CARF Working Paper

CARF-F-215

Stochastic Volatility Model with Leverage and Asymmetrically Heavy-Tailed Error Using GH Skew Student's t-Distribution

Jouchi Nakajima
Duke University
Yasuhiro Omori
University of Tokyo

April 2010

CARF is presently supported by Bank of Tokyo-Mitsubishi UFJ, Ltd., Citigroup, Dai-ichi Mutual Life Insurance Company, Meiji Yasuda Life Insurance Company, Nippon Life Insurance Company, Nomura Holdings, Inc. and Sumitomo Mitsui Banking Corporation (in alphabetical order). This financial support enables us to issue CARF Working Papers.

Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution

Jouchi Nakajima[a], Yasuhiro Omori[b,*]

[a]*Department of Statistical Science, Duke University, Durham 27705, U.S.A.*
[b]*Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan.*

Abstract

Bayesian analysis of a stochastic volatility model with a generalized hyperbolic (GH) skew Student's t-error distribution is described where we first consider an asymmetric heavy-tailed error and leverage effects. An efficient Markov chain Monte Carlo estimation method is described that exploits a normal variance-mean mixture representation of the error distribution with an inverse gamma distribution as the mixing distribution. The proposed method is illustrated using simulated data, daily S&P500 and TOPIX stock returns. The models for stock returns are compared based on the marginal likelihood in the empirical study. There is strong evidence in the stock returns high leverage and an asymmetric heavy-tailed distribution. Furthermore, a prior sensitivity analysis is conducted whether the results obtained are robust with respect to the choice of the priors.

Keywords: generalized hyperbolic skew Student's t-distribution, Markov chain Monte Carlo, Mixing distribution, State space model, Stochastic volatility, Stock returns.

---

*Corresponding author. Tel.: +81 3 5841 5516; fax: +81 3 5841 5521.
Email addresses: `jouchi.nakajima@stat.duke.edu` (Jouchi Nakajima), `omori@e.u-tokyo.ac.jp` (Yasuhiro Omori)

April 20, 2010

1. Introduction

It has been argued that financial time series data such as stock returns and foreign exchange returns have several properties that depart from a normality assumption. Major characteristics of the return distributions for financial variables are their skewness, heavy-tailedness and volatility clustering with leverage effects. These properties are crucial not only for describing the return distributions but also for asset allocation, option pricing, forecasting and risk management.

As a promising approach to model flexible skewness and heavy-tailedness, the generalized hyperbolic (GH) distribution proposed by Barndorff-Nielsen (1977) has recently attracted attention in financial econometrics. It includes a very broad parametric class of distributions such as normal, hyperbolic, normal inverse Gaussian (NIG) and skew Student's t-distributions, and it is closed under affine transformations, conditioning and marginalization. Several studies have investigated the skewness and heavy-tailedness of financial market variables, using for the unconditional return distribution various subclasses of the class of GH distributions: hyperbolic distributions (Eberlein et al. (1998)), GH diffusion processes (Rydberg (1999)), GH skew Student's t-distributions (Hansen (1994), Fernández and Steel (1998), Aas and Haff (2006)).

On the other hand, as regards volatility clustering, the stochastic volatility (SV) model has been widely used to model the time-varying variance of time series in financial econometrics (e.g., Ghysels et al. (2002), Shephard (2005)), and various extensions of the simple SV model with a normal error (SV-Normal) have been discussed in the literature. For example, to describe the heavy-tailedness of the asset return distribution in the SV context, heavy-tailed errors are often incorporated using distributions such as Student's t-distribution (Chib et al. (2002), Berg et al. (2004), Yu (2005), Omori et al. (2007), Nakajima and Omori (2009) for discrete-time SV and Eraker et al. (2003) for continuous-time SV models) and the NIG distribution (Barndorff-Nielsen (1997), Andersson (2001)). In addition, the continuous-time SV model with jump diffusions for stock returns has also been considered (Eraker (2004), Chernov et al. (2003) and Raggi and Bordignon (2006)). The comparison of these models by Nakajima and Omori (2009), using S&P500 and TOPIX daily returns, showed that the SV model with symmetric

Student's t-errors (SVt) model performs better than the SV model with jumps or with both jumps and Student's t-errors. From another perspective, Chen et al. (2008) propose the heavy-tailed threshold SV model.

This paper proposes, for the first time in the literature, to the best of our knowledge, an efficient Bayesian estimation method for the SV model incorporating both leverage and an asymmetrically heavy-tailed error, using the GH skew Student's t-distribution. It includes the SVt and SV-Normal models with and without leverage as special cases. The GH skew Student's t-distribution is one of a subclass of GH distributions, and is well studied in literature (e.g., Prause (1999), Jones and Faddy (2003), Aas and Haff (2006)).

Although the GH skew Student's t-density itself can be easily estimated by the maximum likelihood estimation for a time-independent model, it is difficult to implement for the SV model due to the many latent volatility variables. It imposes a heavy computational burden to repeat the particle filtering many times to evaluate the likelihood function for each set of parameters until we find the maximum. Alternatively, we develop a novel Markov chain Monte Carlo (MCMC) algorithm for the precise and efficient estimation of the SV model with leverage and with an asymmetrically heavy-tailed error using the GH skew Student's t-distribution.

There are various types of skew t-distributions in the literature (e.g., Hansen (1994), Fernández and Steel (1998), Prause (1999), Jones and Faddy (2003), Azzalini and Capitanio (2003), Aas and Haff (2006)). Among these, the GH skew Student's t-error distribution is simple, flexible and easily incorporated into the SV model for a Bayesian estimation scheme using the MCMC algorithm that we develop in this paper. The key point in implementing an efficient MCMC algorithm for our proposed model is to express the GH skew Student's t-distribution as a normal variance-mean mixture of the GIG distribution. Specifically, we consider an inverse gamma distribution as a mixing distribution among the class of GIG distributions to nest and extend various existing SV models. We also show that the choice of the parameterization of the mixing distribution is important for an efficient algorithm. The estimation scheme is illustrated using simulated data and daily stock return data.

The rest of this paper is organized as follows. In Section 2, we describe an efficient MCMC algorithm in detail for the SV model with leverage and asymmetrically heavy-tailed error using the GH skew Student's t-distribution. Section 3 illustrates our proposed method using simulated data. We also examine an alternative parameterization for the GH skew Student's t-distribution. In Section 4, the proposed model is applied to S&P500 and TOPIX daily return data and the competing SV models are compared. Finally Section 5 concludes the paper.

2. SV model with GH skew Student's *t*-distribution

2.1. The model

A basic SV model with leverage and a normal error distribution is given by

$$\begin{aligned} y_t &= \varepsilon_t \exp(h_t/2), \quad t = 1, \ldots, n, \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, \ldots, n-1, \\ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} &\sim N(0, \Sigma), \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}, \end{aligned} \tag{1}$$

where y_t is the asset return, and h_t is the unobserved log-volatility. We assume that $|\phi| < 1$, *i.e.*, that the log-volatility process is stationary and that the initial value, h_1, is assumed to follow the stationary distribution by setting $h_0 = \mu$, and $\eta_0 \sim N(0, \sigma^2/(1-\phi^2))$. The parameter ρ measures the correlation between ε_t and η_t. When $\rho < 0$, this indicates a so-called leverage effect, a drop in the return followed by an increase in the volatility (Yu (2005), Omori et al. (2007)).

For a joint model of the leverage and asymmetric heavy-tailedness, we replace the normal random variable ε_t in (1) by a random variable from the GH skew Student's t-distribution, denoted by w_t, which can be written in the form of the normal variance-mean mixture as

$$w_t = \mu_w + \beta z_t + \sqrt{z_t}\epsilon_t, \quad \epsilon_t \sim N(0,1), \text{ and } z_t \sim IG(\nu/2, \nu/2), \tag{2}$$

where IG denotes the inverse gamma distribution. We assume that $\mu_w = -\beta\mu_z$, where $\mu_z \equiv \mathrm{E}(z_t) = \nu/(\nu-2)$, for $E(w_t) = 0$, and $\nu > 4$ for the finite variance of w_t. This GH skew Student's t-distribution is a special case of the more general class of the GH distributions, defined by

$$w_t^* = \mu_w + \beta z_t + \sqrt{z_t^*}\epsilon_t, \quad \epsilon_t \sim N(0,1), \text{ and } z_t^* \sim GIG(\lambda, \delta, \gamma). \tag{3}$$

The GH skew Student's t-distribution of (2) is the case where $\lambda = -\nu/2$ ($\nu > 0$), $\delta = \sqrt{\nu}$ and $\gamma = 0$, which yields $z_t \sim GIG(-\nu/2, \sqrt{\nu}, 0)$, and equivalently $IG(\nu/2, \nu/2)$. As observed in the previous literature (e.g., Prause (1999), Aas and Haff (2006)), the parameters of the GH distribution are difficult to estimate due to the flatness of the likelihood function, and '...some parameters are hard to separate and the likelihood function may have several local maxima' (Aas and Haff (2006)) even for a GH skew Student's t-distribution with $\lambda = -\nu/2$ ($\nu > 0$) and $\gamma = 0$. Therefore, this paper makes the additional assumption that $\delta = \sqrt{\nu}$, as formulated in Equation (2). The validity of this assumption will be discussed in Section 3.3 and in Section 4.4 for comparison with an alternative parameterization. The first four moments of the GH skew Student's t-distribution are provided by Aas and Haff (2006).

