HITOTSUBASHI UNIVERSITY

Global COE Hi-Stat Discussion Paper Series 124

Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

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

Jouchi Nakajima
Yasuhiro Omori

March 2010

Hi-Stat Discussion Paper

Hi-Stat
Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitatchi Tokyo, 186-8601 Japan
http://gcoe.ier.hit-u.ac.jp

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

Jouchi Nakajima

Department of Statistical Science, Duke University, Durham 27705, U.S.A. and Bank of Japan

jouchi.nakajima@stat.duke.edu

Yasuhiro Omori

Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan

omori@e.u-tokyo.ac.jp

December 19, 2009

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-tailness as well as leverage effects. An efficient Markov chain Monte Carlo estimation method is described exploiting a normal variance-mean mixture representation of the error distribution with an inverse gamma distribution as a mixing distribution. The proposed method is illustrated using simulated data, daily TOPIX and S&P500 stock returns. The model comparison for stock returns is conducted based on the marginal likelihood in the empirical study. The strong evidence of the leverage and asymmetric heavy-tailness is found in the stock returns. Further, the prior sensitivity analysis is conducted to investigate whether obtained results are robust with respect to the choice of the priors.

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

1 Introduction

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

As a promising approach to model the flexible skewness and heavy-tailness, the generalized hyperbolic (GH) distribution proposed by Barndorff-Nielsen (1977) has recently attracted attention in financial econometrics since 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 an affine transformation, conditioning and marginalization. Several studies have investigated the skewness and heavy-tailness of financial market variables using the subclass of the GH distribution: the hyperbolic distribution (Eberlein et al. (1998)), the GH diffusion process (Rydberg (1999)), GH skew Student's t-distribution (Hansen (1994), Fernández and Steel (1998), Aas and Haff (2006)) for the unconditional return distribution.

On the other hand, regarding the volatility clustering, the stochastic volatility (SV) model has been widely used to model a 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 past literature. For example, to describe the heavy-tailness of the asset return distribution in the SV context, heavy-tailed errors are often incorporated using such as Student's t-distribution (Chib et al. (2002), Eraker et al. (2003), Berg et al. (2004), Yu (2005), Omori et al. (2007), Nakajima and Omori (2009)) and NIG distribution (Barndorff-Nielsen (1997), Andersson (2001)). Also, the SV model with jump diffusions for stock returns have been considered (Eraker (2004), Chernov et al. (2003) and Raggi and Bordignon (2006)). The comparison of these models using S&P500 and TOPIX daily returns in Nakajima and Omori (2009) showed that the SV model with symmetric Student's t-errors (SVt) model performs better than the SV model with jumps or both jumps and Student's t-errors.

This paper proposes, first in the literature to the best of our knowledge, an efficient Bayesian estimation method of the SV model with both leverage and 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 in the GH distribution, and 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 many latent volatility variables.

It requires a 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 a precise and efficient estimation of the SV model with leverage and 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 them, the GH skew Student's t-error distribution is simple, flexible and easy to be incorporated into the SV model for a Bayesian estimation scheme using the MCMC algorithm that we develop in this paper. The key feature to implement 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 distribution 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 the 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 GH skew Student's t-distribution. Section 3 illustrates our proposed method using simulated data. We also examine an alternative parameterization in the GH skew Student's t-distribution. In Section 4, the proposed model is applied to S&P500 and TOPIX daily returns data and the model comparison is provided among the competing SV models. Finally Section 5 concludes.

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

2.1 The model

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

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

where y_t is an asset return, and h_t is an unobserved log-volatility. We assume $|\phi| < 1$ for a stationarity of the log-volatility process, and 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 $\rho < 0$, it refers to a so-called leverage effect, a drop in the return followed by an increase in the volatility (Yu (2005), Omori et al. (2007)).

