CARF-F-221

Efficient Bayesian Estimation of a Multivariate Stochastic Volatility Model with Cross Leverage and Heavy-Tailed Errors

Tsunehiro Ishihara
University of Tokyo
Yasuhiro Omori
University of Tokyo

May 2010

Efficient Bayesian estimation of a multivariate stochastic volatility model with cross leverage and heavy-tailed errors

Tsunehiro Ishihara[a], Yasuhiro Omori[1,*]

[a] *Graduate School of Economics, University of Tokyo, Japan.*
[b] *Faculty of Economics, University of Tokyo, Bunkyo-Ku, Tokyo 113-0033, Japan.*

Abstract

An efficient Bayesian estimation using a Markov chain Monte Carlo method is proposed in the case of a multivariate stochastic volatility model as a natural extension of the univariate stochastic volatility model with leverage and heavy-tailed errors. Note that we further incorporate cross-leverage effects among stock returns. Our method is based on a multi-move sampler that samples a block of latent volatility vectors. The method is presented as a multivariate stochastic volatility model with cross leverage and heavy-tailed errors. Its high sampling efficiency is shown using numerical examples in comparison with a single-move sampler that samples one latent volatility vector at a time, given other latent vectors and parameters. To illustrate the method, empirical analyses are provided based on five-dimensional S&P500 sector indices returns.

Keywords: Asymmetry, Heavy-tailed error, Leverage effect, Markov chain Monte Carlo, Multi-move sampler, Multivariate stochastic volatility

---

[*] Corresponding author. Tel: +81-3-5841-5516. Email: omori@e.u-tokyo.ac.jp

May 30, 2010

1. Introduction

Univariate stochastic volatility (SV) models are well-known to successfully account for time-varying variance in financial time series (Broto and Ruiz (2004)). Many efficient Bayesian estimation methods that use Markov chain Monte Carlo (MCMC) methods are proposed, because the likelihood functions are difficult to evaluate in the implementation of the maximum likelihood estimation (e.g. Shephard and Pitt (1997), Omori et al. (2007)).

Extending these models to a multivariate SV (MSV) model has recently become a major concern in the investigation of the correlation structure of multivariate financial time series, especially with regard to portfolio optimization, risk management, and derivative pricing. The multivariate factor modeling of stochastic volatilities has been widely introduced to describe the complex dynamic structure of high-dimensional stock returns data (Jacquier et al. (1999), Liesenfeld and Richard (2003), Pitt and Shephard (1999), Lopes and Carvalho (2007), and several efficient MCMC algorithms have been proposed (So and Choi (2009), Chib et al. (2006)). However, efficient estimation methods for MSV models with cross leverage (i.e., a non-zero correlation between the i-th asset return at time t and the j-th log volatility at time $t+1$ for all i and j) or asymmetry have not been well investigated in the literature except for simple bivariate models; see surveys by Asai et al. (2006) and Chib et al. (2009). Chan et al. (2006) considered the Bayesian estimation of MSV models in which correlations existed between measurement errors and state errors, but their framework did not address leverage effects. Asai and McAleer (2006) simplified the MSV model with leverage by assuming no cross leverage effect (*i.e.*, no correlation between the i-th asset return at time t and the j-th log volatility at time $t+1$ for $i \neq j$) and thus described a Monte Carlo likelihood estimation method.

In this paper, we consider a general MSV model with cross leverage and heavy-tailed errors and propose a novel and efficient MCMC algorithm using a multi-move sampler that samples a block of latent volatility vectors simultaneously. To the best of our knowledge, this is the first efficient multi-move sampler proposed in the literature for the general MSV model with cross leverage and heavy-tailed errors. In any MCMC implementation of the SV models, it is critical to efficiently sample the latent volatility (or state) variables from their full conditional posterior distributions. The single-move sampler that draws a single volatility variable at a time given the other volatility variables and parameters is easy to implement, but the resulting MCMC samples are known to have high autocorrelations. This implies that we must iterate the MCMC algorithm a large number of times in order to obtain accurate estimates when using a single-move sampler. Thus, we propose a fast and efficient state-sampling algorithm based on the approximate linear and Gaussian state space model. Such a model is derived by approximating the conditional likelihood function by a multivariate normal density using a Taylor expansion around the mode. Starting with the current sample of state variables, the mode can be easily obtained by repeatedly applying the disturbance smoother (Koopman (1993)) to the approximate auxiliary state space model. The samples from the posterior distribution of latent state variables are obtained from the Metropolis-Hastings (MH) algorithm in which a simulation smoother for the linear and Gaussian state space models (de Jong and Shephard (1995), Durbin and Koopman (2002)) is used to generate a candidate variables.

The rest of the paper is organized as follows. Section 2 discusses the Bayesian estimation of the MSV model using a multi-move sampler of latent state variables. Extended models with heavy-tailed errors are also considered. In Section 3, we provide numerical examples using simulation data, and show that our proposed method outperforms the simple single-move sampler with respect to sampling efficiencies. Section 4 provides empirical analyses based on five-dimensional stock return indices. Section 5 concludes the paper.

2. The MSV model with cross leverage and heavy-tailed errors

2.1. MSV Model

Let y_t denote a stock return at time t. The univariate SV model with leverage is given by

$$\begin{aligned} y_t &= \exp(\alpha_t/2)\varepsilon_t, \quad t=1,\ldots,n, &(1)\\ \alpha_{t+1} &= \phi\alpha_t + \eta_t, \quad t=1,\ldots,n-1, &(2)\end{aligned}$$

and

$$\alpha_1 \sim \mathcal{N}(0, \sigma_\eta^2/(1-\phi^2)), \tag{3}$$

where

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}_2(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon\sigma_\eta \\ \rho\sigma_\varepsilon\sigma_\eta & \sigma_\eta^2 \end{pmatrix}, \tag{4}$$