Using this GH Skew Student's t-distribution, we propose the SV model (SVSKt model, hereafter) formulated as

$$y_t = \{\beta(z_t - \mu_z) + \sqrt{z_t}\varepsilon_t\}\exp(h_t/2), \quad t = 1, \ldots, n, \tag{4}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, \ldots, n-1, \tag{5}$$

and

$$z_t \sim IG(\nu/2, \nu/2), \tag{6}$$

where (ε_t, η_t) are as in (1). The value of $\nu > 4$ is the degree of freedom and unknown to be estimated. When $\beta \equiv 0$, the model reduces to the SV model with the symmetric Student's t-distribution (denoted the SVt model), which has been widely analyzed in the literature (e.g., Chib et al. (2002), Eraker et al. (2003), Yu (2005), Omori et al. (2007)).

To interpret the parameters (β, ν) in relation to the skewness and heavy-tailedness, the GH skew Student's t-densities are plotted using several combinations of the parameter values in Figure 1. In Figure 1(i), the densities are drawn using $\beta = 0$, -1 and -2, with ν fixed equal to 10. As mentioned, $\beta = 0$ corresponds to a symmetric Student's t-density. A lower value of β implies a more negative skewness or left-skewness as well as heavier tails. Figure 1(ii) shows the densities for $\nu = 5$, 10 and 15 with β fixed equal to -2. As ν becomes larger, the density becomes less skewed and has lighter tails. Hence the skewness and heavy-tailedness are determined jointly by the combination of the parameter values of β and ν.

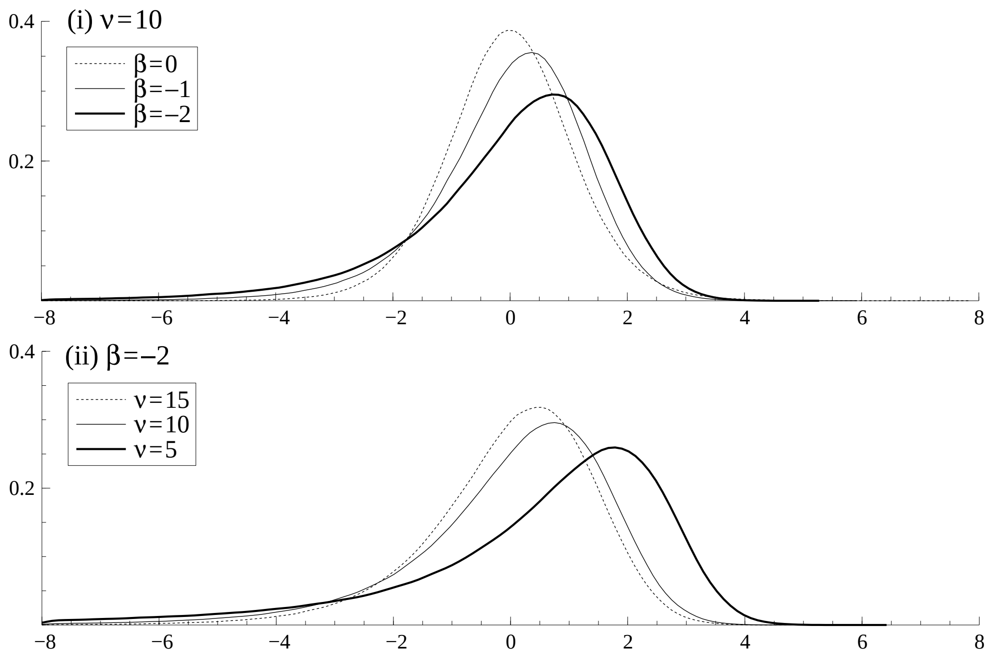

Figure 1: The GH skew Student's t-distribution. (i) $\nu = 10$ fixed; $\beta = 0$ (symmetric t), -1 and -2. (ii) $\beta = -2$ fixed; $\nu = 5$, 10 and 15.

Note that there are several definitions for the skew t-distribution in the literature (e.g., Hansen (1994), Fernández and Steel (1998), Prause (1999), Jones and Faddy (2003), Azzalini and Capitanio (2003)). For example, Aas and Haff (2006) provide an overview of other skew distributions with heavy tails, including several definitions of the skew Student's t-distributions. We could incorporate other skew

Student's t-distributions or a more general class of the GH distribution into the SV model. However, as mentioned above, introducing more parameters would lead to an over-parameterization because the second moment of the return distribution is already modeled as a latent stochastic process in the SV model. Therefore, there is less room to obtain thoughtful estimates from additional parameters.

Our formulation (4) is not only simple but suitable for the Bayesian estimation scheme using the MCMC algorithm that we propose in this paper. The key feature in our formulation of the model is to express the skew Student's t-distribution in the form of the normal variance-mean mixture, as stated in (3). We regard the variable z_t, following the mixing distribution, as a latent variable for a novel implementation of the MCMC algorithm in the context of Bayesian inference. The conditional posterior distribution of each parameter reduces to a much more tractable form conditional on z_t than when the model is considered in the direct likelihood form of the skew Student's t-distribution. Given other parameters, we can draw sample from the conditional posterior distribution of z_t for $t = 1, \ldots, n$. The next section describes our MCMC algorithm in detail.

It is worth noting when $\rho = 0$, the closed form of the density $f(y_t|h_t)$, which is marginalized over z_t, is available (see, e.g., Aas and Haff (2006)). However, in the case $\rho \neq 0$, which we consider in this paper, the closed form of the density $f(y_t|h_t, h_{t+1})$ is not available. Therefore, in our model formulation, the latent variable z_t plays an important role in exploring the posterior distribution using the MCMC algorithm.

2.2. MCMC algorithm

Let $\theta = (\phi, \sigma, \rho, \mu, \beta, \nu)$, $y = \{y_t\}_{t=1}^n$, $h = \{h_t\}_{t=1}^n$, $z = \{z_t\}_{t=1}^n$. For the prior distributions of μ and β, we assume

$$\mu \sim N(\mu_0, v_0^2), \quad \text{and} \quad \beta \sim N(\beta_0, \sigma_0^2), \tag{7}$$

and we let $\pi(\phi)$, $\pi(\vartheta)$ and $\pi(\nu)$ denote the prior probability densities of ϕ, $\vartheta \equiv (\sigma, \rho)'$ and ν respectively. We draw random samples from the posterior distribution of (θ, h, z) given y for the SVSKt model using the MCMC method (e.g., Koop (2003), Geweke (2005), Gamerman and Lopes (2006)), as follows:

1. Initialize θ, h and z.
2. Generate $\phi \,|\, \sigma, \rho, \mu, \beta, \nu, h, z, y$.
3. Generate $(\sigma, \rho) \,|\, \phi, \mu, \beta, \nu, h, z, y$.
4. Generate $\mu \,|\, \phi, \sigma, \rho, \beta, \nu, h, z, y$.
5. Generate $\beta \,|\, \phi, \sigma, \rho, \mu, \nu, h, z, y$.
6. Generate $\nu \,|\, \phi, \sigma, \rho, \mu, \beta, h, z, y$.
7. Generate $z \,|\, \theta, h, y$.
8. Generate $h \,|\, \theta, z, y$.
9. Go to 2.

In the following subsections, we present each sampling step in detail.

2.2.1. Generation of the parameters $(\phi, \sigma, \rho, \mu)$ (Steps 2-4)

Step 2. The conditional posterior probability density $\pi(\phi|\sigma, \rho, \mu, \beta, \nu, h, z, y)$ $(\equiv \pi(\phi|\cdot))$ is

$$\begin{aligned} \pi(\phi|\cdot) &\propto \pi(\phi)\sqrt{1-\phi^2}\exp\left\{-\frac{(1-\phi^2)\bar{h}_1^2}{2\sigma^2} - \sum_{t=1}^{n-1}\frac{(\bar{h}_{t+1}-\phi\bar{h}_t-\bar{y}_t)^2}{2\sigma^2(1-\rho^2)}\right\} \\ &\propto \pi(\phi)\sqrt{1-\phi^2}\exp\left\{-\frac{(\phi-\mu_\phi)^2}{2\sigma_\phi^2}\right\}, \end{aligned} \tag{8}$$

where $\bar{h}_t = h_t - \mu$, $\bar{y}_t = \rho\sigma(y_t e^{-h_t/2} - \beta\bar{z}_t)/\sqrt{z_t}$, $\bar{z}_t = z_t - \mu_z$,

$$\mu_\phi = \frac{\sum_{t=1}^{n-1}(\bar{h}_{t+1}-\bar{y}_t)\bar{h}_t}{\rho^2\bar{h}_1^2 + \sum_{t=2}^{n-1}\bar{h}_t^2}, \quad \text{and} \quad \sigma_\phi^2 = \frac{\sigma^2(1-\rho^2)}{\rho^2\bar{h}_1^2 + \sum_{t=2}^{n-1}\bar{h}_t^2}.$$

To sample from this conditional posterior distribution, we implement the Metropolis-Hastings (MH) algorithm (see, e.g., Chib and Greenberg (1995)). We propose a candidate, $\phi^* \sim TN_{(-1,1)}(\mu_\phi, \sigma_\phi^2)$, where $TN_{(a,b)}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2 truncated on the interval (a, b). Then, we accept it with probability

$$\min\left\{\frac{\pi(\phi^*)\sqrt{1-\phi^{*2}}}{\pi(\phi)\sqrt{1-\phi^2}}, 1\right\}.$$