To model the leverage and asymmetric heavy-tailness jointly, we replace a normal random variable ε_t in (1) by a GH random variable w_t. The random variable, w_t, can be written in the form of the normal variance-mean mixture as

$$w_t = \mu_w + \beta z_t + \sqrt{z_t}\epsilon_t, \quad \epsilon_t \sim N(0,1), \quad z_t \sim GIG(\lambda, \delta, \gamma). \tag{2}$$

As pointed out 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 subclass, a GH skew Student's t-distribution, where $\lambda = -\nu/2$ $(\nu > 0)$ and $\gamma = 0$. Thus we assume $\delta = \sqrt{\nu}$, which yields $z_t \sim GIG(-\nu/2, \sqrt{\nu}, 0)$, or equivalently $IG(\nu/2, \nu/2)$ where IG denotes the inverse gamma distribution. Further, we assume $E(w_t) = 0$, $E(w_t^2) < \infty$ by setting $\mu_w = -\beta\mu_z$ where $\mu_z \equiv \mathrm{E}(z_t) = \nu/(\nu-2)$ and $\nu > 4$. The validity of this assumption will be discussed in Section 3.3 in comparison with an alternative parameterization.

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

$$y_t = \{\beta(z_t - \mu_z) + \sqrt{z_t}\varepsilon_t\}\exp(h_t/2), \quad t = 1, \ldots, n, \tag{3}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, \ldots, n-1, \tag{4}$$

$$z_t \sim IG(\nu/2, \nu/2), \tag{5}$$

where (ε_t, η_t) follows (1). The $\nu > 4$ is the degree of freedom and unknown to be estimated. When $\beta \equiv 0$, the model reduces to the SV model with symmetric Student's t-distribution (denoted SVt model), which is widely analyzed in 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-tailness, 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. The lower value of β implies a more negative skewness or left-skewness as well as heavier tails. Figure 1(ii) shows the densities using $\nu = 5$, 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-tailness are determined jointly by the combination of the parameter values of β and ν.

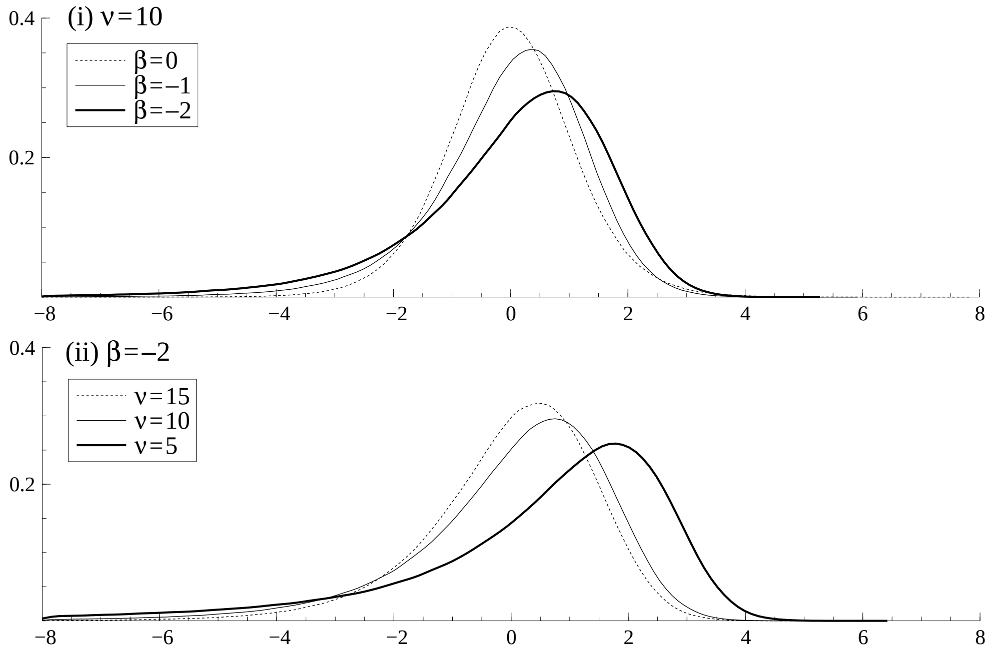

Figure 1: The GH skew Student's t-distribution. (i) $\nu = 10$ fixed; $\beta = 0$ (symmetric t), -1 and -2. (ii) $\beta = -2$ fixed; $\nu = 5$, 10 and 15.

We 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) overviews 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 we mentioned, introducing more parameters would lead to an overparameterization since the second moment of the return distribution is already modeled as a latent stochastic process in the SV model. Therefore, there would be not so much room to obtain thoughtful estimates from additional parameters.

Our formulation (3) 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 (2). We regard the variable z_t following the mixing distribution as a latent variable to accomplish a novel construction of the MCMC algorithm in the context of Bayesian inference. The conditional posterior distribution of each parameter reduces to much more tractable form conditional on z_t than the case 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.

