BAYESIAN INFERENCE FOR NONLINEAR AND NON-GAUSSIAN STOCHASTIC VOLATILITY MODEL WITH LEVERAGE EFFECT

Tomohiro Ando*

Stochastic volatility (SV) models provide useful tools to describe the evolution of asset returns, which exhibit time-varying volatility. This paper extends a basic SV model to capture a leverage effect, a fat-tailed distribution of asset returns and a nonlinear relationship between the current volatility and the previous volatility process. The Bayesian approach with the Markov chain Monte Carlo method is employed to estimate model parameters. To assess the goodness of the estimated model, we calculated several Bayesian model selection criteria that include the Bayes factor, the Bayesian predictive information criterion and the deviance information criterion. The proposed method is tested on simulated data and then applied to daily returns on the Nikkei 225 index where several SV models are formally compared.

Key words and phrases: Bayesian model selection, B-splines, fat tail, leverage effect, Markov chain Monte Carlo.

1. Introduction

It is a well-known fact that the volatility of financial asset return changes randomly over time (Clark (1973)). The stochastic volatility (SV) model pioneered by Taylor (1982) has recently attracted the attention of financial economics researchers and practitioners. It has been shown that an European call option pricing based on the SV model is more accurate than that based on the Black-Scholes model (Melino and Turnbull (1990)). In the context of the basic SV model, the observation equation and the system equation are specified as follows:

\[
\begin{align*}
    y_t &= \exp(h_t/2)u_t, \\
    h_t &= \mu + \phi(h_{t-1} - \mu) + \tau v_t, \\
    (t = 1, \ldots, n).
\end{align*}
\]

Here \( y_n = (y_1, \ldots, y_n)' \) is a time series of asset returns, \( h_t \) is an unobserved log-volatility of \( y_t \) and \( u_t \sim N(0, 1) \) and \( v_t \sim N(0, 1) \) are uncorrelated Gaussian white noise sequences. The scaling factors \( \exp(h_t/2) \) and \( \exp(\mu/2) \) specify the amount of volatility on day \( t \) and the model volatility, \( \tau \) determines the volatility of log-volatilities and \( \phi \) measures the autocorrelation (Kim et al. (1998), Meyer and Yu (2000) and Berg et al. (2004)).

After observing that the basic SV model is too restrictive for many financial time series, numerous studies have been conducted to estimate the volatility process more accurately that include covariate effects (Chib et al. (2002)), leverage

Received November 1, 2005. Revised January 11, 2006. Accepted May 8, 2006.

*Graduate School of Business Administration, Keio University, 2-1-1 Hiyoshi-Honcho, Kohoku-ku, Yokohama-shi, Kanagawa 223-8523, Japan. Email: andoh@kbs.keio.ac.jp

It should be noted that the assumption that the volatility follows a linear AR(1) process given in (1.1) is not always guaranteed. Since different functional forms of innovation lead to different results for option pricing and value at risk evaluation, it is important to model the correct volatility innovations. When focusing on the fact that the distribution of the current volatility depends on the functional form of the previous volatility process, one way to extend the basic SV model is to introduce the nonlinearities in volatility dynamics.

To achieve this, Meddahi (2001) recently introduced a series expansion of the eigen-functions of the conditional expectation operator. One particular model is

\[
\begin{aligned}
  y_t &= h_{t-1} \varepsilon_t, \\
  h_{t-1}^2 &= \sum_{j=0}^{p} a_j H_j(f_{t-1}), \\
  f_t &= \gamma f_{t-1} + \sqrt{1 - \gamma^2 v_t}, \quad |\gamma| < 1,
\end{aligned}
\]

where \( f_t \) is the state variable that governs the volatility and \( H_j(\cdot) \) are the Hermite polynomials. Unfortunately, as pointed out by Meddahi (2001), the positivity of the volatility of asset return is not always guaranteed without imposing some constrains on the model parameters. From practical aspects, it is not convenient to consider the positivity conditions in estimating model parameters.

In this paper, to overcome this problem, we introduce a smoothing spline approach (Green and Silverman (1994)) to capture a nonlinear relationship between the current volatility and the previous volatility process. In particular, we introduce the \( B \)-spline basis expansion approach (Eilers and Marx (1996)). An advantage of the proposed model is that it always guarantees the positivity of the volatility of the asset return. Furthermore, it is useful to capture the nonlinear smooth effects of the previous volatility process with the current volatility.

It is a well known fact that the maximum likelihood estimation of the SV type models is very difficult since the likelihood function depends upon high-dimensional integrals (Watanabe (1999)). Various parameter estimation procedures have been proposed that include an efficient method of moments (Andersen et al. (1999), Gallant and Tauchen (1996)), a simulated method of moments (Duffe and Singleton (1993)), a quasi maximum likelihood (Harvey et al. (1994)), the Monte Carlo maximum likelihood method (Sandmann and Koopman (1998)) and so on.

In this paper, the Bayesian approach with Markov Chain Monte Carlo (MCMC) method is employed for estimating model parameters. The MCMC method has played a major role in the recent advances in Bayesian analyses of SV models (Shephard (1993), Jacquier et al. (1994), Kim et al. (1998), Chib et al. (2002), Omori et al. (2004)). This allows us to estimate various types of
SV models easily because the model parameter estimation can be done without evaluating the likelihood function.

After estimating model parameters, we have to determine the best model from a set of candidate models. Although the progress in MCMC simulation methods has made SV modeling popular, the assessment of the goodness of its extension is still ongoing. In this paper, to evaluate the goodness of the estimated models, we employ several Bayesian model selection criteria that include the Bayes factor (Kass and Raftery (1995), Kim et al. (1998), Chib et al. (2002)), the Bayesian predictive information criterion (Ando (2004, 2006c)) and the deviance information criterion (Spiegelhalter et al. (2002), Berg et al. (2004)). To examine the goodness of the constructed model from various angles, three kinds of model selection criteria are used.

The remainder of the paper is organized as follows. Section 2 introduces a nonlinear and non-Gaussian SV model with leverage effect. The Bayesian MCMC algorithm is proposed for estimating model parameters. Section 3 describes several Bayesian model selection criteria. Section 4 illustrates the proposed method through the Monte Carlo simulation and real data analysis. Summary and conclusions are given in Section 5.

2. Extension of the basic SV model and Bayesian inference

2.1. Model description

The proposed model includes both fat-tailed error distributions and the leverage effect, and captures the nonlinear relationship between the current volatility and the previous volatility process:

\[
\begin{align*}
\{ & y_t = \exp(h_t/2)\varepsilon_t, \\
& h_t = g(h_{t-1}) + \tau v_t, \quad (t = 1, \ldots, n),
\end{align*}
\]

where \( h_{t-1} = (h_0, \ldots, h_{t-1})' \) is the previous volatility process vector. The errors \( \varepsilon_t \) and \( v_t \) follow the student-\( t \) distribution with mean zero and unknown degrees of freedom \( \nu \) and the standard normal distribution \( N(0, 1) \), respectively. Using the fact that the student-\( t \) distribution with mean zero and unknown degrees of freedom \( \nu \) can be expressed as a Gamma scale mixture of normal distributions \( \varepsilon_t = \lambda_t^{-1/2} u_t \) with \( u_t \sim N(0, 1) \) and \( \lambda_t \sim \text{Gamma}(\nu/2, \nu/2) \), the proposed model produces the leverage effect by introducing the correlation \( \rho \) between \( u_t \) and \( v_{t+1} \) (Omori et al. (2004)). Negative correlations can produce the leverage effect, in which the negative impact on \( y_t \) increases of the volatility \( h_{t+1} \). Jacquier et al. (2004) and Yu (2005) considered the correlation between \( u_t \) and \( v_t \). Yu (2005) further considered the correlation between \( u_t \) and \( v_{t+1} \).