Step 3. Because the joint conditional posterior probability density $\pi(\vartheta|\phi,\mu,\nu,h,z,y)$ $(\equiv \pi(\vartheta|\cdot))$ of $\vartheta = (\sigma,\rho)'$ is given by

$$\pi(\vartheta|\cdot) \propto \pi(\vartheta)\sigma^n(1-\rho^2)^{\frac{n-1}{2}} \exp\left\{-\frac{(1-\phi^2)\bar{h}_1^2}{2\sigma^2} - \sum_{t=1}^{n-1}\frac{(\bar{h}_{t+1}-\phi\bar{h}_t-\bar{y}_t)^2}{2\sigma^2(1-\rho^2)}\right\},$$

a probability density from which it is not easy to sample, we apply the MH algorithm based on a normal approximation of the density around the mode. Because we have a constraint, $R = \{\vartheta : \sigma > 0, |\rho| < 1\}$, on the parameter space of the posterior distribution, we consider the transformation ϑ to $\omega = (\omega_1,\omega_2)'$, where $\omega_1 = \log\sigma$, and $\omega_2 = \log(1+\rho) - \log(1-\rho)$, to generate a candidate using a normal distribution. We first search for $\hat{\vartheta}$ that maximizes (or approximately maximizes) $\pi(\vartheta|\cdot)$, and obtain its transformed value $\hat{\omega}$. We next generate a candidate $\omega^* \sim N(\omega_*,\Sigma_*)$, where

$$\omega_* = \hat{\omega} + \Sigma_* \left.\frac{\partial \log\tilde{\pi}(\omega|\cdot)}{\partial\omega}\right|_{\omega=\hat{\omega}} \quad \text{and} \quad \Sigma_*^{-1} = -\left.\frac{\partial \log\tilde{\pi}(\omega|\cdot)}{\partial\omega\partial\omega'}\right|_{\omega=\hat{\omega}},$$

where $\tilde{\pi}(\omega|\cdot)$ is a transformed conditional posterior density. Then, we accept the candidate ω^* with probability

$$\min\left\{\frac{\pi(\vartheta^*|\cdot)f_N(\omega|\omega_*,\Sigma_*)|J(\vartheta)|}{\pi(\vartheta|\cdot)f_N(\omega^*|\omega_*,\Sigma_*)|J(\vartheta^*)|}, 1\right\},$$

where $f_N(x|\mu,\Sigma)$ denotes the probability density function of a normal distribution with mean μ and covariance matrix Σ, and $J(\cdot)$ is the Jacobian for the transformation. The values of (ϑ,ϑ^*) are evaluated at (ω,ω^*), respectively.

Step 4. The conditional posterior probability density $\pi(\mu|\phi,\sigma,\rho,\beta,\nu,h,z,y)$ $(\equiv \pi(\mu|\cdot))$ is given by

$$\pi(\mu|\cdot) \propto \exp\left\{-\frac{(\mu-\mu_0)^2}{2v_0^2} - \frac{(1-\phi^2)\bar{h}_1^2}{2\sigma^2} - \sum_{t=1}^{n-1}\frac{\{(h_{t+1}-\mu)-\phi(h_t-\mu)-\bar{y}_t\}^2}{2\sigma^2(1-\rho^2)}\right\},$$

from which we generate $\mu|\cdot \sim N(\hat{\mu}, \sigma_\mu^2)$, where

$$\begin{aligned}\sigma_\mu^2 &= \left\{\frac{1}{v_0^2} + \frac{(1-\rho^2)(1-\phi^2)+(n-1)(1-\phi)^2}{\sigma^2(1-\rho^2)}\right\}^{-1}, \text{ and} \\ \hat{\mu} &= \sigma_\mu^2 \left\{\frac{\mu_0}{v_0^2} + \frac{(1-\rho^2)(1-\phi^2)h_1 + (1-\phi)\sum_{t=1}^{n-1}(h_{t+1}-\phi h_t - \bar{y}_t)}{\sigma^2(1-\rho^2)}\right\}.\end{aligned}$$

2.2.2. Generation of skew-t parameters (β, ν, z) (Steps 5-7)

Step 5. The posterior probability density $\pi(\beta|\phi, \sigma, \rho, \mu, \nu, h, z, y)$ ($\equiv \pi(\beta|\cdot)$ is given by

$$\begin{aligned}\pi(\beta|\cdot) \propto \exp\Bigg\{ &-\frac{(\beta-\beta_0)^2}{2\sigma_0^2} - \sum_{t=1}^{n} \frac{(y_t - \beta\bar{z}_t e^{h_t/2})^2}{2z_t e^{h_t}} \\ &- \sum_{t=1}^{n-1} \frac{\{\bar{h}_{t+1} - \phi\bar{h}_t - \rho\sigma(y_t e^{-h_t/2} - \beta\bar{z}_t)/\sqrt{z_t}\}^2}{2\sigma^2(1-\rho^2)}\Bigg\},\end{aligned}$$

from which we generate $\beta|\cdot \sim N(\mu_\beta, \sigma_\beta^2)$ where

$$\begin{aligned}\sigma_\beta^2 &= \left\{\frac{1}{\sigma_0^2} + \frac{1}{1-\rho^2}\sum_{t=1}^{n-1}\frac{\bar{z}_t^2}{z_t} + \frac{\bar{z}_n^2}{z_n}\right\}^{-1}, \text{ and} \\ \mu_\beta &= \sigma_\beta^2\left\{\frac{\beta_0}{\sigma_0^2} + \frac{1}{1-\rho^2}\sum_{t=1}^{n-1}\frac{y_t\bar{z}_t}{z_t e^{h_t/2}} + \frac{y_n\bar{z}_n}{z_n e^{h_n/2}} - \frac{\rho}{\sigma(1-\rho^2)}\sum_{t=1}^{n-1}\frac{(\bar{h}_{t+1}-\phi\bar{h}_t)\bar{z}_t}{\sqrt{z_t}}\right\}.\end{aligned}$$

Step 6. Because, as in Step 3, it is not easy to sample from directly from the posterior probability density of ν,

$$\begin{aligned}\pi(\nu|\cdot) \propto\ &\pi(\nu)\prod_{t=1}^{n}\frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} z_t^{-\nu/2}\exp\left(-\frac{\nu}{2z_t}\right) \\ &\times \exp\left\{-\sum_{t=1}^{n}\frac{(y_t-\beta\bar{z}_t e^{h_t/2})^2}{2z_t e^{h_t}} - \sum_{t=1}^{n-1}\frac{(\bar{h}_{t+1}-\phi\bar{h}_t-\bar{y}_t)^2}{2\sigma^2(1-\rho^2)}\right\}, \quad \nu > 4,\end{aligned}$$

we draw a sample of ν using the MH algorithm based on the normal approximation of the posterior probability density. We generate a candidate ν^* using a normal distribution truncated on $(4, \infty)$.

Step 7. The conditional posterior probability density of the latent variable z_t is

$$\begin{aligned}\pi(z_t|\theta, h, y) &\propto g(z_t) \times z_t^{-\left(\frac{\nu+1}{2}+1\right)} \exp\left(-\frac{\nu}{2z_t}\right), \text{ and} \\ g(z_t) &= \exp\left\{-\frac{(y_t - \beta\bar{z}_t e^{h_t/2})^2}{2z_t e^{h_t}} - \frac{(\bar{h}_{t+1} - \phi\bar{h}_t - \bar{y}_t)^2}{2\sigma^2(1-\rho^2)} I(t < n)\right\},\end{aligned}$$

where $I(\cdot)$ is an indicator function. Using the MH algorithm, we generate a candidate $z_t^* \sim IG((\nu+1)/2, \nu/2)$ and accept it with probability $\min\{g(z_t^*)/g(z_t), 1\}$.

2.2.3. Generation of volatility latent variable h (Step 8)

Step 8. An efficient strategy is to sample from the conditional posterior distribution of $h = \{h_t\}_{t=1}^n$ by dividing it into several blocks and sampling each block given the other blocks. This idea, called the block sampler or multi-move sampler, is developed by Shephard and Pitt (1997), and Watanabe and Omori (2004) in the context of state space modeling. They show that the sampler can produce efficient draws from the target conditional posterior distribution in comparison with a single-move sampler which primitively samples one state, say h_t, at a time given the others, h_s $(s \neq t)$. For the SV model with leverage, Omori and Watanabe (2008) develop the associated multi-move sampler and show that it produces efficient samples (see also Takahashi et al. (2009)). We extend their method for sampling h in the SVSKt model. The details of the multi-move sampler are described in the Appendix.

3. Simulation study

3.1. Setup

To illustrate our proposed estimation method, we estimate the SVSKt model using simulated data. We generate 3,000 observations from the SVSKt model given by Equations (1) and (4)–(6) with fixed parameter values $\phi = 0.95$, $\sigma = 0.15$, $\rho = -0.5$, $\mu = -9$, $\beta = -0.5$, and $\nu = 15$. The following prior distributions are assumed:

$$\frac{\phi+1}{2} \sim \text{Beta}(20, 1.5), \qquad \sigma^{-2} \sim \text{Gamma}(2.5, 0.025), \qquad \rho \sim U(-1,1),$$
$$\mu \sim N(-10,1), \qquad \beta \sim N(0,1) \text{ and } \nu \sim \text{Gamma}(16, 0.8)\, I(\nu > 4),$$

The beta prior distribution for $(\phi + 1)/2$ implies that the mean and standard deviation are (0.86, 0.11) for ϕ. The means and standard deviations of Gamma(2.5, 0.025) and Gamma(16,0.8) are (100, 63.2) and (20, 5), respectively. We use these prior distributions to reflect empirical results from the literature.