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 prior distributions of μ and β, we assume

$$\mu \sim N(\mu_0, v_0^2), \quad \beta \sim N(\beta_0, \sigma_0^2), \tag{6}$$

and we let $\pi(\phi)$, $\pi(\vartheta)$, $\pi(\nu)$ denote 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., Chib and Greenberg (1995), Tierney (1994)) 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 show 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

$$\begin{aligned} \pi(\phi|\cdot) &\propto \pi(\phi)\sqrt{1-\phi^2}\exp\left\{-\frac{(1-\phi^2)\bar{h}_1^2}{2\sigma^2} - \sum_{t=1}^{n-1}\frac{(\bar{h}_{t+1} - \phi\bar{h}_t - \bar{y}_t)^2}{2\sigma^2(1-\rho^2)}\right\} \\ &\propto \pi(\phi)\sqrt{1-\phi^2}\exp\left\{-\frac{(\phi-\mu_\phi)^2}{2\sigma_\phi^2}\right\}, \end{aligned} \tag{7}$$

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 a normal distribution with mean μ and variance σ^2 truncated on an 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. Since 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

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

which is not easy to sample from, we conduct the MH algorithm based on a normal approximation of the density around the mode. Since we have a constraint, $R = \{\vartheta : \sigma > 0,\ |\rho| < 1\}$, for the parameter space of the posterior distribution, we consider a 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 $\hat{\vartheta}$ which maximizes (or approximately maximizes) $\pi(\vartheta|\cdot)$, and obtain its transformed value $\hat{\omega}$. We generate a candidate $\omega^* \sim N(\omega_*, \Sigma_*)$, where

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

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 a probability density function of a normal distribution with mean μ and covariance matrix Σ, and $J(\cdot)$ is a 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

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

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

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

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

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

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

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

Step 6. Since, as in Step 3, the posterior probability density of ν

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

is not easy to sample from, 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

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

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 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 the state space modeling. They show 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 others, h_s $(s \neq t)$. For the SV model with leverage, Omori and Watanabe (2008) develop the associated multi-move sampler and show it produces efficient samples (see also Takahashi et al. (2009)). We extend their method for sampling h in the SVSKt model. The detail of the multi-move sampler is described in 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 (3)–(5) with specified 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:

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

The beta prior distribution for $(\phi+1)/2$ implies that 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 in the past literature.

We draw 20,000 samples after the initial 2,000 samples are discarded as a burn-in period, which is selected using time series plots of the marginal averages of samples for each parameter. We compute the inefficiency factor to check the efficiency of the MCMC algorithm. The inefficiency factor is defined as $1 + 2\sum_{s=1}^{\infty} \rho_s$ 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 Parzen window with a 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 look stable and the sample autocorrelations decay quickly, which implies that our sampling method is efficient.

Table 1 shows posterior means, standard deviations, the 95% credible intervals and inefficiency factors. All the posterior means are close enough to the true values such that the corresponding 95% credible intervals include true values. The inefficiency factors in Table 1 are found to be 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.

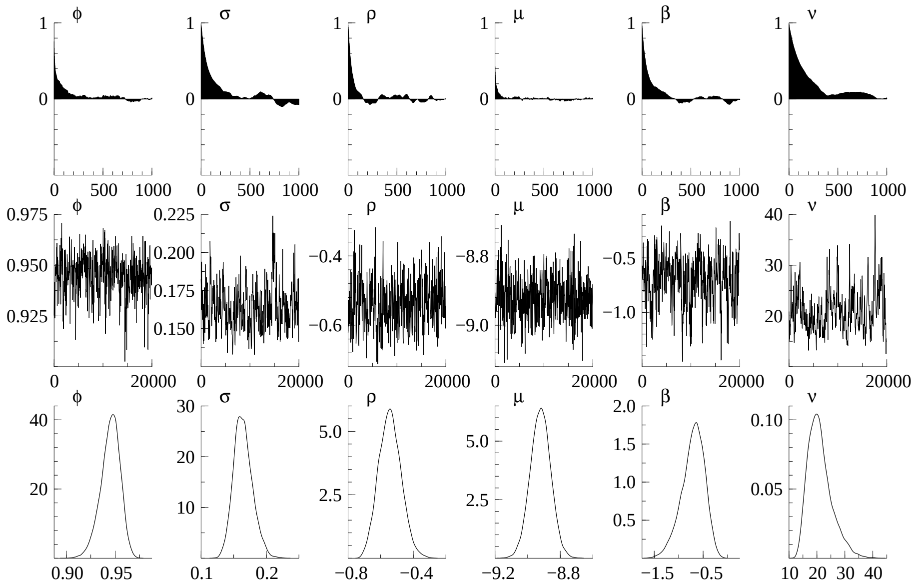

Figure 2: MCMC estimation result of the SVSKt model for simulated data. Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