α_t is a latent variable for the log-volatility, and $\mathcal{N}_m(\boldsymbol{\mu}, \mathbf{\Sigma})$ denotes an m-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. To extend this model to the MSV model, let $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{pt})$' denote a p-dimensional stock returns vector, and let $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{pt})'$ denote the corresponding log volatility vectors. We consider the MSV model given by

$$\begin{aligned}
\boldsymbol{y}_t &= \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, n, &(5)\\
\boldsymbol{\alpha}_{t+1} &= \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \quad t = 1, \ldots, n-1, &(6)
\end{aligned}$$

and

$$\boldsymbol{\alpha}_1 \sim \mathcal{N}_p\left(\mathbf{0}, \boldsymbol{\Sigma}_0\right), \tag{7}$$

where

$$\begin{aligned}
\mathbf{V}_t &= \mathrm{diag}\left(\exp(\alpha_{1t}), \ldots, \exp(\alpha_{pt})\right), &(8)\\
\boldsymbol{\Phi} &= \mathrm{diag}(\phi_1, \ldots, \phi_p), &(9)
\end{aligned}$$

and

$$\left(\begin{array}{c} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{array}\right) \sim \mathcal{N}_{2p}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \left(\begin{array}{cc} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & \boldsymbol{\Sigma}_{\varepsilon\eta} \\ \boldsymbol{\Sigma}_{\eta\varepsilon} & \boldsymbol{\Sigma}_{\eta\eta} \end{array}\right). \tag{10}$$

The (i, j)-th element of $\boldsymbol{\Sigma}_0$ is the (i, j)-th element of $\boldsymbol{\Sigma}_{\eta\eta}$ divided by $1 - \phi_i\phi_j$ to satisfy the stationarity condition $\boldsymbol{\Sigma}_0 = \boldsymbol{\Phi}\boldsymbol{\Sigma}_0\boldsymbol{\Phi} + \boldsymbol{\Sigma}_{\eta\eta}$ such that

$$\mathrm{vec}(\boldsymbol{\Sigma}_0) = \left(\mathbf{I}_{p^2} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi}\right)^{-1} \mathrm{vec}(\boldsymbol{\Sigma}_{\eta\eta}).$$

The expected value of the volatility evolution processes $\boldsymbol{\alpha}_t$ is set equal to $\mathbf{0}$ for identifiability. Let $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\Sigma})$, where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$, and let $\mathbf{1}_p$ denote a $p \times 1$ vector with all elements equal to one. Then the likelihood function of the MSV model outlined in equations (5) to (7) is given by

$$\begin{aligned}
& f(\boldsymbol{\alpha}_1|\boldsymbol{\theta}) \prod_{t=1}^{n-1} f(\boldsymbol{y}_t, \boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t, \boldsymbol{\theta}) f(\boldsymbol{y}_n|\boldsymbol{\alpha}_n, \boldsymbol{\theta}) \\
& \quad \propto \exp\left\{\sum_{t=1}^{n} l_t - \frac{1}{2}\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\alpha}_1 - \frac{1}{2}\sum_{t=1}^{n-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t)'\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t)\right\} \\
& \qquad \times |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}}|\boldsymbol{\Sigma}|^{-\frac{n-1}{2}}|\boldsymbol{\Sigma}_{\varepsilon\varepsilon}|^{-\frac{1}{2}}, \qquad (11)
\end{aligned}$$

where

$$
\begin{aligned}
l_t &= \text{const} - \frac{1}{2}\mathbf{1}_p'\boldsymbol{\alpha}_t - \frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{\mu}_t)'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_t), &(12)\\
&\boldsymbol{\mu}_t = \mathbf{V}_t^{1/2}\boldsymbol{m}_t, \quad \boldsymbol{\Sigma}_t = \mathbf{V}_t^{1/2}\mathbf{S}_t\mathbf{V}_t^{1/2}, &(13)\\
&\boldsymbol{m}_t = \begin{cases} \boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\boldsymbol{\alpha}_{t+1} - \Phi\boldsymbol{\alpha}_t), & t < n, \\ \mathbf{0} & t = n, \end{cases} &(14)
\end{aligned}
$$

and

$$
\mathbf{S}_t = \begin{cases} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} - \boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}, & t < n, \\ \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & t = n. \end{cases} \tag{15}
$$

2.2. Bayesian analysis and MCMC implementation

Since there are many latent volatility vectors $\boldsymbol{\alpha}_t$s, it is difficult to integrate them out in order to evaluate the likelihood function of $\boldsymbol{\theta}$ analytically or to use high-dimensional numerical integration. In this paper, we take a Bayesian approach and we employ a simulation method, namely, the MCMC method, to generate samples from the posterior distribution to conduct statistical inference with respect to the model parameters.

For prior distributions of $\boldsymbol{\theta}$, we assume

$$
\frac{\phi_j + 1}{2} \sim \mathcal{B}(a_j, b_j), \quad j = 1, \ldots, p, \quad \boldsymbol{\Sigma} \sim \mathcal{IW}(n_0, \mathbf{R}_0),
$$

where $\mathcal{B}(a_j, b_j)$ and $\mathcal{IW}(n_0, \mathbf{R}_0)$ respectively denote Beta and inverse Wishart distributions with probability density functions

$$
\begin{aligned}
\pi(\phi_j) &\propto (1 + \phi_j)^{a_j - 1}(1 - \phi_j)^{b_j - 1}, \quad j = 1, 2, \ldots, p, &(16)\\
\pi(\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-\frac{n_0 + p + 1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_0^{-1}\boldsymbol{\Sigma}^{-1}\right)\right\}. &(17)
\end{aligned}
$$

Using Equations (11), (16), and (17), we obtain the joint posterior density function of $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ given by

$$
\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}|Y_n) \propto f(\boldsymbol{\alpha}_1|\boldsymbol{\theta}) \prod_{t=1}^{n-1} f(\boldsymbol{y}_t, \boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t, \boldsymbol{\theta}) f(\boldsymbol{y}_n|\boldsymbol{\alpha}_n, \boldsymbol{\theta}) \prod_{j=1}^{p} \pi(\phi_j)\pi(\boldsymbol{\Sigma}), \tag{18}
$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_n')'$ and $Y_n = \{\boldsymbol{y}_t\}_{t=1}^n$. We implement the MCMC algorithm in three stages:

1. Generate $\boldsymbol{\alpha}|\boldsymbol{\phi}, \boldsymbol{\Sigma}, Y_n$.

2. Generate $\boldsymbol{\Sigma}|\boldsymbol{\phi}, \boldsymbol{\alpha}, Y_n$.
3. Generate $\boldsymbol{\phi}|\boldsymbol{\Sigma}, \boldsymbol{\alpha}, Y_n$.

First, we discuss two methods used to sample $\boldsymbol{\alpha}$ based on its conditional posterior distribution in Step 1. One method is a so-called single-move sampler that samples one $\boldsymbol{\alpha}_t$ at a time given the other $\boldsymbol{\alpha}_j$s, while the other method is a multi-move sampler that samples a block of state vectors, such as, $(\boldsymbol{\alpha}_t, \ldots, \boldsymbol{\alpha}_{t+k})$, given the other state vectors.