The linear function specification \( g(h_{t-1}; \phi_1, \mu) = \mu + \phi_1(h_{t-1} - \mu) \) in (2.1) reduces the previously proposed model (Jacquier et al. (2004), Omori et al. (2004)), where \( \phi_1 \) determines the autocorrelation. We can also use the higher order AR structures. For example, the AR(2) structure \( g(h_{t-1}, h_{t-2}; \phi_1, \phi_2, \mu) = \mu + \phi_1(h_{t-1} - \mu) + \phi_2(h_{t-2} - \mu) \) without a fat-tailed error or leverage effect is considered by Meyer and Yu (2000). These models assume that the previous
volatility process has a linear effect on the current volatility, though there is no guarantee that they have a linear relationship.

To express the nonlinear relationship between the volatility and previous volatility process \(g(h_{t-1}; \mu, \gamma)\) in this paper, we introduce the following AR(1) system:

\[
(2.2) \quad g(h_t; \mu, \gamma) = \mu + \sum_{k=1}^{m} \gamma_k B_k(h_t),
\]

where \(\{B_1(\cdot), \ldots, B_m(\cdot)\}\) is a set of B-spline basis functions with degree \(p\) and \(\gamma = (\gamma_1, \ldots, \gamma_m)'\) are unknown coefficients with \(\gamma'\gamma = \sum_{k=1}^{m} \gamma_k^2 < \infty\). To ensure the model identifiability, the following restriction \(\int \{\sum_{k=1}^{m} \gamma_k B_k(h)\} dh = 0\) is generally imposed (see e.g., Lang and Brezger (2004)).

A B-spline basis expansion approach provides an attractive tool for modeling nonlinear smooth effects. The smoothness of the B-spline curve is controlled by a smoothing parameter (Eilers and Marx (1996)). Konishi et al. (2004) pointed out that the use of a singular normal distribution given in (2.6) can control the smoothness of the curve. If we use the large smoothing parameter \(\beta\) for the prior distribution, the B-spline function reduces to a linear function. Therefore, the proposed model (2.2) implicitly includes the linear function specification \(g(h_{t-1}; \phi_1, \mu) = \mu + \phi_1(h_{t-1} - \mu)\) in (2.1).

Given \(m + p + 1\) knots \(t_1 < \cdots < t_{m+p+1}\), each B-spline basis with the degree \(p\), \(B_k(h; p)\), can be calculated using de Boor’s recursion formula (de Boor (1978)):

\[
B_k(h, 0) = \begin{cases} 1, & t_k \leq h < t_{k+1}, \\ 0, & \text{otherwise}, \end{cases}
\]

\[
B_k(h; p) = \frac{h - t_k}{t_{k+p} - t_k} B_k(h; p - 1) + \frac{t_{k+p+1} - h}{t_{k+p+1} - t_{k+1}} B_{k+1}(h; p - 1).
\]

Since a zero-degree B-spline basis is just a constant on one interval between two knots, it is simple to compute the B-spline basis of any degree. Hereafter, we denote \(B_k(h; p)\) by \(B_k(h)\) for the simplicity of presentation. Figure 1 shows a set of 10 B-spline basis with degree \(p = 3\). The choice of \(m\) and \(p\) will be discussed later.

A critical difference between the SV type models and ARCH (Engle (1982)) type models is the difficulty level of the likelihood evaluation. Since ARCH type models specify the volatility of the current return as a non-stochastic function of past observations, it is easy to evaluate the likelihood function. On the other hand, the SV type models generally specify the volatility as a stochastic process and the likelihood function depends upon high-dimensional integrals:

\[
(2.3) \quad L(y_n \mid \theta) = \prod_{t=1}^{n} f(y_t \mid F_{t-1}, \theta) = \prod_{t=1}^{n} \int f(y_t \mid h_t, \theta) f(h_t \mid F_{t-1}, \theta) dh_t,
\]

where \(F_{t-1}\) denotes the history of the information sequence up to time \(t-1\), and \(\theta\) is an unknown parameter vector. The conditional density functions \(f(y_t \mid h_t, \theta)\)
Figure 1. A set of 10 B-spline basis with degree $p = 3$.

and $f(h_t \mid F_{t-1}, \theta)$ are specified by the observation equation and the system equation, respectively. Thus it is not straightforward to construct the likelihood function of the SV type models and to implement the maximum likelihood method.

Even though the integration in (2.3) cannot be solved analytically, the MCMC method allows us to estimate various types of SV models easily (see e.g., Shephard (1993), Jacquier et al. (1994), Kim et al. (1998), Chib et al. (2002), Omori et al. (2004)). In Subsection 2.3, the unknown parameter vector $\theta$ is estimated by the Bayesian approach with MCMC simulations.

2.2. Remarks on the proposed model

(a): Generally, it is not easy to derive necessary and sufficient conditions for stationarity of nonlinear time series models. It is therefore difficult to derive the analytical stationary conditions of the volatility $h_t$ since the system equation has a complicated form. This is also true for the Meddahi (2001)'s model. One way to investigate the stationarity of the constructed model is to use simulation. By generating a set of long observations repeatedly, the stationarity of the constructed model can be investigated. If the constructed models are nonstationary, then the dynamics can escape to infinity.

(b): Following Meyer and Yu (2000), an initial variance $h_0$ is assumed to be equal to the conditional variance of the volatility ($h_0 \sim N(\mu, \tau^2)$), and it is estimated by the MCMC algorithm described in Subsection 2.3. Though there is no justification for this assumption, the Monte Carlo simulation study shows that this assumption does not affect the model estimation results. As an alternative approach, the initial variance $h_0$ can be determined by model selection criteria. In this paper, we used the former approach.

(c): In a practical situation, a known distributional range of $h_t$ is needed.
In this paper, a set of B-spline basis functions $B_k(\cdot)$ are constructed by using the volatility information $h_t$ obtained from the fitting results, which come from the linear function specification $g(h_{t-1}; \phi_1, \mu) = \mu + \phi_1(h_{t-1} - \mu)$ in (2.1). In detail, the distributional range of posterior samples of the volatilities $[\text{min}_t(h_t), \text{max}_t(h_t)]$ are divided into $m - 1$ equal intervals $s$ to obtain $m$ equally spaced knots. When adding four equally spaced knots out of the interval, say $\text{min}_t(h_t) - 2s, \text{min}_t(h_t) - s, \text{max}_t(h_t) + s$ and $\text{max}_t(h_t) + 2s$, each B-spline basis is constructed using de Boor’s recursion formula. It should be noted that this procedure would work only when a linear model can roughly approximate the true structure. However, it is natural to consider that the differences of the distributional ranges of $h_t$ between the linear model and the proposed model are not so significant. Therefore, it is expected that the knots for $h_t$ in nonlinear models are selected appropriately. Furthermore, as shown in numerical studies, this method works well.

(d): Since the degree of B-spline basis is generally taken as $p = 3$ (Ando and Yamashita (2005), Imoto and Konishi (2003), Eilers and Marx (1996), Lang and Brezger (2004), Yamashita and Ando (2006)), the degree of the B-spline is set to be $p = 3$. In a practical situation, Eilers and Marx (1996) suggested the use of a moderately large number of basis functions to ensure enough flexibility, and then the smoothing parameter on coefficients $\gamma$ is optimized. We therefore set the number of basis functions $m = 10$ and consider the smoothing parameter selection problem. We should point out that the number of basis functions $m$ and the degree of B-spline basis $p$ can be determined by a model selection criterion. However, it is nearly impossible to optimize these parameters from a practical aspect. Furthermore, numerical studies show that it is enough to consider the optimization of the smoothing parameter. Ando (2006b) applied the proposed model for pricing options on the Nikkei 225 index. The major finding is that the proposed model clearly dominates the benchmark Black-Scholes (1973) model in the sense that it gives better option pricing results.