If we assume certain classes of improper priors, then the posterior distribution may be improper (see, e.g., Bauwens and Lubrano (1998)). This problem is well known for the symmetric t-distribution, and evidently, the same problem may arise here. Therefore we use proper priors for the SVSKt model, and further, we provide a prior sensitivity analysis in Section 4.5.

We draw 20,000 samples after discarding the initial 2,000 samples as a burn-in period, which are selected using the time series plots of the marginal averages of the samples for each parameter. We compute the inefficiency factor to check the efficiency of the MCMC algorithm. The inefficiency factor is defined by $1 + 2\sum_{s=1}^{\infty} \rho_s$ where ρ_s is the sample autocorrelation at lag s. It measures how well the MCMC chain mixes (see, e.g., Chib (2001)). It is the estimated ratio of the numerical variance of the posterior sample mean to the variance of the sample mean from uncorrelated draws. When the inefficiency factor is equal to m, we need to draw MCMC samples m times as many as uncorrelated samples. In the following analyses, we compute the inefficiency factor using a Parzen window with bandwidth $b_w = 1{,}000$.

3.2. Estimation results

Figure 2 shows the sample autocorrelation functions, the sample paths and the posterior densities for each parameter. The sample paths appear to be stable and the sample autocorrelations decay quickly, which implies that our sampling method is efficient.

Table 1 shows the posterior means, the standard deviations, the 95% credible intervals and the inefficiency factors. All the posterior means are close enough to the true values that the corresponding 95% credible intervals include the true values. The inefficiency factors in Table 1 are found to be of almost the same magnitude as those in Omori and Watanabe (2008) for the basic SV model with leverage using a multi-move sampler. This suggests that we are successful in extending their method to the SVSKt model without a loss of sampling efficiencies.

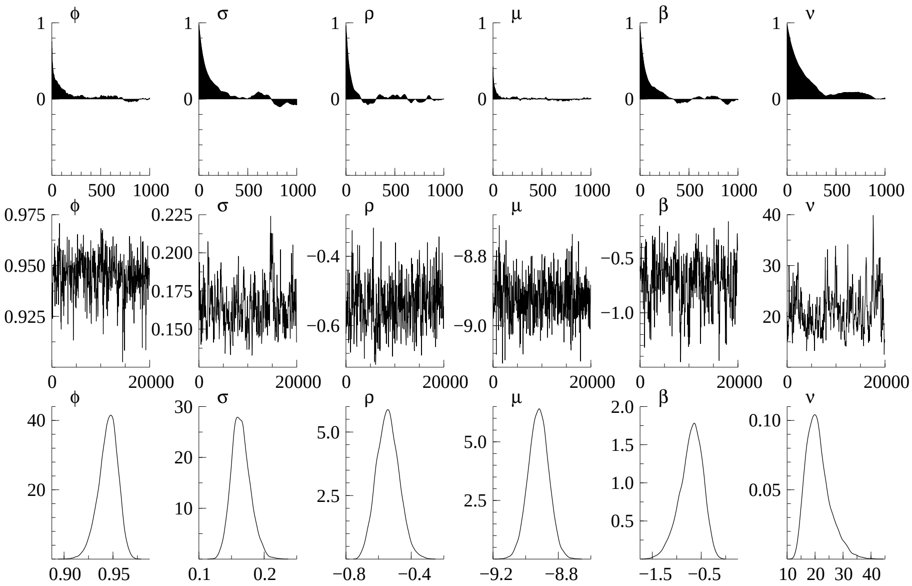

Figure 2: MCMC estimation results of the SVSKt model for simulated data. Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

Parameter	True	Mean	Stdev.	95% interval	Inefficiency
ϕ	0.95	0.9450	0.0099	[0.9233, 0.9624]	79.5
σ	0.15	0.1644	0.0146	[0.1386, 0.1958]	168.5
ρ	-0.5	-0.5425	0.0680	[-0.6694, -0.4042]	75.3
μ	-9.0	-8.9209	0.0620	[-9.0434, -8.8003]	22.5
β	-0.5	-0.7059	0.2349	[-1.2268, -0.3048]	122.2
ν	15.0	21.104	4.2843	[14.682, 31.325]	254.4

Table 1: MCMC estimation results of the SVSKt model for simulated data.

In the MH algorithms, the average acceptance rates are 97.6% for ϕ, 97.5% for (σ, ρ), 99.0% for ν and 86.4% for z_t in this experiment. The acceptance rates of the AR-MH algorithm in the multi-move sampler for the volatility h are 90.0%

and 90.6% in the AR step and the MH step respectively. These results suggest that our proposed algorithm would work well in practice.

3.3. Alternative parameterization

As already mentioned in Section 2.1, we investigate whether our proposed parameterization for the GH skew Student's t-distribution is appropriate. An alternative parameterization is explored in the following example using simulated data. The model is formulated by (1), (4) and (5) but we replace (6) by

$$z_t \sim GIG(-\nu/2, \delta, 0), \quad \text{or} \quad z_t \sim IG(\nu/2, \delta^2/2),$$

where $\delta > 0$. We generate 3,000 observations from the alternative model with parameter values $\phi = 0.95$, $\sigma = 0.15$, $\rho = -0.5$, $\mu = -9.0$, $\beta = -0.5$, $\nu = 15$ and $\delta = 4.0$. In addition to the previous experiment, we assume that the prior distribution as $\delta \sim \text{Gamma}(4, 0.4)$, which implies that the mean and standard deviation are (10.0, 15.8).

Table 2 reports the correlations of the posterior samples, and Figure 3 shows scatter plots of the posterior samples of (β, ν) for the SVSKt model and (δ, ν) for the alternative model. Evidently, the correlation between δ and ν is extremely high (0.99), while that between β and ν is moderate (−0.63). This suggests that we need to sample under the narrow state space when we use the alternative parameterization, which would result in inefficient sampling. Thus, although we could model the GH skew Student's t-distribution in other ways, the alternative models could lead either to inefficient MCMC sampling or to over-parameterization. This example shows that our proposed parameterization is appropriate for the SV model with the GH skew Student's t-distribution. In related work, Strickland et al. (2008) provide an efficiency comparison among different parameterizations of the SV model in the MCMC estimation context.

(i) SVSKt model

	ϕ	σ	ρ	μ	β	ν
ϕ	1	-.60	-.15	.04	-.01	-.03
σ		1	.08	-.05	-.07	.07
ρ			1	.04	.16	.04
μ				1	.28	.09
β					1	-.63
ν						1

(ii) Alternative model

	ϕ	σ	ρ	μ	β	ν	δ
ϕ	1	-.59	-.14	.04	-.03	.01	.01
σ		1	.07	-.07	-.06	.03	.03
ρ			1	.04	.17	-.01	-.01
μ				1	.38	-.06	-.07
β					1	-.66	-.66
ν						1	.99

Table 2: Correlation matrix of posterior samples of (i) the SVSKt model and (ii) the alternative model for simulated data.

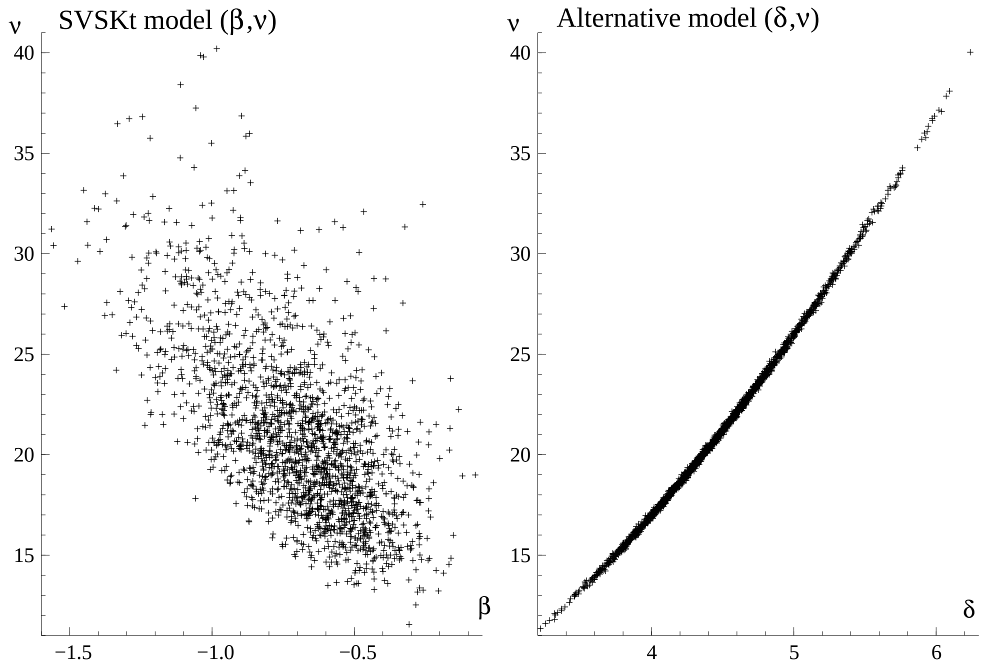

Figure 3: Scatter plots of posterior samples of (β, ν) for the SVSKt model (left) and (δ, ν) for the alternative model (right) using simulated data.