2.2.1. The generation of $\boldsymbol{\alpha}$

Single-move sampler. A simple but inefficient method is to sample one $\boldsymbol{\alpha}_t$ at a time given $\boldsymbol{\alpha}_j$s and other parameters. The conditional posterior probability density function of $\boldsymbol{\alpha}_t$ is

$$\pi(\boldsymbol{\alpha}_t|\{\boldsymbol{\alpha}_s\}_{s\neq t}, \boldsymbol{\phi}, \boldsymbol{\Sigma}, Y_n) \quad \propto \quad \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha}_t - \boldsymbol{m}_{\boldsymbol{\alpha}_t})'\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_t}^{-1}(\boldsymbol{\alpha}_t - \boldsymbol{m}_{\boldsymbol{\alpha}_t}) + g(\boldsymbol{\alpha}_t)\right\}$$

where

$$\boldsymbol{m}_{\boldsymbol{\alpha}_t} \quad = \quad \begin{cases} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_1}\left(-\frac{1}{2}\mathbf{1}_p + \boldsymbol{\Phi}\mathbf{M}_1\boldsymbol{\alpha}_2\right), & t=1, \\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_t}\left(-\frac{1}{2}\mathbf{1}_p + \boldsymbol{\Phi}\mathbf{M}_t\boldsymbol{\alpha}_{t+1} + \mathbf{M}_{t-1}\boldsymbol{\Phi}\boldsymbol{\alpha}_{t-1} + \mathbf{N}_{t-1}\right), & 1<t<n, \\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_n}\left(-\frac{1}{2}\mathbf{1}_p + \mathbf{M}_{n-1}\boldsymbol{\Phi}\boldsymbol{\alpha}_{n-1} + \mathbf{N}_{n-1}\right), & t=n, \end{cases}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_t} \quad = \quad \begin{cases} \left(\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Phi}\mathbf{M}_1\boldsymbol{\Phi}\right)^{-1}, & t=1, \\ \left(\mathbf{M}_{t-1} + \boldsymbol{\Phi}\mathbf{M}_t\boldsymbol{\Phi}\right)^{-1}, & 1<t<n, \\ \mathbf{M}_{n-1}^{-1}, & t=n, \end{cases}$$

$$\mathbf{M}_t = \boldsymbol{\Sigma}_{\eta\eta}^{-1} + \boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}, \quad \mathbf{N}_t = \boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\mathbf{V}_t^{-1/2}\boldsymbol{y}_t,$$

and

$$g(\boldsymbol{\alpha}_t) = -\frac{1}{2}\boldsymbol{y}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{y}_t + \boldsymbol{y}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t.$$

Thus, in order to sample from the conditional posterior distribution using the Metropolis-Hastings (MH) algorithm, we generate a candidate $\boldsymbol{\alpha}_t^{\dagger} \sim \mathrm{N}(\boldsymbol{m}_{\boldsymbol{\alpha}_t}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_t})$ and accept it with probability

$$\min\left\{\exp\{g(\boldsymbol{\alpha}_t^{\dagger}) - g(\boldsymbol{\alpha}_t)\}, 1\right\},$$

for $t = 1, \ldots, n$, where $\boldsymbol{\alpha}_t$ is a current value.

Multi-move sampler. As an alternative method, we propose an efficient method to sample a block of $\boldsymbol{\alpha}_t$s from the posterior distribution. As such, we extend Omori and Watanabe (2008), who considered the univariate SV model with leverage (see also Takahashi et al. (2009)). First, we divide $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_n')'$ into $K+1$ blocks $(\boldsymbol{\alpha}_{k_{i-1}+1}', \ldots, \boldsymbol{\alpha}_{k_i}')'$ using $i = 1, \ldots, K+1$, with $k_0 = 0$, $k_{K+1} = n$, and $k_i - k_{i-1} \geq 2$. K knots $(k_1, \ldots, k_K)$ are generated randomly using

$$k_i = \text{int}[n \times (i + U_i)/(K+2)], \quad i = 1, \ldots, K,$$

where U_is are independent uniform random variables on $(0,1)$ (Shephard and Pitt (1997)). These stochastic knots have an advantage in that they allow the points of conditioning to change over the MCMC iterations, as K is a tuning parameter used to obtain MCMC samples with relatively low autocorrelation.

Suppose that $k_{i-1} = s$ and $k_i = s+m$ for the i-th block. Consider sampling this block from its conditional posterior distribution given other state vectors and parameters. We denote $\boldsymbol{x}_t = \mathbf{R}_t^{-1}\boldsymbol{\eta}_t$, where the matrix $\mathbf{R}_t$ denotes a Choleski decomposition of $\boldsymbol{\Sigma}_{\eta\eta} = \mathbf{R}_t\mathbf{R}_t'$ for $t > 0$, and $\boldsymbol{\Sigma}_0 = \mathbf{R}_0\mathbf{R}_0'$ for $t = 0$. To construct a proposal distribution for the MH algorithm, we focus on the distribution of the disturbance $\boldsymbol{x} \equiv (\boldsymbol{x}_s', \ldots, \boldsymbol{x}_{s+m-1}')'$, which is fundamental in the sense that it derives the distribution of $\boldsymbol{a} \equiv (\boldsymbol{\alpha}_{s+1}', \ldots, \boldsymbol{\alpha}_{s+m}')'$. Then, the logarithm of the full conditional joint density distribution of $\boldsymbol{x}$, excluding constant terms, is given by

$$\log f(\boldsymbol{x}|\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+m+1}, \boldsymbol{y}_s, \ldots, \boldsymbol{y}_{s+m}) = -\frac{1}{2}\sum_{t=s}^{s+m-1} \boldsymbol{x}_s'\boldsymbol{x}_s + L, \tag{19}$$

where

$$L = \sum_{t=s}^{s+m} l_s - \frac{1}{2}(\boldsymbol{\alpha}_{s+m+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_{s+m})'\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\boldsymbol{\alpha}_{s+m+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_{s+m})I(s+m < n).$$

Using the second order Taylor expansion of (19) around the mode $\hat{\boldsymbol{x}}$, we obtain the approximate normal density f^*, which is used for the Acceptance-Rejection MH(ARMH) algorithm as follows.