2.3. The Bayesian estimation via the Markov chain Monte Carlo method

In the Bayesian approach via the Markov chain Monte Carlo method, the latent volatility vector $h = (h_0, \ldots, h_n)'$ and $\lambda = (\lambda_1, \ldots, \lambda_n)'$ in the proposed model (2.1) with a nonlinear structure (2.2) are considered as model parameters. An inference on the parameters is conducted by producing a sample from the posterior distribution

\begin{equation}
\pi(\theta, \lambda, h | y_n) \propto \pi(\theta) \times \prod_{t=1}^{n} f(y_t \mid h_t, \lambda_t, \theta) \times f(h_0 \mid \theta) \times f(h_1 \mid h_0, \theta) \\
\times \prod_{t=2}^{n} f(h_t \mid h_{t-1}, \lambda_{t-1}, y_{t-1}, \theta) \times \prod_{t=1}^{n} f(\lambda_t \mid \theta),
\end{equation}

where $\theta = (\tau, \mu, \rho, \nu, \gamma)'$, $f(y_t \mid h_t, \lambda_t, \theta)$ and $f(h_t \mid h_{t-1}, \lambda_{t-1}, y_{t-1}, \theta)$ denote the conditional density functions specified by the observation equations and the
Considering the model (2.1) specifies the correlation \( \rho \) from the posterior density. The algorithm is summarized as follows. In this paper, we also simply update each of the elements of the parameter vector and the latent volatility one at a time to sample produce the posterior samples. In this paper, we assume a prior independence of the parameters \( (2.6) \) and \( (2.7) \) and \( (2.8) \), respectively. The use of the prior (2.6) for \( \tau \) and \( (2.7) \) for \( \rho \) and \( (2.8) \) for \( \nu \). A prior distribution on the parameter vector \( \gamma \) is a singular \( m \) variate normal distribution given by

\[
\pi(\gamma | \beta) = \left( \frac{2\pi}{n\beta} \right)^{-\frac{m}{2}} |K|^{-1/2} \exp \left( -\frac{n\beta}{2} \gamma'K\gamma \right),
\]

where \( \beta \) is a smoothing parameter, \( K \) is the \( m \times m \) matrix that represents \( \gamma'K\gamma = \sum_{l=1}^{m}(\gamma_l - 2\gamma_{l-1} + \gamma_{l-2})^2 \) and \( |K|_+ \) is the product of \( m - 2 \) nonzero eigen-values of \( K \), respectively. The use of the prior (2.6) for \( B \)-spline function has been investigated by Ando (2004), Konishi et al. (2004) and Lang and Brezger (2004).

In the Bayesian MCMC estimation of the basic SV model (1.1), Jacquier et al. (1994) utilized the single-move Metropolis-Hastings (MH) algorithm to produce the posterior samples. In this paper, we also simply update each of the elements of the parameter vector and the latent volatility one at a time to sample from the posterior density. The algorithm is summarized as follows.

Step 1. Initialize \( \theta, h \) and \( \lambda \).
Step 2. Sample \( h_t \) from \( h_t | \theta, h_{-t}, \lambda, y_n \), for \( t = 0, \ldots, n \).
Step 3. Sample \( \lambda_t \) from \( \lambda_t | \theta, h, \lambda_{-t}, y_n \), for \( t = 1, \ldots, n \).
Step 4. Sample \( \gamma_j \) from \( \gamma_j | \theta_{-\gamma_j}, h, \lambda, y_n \), for \( j = 1, \ldots, m \).
Step 5. Sample \( \tau \) from \( \tau | \theta_{-\tau}, h, \lambda, y_n \).
Step 6. Sample \( \mu \) from \( \mu | \theta_{-\mu}, h, \lambda, y_n \).
Step 7. Sample \( \rho \) from \( \rho | \theta_{-\rho}, h, y_n \).
Step 8. Sample \( \nu \) from \( \nu | \theta_{-\nu}, h, y_n \).
Step 9. Repeat steps 2 ~ 8 for a sufficiently long time.
Here $h_{-ht}$ denotes the rest of the $h$ vector other than $ht$. The Metropolis-Hastings (MH) algorithm is utilized to implement steps 2 ~ 8. For instance, in step 2, the conditional posterior density function of $ht$ is given as follows:

$$
\pi(ht \mid \theta, h_{-ht}, \lambda, y_n) = \left\{
\begin{array}{ll}
    f(h_1 \mid h_0, \theta)f(h_0 \mid \theta), & (t = 0), \\
    f(h_2 \mid h_1, \lambda_1, y_1, \theta)f(h_1 \mid h_0, \theta)f(y_1 \mid h_1, \lambda_1, \theta), & (t = 1), \\
    f(h_{t+1} \mid h_t, \lambda_t, y_t, \theta)f(h_t \mid h_{t-1}, \lambda_{t-1}, y_{t-1}, \theta)f(y_t \mid h_t, \lambda_t, \theta), & (t = 2, \ldots, n - 1), \\
    f(h_n \mid h_{n-1}, \lambda_{n-1}, y_{n-1}, \theta)f(y_n \mid h_n, \lambda_n, \theta), & (t = n).
\end{array}
\right.
$$

At the $(k+1)$-th iteration, we make a candidate draw of $h^{(k+1)}_t$ using the Gaussian proposal density function $p(h_t \mid h^{(k)}_t)$ centered at the current value $h^{(k)}_t$ with the variance $\sigma^2$. Then we accept a candidate draw with the probability

$$
\alpha = \min \left\{ 1, \frac{\pi(h^{(k+1)}_t \mid \theta, h_{-ht}, \lambda, y_n)/p(h^{(k+1)}_t \mid h^{(k)}_t)}{\pi(h^{(k)}_t \mid \theta, h_{-ht}, \lambda, y_n)/p(h^{(k)}_t \mid h^{(k+1)}_t)} \right\}.
$$

Let us define

$$
\ell(h \mid y_n, \lambda, \theta) := f(h_0 \mid \theta) \times f(h_1 \mid h_0, \theta) \times \prod_{t=2}^{n} f(h_t \mid h_{t-1}, \lambda_{t-1}, y_{t-1}, \theta).
$$

Then the remaining conditional posterior density functions are given as follows

$$
\pi(\lambda_t \mid \theta, h, \lambda_{-\lambda_t}, y_n) \propto \left\{
\begin{array}{ll}
    f(y_t \mid h_t, \lambda_t, \theta)f(h_{t+1} \mid h_t, \lambda_t, y_t, \theta)f(\lambda_t \mid \nu), & (t \neq n) \\
    f(y_t \mid h_t, \lambda_t, \theta)f(\lambda_t \mid \nu), & (t = n)
\end{array}
\right.
$$

$$
\pi(\tau \mid \theta_{-\tau}, h, \lambda, y_n) \propto \ell(h \mid y_n, \theta)\pi(\tau),
$$

$$
\pi(\mu \mid \theta_{-\mu}, h, \lambda, y_n) \propto \ell(h \mid y_n, \theta)\pi(\mu),
$$

$$
\pi(\rho \mid \theta_{-\rho}, h, \lambda, y_n) \propto \ell(h \mid y_n, \theta)f(h_0 \mid \theta)^{-1}f(h_1 \mid h_0, \theta)^{-1}\pi(\rho),
$$

$$
\pi(\nu \mid \theta_{-\nu}, h, \lambda, y_n) \propto \prod_{t=1}^{n} f(\lambda_t \mid \nu)\pi(\nu),
$$

$$
\pi(\gamma \mid \theta_{-\gamma}, h, \lambda, y_n) \propto \ell(h \mid y_n, \theta)f(h_0 \mid \theta)^{-1}\pi(\gamma \mid \beta).
$$

As well as step 2, steps 3 and 4 are also implemented by making a proposal draw from a random walk sampler.