4. Application to stock return data

4.1. Data

This section applies our proposed model to daily stock return data. We consider the S&P500 index from January 1, 1970 to December 31, 2003, and the TOPIX (Tokyo stock price index) from January 5, 1970 to December 30, 2004. The returns are computed as the log-difference $y_t = \log P_t - \log P_{t-1}$, where P_t is the closing price on day t. The sample size is 8,869 for S&P500 and 9,376 for TOPIX.

Figure 4 shows the time series plots of the stock returns, and Table 3 summarizes the descriptive statistics. Both series are negatively skewed where the skewness is -1.3778 for S&P500 and -0.4833 for TOPIX. The kurtosis is as large as 37 for S&P500 and 16 for TOPIX. This is partly due to the significant negative return corresponding to the crash in October, 1987. If we remove it from the observations, the skewness and kurtosis reduce to (-0.0642, 7.9835) for S&P500 and (-0.0633, 10.404) for TOPIX. However, these figures still imply the negative skewness and heavy-tailedness of empirical returns distribution of the data.

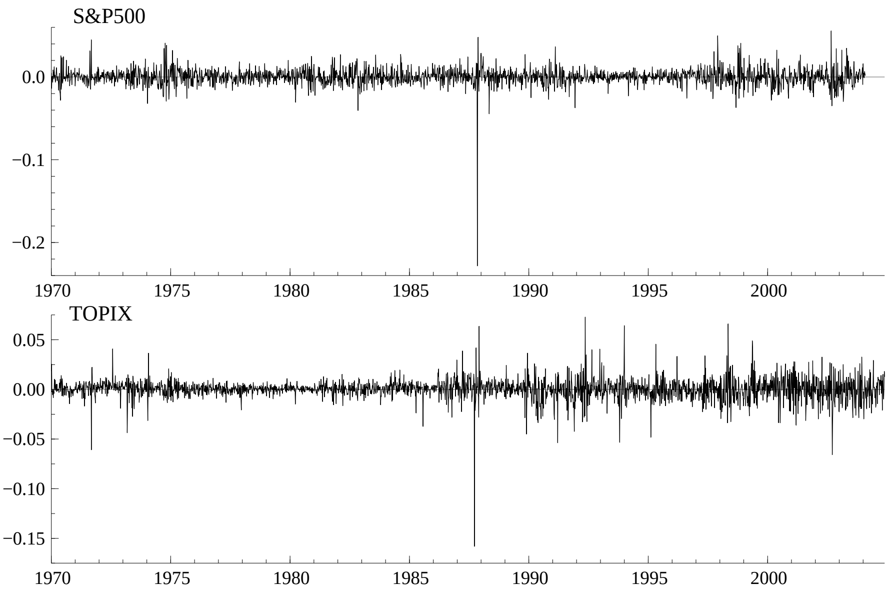

Figure 4: Time series plots for S&P500 (1970/1/1 - 2003/12/31) and TOPIX (1970/1/5 - 2004/12/30) daily returns.

S&P500 (1970/1/1 - 2003/12/31)						
Obs.	Mean	Stdev.	Skewness	Kurtosis	Min.	Max.
8,869	0.0003	0.0101	-1.3778	37.246	-0.2283	0.0871

TOPIX (1970/1/5 - 2004/12/30)						
Obs.	Mean	Stdev.	Skewness	Kurtosis	Min.	Max.
9,376	0.0002	0.0100	-0.4833	16.644	-0.1581	0.0912

Table 3: Summary statistics for S&P500 and TOPIX returns.

4.2. Parameter estimates

We assume the same prior distributions as in Section 3 for the parameters. The number of MCMC iterations and discarded initial samples are also as in Section 3. Figure 5 shows the estimation results for the S&P500 data, where the sample paths appear to stable and the proposed estimation scheme works well.

Table 4 reports the estimation result of the posterior estimates for the S&P500 and TOPIX data. The posterior means of ϕ are close to one, which indicates the well-known high persistence of volatility in stock returns. The ρ values are estimated to be negative, implying that there exist leverage effects. Regarding the skewness, the posterior means of β are -0.0946 for the S&P500 and -0.3901 for TOPIX data. Although the 95% credible interval of β barely contains zero for S&P500 data, its posterior distribution is primarily located in the negative range as shown in Figure 5. For the TOPIX data, the posterior probability that β is negative is greater than 0.95, and the negativity of β is credible. This supports the strong evidence of skewnesses in both data. On the other hand, the posterior means of ν's are around 13 for the S&P500 and 30 for the TOPIX returns, which indicates a heavy-tailedness in the stock return distributions especially for the S&P500 data.

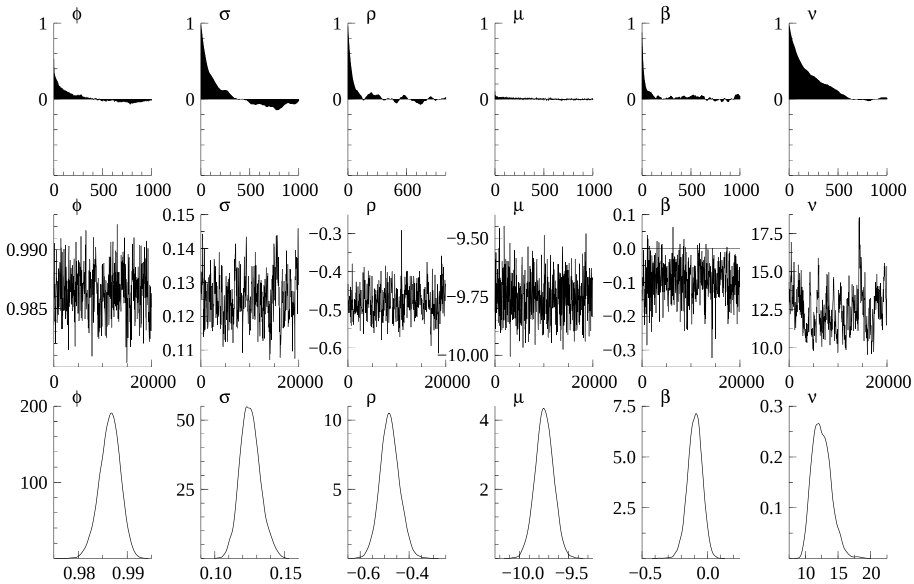

Figure 5: MCMC estimation results of the SVSKt model for S&P500 data. Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

(i) S&P500

Parameter	Mean	Stdev.	95% interval	Inefficiency
ϕ	0.9865	0.0021	[0.9821, 0.9904]	64.6
σ	0.1253	0.0072	[0.1117, 0.1407]	162.6
ρ	-0.4786	0.0397	[-0.5548, -0.3975]	86.2
μ	-9.7455	0.0929	[-9.9287, -9.5637]	11.2
β	-0.0946	0.0558	[-0.2093, 0.0097]	55.6
ν	12.513	1.4522	[10.122, 15.623]	292.2

(ii) TOPIX

Parameter	Mean	Stdev.	95% interval	Inefficiency
ϕ	0.9742	0.0032	[0.9675, 0.9802]	123.6
σ	0.2641	0.0149	[0.2396, 0.2945]	272.1
ρ	-0.3577	0.0315	[-0.4186, -0.2966]	25.3
μ	-9.8653	0.1057	[-10.241, -9.6263]	9.4
β	-0.3901	0.1225	[-0.6517, -0.1615]	42.6
ν	29.791	4.4430	[21.766, 38.512]	269.2

Table 4: Estimation results of the SVSKt model for stock return data.

4.3. Model comparison

In this subsection, we compare the SVSKt model with two alternative models discussed in the existing literature:

(i) Model SV: the basic SV model with a normal error distribution ($z_t \equiv 1$ for all t and $\beta = 0$).

(ii) Model SVt: the SV model with a symmetric Student's t error distribution ($\beta = 0$).

Note that all models are allowed to include leverage effects (ρ is not set equal to 0 in Equation (1)). In a Bayesian framework, we compare several competing models using their posterior probabilities to select the one that is best supported by the data. The posterior probability of each model is proportional to the product of prior probability of the model and the marginal likelihood. The ratio of two posterior probabilities is also well known as a Bayes factor. If the prior probabilities are assumed to be equal, we choose the model that yields the largest marginal likelihood.

The marginal likelihood is defined as the integral of the likelihood with respect to the prior density of the parameter. Following Chib (1995), we estimate the logarithm of the marginal likelihood $m(y)$, as

$$\log m(y) = \log f(y|\Theta) + \log \pi(\Theta) - \log \pi(\Theta|y), \tag{9}$$

where Θ is a parameter, $f(y|\Theta)$ is a likelihood, $\pi(\Theta)$ is a prior probability density and $\pi(\Theta|y)$ is a posterior probability density. The equality holds for any values of Θ, but we usually use the posterior mean of Θ to obtain a stable estimate of $m(y)$. The prior probability density is easily calculated, although the likelihood and posterior part must be evaluated by simulation.

The likelihood is estimated using the auxiliary particle filter (see, e.g., Pitt and Shephard (1999), Chib et al. (2002), Omori et al. (2007)) with 10,000 particles. It is replicated 10 times to obtain the standard error of the likelihood estimate. The posterior probability density at Θ is evaluated by the method of Chib (1995) and Chib and Jeliazkov (2001) through additional but reduced MCMC runs. The number of iterations for the reduced runs is set 5,000.