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 result 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. As for 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 we 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), (3) and (4) but we replace (5) by

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

where $\delta > 0$. Since there are two parameters, (ν, δ), to determine first and second moments of z_t, we set $\mu \equiv 0$ to identify the parameter for the second moment of the return distribution. We generate 3,000 observations from the alternative model with parameter values $\phi = 0.95$, $\sigma = 0.08$, $\rho = -0.4$, $\beta = -0.3$, $\nu = 14$ and $\delta = 4$. In addition to the previous experiment, we assume the prior distribution as $\delta \sim \text{Gamma}(4, 0.4)$, which implies 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 (0.98), while that between β and ν is moderate (-0.69). This suggests that we need to sample under the narrow state space when we use the alternative parameterization, which would result in the inefficient sampling. Thus, although we could model the GH skew Student's t-distribution in other ways, alternative models could lead to either the inefficient MCMC sampling or the over-parameterization. This example shows that our proposed parameterization is appropriate for the SV model with the GH skew Student's t-distribution.

(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	β	ν	δ
ϕ	1	-.33	-.10	-.06	.04	.05
σ		1	-.13	.01	.01	.01
ρ			1	.01	-.04	-.04
β				1	-.68	-.60
ν					1	.98
δ						1

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

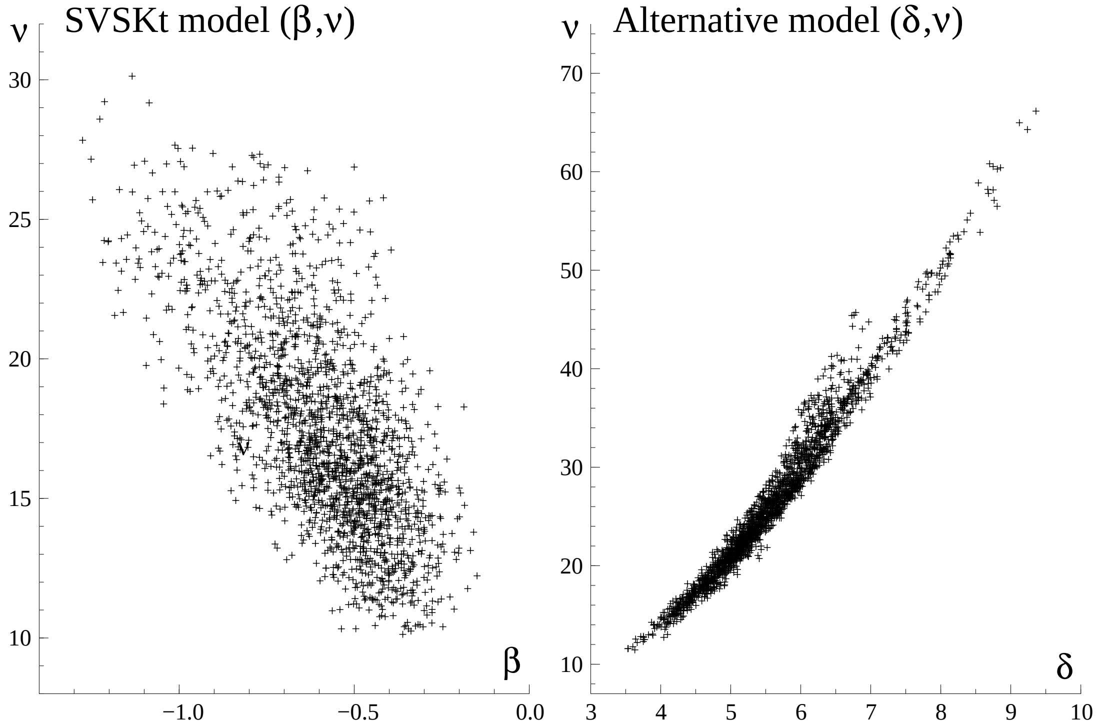

Figure 3: Scatter plots of posterior samples of (β, ν) for the SVSKt model (left) and (δ, ν) for the alternative model (right).