It is well known that the efficiency of the MH algorithm depends on the choice of the proposal density. When we choose a proposal density such that it is close to the posterior probability density, the efficiency of the MH algorithm will be improved. A simple way for constructing a proposal density which is close to the posterior probability density function, is the use of Laplace approximation (Tierney and Kadane (1986)).
We therefore calculate the posterior mode $\hat{\theta}$ and construct the multivariate normal distribution $N(\hat{\theta}, \Sigma(\hat{\theta}))$, where $\Sigma(\hat{\theta})$ is the posterior covariance matrix given as the Hessian matrix of minus the log-posterior at the posterior mode. In steps $5 \sim 8$, a candidate draw is generated from the truncated normal distribution $TN(\hat{\theta}, \Sigma(\hat{\theta}) \mid R)$, where $R = \{ \theta; |\phi| < 1, |\rho| < 1, 0 < \tau, 3 < \nu < 40 \}$. The outcomes from the MH algorithm can be regarded as a sample from the posterior density function after a burn-in period. For more details on the MCMC method, we refer to Carlin and Louis (1996), Gilks et al. (1996), Omori (2001) and Tierney (1994).

We should point out that a more efficient sampling algorithm might be considered. To achieve a better simulation efficiency, Kim et al. (1998), Chib et al. (2002) and Watanabe and Omori (2004) proposed several multi-move sampler algorithms. Although their algorithms are not directly applicable to the asymmetric type SV models (Yu (2005)), an efficient sampling algorithm developed by Omori et al. (2004) may be applicable to our models. When considering that the main aim of this paper is to extend the basic SV model to capture a leverage effect, a fat-tailed distribution of asset returns and a nonlinear relationship between the current volatility and previous volatility process, it is beyond the scope of this paper to go into the details of their approach.

Although the convergence time of the proposed single-move MH algorithm is longer than that of the multi-move sampler, it is not a computationally intensive task thanks to progress in computer technology. Tanizaki (2004) used the single-move algorithm and reported that single-move MH sampling algorithm works well with a sufficiently large number of iterations. Furthermore, the estimation results show that the MH algorithm used here produces reliable results.

3. Model diagnosis

3.1. Likelihood estimation via particle filtering method

Since each Bayesian model selection criterion offers the likelihood evaluation, we first review the particle filtering method (Kitagawa (1996)) that approximates the likelihood function numerically. As shown in equation (2.3), the likelihood function has no analytical form as it is marginalized over the latent volatilities. However, it is possible to approximate the likelihood by using the particle filtering method (see Kitagawa (1996), Pitt and Shephard (1999)). A likelihood estimation procedure is generally given as follows:

Let us take a sample $h_{t-1}, \ldots, h_{t-1} \sim f(h_{t-1} \mid F_{t-1}, \theta)$ from the filtered distribution. It then follows that the one-step ahead predictive distribution $f(h_t \mid F_{t-1}, \theta)$ can be approximated as follows:

$$f(h_t \mid F_{t-1}, \theta) = \int f(h_t \mid h_{t-1}, \theta) f(h_{t-1} \mid F_{t-1}, \theta) dh_{t-1} \approx \frac{1}{M} \sum_{j=1}^{M} f(h_t \mid h_{t-1}^{j}, \theta).$$

The one-step ahead density is then estimated by Monte Carlo averaging of $f(y_t \mid$
probabilities for each model are equal, the Bayes factor for a particular model

\[ f(y_t \mid F_{t-1}, \theta) = \int f(y_t \mid h_t, \theta)f(h_t \mid F_{t-1}, \theta)dh_t \approx \frac{1}{M} \sum_{j=1}^{M} f(y_t \mid h_t^j, \theta). \]

This procedure recursively needs sequences of draws of \( h_t \) from the filtered distribution \( f(h_t \mid F_t, \theta) \). The problem is therefore how to obtain a filtered sample \( h_t^j \sim f(h_t \mid F_t, \theta) \). From Bayes theorem \( f(h_t \mid F_t, \theta) \propto f(y_t \mid h_t, \theta)f(h_t \mid F_{t-1}, \theta) \), a simple way to obtain a sample \( h_1^M, \ldots, h_t^M \) from the filtered distribution \( f(h_t \mid F_t, \theta) \) is to resample the draws of \( h_t^j \sim f(h_t \mid h_{t-1}^j, \theta) \) with probabilities proportional to \( f(y_t \mid h_t^j, \theta) \).

We now summarize the likelihood estimation steps for the proposed model (2.1) based on the particle filtering method (Kitagawa (1996)):

1. Initialize \( \theta \) and generate \( M \) samples \( h_0^1, \ldots, h_0^M \sim N(\mu, \tau^2) \).
2. Repeat the following steps for \( t = 1 \sim n \).
   2-1. Generate \( M \) samples \( h_t^1, \ldots, h_t^M \) from
   \[
   \begin{align*}
   f(h_t \mid h_{t-1}^j, \theta), & \quad (t = 1) \\
   f(h_t \mid h_{t-1}^j, \lambda_{t-1}^j, y_{t-1}, \theta), & \quad (t = 2, \ldots, n)
   \end{align*}
   \]
   for \( j = 1, \ldots, M \) and \( \lambda_t^1, \ldots, \lambda_t^M \) from Gamma(\( \nu \)/2, \( \nu \)/2).
   2-2. Compute the density estimate \( \hat{f}(y_t \mid F_{t-1}, \theta) = \sum_{j=1}^{M} f(y_t \mid h_t^j, \lambda_t^j, \theta)/M \).
   2-3. Resample \( h_t^1, \ldots, h_t^M \) and \( \lambda_t^1, \ldots, \lambda_t^M \) with probabilities proportional to \( f(y_t \mid h_t^j, \lambda_t^j, \theta) \times \exp(-h_t^j/2)\sqrt{\lambda_t^j} \) to produce the filtered samples \( h_1^1, \ldots, h_t^M \sim f(h_t \mid F_t, \theta) \) and \( \lambda_1^1, \ldots, \lambda_t^M \sim f(\lambda_t \mid F_t, \theta) \).
   2-4. Update the filtered sample \( h_{t-1}^j \) and \( \lambda_{t-1}^j \) to \( h_t^j \) and \( \lambda_t^j \), \( j = 1, \ldots, M \).
3. Return the likelihood estimate \( L(y_n \mid \theta) = \prod_{t=1}^{n} \hat{f}(y_t \mid F_{t-1}, \theta) \).

One can also utilize a more efficient particle filter algorithm such as the auxiliary particle filter (Pitt and Shephard (1999)). In the context of SV models, Kim et al. (1998) and Chib et al. (2002) implemented this algorithm. Since it is beyond the scope of this paper to describe the details of the auxiliary particle filter, we refer to Pitt and Shephard (1999) as a key reference.

3.2. Bayesian model selection

In this section, we describe three Bayesian model selection criteria; the Bayes factor (Kass and Raftery (1995), Kim et al. (1998), Chib et al. (2002)), the Bayesian predictive information criterion (Ando (2004, 2006c)) and the deviance information criterion (Spiegelhalter et al. (2002), Berg et al. (2004)).

3.2.1. The Bayes factor

In the Bayesian approach to model selection, the Bayes factor (Kass and Raftery (1995)) has played a major role in the evaluation of the goodness of the SV models (see e.g., Chib et al. (2002)). Under the situation that the prior probabilities for each model are equal, the Bayes factor for a particular model
against another model given the data is defined as the ratio of the marginal likelihood
\[ m(y_n) = \int L(y_n | \theta) \pi(\theta) d\theta. \]
This quantity measures how well the specified prior distributions fit to the observed data. The Bayes factor chooses the model with the largest value of marginal likelihood among a set of candidate models.

In order to compute the marginal likelihood, one can utilize the harmonic mean estimator (Newton and Raftery (1994)). However, a few outlying values with small likelihood values can have a large effect on this estimate. This is because the inverse likelihood does not possess a finite variance (Chib (1995)).