$$
\begin{aligned}
&\log f(\boldsymbol{x}|\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+m+1}, \boldsymbol{y}_s, \dots, \boldsymbol{y}_{s+m}) \\
\approx\ & \text{const.} - \frac{1}{2}\sum_{t=s}^{s+m-1} \boldsymbol{x}_t'\boldsymbol{x}_t + \hat{L} \\
& + \left.\frac{\partial L}{\partial \boldsymbol{x}'}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}} (\boldsymbol{x}-\hat{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x}-\hat{\boldsymbol{x}})'\mathbb{E}\left(\frac{\partial^2 L}{\partial \boldsymbol{x}\partial \boldsymbol{x}'}\right)(\boldsymbol{x}-\hat{\boldsymbol{x}}) \\
=\ & \text{const.} - \frac{1}{2}\sum_{t=s}^{s+m-1} \boldsymbol{x}_t'\boldsymbol{x}_t + \hat{L} + \hat{\boldsymbol{d}}'(\boldsymbol{a}-\hat{\boldsymbol{a}}) - \frac{1}{2}(\boldsymbol{a}-\hat{\boldsymbol{a}})'\hat{\mathbf{Q}}(\boldsymbol{a}-\hat{\boldsymbol{a}}) \qquad (20) \\
=\ & \text{const.} + \log f^*(\boldsymbol{x}|\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+m+1}, \boldsymbol{y}_s, \dots, \boldsymbol{y}_{s+m}), \qquad (21)
\end{aligned}
$$

where $\hat{\mathbf{Q}}$ and $\hat{\boldsymbol{d}}$ are $\mathbf{Q} = -E(\partial^2 L/\partial \boldsymbol{a}\partial \boldsymbol{a}')$, and $\boldsymbol{d} = \partial L/\partial \boldsymbol{a}$, is evaluated at $\boldsymbol{a} = \hat{\boldsymbol{a}}$ (*i.e.*, $\boldsymbol{x} = \hat{\boldsymbol{x}}$). Note that $\boldsymbol{Q}$ is positive definite and invertible. However, when m is large, it is time consuming to invert the $mp \times mp$ Hessian matrix to obtain the covariance matrix of the mp-variate multivariate normal distribution. To overcome this difficulty, we interpret equation (21) as the posterior probability density derived from an auxiliary state space model so that we only need to invert $p \times p$ matrices by using the Kalman filter and the disturbance smoother. It can be shown that f^* is the posterior probability density function of $\boldsymbol{x}$ obtained from the state space model:

$$
\begin{aligned}
\hat{\boldsymbol{y}}_t &= \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{G}_t\boldsymbol{u}_t, \quad t = s+1, \dots, s+m, \qquad (22) \\
\boldsymbol{\alpha}_{t+1} &= \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \mathbf{H}_t\boldsymbol{u}_t, \quad t = s+1, \dots, s+m-1, \qquad (23)
\end{aligned}
$$

where $\boldsymbol{u}_t \sim \mathcal{N}_{2p}(\mathbf{0}, \mathbf{I}_{2p})$, $\hat{\boldsymbol{y}}_t$, $\mathbf{Z}_t$, and $\mathbf{G}_t$ are defined in Appendix A.1, and $\mathbf{H}_t = [\mathbf{0}, \mathbf{R}_t]$. To find a mode $\hat{\boldsymbol{x}}$, we repeat the following three steps until the convergence,

1. Compute $\hat{\boldsymbol{a}}$ at $\boldsymbol{x} = \hat{\boldsymbol{x}}$ using (6).
2. Obtain the approximate linear Gaussian state space model given by (22) and (23).
3. Applying the disturbance smoother by Koopman (1993) to the approximating linear Gaussian state space model in Step 2, compute the posterior mode $\hat{\boldsymbol{x}}$.

Note that these steps are equivalent to the method of scoring used to maximize the conditional posterior density. As an initial value of $\hat{\boldsymbol{x}}$, the current sample of $\boldsymbol{x}$ may be used in MCMC implementation. If the approximate linear Gaussian state space model is obtained using mode $\hat{\boldsymbol{x}}$, we draw a sample

$\boldsymbol{x}$ from the conditional posterior distribution by the AR-MH algorithm as follows.

1. Propose a candidate $\boldsymbol{x}^\dagger$ by sampling from $q(\boldsymbol{x}^\dagger) \propto \min(f(\boldsymbol{x}^\dagger), cf^*(\boldsymbol{x}^\dagger))$ using the AR algorithm, where c can be constructed from a constant term and $\hat{L}$ from (20).
   (a) Generate $\boldsymbol{x}^\dagger \sim f^*$ using a simulation smoother (de Jong and Shephard (1995), Durbin and Koopman (2002)) based on the approximate linear Gaussian state space model in (22) and (23).
   (b) Accept $\boldsymbol{x}^\dagger$ with probability $\min\left\{\frac{f(\boldsymbol{x}^\dagger)}{cf^*(\boldsymbol{x}^\dagger)}, 1\right\}$. If it is rejected, return to (a).
2. Given the current value $\boldsymbol{x}$, accept $\boldsymbol{x}^\dagger$ with probability

$$\min\left\{1, \frac{f(\boldsymbol{x}^\dagger)\min(f(\boldsymbol{x}), cf^*(\boldsymbol{x}))}{f(\boldsymbol{x})\min(f(\boldsymbol{x}^\dagger), cf^*(\boldsymbol{x}^\dagger))}\right\}.$$

   If rejected, accept the current $\boldsymbol{x}$ as a sample.

We investigate the efficiency performance of these two sampling methods in Section 3 using simulated data.

2.2.2. The generation of $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\phi}$

Generation of $\boldsymbol{\Sigma}$. The conditional posterior probability density function of $\boldsymbol{\Sigma}$ is

$$\pi(\boldsymbol{\Sigma}|\boldsymbol{\phi}, \boldsymbol{\alpha}, Y_n) \propto |\boldsymbol{\Sigma}|^{-\frac{n_1+2p+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_1^{-1}\boldsymbol{\Sigma}^{-1}\right)\right\} \times g(\boldsymbol{\Sigma}),$$
$$g(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_{\varepsilon\varepsilon}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\alpha}_1 + \boldsymbol{y}_n'\mathbf{V}_n^{-1/2}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{V}_n^{-1/2}\boldsymbol{y}_n\right)\right\},$$

where $n_1 = n_0 + n - 1$, $\mathbf{R}_1^{-1} = \mathbf{R}_0^{-1} + \sum_{t=1}^{n-1} \boldsymbol{v}_t\boldsymbol{v}_t'$ and

$$\boldsymbol{v}_t = \begin{pmatrix} \mathbf{V}_t^{-1/2}\boldsymbol{y}_t \\ \boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t \end{pmatrix}.$$

Using the MH algorithm, we propose a candidate $\boldsymbol{\Sigma}^\dagger \sim \mathcal{IW}(n_1, \mathbf{R}_1)$ and accept it with probability $\min\{g(\boldsymbol{\Sigma}^\dagger)/g(\boldsymbol{\Sigma}), 1\}$ where $\boldsymbol{\Sigma}$ is a current sample.

