CIRJE-F-514

Leverage, Heavy-Tails and Correlated Jumps in Stochastic Volatility Models

Jouchi Nakajima
Bank of Japan

Yasuhiro Omori
University of Tokyo

September 2007; Revised in January 2008

Leverage, heavy-tails and correlated jumps in stochastic volatility models *

Jouchi Nakajima [a, †] Yasuhiro Omori [b, ‡]

[a] Bank of Japan, Tokyo 103-8660, Japan,
[b] Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan.

Abstract

This paper proposes the efficient and fast Markov chain Monte Carlo estimation methods for the stochastic volatility model with leverage effects, heavy-tailed errors and jump components, and for the stochastic volatility model with correlated jumps. Our method is illustrated using simulated data and analyze daily stock returns data on S&P500 index and TOPIX. Model comparisons are conducted based on the marginal likelihood for various SV models including the superposition model.

Key words: Bayesian analysis, Correlated jumps, Heavy-tailed error, Jumps, Leverage effect, Markov chain Monte Carlo, Marginal likelihood, Mixture sampler, Stochastic volatility, Stock returns.

[*] Views expressed are those of authors and do not necessarily reflect those of the Bank of Japan.
[†] E-mail:jouchi.nakajima-1@boj.or.jp
[‡] Corresponding author: Tel:+81-3-5841-5516, Fax:+81-3-5841-5521, E-mail:omori@e.u-tokyo.ac.jp

1 Introduction

The stochastic volatility (SV) models have been widely used to model a changing variance of time series in financial econometrics (e.g., Ghysels et al. (2002), Shephard (2005)). Various generalizations of the standard SV model have emerged and their model-fittings have been investigated especially in high-frequency financial data. Among such generalizations, the leverage effect, jump components and heavy-tailed errors in asset returns are well-known to be important in the recent literature (Chib et al. (2002), Jacquier et al. (2004), Yu (2005), Omori et al. (2007), Berg et al. (2004)). It has been pointed out in many empirical studies that asset returns data have heavier tails than those of normal distributions. The SV model with Student-t errors (SVt) is one of the most popular basic models to account for heavier tailed returns. However, it has been found insufficient to express the tail fatness of returns, and the jump components, which may be correlated, have recently been introduced to explain the tail behavior (Eraker et al. (2003)). The jump component is considered to be a discretization of a Lévy process which is also widely used in the continuous time modelling of financial asset pricing. Eraker (2004) showed the empirical performance of jump diffusion models of stock price dynamics and applied them to options and returns data. Chernov et al. (2003) and Raggi and Bordignon (2006) compares various different specifications of jump diffusions in the SV model in their empirical studies. We also refer to another remarkable jump specification in GARCH model, discussed by Chan and Maheu (2002) and Maheu and McCurdy (2004), which incorporates the autoregressive conditional jump intensity parameterization.

Focusing the estimation method, Kim et al. (1998) develops a fast and reliable Markov chain Monte Carlo (MCMC) algorithm for the SV model. Their impressive method, called *mixture sampler*, has been widely used in the SV literatures and extended in various ways. In the context of the extension of their method, Chib et al. (2002) estimates the SV model with jumps and Student-t errors (SVJt) (but without leverage effect). The leverage effect refers to the increase in volatility following a previous drop in stock returns, and modelled by the negative correlation coefficient between error terms of stock returns and the volatility (e.g. Black (1976), Nelson (1991), Yu (2005), Omori et al. (2007)). The SV model with leverage effect (SVL) is also called the asymmetric stochastic volatility model. Omori et al. (2007) constructs the efficient MCMC estimation method for the SV model with leverage effect and Student-t errors (SVLt) (but without jumps) and demonstrates some empirical results.

In the line of developing the mixture sampler of Kim et al. (1998) for more suitable models to detect the complicated empirical structure in financial market, this paper discusses the SV models with leverage, jump components and heavy-tailed errors (SVLJt) jointly.

We consider the SV model given by

$$y_t = k_t\gamma_t + \sqrt{\lambda_t}\varepsilon_t \exp(h_t/2), \quad t = 1, \ldots, n, \tag{1}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 1, \ldots, n-1, \tag{2}$$

where y_t is a response, h_t is an unobserved log-volatility, $|\phi| < 1$, $h_1 \sim N(0, \sigma^2/(1-\phi^2))$,

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N(0, \Sigma), \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}.$$

The leverage effect measured by the correlation coefficient ρ is expected to be negative as reported in several empirical studies (Yu (2005), Omori et al. (2007)). The correlation coefficient $\rho = 0$ implies the SV model without leverage effect. The $k_t\gamma_t$ represents a jump component in the measurement equation (1). The γ_t is a jump flag defined as a Bernoulli random variable such that

$$\pi(\gamma_t = 1) = \kappa, \quad \pi(\gamma_t = 0) = 1 - \kappa, \quad 0 < \kappa < 1,$$

and the k_t is a jump size specified by

$$\psi_t \equiv \log(1 + k_t) \sim N(-0.5\delta^2, \delta^2), \tag{3}$$

following Andersen et al. (2002), Chib et al. (2002) where the jump parameter κ and δ are unknown and to be estimated. We denote the SV and SVL models with jumps as the SVJ and SVLJ models respectively.

The measurement error $\sqrt{\lambda_t}\varepsilon_t$ is assumed to follow the heavy-tailed Student-t distribution with unknown degrees of freedom ν by letting

$$\lambda_t^{-1} \sim \text{Gamma}(\nu/2, \nu/2). \tag{4}$$

We may also assume $\log \lambda_t \sim N(-0.5\tau^2, \tau^2)$ to obtain the lognormal scale mixture as in Omori et al. (2007), but we illustrate our algorithm using Gamma scale mixture given by (4). When $\lambda_t \equiv 1$ for all t, the model reduces to the SV or SVL model with normal errors.

The contribution of this paper comprises two parts. First, we develop the efficient and fast MCMC parameter estimation method for the SVLJt model (SV model with leverage, jumps and Student-t errors) extending Chib et al. (2002) and Omori et al. (2007). Second, we extend it to the SV model with correlated jumps, which have recently been popular in financial literatures.

We illustrate our approach using simulated data and apply it to the stock returns data of S&P500 index and TOPIX (Tokyo Stock Price index). Using Bayesian approach of marginal likelihood computation, we compare various candidate models over the class of SV model with jumps, leverage and heavy-tails. The superposition model is also considered.

The rest of paper is organized as follows. In Section 2 we discuss the MCMC estimation for our SV model with jumps, leverage and heavy-tails. Section 3 illustrates our method using simulated data. In Section 4, we extend it to the SV model with correlated jumps. In Section 5, we apply our proposed method to the daily asset returns data of S&P500 and TOPIX. Section 6 concludes the paper.

2 SV model with jumps, leverage and heavy-tails

The well-known difficulty of estimating the discrete-time SV model is that the likelihood function is not easily available. It is possible to compute the likelihood using a simulation-based method for a given set of parameters, which is called a particle filter. But it requires a computational burden since we need to repeat the particle filter many times to evaluate the likelihood function for each set of parameters until we reach the maximum. To overcome this difficulty, we take Bayesian estimation approach and propose the MCMC methods (e.g., Chib and Greenberg (1996)) for a precise and efficient estimation of the SVLJt model.

2.1 Auxiliary mixture sampler

Following Omori et al. (2007), we define $y_t^* = \log(y_t - k_t\gamma_t)^2 - \log\lambda_t$, $d_t = \text{sign}(y_t - k_t\gamma_t) = I(\varepsilon_t > 0) - I(\varepsilon_t \leq 0)$, which rewrites equation (1) as

$$y_t^* = h_t + \xi_t, \tag{5}$$

where $\xi_t = \log\varepsilon_t^2$. Omori et al. (2007) proposes to approximate the bivariate conditional density of $(\xi_t, \eta_t)|d_t$ by a K-components mixture of bivariate Gaussian densities, which is an exhaustive extension Kim et al. (1998) approach for the SV model with leverage effect. The key essence of their approach is that the model (5) and (2) is approximated to a linear Gaussian state space model conditioned on the mixture component indicator $s_t \in \{1, 2, \ldots, K\}$. Given $s = \{s_1, \ldots, s_n\}$, this permits us to sample the latent variable $h = \{h_1, \ldots, h_n\}$ in one block from its joint distribution using the simulation smoother for a linear Gaussian state space model (de Jong and Shephard (1995), Durbin and Koopman (2002a)). We estimate the mixture approximation model

$$\begin{pmatrix} y_t^* \\ h_{t+1} \end{pmatrix} = \begin{pmatrix} h_t \\ \mu + \phi(h_t - \mu) \end{pmatrix} + \begin{pmatrix} \xi_t \\ \eta_t \end{pmatrix}, \tag{6}$$

where

$$\begin{aligned} \left\{\begin{pmatrix} \xi_t \\ \eta_t \end{pmatrix} |d_t, (s_t = i)\right\} &\overset{L}{=} \begin{pmatrix} m_i + v_i z_{1t} \\ d_t\rho\sigma(a_i + b_i v_i z_{1t})\exp(m_i/2) + \sigma\sqrt{1-\rho^2}z_{2t} \end{pmatrix}, \\ \begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} &\sim N(0, I_2), \end{aligned}$$

for $i = 1, 2, \ldots, K$. Omori et al. (2007) proposes the approximation based on $K = 10$ and lists the selection of $p_i \equiv \Pr(s_t = i)$ and the mixture component parameters (m_i, v_i, a_i, b_i) for $i = 1, \ldots, 10$, which we reproduced in Table 1. Note that (m_i, v_i, a_i, b_i) do not depend on model parameters $\theta \equiv (\phi, \sigma, \rho)$ and μ.

i	p_i	m_i	v_i^2	a_i	b_i
1	0.00609	1.92677	0.11265	1.01418	0.50710
2	0.04775	1.34744	0.17788	1.02248	0.51124
3	0.13057	0.73504	0.26768	1.03403	0.51701
4	0.20674	0.02266	0.40611	1.05207	0.52604
5	0.22715	−0.85173	0.62699	1.08153	0.54076
6	0.18842	−1.97278	0.98583	1.13114	0.56557
7	0.12047	−3.46788	1.57469	1.21754	0.60877
8	0.05591	−5.55246	2.54498	1.37454	0.68728
9	0.01575	−8.68384	4.16591	1.68327	0.84163
10	0.00115	−14.65000	7.33342	2.50097	1.25049

Table 1: Selection of $(p_i, m_i, v_i^2, a_i, b_i)$.

2.2 MCMC algorithm

Let $y = \{y_t\}_{t=1}^n,\ y^* = \{y_t^*\}_{t=1}^n,\ d = \{d_t\}_{t=1}^n,\ k = \{k_t\}_{t=1}^n,\ \gamma = \{\gamma_t\}_{t=1}^n,\ \lambda = \{\lambda_t\}_{t=1}^n$ and we set the prior probability density $\pi(\theta), \pi(\mu), \pi(\kappa), \pi(\delta), \pi(\nu)$ for $\theta,\ \mu,\ \kappa,\ \delta,\ \nu$. Then, we draw sample from the posterior distribution

$$\pi(\theta, \mu, \kappa, \delta, \nu, s, h, k, \gamma, \lambda | y)$$

by the MCMC technique. Let us reparameterize k_t by $\psi_t \equiv \log(1 + k_t)$ and denote $\psi = \{\psi_t\}_{t=1}^n,\ \psi^{(0)} = \{\psi_t | t = 1, \ldots, n,\ \text{s.t. } \gamma_t = 0\},\ \psi^{(1)} = \{\psi_t | t = 1, \ldots, n,\ \text{s.t. } \gamma_t = 1\}$. We propose the following sampling algorithm:

1. Initialize $\theta, \mu, \kappa, \delta, \nu, s, h, \psi, \gamma$ and λ.
2. Sample $(\theta, \mu, h)|s, y^*, d$ by
   (a) Sampling $\theta|s, y^*, d$,
   (b) Sampling $(\mu, h)|\theta, s, y^*, d$.
3. Sample $\psi^{(1)}|\theta, \mu, \delta, h, \gamma, \lambda, y$.
4. Sample $(\delta, \psi^{(0)})|\psi^{(1)}, \gamma$ by
   (a) Sampling $\delta|\psi^{(1)}, \gamma$,
   (b) Sampling $\psi^{(0)}|\delta, \gamma$.
5. Sample $(\gamma, s)|\theta, \mu, \kappa, h, \psi, \lambda, y$ by
   (a) Sampling $\gamma|\theta, \mu, \kappa, h, \psi, \lambda, y$,
   (b) Sampling $s|\theta, \mu, h, y^*, d$.
6. Sample $\kappa|\gamma$.