By rearranging the definition of the posterior distribution, Chib (1995) evaluated the log of the marginal likelihood as follows:

\[
\log \{ m(y_n) \} = \log \{ L(y_n | \theta^*) \} + \log \{ \pi(\theta^*) \} - \log \{ \pi(\theta^* | y_n) \},
\]

for any values of \( \theta^* \). Following the suggestion in Chib (1995), we use the posterior mean of the parameter vector. The first term and the second term on the right hand side of equation (3.1) can be evaluated easily. The third term can be evaluated by using a multivariate kernel density estimate based on the posterior sample (see Kim et al. (1998) and Berg et al. (2004)). Hereafter, we call the marginal likelihood evaluated by Chib’s (1995) approach as the Chib’s BF (Chib’s Bayes Factor).

3.2.2. The Bayesian predictive information criterion

Let us take another set of observations \( z_n = (z_1, \ldots, z_n)' \) generated by the same mechanism that gave rise to the observed data \( y_n \) drawn from the true model \( s(z_n) \). Concerning the concept of the future observations and the true model, we refer to Konishi and Kitagawa (1996) and Ando et al. (2006). To evaluate the fitness of the Bayes model to the true model \( s(z_n) \), Spiegelhalter et al. (2002) and Ando (2006c) considered the maximization of the posterior mean of the expected log-likelihood

\[
\eta = \int \left[ \int \log L(z_n | \theta) \pi(\theta | y_n) d\theta \right] s(z_n) dz_n.
\]

It is obvious that the posterior mean of the expected log-likelihood depends on the model, and further depends on the unknown true model \( s(\cdot) \). Therefore, the problem is how to estimate the posterior mean of the expected log-likelihood accurately.

Although the posterior mean of the log-likelihood \( \hat{\eta} \) is a natural estimator of \( \eta \)

\[
\hat{\eta} = \int \log L(y_n | \theta) \pi(\theta | y_n) d\theta,
\]

it provides a positive bias as an estimator of \( \eta \) because the same data \( y_n \) are used both to construct the posterior distributions and to evaluate the posterior mean of the expected log-likelihood (Ando (2006c)). Therefore, the bias correction of the posterior mean of the log-likelihood should be considered.

Defining the bias \( b \) of the posterior mean of the log-likelihood in estimating the posterior mean of the expected log-likelihood as

\[
b = \int [\hat{\eta} - \eta] s(y_n) dy_n,
\]
a real world situation, in which the specified family of probability distributions does not contain the true distribution, Ando (2006c) evaluated the asymptotic bias as follows: 

\[ n \hat{b} \approx E_{\theta | y_n} \left[ \log \left\{ L(y_n | \theta) \pi(\theta) \right\} \right] - \log \left\{ L(y_n | \hat{\theta}_n) \pi(\hat{\theta}_n) \right\} + \text{tr} \left\{ J_n^{-1}(\hat{\theta}_n) I_n(\hat{\theta}_n) \right\} + 0.5q. \]

Here \( q \) is the dimension of \( \theta \), \( E_{\theta | y_n} \left[ \cdot \right] \) denotes the expectation with respect to the posterior distribution, \( \hat{\theta}_n \) is the posterior mode, and the matrices are

\[ I_n(\hat{\theta}_n) = \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\partial \eta_n(y_t, \theta)}{\partial \theta} \frac{\partial \eta_n(y_t, \theta)}{\partial \theta'} \right) \bigg|_{\theta = \hat{\theta}_n}, \]
\[ J_n(\hat{\theta}_n) = \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\partial^2 \eta_n(y_t, \theta)}{\partial \theta \partial \theta'} \right) \bigg|_{\theta = \hat{\theta}_n}, \]

with \( \eta_n(y_t, \theta) = \log f(y_t | F_{t-1}, \theta) + \log \pi(\theta) / n \). Correcting the asymptotic bias of the posterior mean of the log-likelihood, the Bayesian predictive information criterion (BPIC; Ando (2006c)) is

\[ \text{BPIC} = -2E_{\theta | y_n} \left[ \log L(y_n | \theta) \right] + 2n\hat{b}. \]

The best model is chosen as the minimizer of BPIC.

**3.2.3. Deviance information criterion**

Under the assumption that the future observations \( z_n \) are equivalent to the observed data \( y_n \), Spiegelhalter et al. (2002) introduced the deviance information criterion:

\[ \text{DIC} = -2E_{\theta | y_n} \left[ \log L(y_n | \theta) \right] + P_D, \tag{3.2} \]

where the second term is the effective number of parameters, defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters: \( P_D = 2[\log L(y_n | \hat{\theta}) - E_{\theta | y_n} \left[ \log L(y_n | \theta) \right]] \).

Berg et al. (2004) pointed out many drawbacks of the Bayes factor and proposed the use of the deviance information criterion (DIC, Spiegelhalter et al. (2002)) as an alternative, though DIC suffers in some theoretical aspects (see e.g., Stone (2002), Robert and Titterington (2002), Ando (2004, 2006c)). To overcome theoretical problems in DIC, Ando (2004, 2006c) recently proposed the Bayesian predictive information criterion (BPIC), an improved version of DIC.

**4. Numerical results**

In order to assess the performance of the proposed SV modeling procedure, we present numerical results based on both simulated and real datasets. In our application, the total number of MCMC iterations is chosen to be 1,100,000 in which the first 100,000 iterations are discarded as a burn-in period. Due to higher posterior correlations amongst the parameters and thus slower convergence of the MCMC sampling algorithm, we stored every 1,000-th iteration after a burn-in
It is necessary to check whether the generated posterior sample is taken from the stationary distribution. We assessed the convergence of MCMC simulation by calculating the convergence diagnostic (CD) test statistics (Geweke (1992)). Geweke’s (1992) CD test statistic measures the equality of the means of the first and last part of a Markov chain. If the samples are drawn from the stationary distribution, these two means calculated from the first and the last part of a Markov chain are equal. Therefore, CD test statistic has an asymptotically standard normal distribution. All the results we report in this paper are based on samples that have passed the Geweke’s (1992) convergence test at a significance level of 5% for all parameters.

4.1. Simulation study

In this simulation study, a set of \( n = 1,000 \) observations is generated from the model (2.1) with the parameter values \( \mu = -8, \tau = 0.15, \nu = 15, \rho = -0.5, \) respectively. The following function is assumed for the true function \( g(h_t) \) in (2.1)

\[
g(h_t) = 0.35(h_t - \mu)^4 - 0.42(h_t - \mu)^3 + 0.56(h_t - \mu)^2 + 0.7(h_t - \mu).
\]

Four different types of SV models including the proposed SV model are fitted to the generated data. We now describe the observation equation, the state equation, the distributional assumptions and the prior distributions for each model.

Model 1 is equivalent to the basic SV model (1.1) including the leverage effect that allows for correlation \( \rho \) between \( u_t \) and \( v_{t+1} \). An important empirical fact in many financial time series is the leverage effect. We assume that each parameter is an a priori independent \( \pi(\theta) = \pi(\mu)\pi(\phi)\pi(\tau) \). The same prior specifications of Kim et al. (1998) are then used. For the prior densities of \( (\phi + 1)/2, \tau^2 \) and \( \mu \), a beta distribution \( Be(20,1.5) \), an inverse-gamma distribution \( IG(2.5,0.025) \) and a normal distribution \( N(-10,5^2) \) are utilized. We also specify a uniform prior distribution \( U[-1,1] \) for \( \rho \). In Model 2, the normal distribution of \( u_t \) in (1.1) is replaced by independent central Student-\( t \) distributions with \( \nu \) degrees of freedom. We use the same prior for \( \phi, \mu, \tau^2 \) as for model 1 and use the uniform prior distribution \( U[3,40] \) for \( \pi(\nu) \). Model 3 is equivalent to Model 2 including the leverage effect. Model 4 is the proposed SV model (2.1) with the nonlinear AR(1) structure in (2.2). As a fitting result, BPIC selected the value of smoothing parameter \( \beta = 1 \).