Generation of $\boldsymbol{\phi}$. Let $\boldsymbol{\Sigma}^{ij}$ be a $p \times p$ matrix, and denote the (i, j)-th block of $\boldsymbol{\Sigma}^{-1}$. Furthermore, let $\mathbf{A} = \sum_{t=1}^{n-1} \boldsymbol{\alpha}_t\boldsymbol{\alpha}_t'$, $\mathbf{B} = \sum_{t=1}^{n-1}\{\boldsymbol{\alpha}_t\boldsymbol{y}_t'\mathbf{V}_t^{-1/2}\boldsymbol{\Sigma}^{12} +$

$\boldsymbol{\alpha}_t\boldsymbol{\alpha}_{t+1}'\boldsymbol{\Sigma}^{22}\}$, and $\boldsymbol{b}$ denote a vector for which the i-th element is equal to the (i,i)-th element of $\mathbf{B}$. Then the conditional posterior probability density function of $\boldsymbol{\phi}$ is

$$\begin{aligned}\pi(\boldsymbol{\phi}|\boldsymbol{\Sigma},\boldsymbol{\alpha},Y_n) &\propto h(\boldsymbol{\phi})\times\exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{\Sigma}^{22}\boldsymbol{\Phi}\mathbf{A})-2\mathrm{tr}(\boldsymbol{\Phi}\mathbf{B})\right\}\\ &\propto h(\boldsymbol{\phi})\times\exp\left\{-\frac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\mu}_{\boldsymbol{\phi}})'\boldsymbol{\Sigma}_{\boldsymbol{\phi}}(\boldsymbol{\phi}-\boldsymbol{\mu}_{\boldsymbol{\phi}})\right\},\\ h(\boldsymbol{\phi}) &= |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}}\prod_{j=1}^{p}(1+\phi_j)^{a_j-1}(1-\phi_j)^{b_j-1}\exp\left\{-\frac{1}{2}\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\alpha}_1\right\},\end{aligned}$$

where $\boldsymbol{\mu}_{\boldsymbol{\phi}}=\boldsymbol{\Sigma}_{\boldsymbol{\phi}}\boldsymbol{b}$, $\boldsymbol{\Sigma}_{\boldsymbol{\phi}}^{-1}=\boldsymbol{\Sigma}^{22}\odot\mathbf{A}$, and $\odot$ denotes a Hadamard product. To sample $\boldsymbol{\phi}$ based on its conditional posterior distribution using the MH algorithm, we generate a candidate from a truncated normal distribution over the region R, $\boldsymbol{\phi}^\dagger\sim\mathcal{TN}_R(\boldsymbol{\mu}_{\boldsymbol{\phi}},\boldsymbol{\Sigma}_{\boldsymbol{\phi}})$, and $R=\{\boldsymbol{\phi}:|\phi_j|<1, j=1,\ldots,p\}$ and accept it with probability $\min\{h(\boldsymbol{\phi}^\dagger)/h(\boldsymbol{\phi}),1\}$ where $\boldsymbol{\phi}$ is a current sample.

2.3. The associated particle filter

This subsection describes the auxiliary particle filter (Pitt and Shephard (1999)) used to compute the likelihood function ordinate given the parameter $\boldsymbol{\theta}$, which can be used for model comparison. For more detail on the auxiliary particle filter, see Doucet et al. (2001).

Let $f(\boldsymbol{\alpha}_t|Y_t,\boldsymbol{\theta})$ denote the conditional probability density function of $\boldsymbol{\alpha}_t$ given $(Y_t,\boldsymbol{\theta})$, and let $\hat{f}(\boldsymbol{\alpha}_t|Y_t,\boldsymbol{\theta})$ denote the corresponding discrete probability mass function that approximates $f(\boldsymbol{\alpha}_t|Y_t,\boldsymbol{\theta})$. We consider sampling from the conditional joint distribution of $(\boldsymbol{\alpha}_{t+1},\boldsymbol{\alpha}_t)$ given $(Y_{t+1},\boldsymbol{\theta})$, with a probability density function given by

$$f(\boldsymbol{\alpha}_{t+1},\boldsymbol{\alpha}_t|Y_{t+1},\boldsymbol{\theta}) \propto f(\boldsymbol{y}_{t+1}|\boldsymbol{\alpha}_{t+1})f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{y}_t,\boldsymbol{\alpha}_t,\boldsymbol{\theta})f(\boldsymbol{\alpha}_t|Y_t,\boldsymbol{\theta}), \quad (24)$$

where

$$\begin{aligned}&f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t)=\\ &\quad(2\pi)^{-p/2}\,|\mathbf{V}_t^{1/2}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\mathbf{V}_t^{1/2}|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{y}_t'\mathbf{V}_t^{-1/2}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{V}_t^{-1/2}\boldsymbol{y}_t\right\}, \quad (25)\\ &f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{y}_t,\boldsymbol{\alpha}_t,\boldsymbol{\theta})=\\ &\quad(2\pi)^{-p/2}\,|\boldsymbol{\Sigma}_\alpha|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\alpha}_{t+1}-\boldsymbol{\mu}_{\alpha,t+1})'\boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\alpha}_{t+1}-\boldsymbol{\mu}_{\alpha,t+1})\right\}, (26)\end{aligned}$$

and

$$\boldsymbol{\mu}_{\alpha,t+1} = \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\Sigma}_{\eta\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{V}_t^{-1/2}\boldsymbol{y}_t, \quad \boldsymbol{\Sigma}_{\alpha} = \boldsymbol{\Sigma}_{\eta\eta} - \boldsymbol{\Sigma}_{\eta\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}. \tag{27}$$