7. Sample $(\lambda, \nu)|\theta, \mu, s, h, \psi, \gamma, y$ by

   (a) Sampling $\lambda|\theta, \mu, \nu, s, h, \psi, \gamma, y$,

   (b) Sampling $\nu|\lambda$.

8. Go to 2.

Although we draw samples from the posterior distribution for the approximated model (6), they can be reweighted to obtain the moments of the exact posterior distribution for the original SV model (1) and (2) as we will show. In the following subsections, we give a brief description of each sampling step (see Appendix A for the details).

2.2.1 Sampling volatility parameters (θ, μ, h)

The conditional posterior probability density function of (θ, μ, h) is

$$\pi(\theta, \mu, h|s, y^*, d) \quad \propto \quad \pi(\theta|s, y^*, d) \times \pi(\mu, h|\theta, s, y^*, d),$$

where

$$\begin{aligned} \pi(\theta|s, y^*, d) &\propto f(y^*|\theta, s, d)\pi(\theta), \\ \pi(\mu, h|\theta, s, y^*, d) &\propto \pi(\mu|\theta, s, y^*, d)\pi(h|\mu, \theta, s, y^*, d), \end{aligned}$$

and f is the conditional likelihood of the approximated model. Note that the conditional posterior probability density $\pi(\theta|s, y^*, d)$ is marginalized over μ. Integrating μ from the joint posterior density $\pi(\theta, \mu|s, y^*, d)$ provides a good acceleration of the convergence in our procedure. We can evaluate the conditional likelihood $f(y^*|\theta, s, d)$ through the augmented Kalman filter (see Appendix B and de Jong (1991), Durbin and Koopman (2002b)). In Step 2a, we find $\hat{\theta} = (\hat{\phi}, \hat{\sigma}, \hat{\rho})$ which maximizes (or approximately maximizes) the posterior probability density $\pi(\theta|s, y^*, d)$, and generate a candidate θ^* from a normal distribution $N(\theta_*, \Sigma_*)$ truncated over the region $R = \{\theta : |\phi| < 1, \sigma > 0, |\rho| < 1\}$, where

$$\theta_* = \hat{\theta} + \Sigma_* \left.\frac{\partial \log \pi(\theta|s, y^*, d)}{\partial \theta}\right|_{\theta=\hat{\theta}}, \quad \Sigma_*^{-1} = -\left.\frac{\partial \log \pi(\theta|s, y^*, d)}{\partial \theta \partial \theta'}\right|_{\theta=\hat{\theta}}. \tag{7}$$

We use the Metropolis-Hastings (M-H) algorithm (see e.g. Chib and Greenberg (1995)) with this proposal density to accept or reject θ^*. In Step 2b, we straightforwardly sample μ from a normal distribution using the by-products of the augmented Kalman filter, and sample h by the simulation smoother (de Jong and Shephard (1995), Durbin and Koopman (2002a)) given (μ, θ).

2.2.2 Sampling jump parameters (δ, ψ)

In Step 3, we sample from the conditional posterior probability density $\pi(\psi^{(1)}|\delta, \omega_1, \gamma, y)$ where we marginalized the posterior probability density over s to accelerate the convergence, and $\omega_1 = (\theta, \mu, h, \lambda)$. Note that the density does not depend on $\psi^{(0)}$, since $\psi^{(0)}$ and $\psi^{(1)}$ are conditionally independent given $(\omega_1, \delta, \gamma, y)$. To sample $\psi^{(1)}$, we use the M-H algorithm. The reparameterization $\psi_t \equiv \log(1 + k_t)$ yields a useful proposal density. Since we use returns (y_t's) measured in decimals in empirical study, the k_t's are expected to be small. When k_t is small, $k_t = e^{\psi_t} - 1$ may be well approximated by ψ_t. We have exactly $\psi_t \sim N(-0.5\delta^2, \delta^2)$, and approximately

$$y_t \approx \psi_t \gamma_t + \sqrt{\lambda_t} \varepsilon_t \exp(h_t/2). \tag{8}$$

Then a candidate for $\psi_t^{(1)}$ can be drawn from the normal density $N(\hat{\psi}_t, \sigma^2_{\psi_t})$ where

$$\begin{aligned} \hat{\psi}_t &= \sigma^2_{\psi_t} \left(-0.5 + \frac{\gamma_t (y_t - \sqrt{\lambda_t} \rho \exp(h_t/2) \{(h_{t+1} - \mu) - \phi(h_t - \mu)\}/\sigma)}{\lambda_t (1 - \rho^2) e^{h_t}} \right), \qquad (9) \\ \sigma^2_{\psi_t} &= \left(\delta^{-2} + \frac{\gamma_t^2}{\lambda_t (1 - \rho^2) e^{h_t}} \right)^{-1}. \end{aligned}$$

In Step 4, we sample $(\delta, \psi^{(0)})$ in one block conditional on $\psi^{(1)}$. Noting that ψ_t vanishes in the measurement equation (1) if $\gamma_t = 0$, we can express

$$\pi(\delta, \psi^{(0)}|\psi^{(1)}, \omega_1, \gamma, y) \quad \propto \quad \pi(\delta|\psi^{(1)}, \gamma) \times \pi(\psi^{(0)}|\delta, \gamma).$$

In Step 4a, we draw a sample from the posterior distribution

$$\pi(\delta|\psi^{(1)}, \gamma) \propto g(\psi^{(1)}|\delta, \gamma) \pi(\delta),$$

using the Acceptance-Rejection M-H (A-R M-H) algorithm (see e.g. Tierney (1994), Chib and Greenberg (1995)). In Step 4b, $\psi_t^{(0)}|\delta, \gamma$ is directly drawn from the normal distribution,

$$\psi_t^{(0)}|\delta, \gamma \sim N(-0.5\delta^2, \delta^2).$$

Instead of Steps 3 and 4, we may first sample δ given ψ and then sample ψ given δ. However, when sampling in such an order, we found the conditional distribution of δ to produce MCMC samples with larger autocorrelations.

The key feature is that the posterior distribution for $\psi_t^{(1)}$ is marginalized over s_t. We found that this marginalization works to increase the acceptance rate of $\psi_t^{(1)}$ in the M-H step and to decrease the inefficiency in sampling δ. We find that sampling $\psi_t^{(1)}$ is sensitive to the conditioned s_t. When the state s_t whose probability (p_{s_t}) is very small is conditioned, the conditional posterior density for $\psi_t^{(1)}$ is irregularly changed and the draw of $\psi_t^{(1)}$ is overwhelmingly affected.

To remove this effect of the s_t, we provide the marginalization of the conditional posterior distribution for $\psi_t^{(1)}$.

2.2.3 Sampling mixture state and heavy-tailed parameters $(\gamma, s, \kappa, \lambda, \nu)$

Since the γ_t and s_t are conditionally independent for $t = 1, \ldots, n$, we can obtain independent samples from their conditional posterior distribution. The posterior distribution of γ which we draw from is also marginalized over s, similar to δ and ψ in Steps 3 and 4. Step 5a requires only to evaluate Bernoulli distribution $\pi(\gamma_t|\theta, \mu, h, \kappa, \psi, \lambda, y)$ where $\gamma_t = 0, 1$. In Step 5b we evaluate a K-point discrete distribution $\pi(s_t = i|\theta, \mu, h, y^*, d)$ for $i = 1, \ldots, K$. If we use a beta prior for κ,

$$\kappa \sim \text{Beta}(n_{\kappa 1}, n_{\kappa 0}),$$

then we draw a sample $\kappa|q$ from $\text{Beta}(n_{\kappa 1} + n_1, n_{\kappa 0} + n_0)$, where n_0 and n_1 are the numbers of time such that $\gamma_t = 0$ and $\gamma_t = 1$ respectively.

Finally, we sample from the conditional posterior distribution of (λ, ν) where the joint probability density is

$$\pi(\lambda, \nu|\theta, \mu, s, h, \psi, \gamma, y) \propto f(y|\theta, \mu, s, h, \psi, \gamma) g(\lambda|\nu)\pi(\nu).$$

In Step 6a, we sample λ_t by M-H algorithm using a proposal distribution $\lambda_t^{-1} \sim \text{Gamma}(\nu/2, \nu/2)$ independently for $t = 1, \ldots, n$. Step 6b requires the A-R M-H algorithm for ν used in sampling δ.

2.2.4 Reweighting to correct for mixture-approximation error

The normal mixture provides a good approximation as shown in Omori et al. (2007), though we can correct a minor error of approximation and obtain samples from the exact posterior distribution as follows. Let $\vartheta = (\theta, \mu, \kappa, \delta, \nu, h, k, \gamma, \lambda)$ and ϑ^j denote the j-th sample. To obtain sample from the posterior distribution for the original SV model, denoted by $\tilde{\pi}(\vartheta|y)$, we resample the j-th sample drawn from the approximated posterior density $\pi(\vartheta|y^*, d) = \sum_s \pi(\vartheta, s|y^*, d)$ with the weights proportional to

$$w_j = \frac{w_j^*}{\sum_{i=1}^M w_i^*}, \quad w_j^* = \frac{\tilde{\pi}(\vartheta^j|y)}{\pi(\vartheta^j|y^*, d)} = \frac{\tilde{f}(y|\vartheta^j)}{f(y^*|\vartheta^j, d)},$$

for $j = 1, \ldots, M$ where $\tilde{f}$ is a likelihood for the original SV model, f is a likelihood marginalized over s for the approximate mixture model and M is the sample size. To estimate the posterior mean of a function of the parameter $g(\vartheta)$,

$$E\{g(\vartheta)|y\} = \int g(\vartheta)\tilde{\pi}(\vartheta|y)d\vartheta = \int g(\vartheta)\frac{\tilde{\pi}(\vartheta|y)}{\pi(\vartheta|y^*, d)}\pi(\vartheta|y^*, d)d\vartheta,$$

we obtain the reweighted estimate as

$$\hat{E}\{g(\vartheta)|y^*, d\} = \sum_{j=1}^{M} g(\vartheta^j) w_j.$$

3 Illustrative example

This section illustrates our estimation procedure using the simulated data. We generated 3,000 observations from the SVLJt model given by equations (1) and (2) with $\phi = 0.97$, $\sigma = 0.1$, $\rho = -0.3$, $\exp(\mu/2) = 0.01$, $\kappa = 0.005$, $\delta = 0.04$ and $\nu = 15$. These values are based on the empirical estimates for daily returns data (e.g. Chib et al. (2002), Omori et al. (2007)). As suggested by Kim et al. (1998) and Omori et al. (2007), we take $y_t^* = \log((y_t - k_t\gamma_t)^2 + c)$ where c is the offset for the case where $(y_t - k_t\gamma_t)^2$ is too small. We set $c = 10^{-7}$ in this paper throughout. The following prior distributions are assumed:

$$\begin{aligned}
\frac{\phi+1}{2} &\sim \text{Beta}(20, 1.5), \quad \sigma^{-2} \sim \text{Gamma}(2.5, 0.025), \\
\rho &\sim U(-1, 1), \qquad\quad \mu \sim N(-10, 1), \\
\kappa &\sim \text{Beta}(2, 100), \quad \log(\delta) \sim N(-2.5, 0.15), \quad \nu \sim \text{Gamma}(16, 0.8).
\end{aligned}$$

The beta prior distributions for $(\phi+1)/2$ and κ imply means and standard deviations are (0.86, 0.11) for ϕ and (0.02, 0.01) for κ. The gamma priors for (σ^{-2}, ν) and lognormal prior for δ have means and standard deviations (100,63.2) for σ^{-2}, (0.09, 0.04) for δ and (20, 5) for ν. These prior distributions reflect the values obtained in the past literature to a certain extent.

We draw $M = 5,000$ sample after the initial 10,000 samples are discarded. The number of discarding samples is selected using time series plots of the marginal averages of samples for each parameter. Figure 1 shows the sample autocorrelation function, the sample paths and the posterior densities for each parameters. After discarding samples in burn-in period, the sample paths look stable and the sample autocorrelations drop very quickly, indicating our sampling method efficiently produces uncorrelated samples.