To save the computational time, initial values of each parameter, including the latent volatility and model parameters, were set to be their true values. The absolute values of Geweke (1992)’s CD test statistics are less than 1.96 (at a significance level of 5%). Therefore, we can judge that MCMC sampling works well. Figures 2 and 3 show the sample paths, the estimated posterior densities and the sample autocorrelation functions for the proposed model. Similar results are also obtained for the remaining coefficients \( \gamma_2, \gamma_3, \gamma_4, \gamma_7, \gamma_8 \) and \( \gamma_9 \).
Figure 2. Sample paths, estimated posterior densities and sample autocorrelation functions for τ, μ, ρ and ν, respectively. These figures draw the fitting result of Model 4.

Table 1 reports the posterior means, the standard errors, 95% confidence intervals, the inefficiency factor (Kim et al. (1998)) and the acceptance rates. Using 1,000 draws for each of the parameters, we calculated the posterior means, the standard errors and the 95% confidence intervals. The 95% confidence intervals are estimated using the 2.5th and 97.5th percentiles of the posterior samples. The inefficiency factor is useful to measure the efficiency of the MCMC sampling algorithm. This is defined as $1 + 2 \sum_{k=1}^{\infty} \rho(k)$, where $\rho(k)$ is the sample autocorrelation at lag $k$ calculated from the sampled draws. We have used 1,000 lags in the estimation of the inefficiency factors. Although the inefficiency factors exceed one hundred, the posterior inference on the parameter is not a computationally intensive task thanks to the progress in computer technology.

The proposed method is evaluated in terms of how accurately it estimates the true parameter values. It can be seen that the estimated results for the parameters appear quite reasonable. For instance, the true model is estimated with reasonably accurate results based on our proposed model. The 95% confidence intervals include the true parameter values. In Model 1, it is assumed that $y_t$ is normally distributed. This assumption made the posterior mean of $\tau$ to be larger than the true value. This is because the probability distribution function of the
Figure 3. Sample paths, estimated posterior densities and sample autocorrelation functions for \( \gamma_2, \gamma_4, \gamma_6 \) and \( \gamma_8 \), respectively. These figures draw the fitting result of Model 4.

Student \( t \)-distribution with degrees of freedom \( \nu = 15 \) has fatter tails than that of the standard normal.

Figure 4 plots the posterior means of the previous volatility process effects \( \hat{g}(h_{t-1}) \) estimated by our method. The dotted lines are 95% confidence intervals. The fitted curve based on our proposed method give a good representation of the underlying function.

Table 2 reports the values of BPIC, DIC and Chib’s BF. To approximate the likelihood function, we set the number of particles to be \( M = 20,000 \). As shown in Table 2, each of the model selection criterion selects the proposed model (Model 4) as the most adequate model. Following Berg et al. (2004), the computing time to generate 100 iterations for each of the eight models is reported. The CPU time needed to estimate Model 4 is approximately six times larger than that of Model 1. Table 2 also compares the average squared error between the true volatility \( \exp(h_{t}/2) / \sqrt{\lambda_t} \) and the estimated volatility. In Model 1, the true volatility is estimated by the posterior mean of \( \exp(h_{t}/2) \). The other models estimate the true volatility by using the posterior mean of \( \exp(h_{t}/2) / \sqrt{\lambda_t} \). The proposed method is also superior to the other models in the sense that it gives the smallest value of the average squared error. This is natural because the data
Table 1. Comparison of the parameter estimates for the simulated data. The posterior means, the standard errors (SEs), 95% confidence intervals, the inefficiency factors (INEFs) and the acceptance rates (ACRs) are calculated.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Stand. errors</th>
<th>95% Conf. interval</th>
<th>INEFs</th>
<th>ACRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>τ</td>
<td>0.2209</td>
<td>0.0218</td>
<td>[0.1846 0.2567]</td>
<td>872.5</td>
</tr>
<tr>
<td></td>
<td>φ</td>
<td>0.7582</td>
<td>0.0513</td>
<td>[0.6722 0.8327]</td>
<td>834.9</td>
</tr>
<tr>
<td></td>
<td>µ</td>
<td>−7.9458</td>
<td>0.0499</td>
<td>[−8.0293 −7.8625]</td>
<td>685.2</td>
</tr>
<tr>
<td></td>
<td>ρ</td>
<td>−0.5035</td>
<td>0.0786</td>
<td>[−0.5650 −0.2377]</td>
<td>842.9</td>
</tr>
<tr>
<td>2</td>
<td>τ</td>
<td>0.1442</td>
<td>0.0188</td>
<td>[0.1139 0.1770]</td>
<td>859.3</td>
</tr>
<tr>
<td></td>
<td>φ</td>
<td>0.8153</td>
<td>0.0499</td>
<td>[0.7267 0.8921]</td>
<td>786.2</td>
</tr>
<tr>
<td></td>
<td>µ</td>
<td>−8.0436</td>
<td>0.0491</td>
<td>[−8.1285 −7.9669]</td>
<td>658.4</td>
</tr>
<tr>
<td></td>
<td>ρ</td>
<td>16.1005</td>
<td>1.5309</td>
<td>[13.5894 18.6274]</td>
<td>666.6</td>
</tr>
<tr>
<td>3</td>
<td>τ</td>
<td>0.1215</td>
<td>0.0144</td>
<td>[0.1406 0.1817]</td>
<td>871.8</td>
</tr>
<tr>
<td></td>
<td>φ</td>
<td>0.9447</td>
<td>0.0075</td>
<td>[0.9324 0.9568]</td>
<td>186.4</td>
</tr>
<tr>
<td></td>
<td>µ</td>
<td>−8.0983</td>
<td>0.0759</td>
<td>[−8.2275 −7.9797]</td>
<td>420.3</td>
</tr>
<tr>
<td></td>
<td>ρ</td>
<td>−0.4769</td>
<td>0.0759</td>
<td>[−0.6684 −0.2841]</td>
<td>829.7</td>
</tr>
<tr>
<td></td>
<td>ν</td>
<td>15.1499</td>
<td>1.6137</td>
<td>[12.4782 17.7144]</td>
<td>728.9</td>
</tr>
<tr>
<td>4</td>
<td>τ</td>
<td>0.1611</td>
<td>0.0163</td>
<td>[0.1330 0.1812]</td>
<td>808.6</td>
</tr>
<tr>
<td></td>
<td>µ</td>
<td>−8.0037</td>
<td>0.0253</td>
<td>[−8.0539 −8.9624]</td>
<td>768.9</td>
</tr>
<tr>
<td></td>
<td>ρ</td>
<td>−0.4873</td>
<td>0.1034</td>
<td>[−0.6789 −0.4603]</td>
<td>847.7</td>
</tr>
<tr>
<td></td>
<td>ν</td>
<td>15.7552</td>
<td>1.6846</td>
<td>[14.0089 19.4437]</td>
<td>813.6</td>
</tr>
<tr>
<td></td>
<td>γ₁</td>
<td>−0.0458</td>
<td>0.0107</td>
<td>[−0.0629 −0.0281]</td>
<td>470.3</td>
</tr>
<tr>
<td></td>
<td>γ₂</td>
<td>−0.0319</td>
<td>0.0069</td>
<td>[−0.0426 −0.0199]</td>
<td>676.4</td>
</tr>
<tr>
<td></td>
<td>γ₃</td>
<td>−0.0187</td>
<td>0.0067</td>
<td>[−0.0306 −0.0078]</td>
<td>756.7</td>
</tr>
<tr>
<td></td>
<td>γ₄</td>
<td>−0.0070</td>
<td>0.0071</td>
<td>[−0.0192 0.0436]</td>
<td>797.9</td>
</tr>
<tr>
<td></td>
<td>γ₅</td>
<td>−0.0021</td>
<td>0.0061</td>
<td>[−0.0075 0.0125]</td>
<td>725.9</td>
</tr>
<tr>
<td></td>
<td>γ₆</td>
<td>0.0090</td>
<td>0.0057</td>
<td>[−0.0003 0.0185]</td>
<td>718.6</td>
</tr>
<tr>
<td></td>
<td>γ₇</td>
<td>0.0145</td>
<td>0.0065</td>
<td>[0.0436 0.0252]</td>
<td>806.6</td>
</tr>
<tr>
<td></td>
<td>γ₈</td>
<td>0.0195</td>
<td>0.0072</td>
<td>[0.0859 0.0318]</td>
<td>820.4</td>
</tr>
<tr>
<td></td>
<td>γ₉</td>
<td>0.0257</td>
<td>0.0069</td>
<td>[0.0146 0.0369]</td>
<td>720.5</td>
</tr>
<tr>
<td></td>
<td>γ₁₀</td>
<td>0.0325</td>
<td>0.0050</td>
<td>[0.0247 0.0404]</td>
<td>485.9</td>
</tr>
</tbody>
</table>

is generated from the non-Gaussian and asymmetric SV model with a nonlinear AR(1) structure. We therefore can expect that the proposed method can work effectively in real data analysis.