We use the importance probability density function given by

$$\begin{aligned} g(\boldsymbol{\alpha}_{t+1}, \boldsymbol{\alpha}_t^i | Y_{t+1}, \boldsymbol{\theta}) &\propto f(\boldsymbol{y}_{t+1}|\boldsymbol{\mu}_{\alpha,t+1}^i) f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{y}_t, \boldsymbol{\alpha}_t^i, \boldsymbol{\theta})\hat{f}(\boldsymbol{\alpha}_t^i|Y_t, \boldsymbol{\theta}) \\ &\propto f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{y}_t, \boldsymbol{\alpha}_t^i, \boldsymbol{\theta}) g(\boldsymbol{\alpha}_t^i|Y_{t+1}, \boldsymbol{\theta}), \end{aligned}$$

where

$$\begin{aligned} g(\boldsymbol{\alpha}_t^i|Y_{t+1}, \boldsymbol{\theta}) &= \frac{f(\boldsymbol{y}_{t+1}|\boldsymbol{\mu}_{\alpha,t+1}^i)\hat{f}(\boldsymbol{\alpha}_t^i|Y_t, \boldsymbol{\theta})}{\sum_{i=1}^I f(\boldsymbol{y}_{t+1}|\boldsymbol{\mu}_{\alpha,t+1}^i)\hat{f}(\boldsymbol{\alpha}_t^i|Y_t, \boldsymbol{\theta})}, \\ \boldsymbol{\mu}_{\alpha,t+1}^i &= \boldsymbol{\Phi}\boldsymbol{\alpha}_t^i + \boldsymbol{\Sigma}_{\eta\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{V}_t^{i\ -1/2}\boldsymbol{y}_t, \quad \mathbf{V}_t^i = \mathbf{V}_t|_{\boldsymbol{\alpha}_t=\boldsymbol{\alpha}_t^i}. \end{aligned}$$

We derive the auxiliary particle filter as follows:

Step 1. Initialize $t = 1$, and generate $\boldsymbol{\alpha}_1^i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$, $(i = 1, \ldots, I)$.
- (a) Compute $w_i = f(\boldsymbol{y}_1|\boldsymbol{\alpha}_1^i, \boldsymbol{\theta})$, and record $\bar{w}_1 = \frac{1}{I}\sum_{i=1}^I w_i$.
- (b) Let $\hat{f}(\boldsymbol{\alpha}_1^i|Y_1, \boldsymbol{\theta}) = \pi_1^i = w_i/\sum_{j=1}^I w_j$, $(i = 1, \ldots, I)$.

Step 2.
- (a) For each i, generate $\left(\boldsymbol{\alpha}_{t+1}^i, \boldsymbol{\alpha}_t^i\right)$ from $g(\boldsymbol{\alpha}_{t+1}, \boldsymbol{\alpha}_t^j|Y_{t+1}, \boldsymbol{\theta})$, $(i, j = 1, \ldots, I)$, as follows. First, resample $\boldsymbol{\alpha}_t^i$, $i = 1, \ldots, I$, with probability $g(\boldsymbol{\alpha}_t^j|Y_{t+1}, \boldsymbol{\theta})$, $j = 1, \ldots, I$. Then generate $\boldsymbol{\alpha}_{t+1}^i$, $i = 1, \ldots, I$, from the normal distribution with density function $f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{y}_t, \boldsymbol{\alpha}_t^i, \boldsymbol{\theta})$.
- (b) Compute

$$\begin{aligned} w_i &= \frac{f(\boldsymbol{y}_{t+1}|\boldsymbol{\alpha}_{t+1}^i) f(\boldsymbol{\alpha}_{t+1}^i|\boldsymbol{y}_t, \boldsymbol{\alpha}_t^i, \boldsymbol{\theta})\hat{f}(\boldsymbol{\alpha}_t^i|Y_t, \boldsymbol{\theta})}{g\left(\boldsymbol{\alpha}_{t+1}^i, \boldsymbol{\alpha}_t^i|Y_{t+1}, \boldsymbol{\theta}\right)} \\ &= \frac{f(\boldsymbol{y}_{t+1}|\boldsymbol{\alpha}_{t+1}^i)\hat{f}(\boldsymbol{\alpha}_t^i|Y_t, \boldsymbol{\theta})}{g(\boldsymbol{\alpha}_t^i|Y_{t+1}, \boldsymbol{\theta})}, \end{aligned}$$

  for $i = 1, \ldots, I$, and record $\overline{w}_t = \sum_{i=1}^I w_i/I$.
- (c) Let $\hat{f}(\boldsymbol{\alpha}_{t+1}^i|Y_{t+1}, \boldsymbol{\theta}) = \pi_{t+1}^i = w_i/\sum_{j=1}^I w_j$ $(i = 1, \ldots, I)$.

Step 3 Increase t by one and return to Step 2.

Then,

$$\sum_{t=1}^n \log \overline{w}_t \xrightarrow{p} \sum_{t=1}^n \log f(y_t|Y_{t-1}, \boldsymbol{\theta}), \quad \text{as} \quad I \to \infty,$$

is a consistent estimate of the conditional log-likelihood.

2.4. The extension of the MSV model with multivariate-t errors (MSVt model)

The MSV model can be extended to incorporate heavy-tailed errors in stock returns. Although jump components can also be introduced, they are not considered here for the reasons of simplicity. To describe the fat-tailed distributions, we consider the multivariate-t distribution, which is a scale mixture of normal distributions. Let $\mathcal{G}(a,b)$ denote a gamma distribution with mean a/b and variance a/b^2. Using a common scalar gamma random variable λ_t, the multivariate-t random variable with ν degrees of freedom is obtained as

$$\lambda_t^{-1/2}\boldsymbol{\varepsilon}_t, \quad \text{where} \quad \lambda_t \sim \mathcal{G}(\nu/2, \nu/2), \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon}). \tag{28}$$

Thus, we extend the MSV model to a MSV multivariate-t errors (MSVt) model, where the measurement equation is given by

$$\boldsymbol{y}_t = \lambda_t^{-1/2}\mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \tag{29}$$

for $t = 1, \ldots, n$. The prior distribution for ν is assumed to be $\nu \sim \mathcal{G}(m_0^\nu, S_0^\nu)$, and we let $\pi(\nu)$ denote its prior probability density function.