Table 2 gives the estimates for posterior means, standard deviations and the 95% credible intervals. All estimated posterior means are close to the true values and the true values are contained in their corresponding 95% credible intervals. The inefficiency factors are also reported to check the performance of our sampling efficiency. The inefficiency factor is defined as $1 + 2\sum_{s=1}^{\infty}\rho_s$ where ρ_s is the sample autocorrelation function at lag s. It is the ratio of variance of the posterior mean from the correlated draws to the one from the hypothetical uncorrelated sample, which measures the loss of sampling efficiency in our correlated MCMC draws (see e.g. Chib (2001)). Similarly to the result of Kim et al. (1998), Chib et al. (2002) and Omori et al. (2007), the inefficiency factors in Table 2 take very low values except ν, compared with those of conventional MCMC samplers used in the estimation of SV models. This suggests that we are successful in extending their method to the SVLJt model without loss of sampling efficiency.

Thus the proposed algorithm in which we marginalized conditional posterior densities provides the reliable and efficient MCMC sampling.

The relatively high inefficiency factor of ν is assumed to be caused from the A-R M-H algorithms, which provides a high acceptance rate in the M-H step but produces the highly correlated sample as seen in Figure 1. Further the conditional posterior distribution for ν depends only the latent variables λ, which makes the inefficiency factor higher. On the other hand, we also apply the A-R M-H algorithm for δ, but the inefficiency factor for δ in not so high. This is the successful result of the marginalization and the order of sampling as discussed in Section 2.

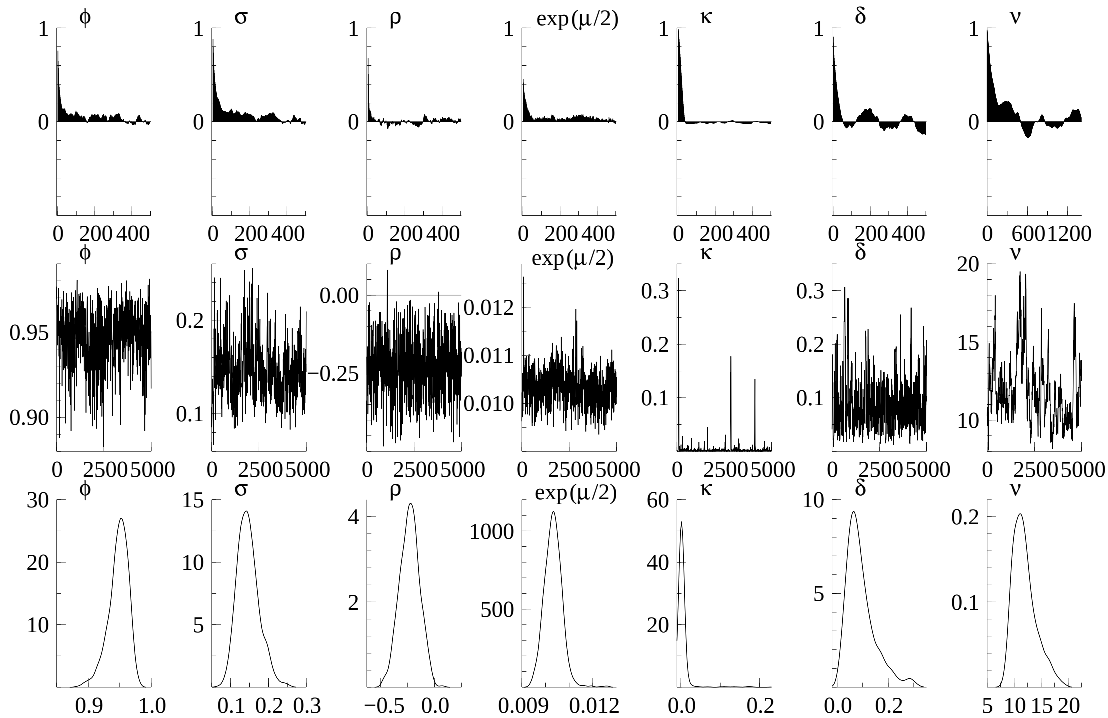

Figure 1: Estimation result of simulation data (SVLJt model). Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

Parameter	True	Mean	Stdev.	95% interval	Inefficiency
ϕ	0.97	0.9477	0.0160	[0.9111, 0.9735]	38.33
σ	0.1	0.1463	0.0292	[0.0969, 0.2114]	56.24
ρ	-0.3	-0.2269	0.0914	[-0.4045, -0.0473]	6.63
$\exp(\mu/2)$	0.01	0.0103	0.0004	[0.0096, 0.0111]	25.35
κ	0.005	0.0054	0.0259	[0.0002, 0.0268]	32.58
δ	0.04	0.0937	0.0564	[0.0207, 0.2425]	37.28
ν	15	11.939	2.1361	[8.9594, 17.230]	159.44

Table 2: Estimation result of simulation data (SVLJt model).

4 SV model with correlated jumps

A correlated-jump SV model has been started to receive widespread attention in the recent literature (e.g. Eraker et al. (2003), Kobayashi (2006)) to investigate the simultaneous jumps in the observations and volatilities. Based on the SV model with leverage and jumps discussed in the previous subsections, we extend it to the SV model with correlated jumps (SVLCJ) formulated as

$$y_t = k_t\gamma_t + \varepsilon_t \exp(h_t/2), \quad t = 1, \ldots, n, \tag{10}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + j_t\gamma_t + \eta_t, \quad t = 0, \ldots, n-1. \tag{11}$$

To model jumps that occur concurrently both in return and in volatility with probability $\pi(\gamma_t = 1) = \kappa$, the joint distribution of jump sizes is assumed to be

$$\begin{aligned} j_t &\sim \text{Exp}(\mu_J), \\ k_t|j_t &\sim N(\mu_k + \beta_J j_t,\, \sigma_k^2), \end{aligned}$$

where Exp denotes the exponential distribution, Gamma$(1, \mu_J)$. The correlation between jump sizes in return and in volatility is considered through the parameter β_J. Additionally to the specification for the SVLJ model, we assume the prior $\pi(\mu_J)$, $\pi(\beta_J)$, $\pi(\mu_k)$, $\pi(\sigma_k)$. We can explore the posterior distribution

$$\pi(\theta, \mu, \kappa, \mu_J, \beta_J, \mu_k, \sigma_k, s, h, k, j, \gamma|y),$$

where $k = \{k_t\}_{t=1}^n$, $j = \{j_t\}_{t=1}^n$, by drawing samples by the corresponding MCMC procedure as below, letting $k^{(1)} = \{k_t|t = 1, \ldots, n, \text{ s.t. } \gamma_t = 1\}$, $k^{(0)} = \{k_t|t = 1, \ldots, n, \text{ s.t. } \gamma_t = 0\}$, $j^{(1)} = \{j_t|t = 1, \ldots, n, \text{ s.t. } \gamma_t = 1\}$ and $j^{(0)} = \{j_t|t = 1, \ldots, n, \text{ s.t. } \gamma_t = 0\}$;

1. Initialize $\theta, \mu, \kappa, \mu_J, \beta_J, \mu_k, \sigma_k, s, h, k, j$ and γ.

2. Sample $(\theta, \mu, h)|s, j, y^*, d$.

3. Sample $(k^{(1)}, j^{(1)})|\theta, \mu, \mu_J, \beta_J, \mu_k, \sigma_k, h, \gamma, y$.

4. Sample $(\mu_J, \beta_J, \mu_k, \sigma_k, k^{(0)}, j^{(0)})|k^{(1)}, j^{(1)}, \gamma$ by

   (a) Sampling $\mu_J|j^{(1)}, \gamma$,

   (b) Sampling $(\beta_J, \mu_k)|k^{(1)}, j^{(1)}, \sigma_k, \gamma$,

   (c) Sampling $\sigma_k|k^{(1)}, j^{(1)}, \beta_J, \mu_k, \gamma$,

   (d) Sampling $(k^{(0)}, j^{(0)})|\mu_J, \beta_J, \mu_k, \sigma_k, \gamma$.

5. Sample $(\gamma, s)|\theta, \mu, \kappa, h, k, j, y$ by

(a) Sampling $\gamma|\theta, \mu, \kappa, h, k, j, y$,

(b) Sampling $s|\theta, \mu, h, j, y^*, d$.

6. Sample $\kappa|\gamma$.

7. Go to 2.

Throughout, this algorithm is a straightforward extension from the one of the SVLJt model. We remark several points in details. In Step 3, we sample (k_t, j_t) for time t such that $\gamma_t = 1$ from its conditional density by the M-H algorithm with the proposal distribution based on the original equation (10) and (11), which produces the candidate $(k_t^*, j_t^*)' \sim N(\mu_{kj_t}, \Sigma_{kj_t})_{[j_t^*>0]}$ where

$$\mu_{kj_t} = \begin{pmatrix} y_t \\ (h_{t+1} - \mu) - \phi(h_t - \mu) \end{pmatrix}, \quad \Sigma_{kj_t} = \begin{pmatrix} \exp(h_t) & \rho \exp(h_t/2)\sigma \\ \rho \exp(h_t/2)\sigma & \sigma^2 \end{pmatrix}.$$

In Step 4a, we sample μ_J by M-H algorithm with the proposal point drawn as $\mu_J^* \sim \text{Exp}(\frac{1}{n}\sum_{\gamma_t=1} j_t)$. In Step 4b, if $\mu_k \sim N(\mu_{k_0}, \sigma_{k_0}^2)$, $\beta_J \sim N(\mu_{\beta_0}, \sigma_{\beta_0}^2)$, the conditional posterior density for (μ_k, β_J) is the normal distribution. Thus we can sample directly from the posterior distribution. In Step 4c, if $\sigma_k^{-2} \sim \text{Gamma}(v_{k_0}, S_{k_0})$, the conditional posterior density for σ_k^2 is the inverse-Gamma distribution, which also enables us to have a direct draw.

5 Application to stock returns data

5.1 Data

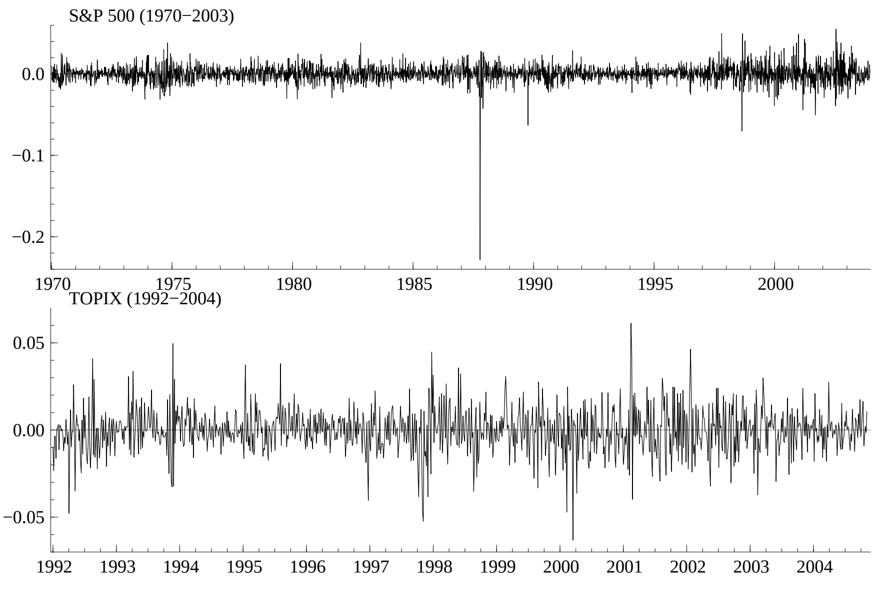

Figure 2: Return data for S&P500 and TOPIX.

We apply our MCMC estimation method to daily stock returns data, the S&P500 index from January 1, 1970 to December 31, 2003, and the TOPIX (Tokyo stock price index) from January 6, 1992 to December 30, 2004. The log-difference returns are computed as $y_t = \log P_t - \log P_{t-1}$ where P_t is the closing price on day t. The sample size is 8,869 for S&P500 and 3,203 for TOPIX. Table 3 summarizes the descriptive statistics of the returns data including the count of positive and negative returns, and Figure 2 plots the series of returns for S&P500 and TOPIX. The largest negative impact on S&P500 corresponds to the crash of October 1987.