We use eight series of daily return data for the model comparison, as considered in Nakajima and Omori (2009). In addition to the datasets used for the previous estimation, we use the datasets of the S&P500 series from 1970 to 1985, from 1990 to 2003, from 2004 to 2009, and the TOPIX series from 1970 to 1985, from 1990 to 2004, and from 2005 to 2009, *i.e.*, we consider two long-period (about thirty years) data sets and six short-period (about fifteen or recent five years) data sets. We select two short (about fifteen-year) periods to exclude the crash of October 1987 because the huge negative return could affect the model selection among the competing models. Regarding computational time, it takes about six hours and 17 minutes to obtain the marginal likelihood for the S&P500 (1994-2003) data using 2.1 GHz CPU computer.

Table 5 shows the logarithm of the estimated marginal likelihoods and their standard errors. Overall, the SVSKt model outperforms other models for all datasets regardless of the sample periods. Taking the standard errors into account, we can see that the GH skew Student's t-error distribution in the SV model is clearly successful in describing the distribution of the daily stock return data.

We also report the posterior estimates of the skewness parameter β for each dataset in Table 5. It is interesting to observe that the posterior distribution of β is estimated to be negative for the S&P500 of 1970-2003, 1994-2003 and 2004-2009, and for the TOPIX of 1970-2004, 1970-1985 and 2005-2009, while it is centered around zero for the TOPIX of 1992-2004 and is almost certainly positive for the S&P500 of 1970-1985. The result of the recent data (2004-2009 for S&P500 and 2005-2009 for TOPIX) exhibits the largest negative posterior mean value of β, probably due to the recent financial crisis. Although the skewness of the empirical return distributions seems to change depending on the sample periods, we can conclude that the SVSKt model is favoured over other symmetric error SV models for all the sample periods.

S&P500	1970-2003	1970-1985	1994-2003	2004-2009
SV	29605.67 (1.54)	14198.89 (0.39)	8406.06 (0.37)	4840.71 (0.36)
SVt	29657.41 (1.62)	14205.03 (0.47)	8417.55 (0.43)	4849.08 (0.84)
SVSKt	29666.51 (1.42)	14206.97 (0.40)	8419.35 (0.23)	4852.07 (0.86)
Posterior of β				
Mean (Stdev.)	-0.0946 (0.0558)	0.2699 (0.1775)	-0.3942 (0.1977)	-0.4822 (0.2526)
95% interval	[-0.2093, 0.0097]	[-0.0460, 0.6599]	[-0.8165, -0.0460]	[-1.0307, -0.0213]

TOPIX	1970-2004	1970-1985	1992-2004	2005-2009
SV	32461.14 (1.50)	17626.79 (0.54)	9738.27 (0.22)	3583.16 (0.34)
SVt	32483.03 (1.55)	17641.75 (0.49)	9743.49 (0.32)	3591.92 (0.39)
SVSKt	32490.13 (0.72)	17665.91 (0.52)	9746.98 (0.31)	3598.67 (0.35)
Posterior of β				
Mean (Stdev.)	-0.3901 (0.1225)	-0.5979 (0.1790)	-0.0163 (0.1109)	-0.6195 (0.2908)
95% interval	[-0.6517, -0.1615]	[-0.9643,-0.2730]	[-0.2068, 0.2344]	[-1.2513, -0.0961]

*Standard errors of the log-ML in parentheses.

Table 5: Estimated marginal likelihoods on a logarithmic scale (log-ML) and the parameter estimates of β for S&P500 (top) and TOPIX (bottom) returns data.

4.4. Alternative model

As shown in Section 3.3, the alternative model with the additional parameter for the GH skew Student's t-distribution can be considered, although it runs the risk of serial over-identification. To confirm this point, we estimate the alternative model introduced in Section 3.3 for the S&P500 return data (1994-2003).

Table 6 reports the correlations of the posterior samples, and Figure 6 shows scatter plots of the posterior samples of (β, ν) for the SVSKt model and (δ, ν) for the alternative model. Again, the posterior correlation between δ and ν is extremely high (0.99), which implies that the additional parameter leads either to inefficient MCMC sampling or to over-parameterization.

(i) SVSKt model

	ϕ	σ	ρ	μ	β	ν
ϕ	1	-.67	-.10	-.01	-.00	-.02
σ		1	.13	-.04	-.17	.16
ρ			1	-.01	.07	.00
μ				1	.26	.15
β					1	-.69
ν						1

(ii) Alternative model

	ϕ	σ	ρ	μ	β	ν	δ
ϕ	1	-.59	.02	-.02	-.01	-.01	-.02
σ		1	-.07	.02	-.02	.00	.01
ρ			1	.13	.22	.02	.02
μ				1	.17	.04	.04
β					1	-.34	-.34
ν						1	.99

Table 6: Correlation matrix of posterior samples of (i) the SVSKt model and (ii) the alternative model for S&P500 data.

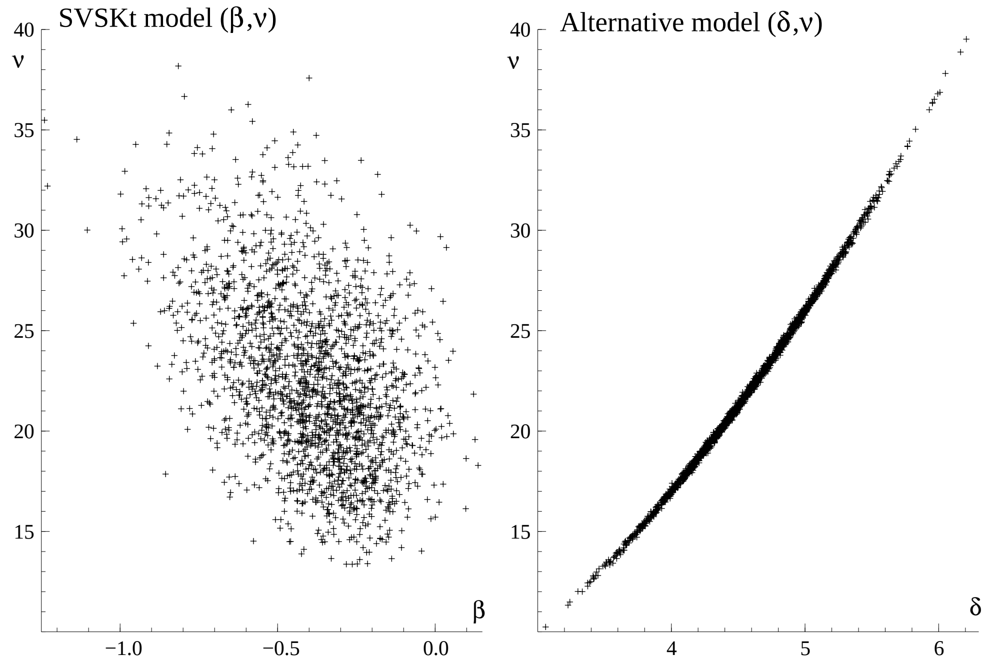

Figure 6: Scatter plots of posterior samples of (β, ν) for the SVSKt model (left) and (δ, ν) for the alternative model (right) using the S&P500 return (1994-2003) data.

4.5. Prior sensitivity analysis

To check the robustness of the model comparison, we assess the sensitivity of our results to the choice of prior distributions. As we have assumed the values commonly used in the previous literature for the prior distributions of $(\phi, \sigma, \rho, \mu)$, we focus on the parameters of the GH skew Student's t-distribution, *i.e.*, on the skewness and heavy-tailedness parameters (β, ν).

Let Prior #1 denote the prior distribution with hyper-parameters assumed in the previous estimation. Three alternative priors are considered:

$$
\begin{aligned}
&\text{Prior \#1:} \quad \beta \sim N(0,1),\ \nu \sim \text{Gamma}(16, 0.8) I(\nu > 4),\\
&\text{Prior \#2:} \quad \beta \sim N(0,4),\ \nu \sim \text{Gamma}(16, 0.8) I(\nu > 4),\\
&\text{Prior \#3:} \quad \beta \sim N(0,1),\ \nu \sim \text{Gamma}(24, 0.6) I(\nu > 4),\\
&\text{Prior \#4:} \quad \beta \sim N(0,4),\ \nu \sim \text{Gamma}(24, 0.6) I(\nu > 4),\\
&\text{Prior \#5:} \quad \beta \sim N(0,1),\ \nu \sim \text{Gamma}(1.2, 0.03) I(\nu > 4),
\end{aligned}
$$

where we note that the mean and standard deviation are (40, 8) for Gamma(24,0.6) and (40, 36.5) for Gamma(1.2, 0.03), respectively. Prior #5 for ν is rather flat compared to Priors #1–#4. First, the SVSKt model is estimated using the S&P500 data (1994-2003) under alternative priors. The estimates for $(\phi, \sigma, \rho, \mu)$ are found to be almost the same under all priors. Table 7 shows the parameter estimates and the inefficiency factors for β and ν. The estimates for (β, ν) are not affected by changing the prior for β from Prior #1 to Prior #2 (or from Prior #3 to Prior #4).

On the other hand, the estimates of (β, ν) are largely affected by altering the prior for ν from Prior #1 to Prior #3 (or from Prior #2 to Prior #4). The estimates of β get smaller (from -0.4 to -0.6) and the posterior means of ν get larger (from 22 to 40), implying greater skewness and less heavy-tailedness. The posterior standard deviations also become larger reflecting the increase in the dispersion of the prior distribution for ν. Also, as suggested by the 95% credible intervals, the posterior distribution of ν (β) moves to right (left). Given less information on ν, as described by Prior #5, the estimate of β is similar to those obtained by using Priors #3 and #4, while the posterior mean of ν is around 36, and its standard deviation and credible intervals indicate the flatter posterior distribution.