4 Application to stock returns data

4.1 Data

This section applies our proposed model to daily stock returns 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 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 huge as 37 for S&P500 and 16 for TOPIX. This is partly due to the largest negative return corresponding 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-tailness of empirical returns distribution of the data.

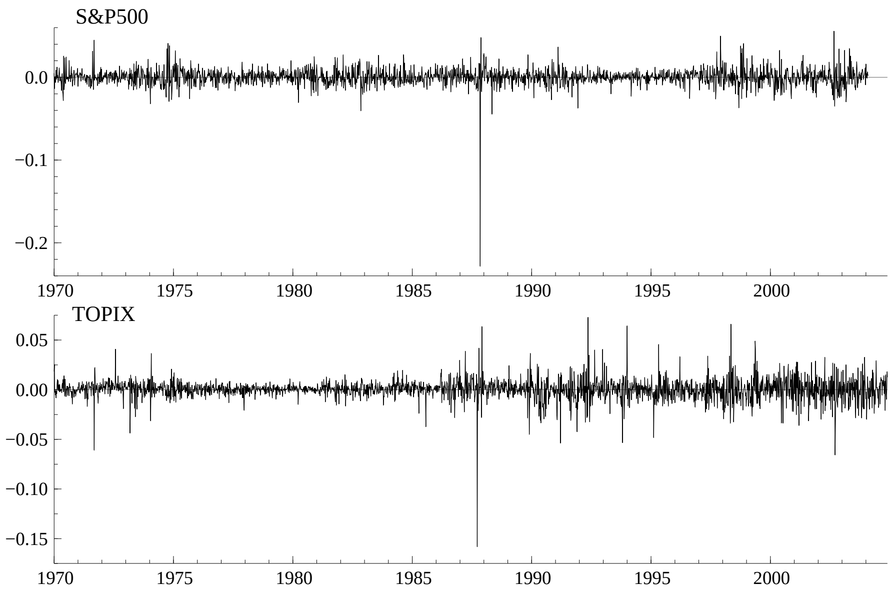

Figure 4: Time series plots for S&P500 (1970/1/1 - 2003/12/31) and TOPIX (1970/1/5 - 2004/12/30) daily returns.

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 discarding initial samples are also taken as in Section 3. Figure 5 shows the estimation results for S&P500 data where the sample paths look 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's are estimated to be negative, implying that there exist the leverage effects. Regarding the skewness, the posterior means of β are -0.0946 for 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 mostly located in the negative range as shown in Figure 5. For 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 that there are 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-tailness in the stock return distributions especially for S&P500 data.

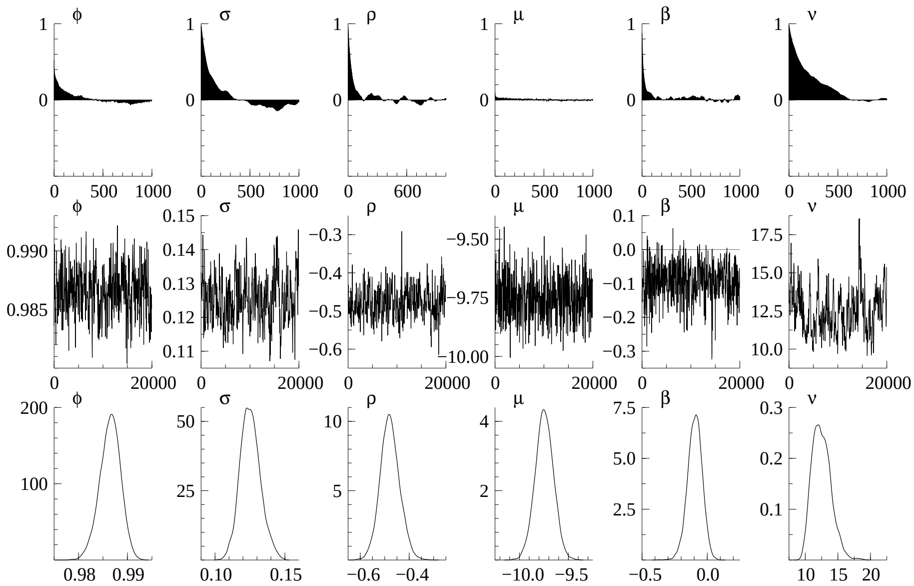

Figure 5: MCMC estimation result of the SVSKt model for S&P500 data. Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

(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 result 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 past 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-t error distribution ($\beta = 0$).

