, we obtain

$$f(y) = g(y) \tilde{E}\{W(h, y)\}, \quad (\text{B.4})$$

where

$$g(y) = \prod_{t=1}^T g(y_t|Y_{t-1}), \quad (\text{B.5})$$

$$W(h, y) = \prod_{t=1}^T \left\{ \frac{f(y_t, h_t|Y_{t-1})}{g(y_t, h_t|Y_{t-1})} \right\}, \quad (\text{B.6})$$

and $\tilde{E}\{\cdot\}$ stands for the expectation with respect to the multivariate normal distribution conditional on y :

$$f_N(h|y) = \prod_{t=1}^T f_N(h_t|h_{t+1}, Y_t). \quad (\text{B.7})$$

Equation (B.6) can be interpreted as a formula of importance sampling with importance density $f_N(h|y)$. However, since $f(y_t, h_t|Y_{t-1})$ is not easy to evaluate, equation(B.6) is not useful for importance sampling in practice.

The LA approximates the weight function as $W(h, y) \simeq 1$. Hence, the marginal density of y can be analytically obtained as $f(y) \simeq g(y)$. If one step ahead prediction density of (y_t, h_t) conditional on Y_{t-1} is sufficiently accurately approximated, the LA method is workable.

References

- Andersen, T.G., (1994), Stochastic autoregressive volatility : A framework for volatility modeling, *Mathematical Finance*, **4**, 75-102.
- Andersen, T. G., (1996), Return Volatility and Trading Volume : An Information Flow Interpretation of Stochastic Volatility, *Journal of Finance*, **51**, No.1, 169-204.
- Andersen, T. G., and B. Sørensen, (1996), GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study, *Journal of Business and Economic Statistics*, **14**, 328-352.
- Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **31**, 307-327.
- Bollerslev, T., (1987), A conditional heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics*, **69**, 542-547.
- Clark, P., (1973), A subordinated stochastic process model with finite variance for speculative process, *Econometrica*, **41**,135-155.
- Danielson, J., (1994), Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood, *Journal of Econometrics*, **61**, 375-400.
- Diebold, R.F., (1988), Empirical Modeling of Exchange Rate Dynamics, *Springer-Verlag, New York*.
- Durbin, J. and S. J. Koopman (1997), Monte Carlo maximum likelihood estimation of non-Gaussian state space model, *Biometrika*, **84**, 669-684.
- Durbin, J. and S. J. Koopman (2000), Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives, *Journal of Royal Statistical Society, B*, **62**, part1, 3-56.
- Engle, R.F., (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, **50**, 987-1007.
- Harvey, A. C., E. Ruiz and N. Shephard, (1994), Multivariate stochastic variance models, *Review of Economic Studies*, **61**, 247-264.
- Fridman, M. and L. Harris, (1998), A Maximum Likelihood Approach for Non-Gaussian stochastic Volatility Models, *Journal of Business and Economic Statistics*, **16**, 3, 284-291

- Hull, J. and A. White, (1987), The pricing options on assets with stochastic volatilities, *Journal of Finance*, **42**, 281-300.
- Jacquier, E., N. G. Polson and P. E. Rossi, (1994), Bayesian analysis of stochastic volatility models, (with discussion) , *Journal of Business and Economic Statistics*, **12**, 371-417.
- Kim, S., N. Shephard and S. Chib, (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies*, **65**, 361-393.
- Kitagawa, G., (1987), Non-Gaussian state-space modeling of nonstationary time series, *Journal of The American Statistical Association*. **82**, 1032-63, (with discussion).
- Lisenfeld, R., (1998), Dynamic Bivariate Mixture Models: Modeling the Behavior of Prices and Trading Volume, *Journal of Business and Economic Statistics*, **16**, 1, 101-109.
- Melino, A. and S. M. Turnbull, (1990), Pricing foreign currency options with stochastic volatility, *Journal of Econometrics*, **45**, 239-265.
- Nelson, D. B. (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, **59**, 347-370.
- Sandmann, G. and S.J. Koopman, (1998), Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood, *Journal of Econometrics*, **87**, No.2, 271-301.
- Shephard, N. and M.K. Pitt, (1997), Likelihood analysis of non-Gaussian measurement time series, *Biometrika*, **84**, 653-667.
- Taylor, S., (1986), Modeling Financial Time Series, *John Wiley & Sons, New York*.
- Taylor, S., (1994), Modeling Stochastic Volatility, *Mathematical Finance*, **4**, 183-204.
- Tauchen, G. and M. Pitt, (1983), The price variability-volume relationship on speculative markets, *Econometrica*, **51**,485-505.
- Watanabe, T., (1999), A non-linear filtering approach to stochastic volatility models with an application to daily stock returns, *Journal of Applied Econometrics*,

14, 101-121.

Watanabe, T., (2000), Bayesian Analysis of Dynamic Bivariate Mixture Models: Can They Explain the Behavior of Returns and Trading Volume?, *Journal of Business and Economic Statistics*, **18**, No.2, 199-210.

White, H., (1980), A heteroskedastic-consistent covariance matrix and a direct test for heteroskedasticity, *Econometrica*, **48**, 817-838.