To implement MCMC simulation, we sample $(\boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{\Sigma})$ as in Section 2.2, but we replace $\boldsymbol{y}_t$ with $\lambda_t^{1/2}\boldsymbol{y}_t$. Thus, we focus on sampling from the conditional posterior distributions for the other parameters $(\nu, \boldsymbol{\lambda})$ where $\boldsymbol{\lambda} = \{\lambda_t\}_{t=1}^n$. Their conditional joint posterior probability density function is given by

$$\begin{aligned}
&\pi(\nu, \boldsymbol{\lambda}|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, Y_n) \\
&\quad \propto \pi(\nu)\left\{\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\right\}^n \prod_{t=1}^n \lambda_t^{\frac{p+\nu}{2}-1} \\
&\qquad \times \exp\left[-\frac{1}{2}\sum_{t=1}^n\left\{\nu\lambda_t + \left(\sqrt{\lambda_t}\boldsymbol{y}_t - \boldsymbol{\mu}_t\right)'\boldsymbol{\Sigma}_t^{-1}\left(\sqrt{\lambda_t}\boldsymbol{y}_t - \boldsymbol{\mu}_t\right)\right\}\right].
\end{aligned} \tag{30}$$

To sample from the posterior distribution, we implement the MCMC simulation in three blocks.

1. Generate $(\boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{\Sigma})$ as in Section 2.2.1, replacing $\boldsymbol{y}_t$ with $\lambda_t^{1/2}\boldsymbol{y}_t$.
2. Generate $\nu \sim \pi(\nu|\boldsymbol{\lambda})$.

3. Generate $\lambda_t \sim \pi(\lambda_t|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}_t, \nu, \boldsymbol{y}_t)$ for $t = 1, \ldots, n$.

Generation of ν. The conditional posterior probability density of ν is given by

$$\pi(\nu|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, Y_n) \propto \pi(\nu) \left\{ \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \right\}^n \left(\prod_{t=1}^n \lambda_t \right)^{\frac{\nu}{2}} \exp\left\{ -\frac{\sum_{t=1}^n \lambda_t}{2} \nu \right\}. \quad (31)$$

To sample from this conditional posterior distribution, we transform ν such that $\vartheta_\nu = \log \nu$ and we let $\hat{\vartheta}_\nu$ denote a conditional mode of $\pi(\vartheta_\nu|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, Y_n)$. Using the MH algorithm, we propose a candidate from the normal distribution $\vartheta_\nu^\dagger \sim \mathcal{N}(\mu_\nu, \sigma_\nu^2)$, where

$$\mu_\nu = \hat{\vartheta}_\nu + \sigma_\nu^2 \left[\left. \frac{\partial \log \pi(\vartheta_\nu|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, Y_n)}{\partial \vartheta_\nu} \right|_{\vartheta_\nu = \hat{\vartheta}_\nu} \right],$$

and

$$\sigma_\nu^2 = \left[- \left. \frac{\partial^2 \log \pi(\vartheta_\nu|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, Y_n)}{\partial \vartheta_\nu^2} \right|_{\vartheta_\nu = \hat{\vartheta}_\nu} \right]^{-1}.$$

We accept the candidate with probability

$$\min \left[\frac{\pi(\vartheta_\nu^\dagger|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, Y_n) f_N(\vartheta_\nu|\mu_\nu, \sigma_\nu^2)}{\pi(\vartheta_\nu|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, Y_n) f_N(\vartheta_\nu^\dagger|\mu_\nu, \sigma_\nu^2)}, 1 \right],$$

where ϑ_ν is a current sample, and $f_N(x|\mu, \sigma^2)$ denotes a probability density function derived from a normal distribution with mean μ and variance σ^2.

Generation of $\boldsymbol{\lambda}$. The conditional posterior probability density function of λ_t is

$$\pi(\lambda_t|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \nu, \boldsymbol{\alpha}, Y_n) \propto \lambda_t^{\frac{\nu+p}{2}-1} \exp\left\{ -\frac{c_t}{2}\lambda_t + d_t\sqrt{\lambda_t} \right\},$$

where $c_t = \nu + \boldsymbol{y}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{y}_t$, and $d_t = \boldsymbol{y}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t$. To sample λ_t using the MH algorithm, we generate a candidate $\lambda_t^\dagger \sim \mathcal{G}((\nu+p)/2, c_t/2)$ and accept it with probability,

$$\min\left[1, \exp\left\{d_t\left(\sqrt{\lambda_t^\dagger} - \sqrt{\lambda_t}\right)\right\}\right],$$

where λ_t is a current sample. Note that we generate $\lambda_n \sim \mathcal{G}((\nu+p)/2, c_n/2)$, since $\boldsymbol{\mu}_n = \mathbf{0}$ implies $d_n = 0$.

3. Illustrative examples using simulated data

This section illustrates our proposed method using simulated data and shows the efficiency of our proposed multi-move sampler in comparison with the single-move sampler. The data are simulated using the MSVt model presented in Section 2.4, with

$$\begin{gathered}\phi_i = 0.97, \quad \sigma_{i,\varepsilon\varepsilon} \equiv \sqrt{Var(\varepsilon_{it})} = 1.2, \quad \sigma_{i,\eta\eta} \equiv \sqrt{Var(\eta_{it})} = 0.2, \\ \rho_{i,\varepsilon\eta} \equiv Corr(\varepsilon_{it}, \eta_{it}) = -0.4, \quad \nu = 15, \quad i = 1, 2, \ldots, 5,\end{gathered}$$

which correspond to typical values for the parameters of the univariate SV models in past empirical studies. The negative value of $\rho_{i,\varepsilon\eta}$ implies the existence of leverage effects. For the correlations among ε_{it}s and η_{jt}s, we set similar values to those obtained in our empirical studies.

$$\begin{aligned}&\rho_{ij,\varepsilon\varepsilon} \equiv Corr(\varepsilon_{it}, \varepsilon_{jt}) = 0.6, \quad \rho_{ij,\eta\eta} \equiv Corr(\eta_{it}, \eta_{jt}) = 0.7, \\ &\rho_{ij,\varepsilon\eta} \equiv Corr(\varepsilon_{it}, \eta_{jt}) = -0.3, \quad \text{for } i \neq j,\end{aligned}$$

where the negative value of $\rho_{ij,\varepsilon\eta}$ indicates cross leverage effects. Using these parameters, we generated $n = 4,000$ observations with $p = 5$. For prior distributions, we assume

$$\frac{\phi_i + 1}{2} \sim \mathcal{B}(20, 1.5), \quad \boldsymbol{\Sigma} \sim \mathcal{IW}(10, (10\boldsymbol{\Sigma}^*)^{-1}), \quad \nu \sim \mathcal{G}(1, 0.05),$$

where $\boldsymbol{\Sigma}^*$ is a true covariance matrix so that $E(\boldsymbol{\Sigma}^{-1}) = \boldsymbol{\Sigma}^{*-1}$. The mean and standard deviation of the prior distribution of ϕ_j are set 0.86 and 0.11 respectively, while the mean and standard deviation of ν are set 20 and 20, respectively. Using the MCMC algorithm described in Section 2.2, we generated 120,000 samples using the multi-move sampler and 550,000 samples using the single-move sampler, discarding the first 20,000 and 50,000 samples as burn-in periods, respectively.