	SVSKt model				
	Prior #1	Prior #2	Prior #3	Prior #4	Prior #5
	-0.3867 (0.1943)	-0.3813 (0.1980)	-0.6046 (0.2999)	-0.6766 (0.3243)	-0.5686 (0.3221)
β	[-0.8167, -0.0460]	[-0.7976, -0.0357]	[-1.2432, -0.0133]	[-1.3762, -0.0991]	[-1.3896, -0.0839]
	76.05	76.16	66.86	91.8	150.65
	21.432 (4.4932)	21.985 (4.4399)	38.492 (7.9765)	40.915 (7.1658)	36.457 (13.847)
ν	[15.316, 33.162]	[14.723, 31.495]	[25.499, 53.776]	[27.192, 57.732]	[16.533, 68.380]
	223.49	209.46	186.16	194.65	285.08

The first row: posterior mean and standard deviation in parentheses.
The second row: 95% credible interval in square brackets.
The third row: inefficiency factor.

Table 7: Prior sensitivity analysis for the SVSKt model. Parameter estimates of β and ν for S&P500 data (1994-2003).

Nakajima and Omori (2009) found that the posterior estimate of ν is rather more sensitive to the choice of the prior distribution for ν than other parameters in the SV model with a symmetric Student's t-error, which is also observed in our prior sensitivity analysis. In addition, our result indicates that the posterior estimate of β is also sensitive to the choice of the prior distribution for ν. This may be because the skewness and heavy-tailedness of the GH skew Student's t-distribution are determined by β and ν simultaneously rather than individually. Our main findings are that the prior distribution of ν with a higher mean value results in its higher posterior means and that it would even lead to a lower posterior mean of β so as to maintain some of the skewness and heavy-tailedness of the empirical return distribution, as shown in Figure 1 of Section 2.1.

Finally, we investigate the prior sensitivity of the marginal likelihoods for the SVt and the SVSKt models using S&P500 data (1994-2003). Table 8 reports the logarithm of estimated marginal likelihoods under alternative priors. For the SVSKt model, all priors yield almost the same marginal likelihoods, which is quite reasonable. Although the marginal likelihoods of Priors #1 and #2 are slightly larger than those of Priors #3–#5 for the SVt model, the SVSKt models are still favoured over the SVt model regardless of the choice of the prior.

Model	Prior #1	Prior #2	Prior #3	Prior #4	Prior #5
SVt	8417.16 (0.35)	8417.77 (0.39)	8413.69 (0.11)	8413.84 (0.12)	8412.46 (0.34)
SVSKt	8420.95 (0.32)	8419.53 (0.25)	8420.03 (0.42)	8418.16 (0.34)	8417.89 (0.36)

*Standard errors of the Log-ML in parentheses.

Table 8: Prior sensitivity analysis. Estimated marginal likelihoods on a logarithmic scale for S&P500 data (1994-2003).

5. Conclusion

This paper proposes a Bayesian estimation of the SV model with leverage and with a GH skew Student's t-error distribution to assess the asymmetrically heavy-tailed distributions of stock returns. The efficient MCMC estimation method is developed using the normal variance-mean mixture representation of the GH skew Student's t-distribution, where the mixing distribution is the inverse gamma distribution. We illustrate our proposed method using simulated data and applied it to daily stock return data. The models are compared on the basis of the marginal likelihood, and the estimation results show strong evidence of skewness and heavy-tailedness. The proposed model is found to outperform other SV models. The prior sensitivity analysis shows that our results are robust, except for the parameter estimates of (β, ν), which are affected by the choice of the prior distribution of ν.

Acknowledgement

The authors would like to thank the anonymous referees and Herman van Dijk for his helpful comments. This work is supported by the Grants-in-Aid for Scientific Research (A) 21243018 from the Japanese Ministry of Education, Science, Sports, Culture and Technology. The computational results were generated using Ox (Doornik (2006)).

Appendix. Multi-move sampler for the SVSKt model

Extending the algorithm of Omori and Watanabe (2008), we describe the multi-move sampler for sampling the volatility variable h in the SVSKt model. Defining $\alpha_t = h_t - \mu$, for $t = 0, \ldots, n$ and $\gamma = \exp(\mu/2)$, we consider the state space model with respect to $\{\alpha_t\}_{t=1}^n$ as

$$\begin{aligned} y_t &= \{\beta \bar{z}_t + \sqrt{z_t}\varepsilon_t\} \exp(\alpha_t/2)\gamma, \quad t = 1, \ldots, n, \quad \text{and} \\ \alpha_{t+1} &= \phi\alpha_t + \eta_t, \quad t = 0, \ldots, n-1. \end{aligned}$$

Let $\tilde{\Theta} = (\theta, \alpha_r, \alpha_{r+d+1}, z_r, \ldots, z_{r+d}, y_r, \ldots, y_{r+d})$. To sample a block $(\alpha_{r+1}, \ldots, \alpha_{r+d})$ from its joint conditional posterior density using MH algorithm, ($r \geq 0$, $d \geq 1$, $r + d \leq n$), we sample disturbances

$$\begin{aligned} (\eta_r, \ldots, \eta_{r+d-1}) &\sim \pi(\eta_r, \ldots, \eta_{r+d-1}|\tilde{\Theta}) \\ &\propto \prod_{t=r}^{r+d} \frac{1}{\sqrt{2\pi}\tilde{\sigma}_t} \exp\left\{-\frac{(y_t - \tilde{\mu}_t)^2}{2\tilde{\sigma}_t^2}\right\} \times \prod_{t=r}^{r+d-1} f(\eta_t) \times f(\alpha_{r+d}), \end{aligned}$$

where

$$\begin{aligned} \tilde{\mu}_t &= \{\beta\bar{z}_t + \rho_t\sqrt{z_t}(\alpha_{t+1} - \phi\alpha_t)/\sigma\} \exp(\alpha_t/2)\gamma, \\ \tilde{\sigma}_t^2 &= (1 - \rho_t^2) z_t \exp(\alpha_t)\gamma^2, \\ f(\alpha_{r+d}) &= \exp\left\{-\frac{(\alpha_{r+d+1} - \phi\alpha_{r+d})^2}{2\sigma^2}\right\} \cdot I[r + d < n], \end{aligned}$$

and $\rho_t = \rho \cdot I[r + d < n]$. To determine the block (r and d), we use the stochastic knots (e.g. Shephard and Pitt (1997)). Let $\underline{\eta} = (\eta_r, \ldots, \eta_{r+d-1})'$ and $\underline{\alpha} = (\alpha_{r+1}, \ldots, \alpha_{r+d})'$. To construct a proposal density based on the normal approximation of the posterior density of $\underline{\eta}$, we first define

$$
\begin{aligned}
L &= \sum_{t=r}^{r+d} \left\{ -\frac{\alpha_t}{2} - \frac{(y_t - \tilde{\mu}_t)^2}{2\tilde{\sigma}_t^2} \right\} + \log f(\alpha_{r+d}), \\
\delta &= (\delta_{r+1}, \ldots, \delta_{r+d})', \quad \delta_t = \frac{\partial L}{\partial \alpha_t}, \\
Q &= -E\left(\frac{\partial^2 L}{\partial \underline{\alpha} \partial \underline{\alpha}'} \right) = \begin{pmatrix} A_{r+1} & B_{r+2} & 0 & \cdots & 0 \\ B_{r+2} & A_{r+2} & B_{r+3} & \cdots & 0 \\ 0 & B_{r+3} & A_{r+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & B_{r+d} \\ 0 & \cdots & 0 & B_{r+d} & A_{r+d} \end{pmatrix}, \\
A_t &= -E\left(\frac{\partial^2 L}{\partial \alpha_t^2} \right), \quad \text{and} \quad B_t = -E\left(\frac{\partial^2 L}{\partial \alpha_t \partial \alpha_{t-1}} \right),
\end{aligned}
$$

for $t = r+2, \ldots, r+d$, and $B_{r+1} = 0$. For the second derivatives, we take the expectations with respect to y_t's and obtain

$$
\begin{aligned}
A_t &= \frac{1}{2} + \frac{1}{\tilde{\sigma}_t^2} \left(\frac{\partial \tilde{\mu}_t}{\partial \alpha_t} \right)^2 + \frac{1}{\tilde{\sigma}_{t-1}^2} \left(\frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t} \right)^2 + \phi^2/\sigma^2 \cdot I[t = r+d < n], \text{ and} \\
B_t &= \frac{1}{\tilde{\sigma}_{t-1}^2} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_{t-1}} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t}.
\end{aligned}
$$

Applying the second-order Taylor expansion to the log of the posterior density around the mode, $\underline{\eta} = \hat{\underline{\eta}}$, we obtain an approximate normal density as follows:

$$
\begin{aligned}
&\log \pi(\underline{\eta}|\tilde{\Theta}) \\
\approx\ & \hat{L} + \left. \frac{\partial L}{\partial \underline{\eta}'} \right|_{\underline{\eta}=\hat{\underline{\eta}}} (\underline{\eta} - \hat{\underline{\eta}}) + \frac{1}{2}(\underline{\eta} - \hat{\underline{\eta}})' E\left(\frac{\partial^2 L}{\partial \underline{\eta} \partial \underline{\eta}'} \right) \Bigg|_{\underline{\eta}=\hat{\underline{\eta}}} (\underline{\eta} - \hat{\underline{\eta}}) + \sum_{t=r}^{r+d-1} \left(-\frac{1}{2} \eta_t^2 \right) + (\text{const.}) \\
=\ & \hat{L} + \hat{\delta}'(\underline{\alpha} - \hat{\underline{\alpha}}) - \frac{1}{2}(\underline{\alpha} - \hat{\underline{\alpha}})' \hat{Q} (\underline{\alpha} - \hat{\underline{\alpha}}) + \sum_{t=r}^{r+d-1} \left(-\frac{1}{2} \eta_t^2 \right) + (\text{const.}) \\
\equiv\ & \log q(\underline{\eta}|\tilde{\Theta}).
\end{aligned}
$$

where $\hat{L}$, $\hat{\delta}$ and $\hat{Q}$ is the value of L, δ and Q at $\underline{\alpha} = \underline{\hat{\alpha}}$ (or, equivalently at $\underline{\eta} = \underline{\hat{\eta}}$). It can be shown that the proposal density $q(\underline{\eta}|\tilde{\Theta})$ is the posterior density of $\underline{\eta}$ for a linear Gaussian state space model given by (10)–(12) below. The mode $\underline{\hat{\eta}}$ can be obtained by repeating the following algorithm until it converges.