Note that all models are allowed to have the leverage effect (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 supported by the data. The posterior probability of each model is proportional to the prior probability of the model times 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 which 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

$$\log m(y) = \log f(y|\Theta) + \log \pi(\Theta) - \log \pi(\Theta|y). \tag{8}$$

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 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, though the likelihood and posterior part need a simulation evaluation.

The likelihood is estimated using the auxiliary particle filter (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 the additional but reduced MCMC runs. The number of iterations for the reduced run is set 5,000.

We use six 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 and from 1990 to 2003, and the TOPIX series from 1970 to 1985 and from 1990 to 2004, *i.e.*, we consider two long-period (about thirty years) data and four short-period (about fifteen years) data. We select these short periods such that the crash of October 1987 is excluded, because the huge negative return could affect the model selection among the competing models.

Table 5 reports the logarithm of estimated marginal likelihoods, their standard errors and rankings. Overall, the SVSKt model outperforms other models for all dataset regardless of sample periods. Taking account of the standard errors, we can see that the GH skew Student's t-error distribution in the SV model clearly is successful to describe the distribution of the daily stock returns 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 negative for S&P500 of 1970-2003 and 1994-2003, TOPIX of 1970-2004 and 1970-1985, while it is centered around zero for the TOPIX of 1992-2004, and moreover, almost positive for the S&P500 of 1970-1985. The skewness of the empirical return distributions seems to change depending on the sample periods. However, the SVSKt model is still favoured over other symmetric error SV models in every period.

S&P500	1970-2003		1970-1985		1994-2003	
Model	Log-ML	Ranking	Log-ML	Ranking	Log-ML	Ranking
SV	29605.67 (1.54)	3	14198.89 (0.39)	3	8406.06 (0.37)	3
SVt	29657.41 (1.62)	2	14205.03 (0.47)	2	8417.55 (0.43)	2
SVSKt	29666.51 (1.42)	1	14206.97 (0.40)	1	8419.35 (0.23)	1
Posterior of β						
Mean (Stdev.)	-0.0946 (0.0558)		0.2699 (0.1775)		-0.3942 (0.1977)	
95% interval	[-0.2093, 0.0097]		[-0.0460, 0.6599]		[-0.8165, -0.0460]	

TOPIX	1970-2004		1970-1985		1992-2004	
Model	Log-ML	Ranking	Log-ML	Ranking	Log-ML	Ranking
SV	32461.14 (1.50)	3	17626.79 (0.54)	2	9738.27 (0.22)	3
SVt	32483.03 (1.55)	2	17641.75 (0.49)	3	9743.49 (0.32)	2
SVSKt	32490.13 (0.72)	1	17665.91 (0.52)	1	9746.98 (0.31)	1
Posterior of β						
Mean (Stdev.)	-0.3901 (0.1225)		-0.5979 (0.1790)		-0.0163 (0.1109)	
95% interval	[-0.6517, -0.1615]		[-0.9643,-0.2730]		[-0.2068, 0.2344]	

*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 Prior sensitivity analysis

To check the robustness of the model comparison, we assess the sensitivity of our results to the choice of prior distributions. Since we assumed the values commonly used in the previous literature for the prior distributions of $(\phi, \sigma, \rho, \mu)$, we focus on the parameters of GH skew Student's t-distribution, *i.e.*, the skewness and heavy-tailness 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. The 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 6 shows the parameter estimates and 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 become smaller (from -0.4 to -0.6) and the posterior means of ν get larger (from 22 to 40), implying more skewness and less heavy-tailness. The posterior standard deviations also become larger reflecting the increase in the dispersion of the prior distribution for ν. Also, as suggested by 95% credible intervals, the posterior distribution of ν (β) moves to right (left). Under less informative situation for 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 6: Prior sensitivity analysis for the SVSKt model. Parameter estimates of β and ν for S&P500 data (1994-2003).