S&P500 (1970/1/1 - 2003/12/31)

Obs.	Mean	Stdev.	Max.	Min.	pos.(+)	neg.(-)
8,869	0.0003	0.010	0.087	-0.227	4,762	4,107

TOPIX (1992/1/6 - 2004/12/30)

Obs.	Mean	Stdev.	Max.	Min.	pos.(+)	neg.(-)
3,203	-0.0001	0.013	0.073	-0.066	1,563	1,640

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

5.2 Estimation results

5.2.1 SV models with jumps, leverage effects and heavy-tails

We first consider the following four candidate SV models with leverage effect to be fitted to the data:

(i) Model SVL: the SV model with leverage effect and no jump. The error terms in the measurement equation (1) is assumed to follow normal distribution ($\lambda_t \equiv 1$ for all t).

(ii) Model SVLt: the SV model with leverage effect and no jump. The error terms in the measurement equation (1) is assumed to Student-t distribution with unknown degrees of freedom.

(iii) Model SVLJ: the SV model with leverage effect and jumps. The error terms in the measurement equation (1) is assumed to follow normal distribution.

(iv) Model SVLJt: the SV model with leverage effect and jumps. The error terms in the measurement equation (1) is assumed to Student-t distribution with unknown degrees of freedom.

The prior specifications are same as the simulation study in the previous section. The number of MCMC iterations is 5,000 and the initial 10,000 samples are discarded. Table 4 and 5 reports the estimation result; the posterior means, standard deviations, 95% intervals and the inefficiency factors for TOPIX data and S&P500 data respectively. Figure 3 plots the sampling result for the Model SVLJt on the S&P500 series.

The estimates of the volatility parameters $(\phi, \sigma, \rho, \exp(\mu/2))$ are consistent with the results of the previous literatures (e.g. Chib et al. (2002), Omori et al. (2007)). The posterior mean of ϕ is close to one, which indicates the well-known high persistence of volatility on asset returns. The estimates of σ for the Models SVLt and SVLJt are slightly lower than those for the Models SVL and SVLJ. The models allowing the heavy-tail errors seem to explain the excess returns as a realization of the disturbance ε_t, which decreases the variance of the volatility process. The parameter ρ is estimated significantly negative, implying that there exists the leverage effect in daily stock returns.

Parameter	**SVL**	**SVLt**	**SVLJ**	**SVLJt**
ϕ	0.9812 (0.0029)	0.9878 (0.0020)	0.9819 (0.0028)	0.9875 (0.0022)
	[0.9752, 0.9866]	[0.9837, 0.9916]	[0.9763, 0.9870]	[0.9828, 0.9913]
	7.07	18.51	7.37	18.98
σ	0.1452 (0.0102)	0.1111 (0.0083)	0.1423 (0.0103)	0.1132 (0.0084)
	[0.1263, 0.1667]	[0.0950, 0.1278]	[0.1227, 0.1635]	[0.0978, 0.1304]
	12.51	32.49	9.85	33.88
ρ	-0.4970 (0.0379)	-0.5718 (0.0402)	-0.5019 (0.0382)	-0.5641 (0.0418)
	[-0.5682, -0.4186]	[-0.6489, -0.4916]	[-0.5757, -0.4275]	[-0.6441, -0.4798]
	3.34	7.83	6.37	12.66
$\exp(\mu/2)$	0.0089 (0.0003)	0.0084 (0.0004)	0.0089 (0.0003)	0.0084 (0.0004)
	[0.0082, 0.0096]	[0.0077, 0.0092]	[0.0083, 0.0096]	[0.0077, 0.0092]
	1.31	3.44	1.29	8.29
κ			0.0013 (0.0049)	0.0008 (0.0027)
			[0.0001, 0.0062]	[0.0000, 0.0037]
			14.44	11.53
δ			0.1076 (0.0695)	0.0971 (0.0619)
			[0.0252, 0.3184]	[0.0223, 0.2988]
			74.86	62.94
ν		11.180 (1.3056)		11.812 (1.3333)
		[9.4423, 13.339]		[9.2692, 14.376]
		164.93		191.20

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 4: Estimation result for S&P500 return.

In the Models SVLJ and SVLJt, the posterior means of the jump probabilities, κ's, are very low, around 0.1% for the S&P500 series and around 0.3% for the TOPIX series. When the jump probability κ is very small, most of the jump sizes ψ_t's vanish from the likelihood. The posterior densities of the jump intensity parameters, δ's, are widely spread, suggesting that we would fail to extract enough information of jump intensities from rare jump events.

The magnitude of tail-fatness is measured by the parameter ν in the Models SVLt and SVLJt. The posterior means of ν's are around 10 for the S&P500 returns and 20 for the TOPIX returns. This indicates that measurement errors of stock returns have heavy-tailed distributions as pointed out in the past literature.

Parameter	**SVL**	**SVLt**	**SVLJ**	**SVLJt**
ϕ	0.9675 (0.0073)	0.9747 (0.0064)	0.9678 (0.0073)	0.9749 (0.0063)
	[0.9518, 0.9800]	[0.9605, 0.9859]	[0.9522, 0.9811]	[0.9611, 0.9857]
	6.25	19.69	7.12	14.15
σ	0.1908 (0.0184)	0.1623 (0.0182)	0.1899 (0.0185)	0.1621 (0.0177)
	[0.1634, 0.2190]	[0.1291, 0.1999]	[0.1564, 0.2281]	[0.1298, 0.1994]
	15.13	35.66	8.98	27.66
ρ	-0.4464 (0.0577)	-0.4832 (0.0626)	-0.4499 (0.0576)	-0.4842 (0.0619)
	[-0.5322, -0.3290]	[-0.5994, -0.3539]	[-0.5324, -0.3382]	[-0.6002, -0.3572]
	6.21	11.70	14.12	7.98
$\exp(\mu/2)$	0.0106 (0.0005)	0.0100 (0.0006)	0.0106 (0.0005)	0.0101 (0.0006)
	[0.0096, 0.0118]	[0.0089, 0.0112]	[0.0096, 0.0111]	[0.0089, 0.0112]
	1.46	11.20	2.11	12.33
κ			0.0031 (0.0091)	0.0029 (0.0125)
			[0.0002, 0.0188]	[0.0001, 0.0161]
			29.02	21.57
δ			0.0921 (0.0538)	0.1019 (0.0659)
			[0.0252, 0.2499]	[0.0248, 0.2667]
			24.33	89.12
ν		18.352 (4.4163)		18.501 (4.2324)
		[11.830, 27.864]		[11.940, 28.223]
		152.76		220.80

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 5: Estimation result for TOPIX return.

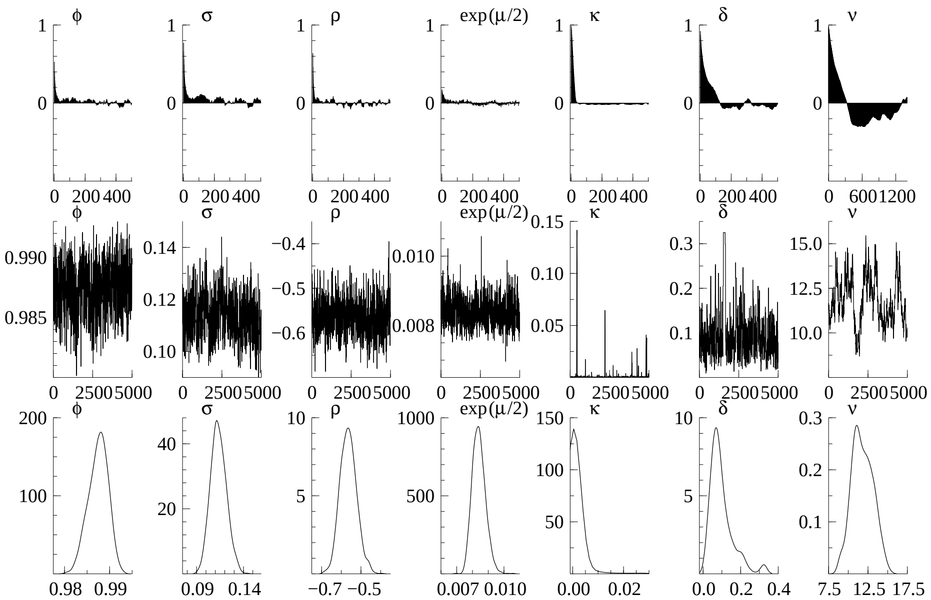

Figure 3: Estimation result for S&P500 returns (SVLJt model). Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

5.2.2 SV model with correlated jumps

For the SV model with correlated jumps (SVLCJ), we additionally specify the prior as

$$\mu_J \sim \text{Exp}(0.2), \quad \beta_J \sim N(0,1),$$
$$\mu_k \sim N(0,1), \quad \sigma_k^{-2} \sim \text{Gamma}(2.5, 0.025).$$

The exponential prior distribution for μ_J implies a mean and a standard deviation is (5, 5). The gamma prior for σ_k^{-2} has a mean and a standard deviation (100, 63.2). Applying the MCMC algorithm to the S&P500 daily return data (from 1970 to 2003), we draw $M = 5,000$ samples after the initial 10,000 samples are discarded. The posterior estimates are reported in Figure 6. Since inefficiency factors are low, all parameters are sampled efficiently. One remark should go to the estimate of β_J. Although the posterior mean of β_J is negative, its 95% confidence interval contains zero. We estimated the SVLCJ models using other 5 series of S&P500 and TOPIX return data in Section 5.3 and found all the estimates of β_J are negative but their 95% confidence interval contain zero, too. This result indicates that the two jump sizes in return and in volatility may be not so strongly correlated.

Parameter	Mean	Stdev.	95% interval	Inefficiency
ϕ	0.9835	0.0025	[0.9783, 0.9882]	5.58
σ	0.1298	0.0093	[0.1122, 0.1485]	8.82
ρ	-0.5243	0.0378	[-0.5960, -0.4462]	6.48
$\exp(\mu/2)$	0.0088	0.0003	[0.0082, 0.0095]	5.40
κ	0.0006	0.0003	[0.0002, 0.0013]	36.42
μ_J	1.3152	0.9171	[0.3338, 3.5807]	12.10
β_J	-0.0978	0.4846	[-1.0685, 0.8861]	1.27
μ_k	-0.0061	0.5235	[-1.0877, 1.0743]	1.08
σ_k	0.4338	0.1501	[0.2474, 0.8297]	9.98

Table 6: Estimation result of the SVLCJ model for S&P500 return.

5.2.3 Superposition model

For model comparisons, we also consider the superposition model which has become popular and discussed as a flexible dynamic volatility model in the SV literatures (e.g. Omori et al. (2007)). It is formulated as

$$\begin{aligned}
y_t &= \varepsilon_t \exp(h_t/2), \quad t = 1, \ldots, n, \\
h_t &= \alpha_{1t} + \alpha_{2t}, \quad t = 1, \ldots, n, \\
\alpha_{1,t+1} &= \mu + \phi_1(\alpha_{1t} - \mu) + \eta_{1t}, \quad t = 0, \ldots, n-1, \\
\alpha_{2,t+1} &= \phi_2 \alpha_{2t} + \eta_{2t}, \quad t = 0, \ldots, n-1,
\end{aligned}$$

where

$$\begin{pmatrix} \varepsilon_t \\ \eta_{1t} \\ \eta_{2t} \end{pmatrix} \sim N(0, \Sigma), \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & \rho_1\sigma_1 & \rho_2\sigma_2 \\ \rho_1\sigma_1 & \sigma_1^2 & 0 \\ \rho_2\sigma_2 & 0 & \sigma_2^2 \end{pmatrix},$$

$|\phi_1| < 1$, $|\phi_2| < 1$, and $\rho_1^2 + \rho_2^2 < 1$. For identifiability, we assume that $\phi_1 > \phi_2$.

Although this model doesn't have the jump and heavy-tail components, the log-volatility consists of two independent autoregressive processes, each with a different persistence level and leverage effect, which is considered to play an effective role to grasp the complicated volatility dynamics in financial time series.

Let $\alpha_t = (\alpha_{1t}, \alpha_{2t})'$ for $t = 0, \ldots, n$, and $\alpha = \{\alpha_t\}_{t=0}^n$. Following Omori et al. (2007), the MCMC implementation for the Bayesian inference of the superposition model is given as follows:

1. Initialize $\theta = (\phi_1, \phi_2, \sigma_1, \sigma_2, \rho_1, \rho_2), \mu, s, \alpha$.
2. Sample $(\theta, \mu, \alpha)|s, y^*, d$ by
   (a) Sampling $\theta|s, y^*, d$,
   (b) Sampling $(\mu, \alpha)|\theta, s, y^*, d$.
3. Sample $s|\theta, \mu, \alpha, y^*, d$.
4. Go to 2.