1. Initialize $\underline{\hat{\eta}}$ and compute $\underline{\hat{\alpha}}$ at $\underline{\eta} = \underline{\hat{\eta}}$ using the state equation (5) recursively.
2. Evaluate $\hat{\delta}_t$'s, $\hat{A}_t$'s and $\hat{B}_t$'s at $\underline{\alpha} = \underline{\hat{\alpha}}$.
3. Let $\hat{D}_{r+1} = \hat{A}_{r+1}$ and $\hat{b}_{r+1} = \hat{\delta}_{r+1}$. Compute the following variables recursively for $t = r+2, \ldots, r+d$:

$$\hat{D}_t = \hat{A}_t - \hat{D}_{t-1}^{-1}\hat{B}_t^2, \qquad \hat{K}_t = \sqrt{\hat{D}_t}, \qquad \hat{b}_t = \hat{\delta}_t - \hat{B}_t\hat{D}_{t-1}^{-1}\hat{b}_{t-1},$$

   and $\hat{B}_{d+r+1} = 0$.
4. Define an auxiliary variable $\hat{y}_t = \hat{\gamma}_t + \hat{D}_t^{-1}\hat{b}_t$, where $\hat{\gamma}_t = \hat{\alpha}_t + \hat{D}_t^{-1}\hat{B}_{t+1}\hat{\alpha}_{t+1}$, for $t = r+1, \ldots, r+d$, and $\hat{\alpha}_{r+d+1} = \alpha_{r+d+1}$.
5. Consider the linear Gaussian state space model formulated by

$$\hat{y}_t = Z_t\alpha_t + G_t\zeta_t, \qquad t = r+1, \ldots, r+d, \tag{10}$$
$$\alpha_{t+1} = \phi\alpha_t + H_t\zeta_t, \qquad t = r, \ldots, r+d, \tag{11}$$

   and

$$\zeta_t \sim N(0, I_2), \tag{12}$$

   where

$$Z_t = 1 + \phi\hat{D}_t^{-1}\hat{B}_{t+1}, \quad G_t = (\hat{K}_t^{-1}, \hat{D}_t^{-1}\hat{B}_{t+1}\sigma), \quad \text{and} \quad H_t = (0, \sigma),$$

   for $t = r+1 \ldots, r+d$ and $H_0 = (0, \sigma/\sqrt{1-\phi^2})$. Apply the Kalman filter and the disturbance smoother to this state space model, and obtain the posterior mode $\underline{\hat{\eta}}$ and $\underline{\hat{\alpha}}$.
6. Go to 2.

In the MCMC sampling procedure, the current sample of $\underline{\eta}$ may be taken as an initial value of the $\underline{\hat{\eta}}$ in Step 1. To sample $\underline{\eta}$ from the conditional posterior density, we implement the AR (Accept-Reject)-MH algorithm via the simulation smoother (e.g., de Jong and Shephard (1995), Durbin and Koopman (2002)) using the mode $\underline{\hat{\eta}}$ to obtain the approximated linear Gaussian state space model (10)–(12). See Omori and Watanabe (2008), Takahashi et al. (2009) for the detail of this AR-MH algorithm.

References

Aas, K., Haff, I.H., 2006. The generalized hyperbolic skew Student's t-distribution. Journal of Financial Econometrics 4, 275–309.

Andersson, J., 2001. On the normal inverse Gaussian stochastic volatility model. Journal of Business and Economic Statistics 19, 44–54.

Azzalini, A., Capitanio, A., 2003. Distributions generated by pertubation of symmetry with emphasis on a multivariate skew t distribution. Journal of the Royal Statistical Society B 65, 579–602.

Barndorff-Nielsen, O.E., 1977. Exponentially decreasing distributions for the logarithm of particle size. Proceedings of the Royal Society of London, Series A 353, 401–419.

Barndorff-Nielsen, O.E., 1997. Normal inverse Gaussian distributions and stochastic volatility modelling. Scandinavian Journal of Statistics 24, 1–13.

Bauwens, L., Lubrano, M., 1998. Bayesian inference on GARCH models using Gibbs sampler. Econometrics Journal 1, c23–c46.

Berg, A., Meyer, R., Yu, J., 2004. DIC as a model comparison criterion for stochastic volatility models. Journal of Business and Economic Statistics 22, 107–120.

Chen, C.W.S., Liu, F.C., Mike, K.P.S., 2008. Heavy-tailed-distributed threshold stochastic volatility models in financial time series. Australian and New Zealand Journal of Statistics 50, 29–51.

Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G., 2003. Alternative models for stock price dynamics. Journal of Econometrics 116, 225–257.

Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321.

Chib, S., 2001. Markov chain Monte Carlo methods: computation and inference, in: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics. North-Holland, Amsterdam. volume 5, pp. 3569–3649.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. American Statistician 49, 327–335.

Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis-Hastings output. Journal of the American Statistical Association 96, 270–291.

Chib, S., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics 108, 281–316.

Doornik, J., 2006. Ox: Object Oriented Matrix Programming. Timberlake Consultants Press, London.

Durbin, J., Koopman, S.J., 2002. Simple and efficient simulation smoother for state space time series analysis. Biometrika 89, 603–616.

Eberlein, E., Keller, U., Prause, K., 1998. New insighs into smile, mispricing and value at risk: the hyperbolic model. Journal of Business 71, 371–405.

Eraker, B., 2004. Do equity prices and volatility jump? Reconciling evidence from spot and option prices. Journal of Finance 59, 1367–1403.

Eraker, B., Johanners, M., Polson, N.G., 2003. The impact of jumps in returns and volatility. Journal of Finance 53, 1269–1330.

Fernández, C., Steel, M.F.J., 1998. On Bayesian modeling of fat tails and skewness. Journal of the American Statistical Association 93, 359–371.

Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference. Chapman & Hall/CRC, Boca Raton, FL. 2 edition.

Geweke, J., 2005. Contemporary Bayesian Econometrics and Statistics. Wiley.

Ghysels, E., Harvey, A.C., Renault, E., 2002. Stochastic volatility, in: Rao, C.R., Maddala, G.S. (Eds.), Statistical Methods in Finance. Amsterdam: North-Holland, pp. 119–191.

Hansen, B.E., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.

Jones, M.C., Faddy, M.J., 2003. A skew extension of the t-distribution, with application. Journal of Royal Statistical Society, Series B 65, 159–174.

de Jong, P., Shephard, N., 1995. The simulation smoother for time series models. Biometrika 82, 339–350.

Koop, G., 2003. Bayesian Econometrics. Wiley, Chichester.

Nakajima, J., Omori, Y., 2009. Leverage, heavy-tails and correlated jumps in stochastic volatility models. Computational Statistics and Data Analysis 53, 2535–2553.

Omori, Y., Chib, S., Shephard, N., Nakajima, J., 2007. Stochastic volatility with leverage: fast likelihood inference. Journal of Econometrics 140, 425–449.

Omori, Y., Watanabe, T., 2008. Block sampler and posterior mode estimation for asymmetric stochastic volatility models. Computational Statistics and Data Analysis 52, 2892–2910.

Pitt, M., Shephard, N., 1999. Filtering via simulation: auxiliary particle filter. Journal of the American Statistical Association 94, 590–599.

Prause, K., 1999. The Generalized Hyperbolic models: estimation, financial derivatives and risk measurement. PhD dissertation, University of Freiburg.

Raggi, D., Bordignon, S., 2006. Comparing stochastic volatility models through Monte Carlo simulations. Computational Statistics and Data Analysis 50, 1678–1699.

Rydberg, T.H., 1999. Generalized hyperbolic diffusion processes with applications in finance. Mathematical Finance 9, 183–201.

Shephard, N., 2005. Stochastic Volatility: Selected Readings. Oxford University Press, Oxford.

Shephard, N., Pitt, M., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika 84, 653–667.

Strickland, C.M., Martin, G.M., Forbes, C.S., 2008. Parameterisation and efficient MCMC estimation of non-Gaussian state space models. Computational Statistics and Data Analysis 51, 2911–2930.

Takahashi, M., Omori, Y., Watanabe, T., 2009. Estimating stochastic volatility models using daily returns and realized volatility simultaneously. Computational Statistics and Data Analysis 53, 2404–2426.

Watanabe, T., Omori, Y., 2004. A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard & Pitt (1997). Biometrika 91, 246–248.

Yu, J., 2005. On leverage in a stochastic volatility model. Journal of Econometrics 127, 165–178.