Nakajima and Omori (2009) found the posterior estimate of ν is rather 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-tailness in 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 the lower posterior mean of β so as to keep some skewness and heavy-tailness 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 7 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 still remains 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 7: 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 the 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 to daily stock returns data. The model comparison is conducted based on the marginal likelihood and the estimation results show the strong evidence of the skewness and heavy-tailness. The proposed model is found to outperform other SV models. The prior sensitivity analysis shows that our results are robust except parameter estimates of (β, ν) which are affected by the choice of the prior distribution of ν.

Acknowledgement

The authors would like to thank 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 are generated using Ox (Doornik (2006)). Views expressed in this paper are those of the authors and do not necessarily reflect the official views of the Bank of Japan.

Appendix. Multi-move sampler for sampling h in the SVSKt model

Extending 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

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

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 consider sampling 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

$$\begin{aligned} \tilde{\mu}_t &= \begin{cases} \left\{\beta\bar{z}_t + \rho\sqrt{z_t}(\alpha_{t+1} - \phi\alpha_t)/\sigma\right\}\exp(\alpha_t/2)\gamma & (\text{if } t < n) \\ \beta\bar{z}_n \exp(\alpha_n/2)\gamma & (\text{if } t = n), \end{cases} \\ \tilde{\sigma}_t^2 &= \begin{cases} (1-\rho^2)z_t \exp(\alpha_t)\gamma^2 & (\text{if } t < n) \\ z_n \exp(\alpha_n)\gamma^2 & (\text{if } t = n), \end{cases} \\ f(\alpha_{r+d}) &= \begin{cases} \exp\left\{-\frac{(\alpha_{r+d+1} - \phi\alpha_{r+d})^2}{2\sigma^2}\right\} & (\text{if } r + d < n) \\ 1 & (\text{if } r + d = n), \end{cases} \end{aligned}$$

and $\tilde{\Theta} = (\theta, \alpha_r, \alpha_{r+d+1}, z_r, \ldots, z_{r+d}, y_r, \ldots, y_{r+d})$. 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

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

For the first derivatives, we have

$$\delta_t \;=\; -\frac{1}{2} + \frac{(y_t - \tilde{\mu}_t)^2}{2\tilde{\sigma}_t^2} + \frac{y_t - \tilde{\mu}_t}{\tilde{\sigma}_t^2} \cdot \frac{\partial \tilde{\mu}_t}{\partial \alpha_t} + \frac{y_{t-1} - \tilde{\mu}_{t-1}}{\tilde{\sigma}_{t-1}^2} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t} + j(\alpha_t),$$

where

$$\begin{aligned}
\frac{\partial \tilde{\mu}_t}{\partial \alpha_t} &= \begin{cases} \left\{ \dfrac{\beta \bar{z}_t}{2} + \rho\sqrt{z_t}\left(-\phi + \dfrac{\alpha_{t+1} - \phi\alpha_t}{2} \right) / \sigma \right\} \exp(\alpha_t/2)\gamma & (\text{if } t < n) \\ 0 & (\text{if } t = n), \end{cases} \\
\frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t} &= \begin{cases} 0 & (\text{if } t = 1) \\ \rho\sqrt{z_{t-1}}\exp(\alpha_{t-1}/2)\gamma/\sigma & (\text{if } t > 1), \end{cases} \\
j(\alpha_t) &= \begin{cases} \dfrac{\phi(\alpha_{t+1} - \phi\alpha_t)}{\sigma^2} & (\text{if } t = r+d < n) \\ 0 & (\text{otherwise}). \end{cases}
\end{aligned}$$

For the second derivatives, we take expectations with respect to y_t's and obtain

$$\begin{aligned}
A_t &= \frac{1}{2} + \frac{1}{\tilde{\sigma}_t^2}\left(\frac{\partial \tilde{\mu}_t}{\partial \alpha_t} \right)^2 + \frac{1}{\tilde{\sigma}_{t-1}^2}\left(\frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t} \right)^2 + j'(\alpha_t), \\
B_t &= \frac{1}{\tilde{\sigma}_{t-1}^2} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_{t-1}} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t},
\end{aligned}$$

where

$$j'(\alpha_t) = \begin{cases} \phi^2/\sigma^2 & (\text{if } t = r+d < n) \\ 0 & (\text{otherwise}). \end{cases}$$

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

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

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 (9)–(11) below. The mode $\underline{\hat{\eta}}$ can be obtained by repeating the following algorithm until it converges.