The algorithm is simple except sampling θ with several constraints. We implement Step 2a with the M-H algorithm. We generate a candidate point from the normal distribution after transforming the parameters $\theta_1 = \log(1+\phi_1) - \log(1-\phi_1)$, $\theta_2 = \log(1+\phi_2) - \log(\phi_1 - \phi_2)$, $\theta_3 = \log \sigma_1^2$, $\theta_4 = \log \sigma_2^2$, $\theta_5 = \log(1+\rho_1) - \log(1-\rho_1)$ and $\theta_6 = \log(\sqrt{1-\rho_1^2} + \rho_2) - \log(\sqrt{1-\rho_1^2} - \rho_2)$. This transformation widens the parameter space to all the domain of R^6 in sampling procedure, which clears the parameter constraints and makes it easier to evaluate the marginal likelihood. Through Step 2, we can construct the linear Gaussian state space form with bivariate state variable α_t and conduct the corresponding augmented Kalman filter. We assume priors

$$(\phi_2 + 1)/2 \sim \text{Beta}(10, 10), \quad (\rho_2 + 1)/2 \sim \text{Beta}(10, 10), \quad \sigma_2^{-2} \sim \text{Gamma}(2.5, 0.025),$$

for additional parameters. The beta prior distribution for $(\phi_2 + 1)/2$ and $(\rho_2 + 1)/2$ has a mean and standard deviation (0, 0.22). The gamma prior for σ_2^{-2} has a mean and standard deviation (100, 62.3).

Table 7 shows the estimation result for the superposition model (SVLSP) using the S&P500 return (from 1970 to 2003). It is remarkable that the first component of volatility, α_{1t}, empirically has the high persistence and highly negative leverage of volatility, and the second one has less persistence and leverage effect. These estimates show the log-volatility process would be divided into the two volatility components: one is the high-persistence autoregressive process and the other is almost the i.i.d. process with a larger variance.

Parameter	Mean	Stdev.	95% interval	Inefficiency
ϕ_1	0.9888	0.0020	[0.9848, 0.9926]	4.45
ϕ_2	0.0169	0.0864	[-0.1554, 0.2258]	15.99
σ_1	0.1058	0.0084	[0.0900, 0.1226]	6.25
σ_2	0.5299	0.0329	[0.4498, 0.5917]	21.09
ρ_1	-0.5401	0.0443	[-0.6173, -0.4479]	3.30
ρ_2	-0.1004	0.0389	[-0.1888, -0.0248]	12.33
$\exp(\mu/2)$	0.0086	0.0004	[0.0079, 0.0095]	1.91

Table 7: Estimation result of the superposition model for S&P500 return.

5.3 Model comparisons

This subsection conducts comparisons of those models we discussed in the previous sections using the marginal likelihoods. Twelve competing models are considered. In addition to four SV models with leverage (SVL, SVLt, SVLJ, SVLJt) in the empirical study above, we have corresponding four SV models without leverage (SV, SVt, SVJ, SVJt). Focusing on the heavy-tail behavior of return, we propose an alternative heavy-tail modelling in relation to Student-t error. As estimated in Omori et al. (2007), we also introduce the Gamma scale mixture SV model with leverage (labeled SVLg, SVLJg), where we assume $\log \lambda_t \sim N(-0.5\tau^2, \tau^2)$ and estimated τ instead of ν by the M-H algorithm as sampling ν. The $\tau^2 \sim \text{Gamma}(1, 1)$ is assumed for a prior density, whose mean and standard deviation is (1, 1). The SV model with correlated jumps (SVLCJ) and the superposition model (SVLSP) are also considered.

In a Bayesian framework we compare several competing models and find the evidence in the data using the posterior probabilities of the models. 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 log of marginal likelihood $m(y)$, as

$$\log m(y) = \log f(y|\vartheta) + \log \pi(\vartheta) - \log \pi(\vartheta|y^*, d). \tag{12}$$

where $f(y|\vartheta)$ is a likelihood, $\pi(\vartheta)$ is a prior probability density and $\pi(\vartheta|y^*, d)$ is a posterior probability density. This 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 requires a simulation evaluation. For the SV class, the likelihood can be estimated by the particle filter (e.g., Pitt and Shephard (1999), Chib et al. (2002), Omori et al. (2007)). We run 10 replications of the particle filter to estimate the standard error of the likelihood. We can evaluate the posterior density at the point ϑ through the additional but reduced MCMC iterations using the method of Chib (1995), Chib and Jeliazkov (2001, 2005). The reduced MCMC sampling for the posterior part is iterated for 5,000 draws.

We use six series of daily return data for the model comparisons. In addition to the datasets used for the previous parameter estimation, we used 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. We considered 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 this critical event could affect the model selection between models with and without jumps.

S&P500	1970-2003		1970-1985		1994-2003	
Model	Log-ML	Ranking	Log-ML	Ranking	Log-ML	Ranking
SV	29543.69 (2.23)	12	14178.18 (0.48)	11	8363.04 (0.44)	11
SVt	29601.06 (1.68)	10	14185.51 (0.98)	8	8373.40 (0.74)	9
SVJ	29561.59 (1.72)	11	14171.88 (1.10)	12	8357.51 (0.89)	12
SVJt	29602.11 (1.30)	9	14178.46 (1.05)	10	8363.13 (0.77)	10
SVL	29605.67 (2.04)	8	14198.89 (0.49)	4	8406.06 (0.37)	5
SVLt	29659.72 (2.89)	2	14204.31 (0.76)	1	8413.49 (0.48)	1
SVLg	29660.83 (1.86)	1	14202.75 (0.50)	2	8413.06 (0.62)	2
SVLJ	29623.93 (2.06)	6	14189.00 (1.19)	7	8402.85 (1.08)	7
SVLJt	29658.21 (1.91)	3	14196.10 (0.43)	5	8406.38 (0.70)	4
SVLJg	29655.07 (1.60)	4	14192.89 (1.00)	6	8403.96 (0.88)	6
SVLCJ	29608.27 (1.83)	7	14184.87 (1.10)	9	8396.74 (1.15)	8
SVLSP	29654.98 (1.75)	5	14199.66 (0.53)	3	8408.75 (0.50)	3

TOPIX	1970-2004		1970-1985		1992-2004	
Model	Log-ML	Ranking	Log-ML	Ranking	Log-ML	Ranking
SV	32385.53 (2.08)	11	17598.87 (1.30)	8	9716.88 (0.22)	11
SVt	32400.13 (1.54)	9	17597.14 (1.66)	9	9723.69 (0.73)	9
SVJ	32382.33 (1.98)	12	17591.68 (1.17)	10	9712.10 (0.55)	12
SVJt	32396.28 (1.69)	10	17589.48 (0.85)	12	9717.35 (0.69)	10
SVL	32461.14 (3.10)	6	17626.79 (0.94)	2	9738.27 (0.32)	4
SVLt	32475.55 (1.53)	2	17624.79 (1.63)	3	9744.55 (1.00)	1
SVLg	32470.36 (0.82)	3	17623.92 (1.13)	4	9742.74 (0.51)	2
SVLJ	32457.96 (1.77)	7	17618.18 (1.43)	6	9734.03 (0.66)	6
SVLJt	32468.35 (2.33)	4	17618.84 (1.57)	5	9735.50 (0.59)	5
SVLJg	32464.09 (2.08)	5	17617.45 (2.00)	7	9732.24 (0.37)	7
SVLCJ	32431.76 (5.40)	8	17590.98 (3.45)	11	9727.26 (0.95)	8
SVLSP	32498.88 (1.91)	1	17627.33 (1.02)	1	9739.46 (0.30)	3

*The values are based on log scale and standard error in parentheses.

Table 8: Marginal likelihood (ML) for S&P500 (top) and TOPIX (bottom) returns data.

Table 8 reports the estimated marginal likelihoods, standard errors and rankings for all competing models. The most adequate model to fit the S&P500 data is the SVLg model for the period 1970-2003, the SVLt model for the period 1970-1985 and 1994-2003, and that of the TOPIX data is the SVLSP model for the period 1970-2004 and 1970-1985, the SVLt model for the period 1992-2004. These three models, the SVLt, SVLg and SVLSP model, are ranked high for all periods, and taking the standard errors into consideration, these models are almost equivalent to explain the daily return series over our datasets.

As for the model fitting of the jump models, the advantage of the jump component can be seen only when the model has the normal-distributed error and the sample data contains the crash of 1987. Between the SV and SVJ model, the SVJ model dominates the SV model for sample data which contain the crash of 1987 (period 1970-2003) of S&P500, on the other hand, the SV model dominates SVJ model for sample data which exclude the crash (period 1970-1985, 1994-2003 of S&P500 and period 1970-1985, 1992-2004 of TOPIX). However, between the SVt and SVJt model, the SVt model outperforms the SVJt model for all sample period except the period 1970-2003 of S&P500. We found the same contribution of the jump component for the models with leverage. For the SVL, SVLJ, SVLt, SVLJt model, the jump model can beat the no-jump model only for the case where the model has normal-distributed error and the sample data contains the crash of 1987. In other cases, the jump models are outperformed by the no-jump models. Thus we conclude that the jump models can have an advantage (in the sense of model-fitting) when the disturbance of the model follows normal distribution and the very large shock exists in the data. If we allow the disturbance of the model to follow the heavy-tailed distribution (e.g. Student-t error), the incorporation of the jump component into the model does not improve marginal likelihoods. The jump component would not be necessary when the model has leverage effects and Student-t errors.

One of our aim in this paper is to investigate whether the heavy-tailed error distribution or jump component can capture the rare excess returns over the sample period in regard with model fitting. We consider the result of model comparison indicates that the jump component doesn't have enough additional information in the empirical study and heavy-tailed distribution for the disturbance in observation equation is favored, even when the jump component is rearranged to correlated jumps for both observation and volatility equations.

Similar to the result of Raggi and Bordignon (2006), the SVLCJ models are not favored over the SVLJ models for all series. On the marginal likelihood estimated for our return data, the additional benefit to model jumps for capturing rare tail events are not so large as the cost to add the parameters for the jump component. It is obvious that a huge excess return such as the crash of 1987 can be easily classified as a jump in the estimation, but relatively smaller excess returns may be captured as a tail event of the disturbance. How to capture the excess return is a relative matter. Our findings may imply that the consideration of jumps is rather costly to separate the jump event from the tail event of the error distribution over our empirical return data. This is probably because the additional new parameters in the SVLJ models were not very useful to increase the likelihood. The SVLSP model mostly outperforms the jump models and some heavy-tailed models. As claimed by Omori et al. (2007), the SVLSP models are considered to explain the heavy-tailness of asset returns by double sequences of volatility.

5.4 Prior sensitivity analysis

We conduct the prior sensitivity analysis to investigate the robustness of our empirical results in the previous subsection. Since we took the commonly used values for the prior hyperparameters of $(\phi, \sigma^2, \rho, \mu)$ in the basic SVL model, we focus on the parameters of our interests in this paper, i.e., the jump parameters (κ, δ) and heavy-tail parameter ν to study the prior sensitivity.