4.2. An empirical study

This section analyzes the daily returns of Nikkei 225 index from August 28, 2001 to September 22, 2005 on which the market was open leading to a set of 1,000 samples. The returns \( y_t \) are defined as the differences in the logarithm of the daily closing value of Nikkei 225 index \( y_t = \log(x_t) - \log(x_{t-1}) \), where \( x_t \) is the closing price on day \( t \). Figure 5 shows the transformed log-difference of Nikkei 225 index. The vertical axis is the differences in the logarithm of the daily closing value of Nikkei 225 index and the horizontal axis is the time. It may be seen from Figure 5 that the transformed Nikkei 225 index data are
exhibiting time-varying volatility. As the time goes, the transformed Nikkei 225 index data seems to be distributed within a relatively small range. The basic statistics, the mean, standard deviation, skewness and kurtosis are given as \(-0.00018, -0.00019, 0.24762\) and \(4.34971\), respectively. Noting that the kurtosis of the returns is above three, the true distribution of the transformed Nikkei 225 index data would be a fat-tailed distribution. In fact, there are some outliers in Figure 5.

Four different types of SV models (Model 1 \(\sim\) Model 4) described in the previous section are fitted. We also fitted the following two models. Model 5 considers a Cauchy distribution as the error distribution \(u_t\) in the basic SV model (1.1). Setting the degrees of freedom in Model 2 to be \(\nu = 1\), we can derive Model 5. Model 6 replaces the linear AR(1) structure in the system equation (1.1) by an ARMA(1, 1) structure \(h_t = \mu + \phi(h_{t-1} - \mu) + \psi v_{t-1} + \tau v_t\). We specify a uniform prior distribution \(U[-1, 1]\) for \(\psi\).

Except for the initial values of model parameters, each of the models are es-

---

**Table 2.** BPIC, DIC, Chib’s BF and the average squared error (ASE) for simulated data. The CPU time (seconds) to generate 100 iterations for each of the eight models is also reported.

<table>
<thead>
<tr>
<th>Model</th>
<th>BPIC</th>
<th>DIC</th>
<th>Chib’s BF</th>
<th>ASE</th>
<th>CPU time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-4922.552</td>
<td>-4926.607</td>
<td>2466.256</td>
<td>0.000020</td>
<td>0.92</td>
</tr>
<tr>
<td>2</td>
<td>-4915.874</td>
<td>-4917.928</td>
<td>2461.607</td>
<td>0.000019</td>
<td>1.16</td>
</tr>
<tr>
<td>3</td>
<td>-4928.222</td>
<td>-4930.252</td>
<td>2473.272</td>
<td>0.000019</td>
<td>1.68</td>
</tr>
<tr>
<td>4</td>
<td>-4934.250</td>
<td>-4945.744</td>
<td>2496.331</td>
<td>0.000015</td>
<td>5.96</td>
</tr>
</tbody>
</table>

---

Figure 4. True effect of the previous volatility process \(g(h_{t-1})\) (---) and its estimate based on our method (---). The 95% confidence intervals (---) calculated using the 2.5th and 97.5th percentiles of the posterior draws are also shown.
Figure 5. Time series plot for Nikkei 225 index data with sample period from August 28, 2001 to September 22, 2005.

Figure 6. Sample paths, estimated posterior densities and sample autocorrelation functions for $\tau$, $\mu$, $\rho$ and $\nu$, respectively. These figures draw the fitting result of Model 4.
Figure 7. Sample paths, estimated posterior densities and sample autocorrelation functions for $\gamma_2$, $\gamma_4$, $\gamma_6$ and $\gamma_8$, respectively. These figures draw the fitting result of Model 4.

Figure 8. The posterior mean of the volatility estimates (---) $\exp(\hat{h}_t/2)/\sqrt{\lambda_t}$ from the proposed model. The 95% confidence intervals (- - -) calculated using the 2.5th and 97.5th percentiles of the posterior draws are also shown.
Table 3. Comparison of the parameter estimates for the daily returns of Nikkei 225 index. The posterior means, the standard errors (SEs), 95% confidence intervals, the inefficiency factors (INEFs) and the acceptance rates (ACRs) are calculated.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Stand. errors</th>
<th>95% Conf. interval</th>
<th>INEFs</th>
<th>ACRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\tau$</td>
<td>0.1443</td>
<td>0.0104</td>
<td>$[0.1282, 0.1623]$</td>
<td>761.6</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>0.9763</td>
<td>0.0053</td>
<td>$[0.9667, 0.9842]$</td>
<td>210.6</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>$-8.6382$</td>
<td>0.0287</td>
<td>$[-8.6835, -8.5885]$</td>
<td>213.2</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>$-0.5035$</td>
<td>0.0786</td>
<td>$[-0.6147, -0.4075]$</td>
<td>755.0</td>
</tr>
<tr>
<td>2</td>
<td>$\tau$</td>
<td>0.1408</td>
<td>0.0135</td>
<td>$[0.1172, 0.1627]$</td>
<td>760.7</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>0.9803</td>
<td>0.0060</td>
<td>$[0.9705, 0.9903]$</td>
<td>237.2</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>$-8.6845$</td>
<td>0.1631</td>
<td>$[-8.9828, -8.4267]$</td>
<td>255.5</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>20.7227</td>
<td>1.4220</td>
<td>$[18.2594, 22.8851]$</td>
<td>519.0</td>
</tr>
<tr>
<td>3</td>
<td>$\tau$</td>
<td>0.1469</td>
<td>0.0126</td>
<td>$[0.1406, 0.1817]$</td>
<td>930.0</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>0.9730</td>
<td>0.0054</td>
<td>$[0.9634, 0.9816]$</td>
<td>553.0</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>$-8.7942$</td>
<td>0.1188</td>
<td>$[-8.9842, -8.5878]$</td>
<td>557.6</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>$-0.5402$</td>
<td>0.0749</td>
<td>$[-0.6632, -0.4189]$</td>
<td>848.2</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>17.8093</td>
<td>1.3645</td>
<td>$[15.6689, 20.1001]$</td>
<td>974.9</td>
</tr>
<tr>
<td>4</td>
<td>$\tau$</td>
<td>0.1576</td>
<td>0.0152</td>
<td>$[0.1413, 0.1760]$</td>
<td>821.1</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>$-9.0139$</td>
<td>0.0273</td>
<td>$[-9.0539, -8.9624]$</td>
<td>745.1</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>$-0.5150$</td>
<td>0.0678</td>
<td>$[-0.6789, -0.4603]$</td>
<td>714.6</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>17.0958</td>
<td>1.7586</td>
<td>$[14.2829, 19.9647]$</td>
<td>692.7</td>
</tr>
<tr>
<td>5</td>
<td>$\tau$</td>
<td>0.1272</td>
<td>0.0191</td>
<td>$[0.0979, 0.1589]$</td>
<td>802.3</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>0.9925</td>
<td>0.0033</td>
<td>$[0.9726, 0.9954]$</td>
<td>197.7</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>$-9.4378$</td>
<td>0.1621</td>
<td>$[-9.6388, -9.2591]$</td>
<td>213.2</td>
</tr>
<tr>
<td>6</td>
<td>$\tau$</td>
<td>0.1549</td>
<td>0.0145</td>
<td>$[0.1328, 0.1789]$</td>
<td>773.6</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>0.9787</td>
<td>0.0066</td>
<td>$[0.9677, 0.9890]$</td>
<td>210.6</td>
</tr>
<tr>
<td></td>
<td>$\psi$</td>
<td>$-0.0138$</td>
<td>0.0054</td>
<td>$[-0.1001, 0.0767]$</td>
<td>145.9</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>$-8.6309$</td>
<td>0.0321</td>
<td>$[-8.6842, -8.5804]$</td>
<td>217.5</td>
</tr>
</tbody>
</table>