Algorithm 1 (Disturbance smoother):

1. Initialize $\underline{\hat{\eta}}$ and compute $\underline{\hat{\alpha}}$ at $\underline{\eta} = \underline{\hat{\eta}}$ using the state equation (4) 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 $t = r+2, \ldots, r+d$:

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

   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

$$\begin{aligned} \hat{y}_t &= Z_t \alpha_t + G_t \zeta_t, \qquad t = r+1, \ldots, r+d, && (9) \\ \alpha_{t+1} &= \phi \alpha_t + H_t \zeta_t, \qquad t = r, \ldots, r+d, && (10) \\ \zeta_t &\sim N(0, I_2), && (11) \end{aligned}$$

where

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

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 then obtain the posterior mode $\hat{\underline{\eta}}$ and $\hat{\underline{\alpha}}$.

6. Go to 2.

In the MCMC sampling procedure, the current sample of $\underline{\eta}$ may be taken as an initial value of the $\hat{\underline{\eta}}$ in Step 1. Next, to sample $\underline{\eta}$ from the conditional posterior density, we implement the following AR(Accept-Reject)-MH algorithm via the simulation smoother (e.g., de Jong and Shephard (1995), Durbin and Koopman (2002)).

Algorithm 2 (AR-MH algorithm and simulation smoother):

1. Let $\underline{\eta}_0$ denote the current value. Find the mode $\hat{\underline{\eta}}$ using Algorithm 1.

2. Proceed Steps 2–4 in Algorithm 1 to obtain the approximated linear Gaussian state space model (9)–(11).

3. Propose a candidate $\underline{\eta}^*$ by sampling from the density $\tilde{q}(\underline{\eta}^*) \propto \min\{\pi(\underline{\eta}^*|\tilde{\Theta}), cq(\underline{\eta}^*|\tilde{\Theta})\}$ using the Accept-Reject algorithm as follows:

   (a) Generate $\underline{\eta}^*$ using the simulation smoother for the approximated state space model (9)–(11).

   (b) Accept $\underline{\eta}^*$ with a probability

   $$\frac{\min\{\pi(\underline{\eta}^*|\tilde{\Theta}), cq(\underline{\eta}^*|\tilde{\Theta})\}}{cq(\underline{\eta}^*|\tilde{\Theta})}.$$

   If it is rejected, go to (a).

4. Conduct the MH algorithm using the candidate $\underline{\eta}^*$. The MH acceptance probability is given by

$$\min\left\{\frac{\pi(\underline{\eta}^*|\tilde{\Theta}) \min\{\pi(\underline{\eta}_0|\tilde{\Theta}), cq(\underline{\eta}_0|\tilde{\Theta})\}}{\pi(\underline{\eta}_0|\tilde{\Theta}) \min\{\pi(\underline{\eta}^*|\tilde{\Theta}), cq(\underline{\eta}^*|\tilde{\Theta})\}}\right\}.$$

Finally, we note that $\alpha = (\alpha_1, \ldots, \alpha_n)'$ is divided into $K+1$ blocks at random, say, $(\alpha_{k_{i-1}+1}, \ldots, \alpha_{k_i})$ for $i = 1, \ldots, K+1$ with $k_0 = 0$ and $k_{K+1} = n$. We use the stochastic knots given by $k_i = \text{int}\,[n(i+U_i)/(K+2)]$, for $i = 1, \ldots, K$, where U_i is a random sample from a uniform distribution $U[0,1]$ (e.g. Shephard and Pitt (1997)) to make our sampling step for α (equivalently, h) more efficient.

References

Aas, K. and I. H. Haff (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. and A. Capitanio (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.

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

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen (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 J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, pp. 3569–3649. Amsterdam: North-Holland.

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician 49*, 327–335.

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

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

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

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

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

Eberlein, E., U. Keller, and K. Prause (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., M. Johanners, and N. G. Polson (2003). The impact of jumps in returns and volatility. *Journal of Finance 53*, 1269–1330.

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

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

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

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

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

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

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

Pitt, M. and N. Shephard (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. and S. Bordignon (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. (Ed.) (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.

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

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

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics 21*, 1701–1762.

Watanabe, T. and Y. Omori (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.