The following alternative priors are considered in addition to the prior in the previous estimations (Prior #1).

$$\begin{array}{llll}
\text{Prior \#1:} & \kappa \sim \text{Beta}(2,100), & \log(\delta) \sim N(-2.5, 0.15), & \nu \sim \text{Gamma}(16, 0.8), \\
\text{Prior \#2:} & \kappa \sim \text{Beta}(2,100), & \log(\delta) \sim N(-3, 0.5), & \nu \sim \text{Gamma}(16, 0.8), \\
\text{Prior \#3:} & \kappa \sim \text{Beta}(2,40), & \log(\delta) \sim N(-2.5, 0.15), & \nu \sim \text{Gamma}(16, 0.8), \\
\text{Prior \#4:} & \kappa \sim \text{Beta}(2,40), & \log(\delta) \sim N(-3, 0.5), & \nu \sim \text{Gamma}(16, 0.8), \\
\text{Prior \#5:} & \kappa \sim \text{Beta}(2,40), & \log(\delta) \sim N(-3, 0.5), & \nu \sim \text{Gamma}(24, 0.6).
\end{array}$$

The prior means and standard deviations are

$$\begin{array}{llll}
\text{Prior \#1:} & \kappa : (0.02, 0.01), & \delta : (0.09, 0.04), & \nu : (20, 5), \\
\text{Prior \#2:} & \kappa : (0.02, 0.01), & \delta : (0.06, 0.05), & \nu : (20, 5), \\
\text{Prior \#3:} & \kappa : (0.05, 0.03), & \delta : (0.09, 0.04), & \nu : (20, 5), \\
\text{Prior \#4:} & \kappa : (0.05, 0.03), & \delta : (0.06, 0.05), & \nu : (20, 5), \\
\text{Prior \#5:} & \kappa : (0.05, 0.03), & \delta : (0.06, 0.05), & \nu : (40, 8),
\end{array}$$

respectively. We estimated parameters by the proposed MCMC algorithm and computed marginal likelihoods for TOPIX data (1992/1/6 - 2004/12/30) using these alternative priors. Tables 9–11 report the result of parameter estimates and Table 12 listed the model comparison for the alternative priors among the major models (SVL, SVLt, SVLJ, SVLJt, SVLSP). Overall, the estimates of the parameters, $(\phi, \sigma, \rho, \exp(\mu/2))$, are basically unchanged, while the jump parameters and heavy-tail parameters appear to be more or less affected by the priors.

Table 9 shows the estimation result for the SVLJ model. To see the effect of the prior for the jump probability κ, we compare the the posterior estimates for Priors #1 and #3 (or Priors #2 and #4) where the prior mean is larger for Priors #3 (Prior #4). The posterior mean of κ becomes larger for Prior #3 (Prior #4) than the one for Prior #1 (Prior #2) . The posterior mean of δ becomes larger for Prior #3 (Prior #4) as well. The larger prior mean for κ seems to result in the larger posterior means of κ and δ, but their magnitude of changes might be small as in Prior #4 depending on the prior distribution of δ.

With respect δ, we compare the results for Priors #1 nd #2 (or Priors #3 and #4). There is no clear direction of the changes corresponding to the changes in prior distributions of δ. It seems to depend on the combination of prior distributions of κ and δ.

	SVLJ model			
Parameter	Prior #1	Prior #2	Prior #3	Prior #4
ϕ	0.9678 (0.0073)	0.9679 (0.0075)	0.9687 (0.0075)	0.9680 (0.0073)
	[0.9522, 0.9811]	[0.9519, 0.9809]	[0.9522, 0.9818]	[0.9523, 0.9808]
	7.12	4.62	5.36	4.31
σ	0.1899 (0.0185)	0.1885 (0.0188)	0.1860 (0.0203)	0.1890 (0.0188)
	[0.1564, 0.2281]	[0.1523, 0.2262]	[0.1468, 0.2287]	[0.1552, 0.2270]
	8.98	8.36	11.21	11.00
ρ	-0.4499 (0.0576)	-0.4457 (0.0588)	-0.4494 (0.0603)	-0.4476 (0.0589)
	[-0.5324, -0.3382]	[-0.5570, -0.3282]	[-0.5614, -0.3263]	[-0.5595, -0.3595]
	14.12	5.34	4.27	5.08
$\exp(\mu/2)$	0.0106 (0.0005)	0.0106 (0.0006)	0.0107 (0.0006)	0.0106 (0.0006)
	[0.0096, 0.0111]	[0.0096, 0.0117]	[0.0096, 0.0119]	[0.0096, 0.0117]
	2.11	2.39	3.67	2.30
κ	0.0031 (0.0091)	0.0043 (0.0210)	0.0100 (0.0385)	0.0047 (0.0228)
	[0.0002, 0.0188]	[0.0001, 0.0238]	[0.0001, 0.1309]	[0.0001, 0.0297]
	29.02	18.36	21.26	19.28
δ	0.0921 (0.0538)	0.1151 (0.0683)	0.1944 (0.1609)	0.1422 (0.0887)
	[0.0252, 0.2499]	[0.0245, 0.2621]	[0.0156, 0.6883]	[0.0211, 0.3145]
	24.33	32.49	24.59	81.60

Table 9: Prior sensitivity for the SVLJ model (estimation result for TOPIX return data).

	SVLt model	
Parameter	Prior #1	Prior #5
ϕ	0.9747 (0.0064)	0.9716 (0.0069)
	[0.9605, 0.9859]	[0.9566, 0.9839]
	19.69	7.28
σ	0.1623 (0.0182)	0.1756 (0.0186)
	[0.1291, 0.1999]	[0.1420, 0.2146]
	35.66	15.98
ρ	-0.4832 (0.0626)	-0.4715 (0.0582)
	[-0.5994, -0.3539]	[-0.5810, -0.3554]
	11.70	8.56
$\exp(\mu/2)$	0.0100 (0.0006)	0.0103 (0.0006)
	[0.0089, 0.0112]	[0.0092, 0.0114]
	11.20	3.71
ν	18.352 (4.4163)	34.3792 (8.1372)
	[11.830, 27.864]	[22.062, 52.523]
	152.76	169.46

Table 10: Prior sensitivity for SVLt model (estimation result for TOPIX return data).

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

For the SVLt model in Table 10, the posterior mean of ν for Prior #5 is estimated larger than the one for Prior #1 because the prior mean of Prior #5 for ν is 40 which is as twice as that of Prior #1. Such a sensitivity of the tight prior on ν is also reported in Chib et al. (2002) in the case of no-leverage model. As an additional estimation regarding this point, we estimated the SVLt model using a very flat prior, $\nu \sim \text{Gamma}(1, 0.05)$, whose mean and standard deviation are (20, 20) respectively. Using this prior, the posterior mean of ν is around 17, which implies that the density of Prior #1 for ν is matched to be centered around the posterior mean using the less informed prior. Prior #5 for ν is rather concentrated around 40 for the prior mean, which results in the large posterior mean as shown in Table 10.

			SVLJt model		
Parameter	Prior #1	Prior #2	Prior #3	Prior #4	Prior #5
ϕ	0.9749 (0.0063)	0.9750 (0.0061)	0.9748 (0.0063)	0.9746 (0.0065)	0.9722 (0.0065)
	[0.9611, 0.9857]	[0.9617, 0.9858]	[0.9607, 0.9856]	[0.9604, 0.9860]	[0.9520, 0.9807]
	14.15	11.66	10.77	13.24	5.05
σ	0.1621 (0.0177)	0.1612 (0.0169)	0.1631 (0.0178)	0.1630 (0.0191)	0.1736 (0.0185)
	[0.1298, 0.1994]	[0.1286, 0.1956]	[0.1301, 0.2005]	[0.1287, 0.2021]	[0.1561, 0.2283]
	27.66	30.62	18.07	25.99	12.84
ρ	-0.4842 (0.0619)	-0.4827 (0.0603)	-0.4842 (0.0613)	-0.4883 (0.0633)	-0.4687 (0.0590)
	[-0.6002, -0.3572]	[-0.5935, -0.3579]	[-0.6005, -0.3592]	[-0.6048, -0.3550]	[-0.5814, -0.3485]
	7.98	8.30	3.36	11.40	4.13
$\exp(\mu/2)$	0.0101 (0.0006)	0.0101 (0.0006)	0.0101 (0.0006)	0.0102 (0.0007)	0.0104 (0.0006)
	[0.0089, 0.0112]	[0.0090, 0.0112]	[0.0090, 0.0114]	[0.0072, 0.0121]	[0.0093, 0.0117]
	12.33	14.94	8.21	10.24	8.21
κ	0.0029 (0.0125)	0.0021 (0.0044)	0.0082 (0.0323)	0.0099 (0.0460)	0.0139 (0.0475)
	[0.0001, 0.0161]	[0.0001, 0.0110]	[0.0001, 0.0862]	[0.0001, 0.1203]	[0.0001, 0.1754]
	21.57	15.21	19.63	34.94	22.61
δ	0.1091 (0.0659)	0.0957 (0.0511)	0.1393 (0.0914)	0.1609 (0.1118)	0.1537 (0.1214)
	[0.0248, 0.2667]	[0.0262, 0.2110]	[0.0142, 0.3616]	[0.0149, 0.4294]	[0.0138, 0.4434]
	89.12	39.73	116.38	109.69	125.65
ν	18.501 (4.2324)	19.252 (4.2013)	19.764 (3.8080)	21.221 (5.3824)	39.212 (7.9436)
	[11.940, 28.223]	[10.584, 28.517]	[13.914, 28.591]	[14.532, 35.102]	[26.039, 57.026]
	220.80	214.51	177.80	150.33	143.53

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 11: Prior sensitivity for SVLJt model (estimation result for TOPIX return data).

In the SVLJt model, the effects of the prior distributions (Priors #1∼#4) of κ and δ in Table 11 are similar to those for the SVLJ model in Table 9. The posterior distributions of δ seem to be a little affected also by introducing the parameter ν in the SVLJt model. The posterior distributions of the jump probability, κ, seem to shift to the left in the SVLJt model for Priors #1∼#3, and the posterior means tend to be large as those of ν become large.

For Priors #4 and #5, when the prior mean of the ν is larger (assuming less fat-tailed errors), the posterior mean of ν becomes larger, and at the same time, the posterior mean of κ becomes larger. This is probably because the Student-t error distribution with large ν cannot describe excess observations sufficiently and the jump component is used instead to explain such a tail behaviour. This result indicates that there would be an essential trade-off between the roles of the heavy-tailed error distribution and the jump component to explain excess returns in the SV model. We further estimated the SVLJt model replacing the prior for ν in Prior #1 by the prior $\nu \sim \text{Gamma}(1, 0.05)$ again. Similar to the estimation result for Prior #4, the posterior mean of ν is found to be 23.8.

	Prior #1		Prior #2		Prior #3	
Model	Log-ML	Ranking	Log-ML	Ranking	Log-ML	Ranking
SVL	9738.27 (0.32)	3	9738.27 (0.32)	3	9738.27 (0.32)	3
SVLt	9744.55 (1.00)	1	9744.55 (1.00)	1	9744.55 (0.79)	1
SVLJ	9734.03 (0.66)	5	9727.90 (0.82)	5	9713.87 (1.17)	5
SVLJt	9735.50 (0.59)	4	9736.93 (0.99)	4	9725.64 (1.00)	4
SVLSP	9739.46 (0.30)	2	9739.46 (0.30)	2	9739.46 (0.30)	2

	Prior #4		Prior #5	
Model	Log-ML	Ranking	Log-ML	Ranking
SVL	9738.27 (0.32)	3	9738.27 (0.32)	3
SVLt	9744.55 (1.00)	1	9742.25 (0.79)	1
SVLJ	9726.09 (2.25)	4	9726.09 (2.25)	4
SVLJt	9719.45 (2.36)	5	9710.40 (2.14)	5
SVLSP	9739.46 (0.30)	2	9739.46 (0.30)	2

*The values are based on log scale and standard error in parentheses.

Table 12: Prior sensitivity: marginal likelihoods (ML) for TOPIX return data.

We investigate the prior sensitivity for marginal likelihoods in Table 12. For the jump models (SVLJ and SVLJt models) the marginal likelihoods decrease under Priors #4 and #5, which would be affected by the prior distributions for δ. Except these two models the rankings of the models are not altered. In the comparison of the SVLJt model under Priors #4 and #5, we presume that the higher posterior mean of ν for Prior #5 is related to its lower marginal likelihood because there would be less chances for the heavy-tail error distribution to capture the excess returns.

In conclusion, we observed: (i) the jump probability κ tends to become lower when Student-t error is incorporated, (ii) the posterior density of ν may be affected by the tight prior such as Prior #5, though the posterior estimates based on Prior #1 are stable even when the fairly flat prior is used, (iii) when the prior mean of the jump probability κ is set larger, the posterior mean also becomes larger and the excess return is explained more by the jump component, which tends to make the heavy-tail parameter ν to be stochastically larger. However, the effect of the prior density of κ is small regarding the ranking of models based on the marginal likelihoods.

5.5 Alternative jump size specification

The result in the previous sections indicates that the jump component in the SV model does not contribute much to the improvement of the marginal likelihood. In this section, we consider the alternative specification of the jump size in more detail and discuss whether such a specification could improve the marginal likelihood. The specification for jump size (k_t) given by equation (3) depends on just one parameter (δ). To investigate the effect of this parameterization, we consider a jump size distribution with two parameters: $\psi_t \equiv \log(1 + k_t) \sim N(\mu_\psi, \sigma_\psi^2)$. We denote such models by the SVLNJ and SVLNJt models. Using the priors $\mu_\psi \sim N(0, 1)$ and $\sigma_\psi^{-2} \sim \text{Gamma}(2.5, 0.025)$, we estimated the parameters and marginal likelihoods of these models for S&P500 (1970-2003) and TOPIX (1992-2004) data.