estimated by using the same MCMC sampling scheme as described in the previous section. To save the computational time, the initial values of model parameters $\theta$ are chosen to be their posterior modes. Using 1,000 draws for each of the parameters, the posterior means, the standard errors, the 95% confidence intervals, the inefficiency factors (INEFs) and the acceptance rates (ACRs) are calculated. Table 3 reports these results. As shown in Table 3, the posterior mean of $\rho$ in
Table 4. BPIC, DIC and Chib’s BF for the Nikkei 225 index data. Figures in parentheses give the smallest and largest values for each criterion.

<table>
<thead>
<tr>
<th>Model</th>
<th>BPIC</th>
<th>DIC</th>
<th>Chib’s BF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-5708.229$</td>
<td>$-5711.222$</td>
<td>$2855.418$</td>
</tr>
<tr>
<td></td>
<td>$[-5710.213, -5707.801]$</td>
<td>$[-5712.235, -5710.252]$</td>
<td>$[2854.592, 2856.174]$</td>
</tr>
<tr>
<td>2</td>
<td>$-5682.096$</td>
<td>$-5683.323$</td>
<td>$2844.572$</td>
</tr>
<tr>
<td></td>
<td>$[-5683.697, -5680.195]$</td>
<td>$[-5684.507, -5682.239]$</td>
<td>$[2842.281, 2845.629]$</td>
</tr>
<tr>
<td>3</td>
<td>$-5703.578$</td>
<td>$-5704.563$</td>
<td>$2856.451$</td>
</tr>
<tr>
<td></td>
<td>$[-5704.986, -5702.286]$</td>
<td>$[-5706.292, -5703.610]$</td>
<td>$[2855.053, 2857.451]$</td>
</tr>
<tr>
<td>4</td>
<td>$-5711.151$</td>
<td>$-5713.077$</td>
<td>$2864.380$</td>
</tr>
<tr>
<td></td>
<td>$[-5713.269, -5708.848]$</td>
<td>$[-5714.009, -5712.172]$</td>
<td>$[2862.437, 2865.385]$</td>
</tr>
<tr>
<td>5</td>
<td>$-5452.758$</td>
<td>$-5451.940$</td>
<td>$2734.539$</td>
</tr>
<tr>
<td></td>
<td>$[-5454.545, -5451.166]$</td>
<td>$[-5453.483, -5450.249]$</td>
<td>$[2733.152, 2735.521]$</td>
</tr>
<tr>
<td>6</td>
<td>$-5675.013$</td>
<td>$-5679.523$</td>
<td>$2843.012$</td>
</tr>
<tr>
<td></td>
<td>$[-5677.170, -5673.937]$</td>
<td>$[-5681.125, -5678.122]$</td>
<td>$[2841.319, 2844.445]$</td>
</tr>
</tbody>
</table>

Models 2 ~ 4 is significantly below 0. We further observed that the upper limits of the 95% posterior credibility interval of $\rho$ were below 0. This result is consistent with past research (Watanabe (1999), Tanizaki (2004)), where the Tokyo stock price index is analyzed. Noting that the posterior means of $\nu$ are around 17, the transformed Nikkei 225 index data have a fat-tailed distribution. Figures 6 and 7 show sample paths of 1,000 draws, sample autocorrelation functions and estimated posterior densities. These figures draw the fitting result of Model 4. Figure 8 plots the posterior means of volatility $\exp(\frac{h_t}{2})/\sqrt{\lambda_t}$ estimated by the proposed method. The 95% confidence intervals are also shown. The posterior means of volatility exhibit time-varying movements.

In Table 4, we report the scores of BPIC, DIC and Chib’s BF for each of the six models. The model selection scores for the proposed model are based on the value of smoothing parameter $\beta = 0.1$, which is selected by BPIC. The selected value of $\beta$ is relatively small and indicates the possibility of a nonlinear structure in the data. Table 4 indicates that the most adequate model to describe the Nikkei 225 index is the proposed model (Model 4), followed by the model with a leverage effect (Model 1). Following Berg et al. (2004), we also investigated the fluctuations of each criterion and obtained the scores for ten runs for each of the six models. The values in parenthesis shown in Table 4 give the smallest and largest scores for each criterion. It indicates that the fluctuations of each criterion are not so large. Although the ranges of each model selection score for the proposed model overlap with Model 1, it can still be seen that the proposed model obtained better scores than any other model. Since the scores of Models 1, 3 and 4 are much better than that of Model 2, we can determine that the leverage effect is an important feature.

Figure 9 plots the posterior means of the previous volatility process effects $g(h_{t-1}; \mu, \gamma)$ in (2.2). As shown in Figure 9, it shows a nonlinear relationship between the current volatility and the previous volatility effect of around $h_t = 10$. 
This is also supported by each Bayesian model selection criterion. It is therefore important to construct a flexible model to capture the nonlinear relationship between the current volatility and the previous volatility process. These fitting results described above indicate that the proposed method is useful to extract useful information from the observed data.

5. Summary and conclusions

Stochastic volatility models provide useful alternatives to ARCH-type models for describing time-varying volatility exhibited in many financial asset returns. Although numerous studies have been conducted on the generalization of the basic SV model, the research for capturing the nonlinear relationship between the current volatility and the previous volatility process is still ongoing. This paper considered a class of a generalized SV model defined by the fat-tails, the leverage effect on the volatility and $B$-spline basis expansion in the system equation for modeling nonlinear smooth effects of the previous volatility process on the current volatility.

Since the implementation of maximum likelihood method contains practical difficulties, the Bayesian Markov chain Monte Carlo simulation method is used to estimate model parameters. The crucial issue in the model building process is the determination of the best fitting model from a set class of candidate models. We investigated this model selection problem from Bayesian and information theoretical points of view and used three Bayesian model selection criteria.

The proposed method is highly validated on simulated data and then applied to the Nikkei 225 index data. The simulation results show that our SV modeling procedure performs well. Fitting the model to Nikkei 225 index returns, we obtained the following results. According to the scores of model selection criteria, we find that the proposed SV model fits the Nikkei 225 index data better.
than the other models. The volatility of the Nikkei 225 index return is negatively correlated with returns. The normal assumption on the asset return is too restrictive for describing the Nikkei 225 index data. There is a nonlinear relationship between the current volatility and the previous volatility process.

For further research, the following topics are considered. First, we use a simple MCMC algorithm for estimating the model parameters. However, this algorithm needs much time. Therefore, the development of a more efficient algorithm is needed. Second, we estimated the volatility of financial asset return changes without a sudden structural change. Recently, the SV model with jumps (Barndorff-Nielsen and Shephard (2001), Chib (2002)) and the regime switching models (So et al. (1998), Shibata and Watanabe (2005), Ando (2006a)) have received considerable attention. We can extend the proposed model by considering these properties. Third, we employed three kinds of model selection criteria to examine the goodness of the model from various angles. It is also important to assess the goodness of each model selection criterion. We would like to investigate these subjects in a future paper.

Acknowledgements

The author is grateful to Dr. Yoshihiro Omori for helpful suggestions and comments. The author would like to acknowledge two anonymous referees for careful reviews and constructive comments that have substantially improved the article. This study was supported in part by a Grant-in-Aid for Young Scientists (B), 18700273.

References