	S&P500 (1970-2003)		TOPIX (1992-2004)	
Parameter	**SVLNJ**	**SVLNJt**	**SVLNJ**	**SVLNJt**
ϕ	0.9818 (0.0029)	0.9877 (0.0020)	0.9680 (0.0071)	0.9753 (0.0062)
	[0.9756, 0.9870]	[0.9835, 0.9915]	[0.9527, 0.9805]	[0.9622, 0.9862]
	5.38	21.67	6.68	11.54
σ	0.1422 (0.0103)	0.1121 (0.0085)	0.1889 (0.0183)	0.1605 (0.0175)
	[0.1231, 0.1629]	[0.0960, 0.1291]	[0.1557, 0.2268]	[0.1284, 0.1987]
	7.69	36.95	7.23	25.43
ρ	-0.4969 (0.0376)	-0.5711 (0.0415)	-0.4458 (0.0570)	-0.4868 (0.0618)
	[-0.5672, -0.4212]	[-0.6469, -0.4855]	[-0.5549, -0.3299]	[-0.6052, -0.3601]
	6.00	16.01	7.53	8.22
$\exp(\mu/2)$	0.0089 (0.0004)	0.0084 (0.0004)	0.0106 (0.0005)	0.0100 (0.0006)
	[0.0082, 0.0096]	[0.0077, 0.0092]	[0.0096, 0.0117]	[0.0088, 0.0111]
	1.81	8.54	1.05	6.64
κ	0.0003 (0.0002)	0.0002 (0.0002)	0.0007 (0.0005)	0.0007 (0.0005)
	[0.0000, 0.0008]	[0.0000, 0.0007]	[0.0001, 0.0019]	[0.0001, 0.0019]
	1.74	1.68	0.99	1.14
μ_ψ	-0.0813 (0.7854)	-0.0298 (0.9361)	-0.0088 (0.9602)	-0.0031 (0.9664)
	[-1.6917, 1.7824]	[-1.9114, 1.9173]	[-1.9292, 1.8591]	[-1.9396, 1.9452]
	0.99	2.38	1.32	1.54
σ_ψ	0.1161 (0.0484)	0.1197 (0.0519)	0.1187 (0.0522)	0.1200 (0.0532)
	[0.0608, 0.2348]	[0.0633, 0.2523]	[0.0627, 0.2546]	[0.0623, 0.2472]
	1.26	0.76	0.72	1.30
ν		11.3340 (1.5412)		15.9215 (2.7006)
		[9.0096, 14.883]		[11.237, 21.615]
		275.77		198.67

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 13: Estimation result of the SVLNJ and SVLNJt model.

The results are reported in Table 13. The estimation results for $(\phi, \sigma, \rho, \exp(\mu/2))$ are basically unchanged for both the SVLNJ and SVLNJt models. The posterior means of μ_ψ for the SVLNJ model are slightly lower than those for the SVLNJt model, but the differences are small when taking the large posterior standard deviations into consideration. TFhe estimates of the standard deviations of the jump size, σ_ψ, are found to be about the same for both models. As a prior sensitivity analysis, we also estimated these models under alternative priors, $\mu_\psi \sim N(0,1)$ and $\sigma_\psi^{-2} \sim \text{Gamma}(5, 0.05)$, and obtained results of jump parameters are found to be about the same.

Table 14 reports the marginal likelihoods for the S&P500 and TOPIX return data. In fact, the SVLNJ and SVLNJt models are outperformed by the SVLJ and SVLJt models respectively except the SVLNJt model for TOPIX return data whose marginal likelihood is still smaller than that of SVLt model in Table 8.

Model	S&P500 (1970-2003)	TOPIX (1992-2004)
SVLJ	29623.93 (2.06)	9734.03 (0.66)
SVLNJ	29605.80 (3.26)	9732.77 (0.67)
SVLJt	29658.21 (1.91)	9735.50 (0.59)
SVLNJt	29650.13 (2.89)	9738.57 (1.13)

*The values are based on log scale and standard error in parentheses.

Table 14: Marginal likelihood (ML) for S&P500 and TOPIX returns data.

6 Conclusion

In this paper, we developed a fast and efficient MCMC sampling procedure for a Bayesian inference of the SV model with jumps, leverage and heavy-tail, and the SV model with correlated jumps. Our proposed method is illustrated using a simulation data and applied to daily stock returns data, the S&P500 index and the TOPIX. We further provided the overall model comparisons using the marginal likelihood of the nested candidate SV models with jumps and heavy-tails. The empirical result implies that the heavy-tailed SV model with leverage (SVLt and SVLg) and superposition model (SVLSP) fits to data better than other models during our data period.

Acknowledgement

The authors would like to thank anonymous referees, Siddhartha Chib, Herman van Dijk, Teruo Nakatsuma, Sylvia Frühwirth-Schnatter, Alan Gelfand, James LeSage and Wolfgang Polasek for their helpful comments and suggestions. The computational results are generated using Ox version 3.40 (see Doornik (2002)). This work is partially supported by Grants-in-Aid for Scientific Research 18330039 from the Japanese Ministry of Education, Science, Sports, Culture and Technology.

Appendix A. MCMC algorithm for SVLJt model

We show the details of sampling scheme for the SVLJt model in this appendix. Our proposed algorithm is as follows.

1. Initialize $\theta, \mu, \kappa, \delta, \nu, s, h, \psi, \gamma$ and λ
2. Sample $(\theta, \mu, h)|s, y^*, d$ by
   (a) Sampling $\theta|s, y^*, d$
   To sample θ from the posterior distribution $\pi(\theta|s, y^*, d) \propto f(y^*|\theta, s, d)\pi(\theta)$ by M-H algorithm, we evaluate $f(y^*|\theta, s, d)$ using an augmented Kalman filter as shown in Appendix B. We compute θ_* and Σ_* in equation (7) and generate a candidate θ^* from the distribution $N(\theta_*, \Sigma_*)$ truncated over $R = \{\theta : |\phi| < 1, \sigma > 0, |\rho| < 1\}$. Let θ_0 denote the current point of θ. We accept the candidate θ^* with probability
   $$\alpha(\theta_0, \theta^*|s, y^*, d) = \min\left\{\frac{\pi(\theta^*|s, y^*, d) f_N(\theta_0|\theta_*, \Sigma_*)}{\pi(\theta_0|s, y^*, d) f_N(\theta^*|\theta_*, \Sigma_*)}, 1\right\},$$
   where f_N denotes the density of the truncated normal distribution for the proposal above. If the candidate θ^* is rejected, we take the current value θ_0 as the next draw.
   (b) Sampling $(\mu, h)|\theta, s, y^*, d$
   First we generate
   $$\mu|\theta, s, y^*, d \ \sim \ N(Q_{n+1}^{-1} q_{n+1}, \ Q_{n+1}^{-1}),$$
   where q_{n+1} and Q_{n+1} are computed using the by-products of the augmented Kalman filter (see Appendix B). Next we can sample $h|\mu, \theta, s, y^*, d$ in one block using the simulation smoother (de Jong and Shephard (1995), Durbin and Koopman (2002a)). Given $s_t = i$, our approximating linear Gaussian state space model is formed by
   $$\begin{aligned} y_t^* &= m_i + h_t + G_t u_t, \\ h_{t+1} &= d_t \rho\sigma a_i \exp(m_i/2) + (1-\phi)\mu + \phi h_t + H_t u_t, \end{aligned}$$
   where $u_t \sim N(0, I_2)$,
   $$G_t = (v_i, \ 0), \quad H_t = \left(d_t \rho\sigma b_i v_i \exp(m_i/2), \ \sigma\sqrt{1-\rho^2}\right). \qquad (13)$$
3. Sample $\psi^{(1)}|\theta, \mu, \delta, h, \gamma, \lambda, y$
   The conditional posterior density for $\psi_t^{(1)}$ is given by
   $$\pi(\psi_t^{(1)}|\theta, \mu, \delta, h, \gamma, \lambda, y)$$

$$\propto \sum_{i=1}^{K} q_i \cdot \frac{1}{v_i} \exp\left\{-\frac{(y_t^* - h_t - m_i)^2}{2v_i^2}\right\}$$
$$\times \exp\left[-\frac{\{(h_{t+1}-\mu) - \phi(h_t-\mu) - D_i(y_t^*)\}^2}{2\sigma^2(1-\rho^2)}\right] \exp\left\{-\frac{(\psi_t + 0.5\delta^2)^2}{2\delta^2}\right\},$$

where $y_t^* = \log(y_t - (e^{\psi_t} - 1)\gamma_t)^2 - \log\lambda_t$ and

$$D_i(y_t^*) \;=\; d_t\rho\sigma\left\{a_i + b_i(y_t^* - h_t - m_i)\right\}\exp\left(\frac{m_i}{2}\right).$$

In the case of $t = n$, the second $\exp[\cdot]$ term is omitted. We marginalize the posterior density over $s_t = i$ $(i = 1, \ldots, K)$ and conduct the M-H algorithm for sampling $\psi_t^{(1)}$. Under the assumption that k_t is small, $k_t = e^{\psi_t} - 1$ may be well approximated by ψ_t. We generate the candidate in the M-H algorithm based on the equation (8) which produces the proposal density $N(\hat{\psi}_t, \sigma^2_{\psi_t})$ where $\hat{\psi}_t$ and $\sigma^2_{\psi_t}$ is given by the equation (9). We accept or reject the candidate using the M-H algorithm as in sampling θ.

4. Sample $(\delta, \psi^{(0)})|\psi^{(1)}, \gamma$ by

   (a) Sampling $\delta|\psi^{(1)}, \gamma$

   The conditional posterior distribution for δ is given by

$$\pi(\delta|\psi^{(1)}, \gamma) \;\propto\; \prod_{\gamma_t=1} \frac{1}{\sqrt{2\pi}\delta} \exp\left\{-\frac{(\psi_t^{(1)} + 0.5\delta^2)^2}{2\delta^2}\right\}\pi(\delta).$$

   We sample δ by A-R M-H algorithm (Tierney (1994), Chib and Greenberg (1995)). Let $\hat{\delta}$ denote the mode (or approximate mode) of the conditional posterior density $\pi(\delta|\psi^{(1)}, \gamma)$, and let $\ell(\delta) = \log\pi(\delta|\psi^{(1)}, \gamma)$. Applying Taylor expansion to $\ell(\delta)$ around $\hat{\delta}$ as

$$\ell(\delta) \;\approx\; \ell(\hat{\delta}) + \ell'(\hat{\delta})(\delta - \hat{\delta}) + \frac{1}{2}\ell''(\hat{\delta})(\delta - \hat{\delta})^2 \;\equiv\; h(\delta),$$

   where $\ell'(\hat{\delta})$ and $\ell''(\hat{\delta})$ are the first and second derivative of $\ell(\delta)$ evaluated at $\delta = \hat{\delta}$. We construct the approximating density $N(\mu_\delta, \sigma_\delta^2)$ truncated over $(0, \infty)$, where $\mu_\delta = \hat{\delta} - \ell'(\hat{\delta})/\ell''(\hat{\delta})$ and $\sigma_\delta^2 = -1/\ell''(\hat{\delta})$. We sample δ by the following two steps.

   i. A-R step

      Generate a candidate $\delta^* \sim N(\mu_\delta, \sigma_\delta^2)$ truncated over $(0, \infty)$ and accept δ^* with probability $\min(1, \exp\{\ell(\delta^*) - h(\delta^*)\})$. If it is rejected, generate δ^* again till the candidate is accepted.

   ii. M-H step

Let δ_0 denote the current point of δ. Accept δ^* with probability

$$\min\left\{\frac{\exp(\ell(\delta^*))\min\{\exp(\ell(\delta_0)),\,\exp(h(\delta_0))\}}{\exp(\ell(\delta_0))\min\{\exp(\ell(\delta^*)),\,\exp(h(\delta^*))\}},\,1\right\}.$$

If δ^* is rejected here, δ_0 is retained as the next value.

(b) Sampling $\psi^{(0)}|\delta,\gamma$

We sample simply as

$$\psi_t^{(0)}|\delta,\gamma \;\sim\; N(-0.5\delta^2,\delta^2),$$

independently for $t = 1,\ldots,n$ when $\gamma_t = 0$.

5. Sample $(\gamma, s)|\theta,\mu,\kappa,h,\psi,\lambda,y$ by

(a) Sampling $\gamma|\theta,\mu,\kappa,h,\psi,\lambda,y$

We sample γ_t using the probability mass function

$$\begin{aligned}
\Pr(\gamma_t = 1|\theta,\mu,\kappa,h,\psi,\lambda,y) &\propto \kappa\sum_{i=1}^K q_i\frac{1}{v_i}\exp\left\{-\frac{(y_t^{*(1)}-h_t-m_i)^2}{2v_i^2}\right\}\\
&\quad\times\exp\left[-\frac{\{(h_{t+1}-\mu)-\phi(h_t-\mu)-D_i(y_t^{*(1)})\}^2}{2\sigma^2(1-\rho^2)}\right],\\
\Pr(\gamma_t = 0|\theta,\mu,\kappa,h,\psi,\lambda,y) &\propto (1-\kappa)\sum_{i=1}^K q_i\frac{1}{v_i}\exp\left\{-\frac{(y_t^{*(0)}-h_t-m_i)^2}{2v_i^2}\right\}\\
&\quad\times\exp\left[-\frac{\{(h_{t+1}-\mu)-\phi(h_t-\mu)-D_i(y_t^{*(0)})\}^2}{2\sigma^2(1-\rho^2)}\right],
\end{aligned}$$

for $t = 1,\ldots,n-1$, and for $t = n$ the second $\exp[\cdot]$ term is omitted, where

$$\begin{aligned}
y_t^{*(1)} &= \log(y_t-(e^{\psi_t}-1))^2-\log\lambda_t,\\
y_t^{*(0)} &= \log y_t^2-\log\lambda_t.
\end{aligned}$$

This posterior density is also marginalized over s_t as sampling $\psi_t^{(1)}$ in step 3.

(b) Sampling $s|\theta,\mu,h,y^*,d$

To sample s_t, we compute

$$\begin{aligned}
&\pi(s_t = i|\theta,\mu,h,y^*,d)\\
&\propto\; q_i\cdot\frac{1}{v_i}\exp\left\{-\frac{(y_t^*-h_t-m_i)^2}{2v_i^2}\right\}\exp\left[-\frac{\{(h_{t+1}-\mu)-\phi(h_t-\mu)-D_i(y_t^*)\}^2}{2\sigma^2(1-\rho^2)}\right],
\end{aligned}$$

for $i = 1,\ldots,K$. We sample s_t the K-point discrete distribution independently for $t = 1,\ldots,n$. In the case of $t = n$, the second $\exp[\cdot]$ term is omitted.

6. Sample $\kappa|\gamma$.

   We sample κ by

$$\kappa|\gamma \sim \text{Beta}(n_{\kappa 1} + n_1, n_{\kappa 0} + n_0),$$

   where n_0 and n_1 are the numbers of time such that $\gamma_t = 0$ and $\gamma_t = 1$ respectively.

7. Sample $(\lambda, \nu)|\theta, \mu, s, h, \psi, \gamma, y$ by

   (a) Sampling $\lambda|\theta, \mu, \nu, s, h, \psi, \gamma, y$

   We assume $\lambda_t^{-1} \sim \text{Gamma}(\nu/2, \nu/2)$, then the conditional posterior density for (λ, ν) is given by

$$\pi(\lambda, \nu|\theta, \mu, s, h, \psi, \gamma, y) \propto \pi(\nu) \prod_{t=1}^{n} \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \lambda_t^{-\left(\frac{\nu}{2}+1\right)} \exp\left\{ -\frac{\nu}{2\lambda_t} - \frac{(\log \lambda_t - \mu_{\lambda_t})^2}{2\sigma_{\lambda_t}^2} \right\},$$

   where

$$\begin{aligned}
\mu_{\lambda_t} &= \log(y_t - k_t\gamma_t)^2 - m_i - h_t \\
&\quad - \frac{d_t \rho b_i v_i^2 \exp(m_i/2)\{h_{t+1} - \phi h_t - (1-\phi)\mu - d_t \rho \sigma a_i \exp(m_i/2)\}}{\sigma\{(1-\rho^2) + \rho^2 b_i^2 v_i^2 \exp(m_i)\}}, \\
\sigma_{\lambda_t}^2 &= \frac{v_i^2(1-\rho^2)}{1 - \rho^2 + \rho^2 b_i^2 v_i^2 \exp(m_i)},
\end{aligned}$$

   given $s_t = i$, for $t = 1, \ldots, n-1$ and $\mu_{\lambda_n} = \log(y_n - k_n\gamma_n)^2 - m_i - h_n$, $\sigma_{\lambda_n}^2 = v_i^2$. The conditional posterior distribution for λ_t is given by

$$\pi(\lambda|\theta, \mu, \nu, s, h, \psi, \gamma, y) \propto \lambda_t^{-\left(\frac{\nu}{2}+1\right)} \exp\left\{ -\frac{\nu}{2\lambda_t} - \frac{(\log \lambda_t - \mu_{\lambda_t})^2}{2\sigma_{\lambda_t}^2} \right\},$$

   We sample λ_t using the M-H algorithm with the candidate drawn by $(\lambda_t^*)^{-1} \sim \text{Gamma}(\nu/2, \nu/2)$.

   (b) Sampling $\nu|\lambda$

   The conditional posterior distribution for ν is given by

$$\pi(\nu|\lambda) \propto \pi(\nu) \frac{\left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)^n} \prod_{t=1}^{n} \lambda_t^{-\frac{\nu}{2}} \exp\left(-\frac{\nu}{2} \sum_{t=1}^{n} \lambda_t^{-1} \right).$$

   We sample ν by the A-R M-H algorithm as in sampling δ in step 4.

Appendix B. Augmented Kalman filter

To find θ that maximizes $\pi(\theta|s, y^*, d)$, we need to evaluate a likelihood $f(y^*|\theta, s, d)$. In the following, we assume $\theta = (\phi, \sigma, \rho)$ is fixed. Based on de Jong (1991), we conduct an augmented Kalman filter and calculate the likelihood function.

Consider a state space model

$$\begin{aligned} y_t &= X_t\beta + Z_t\alpha_t + G_t u_t, \quad t = 1, \ldots, n, \\ \alpha_{t+1} &= W_t\beta + T_t\alpha_t + H_t u_t, \quad t = 0, 1, \ldots, n, \end{aligned}$$

where $X_t, Z_t, W_t, T_t, G_t, H_t$ $(t = 1, \ldots, n), X_0, H_0$ are constants, and

- $\beta = b + B\mu$, $\mu \sim N(c, C)$. b is fixed and B has full column rank. μ is an $m \times 1$ vector. C is nonsingular unless $C = O$. $Cov(y) = \Sigma$ $(y = (y_1', y_2', \ldots, y_n')')$ is nonsingular if $C = O$.
- $u_t \sim NID(0, I)$ for $t = 0, 1, \ldots, n$. $\alpha_0 = 0$. μ and u_t's are uncorrelated.

In our approximated state space form of the SVLJt model, we set $Z_t = 1, T_t = \phi$, $\mu \sim N(\mu_0, \sigma_{\mu 0}^2)$, $b = (1, 1, 0)', B = (0, 0, 1-\phi)'$, $X_t = (m_i, 0, 0)$, $W_t = (0, d_t\rho\sigma a_i \exp(m_i/2), 1)$ for $t = 1, \ldots, n$, and $W_0 = \left(0, 0, \frac{1}{1-\phi}\right)$, where μ_0 and $\sigma_{\mu 0}^2$ are hyperparamters of the prior for μ. The G_t, H_t's are given in equation (13) and y_t corresponds to y_t^* here.

When μ is fixed, the Kalman filter is the recursion

$$\begin{aligned} D_t &= Z_t P_{t|t-1} Z_t' + G_t G_t', \quad K_t = (T_t P_{t|t-1} Z_t' + H_t G_t') D_t^{-1}, \\ P_{t+1|t} &= T_t P_{t|t-1} L_t' + H_t J_t', \quad L_t = T_t - K_t Z_t, \\ e_t &= y_t - X_t\beta - Z_t a_{t|t-1}, \quad a_{t+1|t} = W_t\beta + T_t a_{t|t-1} + K_t e_t, \end{aligned}$$

for $t = 1, \ldots, n$, where $J_t = H_t - K_t G_t$ and $a_{1|0} = W_0\beta$, $P_1 = H_0 H_0'$. Further, we consider additional equations

$$\begin{aligned} &f_t = y_t - X_t b - Z_t a_{t|t-1}^*, \quad a_{t+1|t}^* = W_t b + T_t a_{t|t-1}^* + K_t f_t, \\ &F_t = X_t B - Z_t A_{t|t-1}^*, \quad A_{t+1|t}^* = -W_t B + T_t A_{t|t-1}^* + K_t F_t, \end{aligned}$$

for $t = 1, \ldots, n$, where $a_{1|0}^* = W_0 b$, $A_{1|0}^* = -W_0 B$. Note that

$$e_t = f_t - F_t\mu, \quad a_{t+1|t} = a_{t+1|t}^* - A_{t+1|t}^*\mu.$$

Then the log likelihood given μ is

$$\log f(y|\mu) = -\frac{1}{2}\left\{n \log 2\pi + \log|\Sigma| + (y - Xb)'\Sigma^{-1}(y - Xb) - 2q'\mu + \mu' Q\mu\right\},$$

where $\log|\Sigma| = \sum_{t=1}^n \log|D_t|$, $(y - Xb)'\Sigma^{-1}(y - Xb) = \sum_{t=1}^n f_t' D_t^{-1} f_t$, $q = \sum_{t=1}^n F_t' D_t^{-1} f_t$ and $Q = \sum_{t=1}^n F_t' D_t^{-1} F_t$. On the other hand, the posterior distribution of μ given y is $N(Q_{n+1}^{-1} q_{n+1}, Q_{n+1}^{-1})$

where

$$q_{t+1} = q_t + F_t' D_t^{-1} f_t, \quad q_1 = C^{-1} c,$$
$$Q_{t+1} = Q_t + F_t' D_t^{-1} F_t, \quad Q_1 = C^{-1},$$

for $t = 1, \ldots, n$. Thus we obtain the likelihood of y as

$$\begin{aligned} \log f(y) &= \log f(y|\mu) + \log \pi(\mu) - \log \pi(\mu|y) \\ &= \text{const.} - \frac{1}{2}\left\{ \sum_{t=1}^{n} \log |D_t| + \log |Q_{n+1}| + \sum_{t=1}^{n} f_t' D_t^{-1} f_t + c' C^{-1} c - q_{n+1}' Q_{n+1}^{-1} q_{n+1} \right\}. \end{aligned}$$

References

Andersen, T., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time models for equity returns. *Journal of Finance 57*, 1239–1284.

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.

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

Chan, W. H. and J. M. Maheu (2002). Conditional jump dynamics is stock market returns. *Journal of Business and Economic Statistics 20*, 377–389.

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 E. Greenberg (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory 12*, 409–431.

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

Chib, S. and I. Jeliazkov (2005). Accept-Reject Metropolis-Hastings sampling and marginal likelihood estimation. *Statistica Neerlandica 59*, 30–44.

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. (1991). The diffuse Kalman filter. *Annals of Statistics 19*, 1073–1083.

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

Doornik, J. (2002). *Object-Oriented Matrix Programming Using Ox* (3rd ed.). London: Timberlake Consultants Press Oxford: www.nuff.ox.ac.uk/Users/Doornik.

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

Durbin, J. and S. J. Koopman (2002b). *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

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.

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.

Jacquier, E., N. Polson, and P. Rossi (2004). Bayesian analysis of stochastic volatility with fat-tails and correlated errors. *Journal of Econometrics 122*(1), 185–212.

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

Kobayashi, M. (2006). Testing for volatility jumps in the stochastic volatility process. *Asia-Pacific Financial Markets 12*, 143–157.

Maheu, J. M. and T. H. McCurdy (2004). News arrival, jump dynamics, and volatility components for individual stock returns. *Journal of Finance 59*, 755–793.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica 59*, 347–370.

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

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

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

Shephard, N. (Ed.) (2005). *Stochastic Volatility: Selected Readings.* Oxford: Oxford University Press.

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

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