Estimation results. Tables 1 and 2 show the estimation results using the multi-move sampler with tuning parameter $K = 200$ for $(\phi_i, \sigma_{i,\varepsilon\varepsilon}, \sigma_{i,\eta\eta}, \rho_{i,\varepsilon\eta}, \nu)$ and $(\rho_{ij,\varepsilon\varepsilon}, \rho_{ij,\eta\eta}, \rho_{ij,\varepsilon\eta})$, respectively. The results using the single-move sampler are omitted, except the inefficiency factors (shown in the brackets in Table 1), since they are similar to those obtained for the multi-move sampler. The posterior means and 95% credible intervals suggest that the estimates are sufficiently close to true values, which indicates that our proposed estimation algorithm works well.

The inefficiency factor is defined as $1+2\sum_{s=1}^{\infty}\rho_s$, where ρ_s is the sample autocorrelation at lag s; this factor is computed to measure how well the MCMC mixes (Chib (2001)). It is the ratio of the numerical variance of the posterior sample mean to the variance of the sample mean based on uncorrelated draws. When the inefficiency factor is equal to m, we must draw m times as many as the number of uncorrelated samples.

Table 1: Estimation results for $\phi_i, \sigma_{i,\varepsilon\varepsilon}, \sigma_{i,\eta\eta}$, $\rho_{i,\varepsilon\eta}$ and ν.
Posterior means, 95% credible intervals, and inefficiency factors.

	True	i	Mean	95% interval	Inefficiency multi	[single]
		1	0.965	[0.956, 0.973]	79	[786]
		2	0.973	[0.965, 0.980]	79	[624]
ϕ_i	0.97	3	0.971	[0.963, 0.978]	78	[298]
		4	0.964	[0.954, 0.972]	58	[563]
		5	0.969	[0.960, 0.977]	111	[449]
		1	1.191	[1.093, 1.297]	180	[4083]
		2	1.177	[1.068, 1.304]	236	[6732]
$\sigma_{i,\varepsilon\varepsilon}$	1.2	3	1.195	[1.089, 1.310]	210	[6327]
		4	1.261	[1.163, 1.370]	163	[2830]
		5	1.187	[1.081, 1.300]	198	[4450]
		1	0.207	[0.181, 0.235]	128	[827]
		2	0.192	[0.168, 0.218]	152	[788]
$\sigma_{i,\eta\eta}$	0.2	3	0.189	[0.167, 0.214]	136	[448]
		4	0.208	[0.182, 0.236]	132	[901]
		5	0.209	[0.183, 0.236]	181	[676]
		1	-0.412	[-0.503,-0.316]	58	[294]
		2	-0.400	[-0.496,-0.298]	62	[266]
$\rho_{i,\varepsilon\eta}$	-0.4	3	-0.335	[-0.438,-0.226]	62	[297]
		4	-0.377	[-0.476,-0.275]	42	[221]
		5	-0.396	[-0.490,-0.296]	81	[253]
ν	15		15.7	[13.0,19.1]	48	[151]

Table 2: Estimation results for $\rho_{ij,\varepsilon\varepsilon}$, $\rho_{ij,\eta\eta}$ and $\rho_{ij,\varepsilon\eta}$.
Posterior means, 95% credible intervals and inefficiency factors.

	True	ij	Mean	95% interval	Inefficiency multi	[single]
$\rho_{ij,\varepsilon\varepsilon}$	0.6	12	0.596	[0.574, 0.617]	3	[16]
		13	0.606	[0.585, 0.627]	2	[18]
		14	0.610	[0.589, 0.631]	2	[13]
		15	0.586	[0.564, 0.608]	2	[18]
		23	0.582	[0.559, 0.603]	3	[21]
		31	0.586	[0.564, 0.608]	3	[27]
		32	0.597	[0.575, 0.618]	3	[20]
		34	0.599	[0.577, 0.620]	3	[27]
		35	0.590	[0.567, 0.611]	3	[24]
		45	0.607	[0.585, 0.628]	3	[24]
$\rho_{ij,\eta\eta}$	0.7	12	0.693	[0.599, 0.776]	114	[473]
		13	0.677	[0.575, 0.765]	81	[724]
		14	0.692	[0.592, 0.777]	115	[726]
		15	0.727	[0.637, 0.800]	137	[686]
		23	0.676	[0.576, 0.760]	113	[794]
		31	0.645	[0.531, 0.738]	115	[676]
		32	0.648	[0.543, 0.739]	119	[611]
		34	0.659	[0.554, 0.749]	147	[882]
		35	0.749	[0.656, 0.825]	160	[937]
		45	0.635	[0.529, 0.729]	106	[527]
$\rho_{ij,\varepsilon\eta}$	-0.3	12	-0.282	[-0.389,-0.170]	60	[279]
		13	-0.242	[-0.351,-0.129]	68	[248]
		14	-0.297	[-0.399,-0.190]	48	[236]
		15	-0.253	[-0.357,-0.148]	57	[432]
		21	-0.275	[-0.370,-0.174]	55	[195]
		23	-0.273	[-0.377,-0.165]	66	[209]
		24	-0.281	[-0.382,-0.176]	41	[217]
		25	-0.229	[-0.330,-0.127]	74	[509]
		31	-0.316	[-0.412,-0.214]	43	[250]
		32	-0.264	[-0.372,-0.152]	74	[227]
		34	-0.262	[-0.365,-0.155]	37	[268]
		35	-0.237	[-0.343,-0.129]	52	[462]
		41	-0.337	[-0.434,-0.236]	49	[254]
		42	-0.368	[-0.467,-0.261]	57	[282]
		43	-0.289	[-0.395,-0.178]	68	[254]
		45	-0.237	[-0.340,-0.131]	61	[455]
		51	-0.382	[-0.477,-0.283]	47	[204]
		52	-0.352	[-0.457,-0.244]	67	[241]
		53	-0.372	[-0.481,-0.263]	90	[190]
		54	-0.280	[-0.386,-0.174]	47	[218]

As shown in Table 1, the inefficiency factors for the parameters obtained from the single-move sampler are much larger than those from the multi-move sampler. Especially, for $\sigma_{i,\varepsilon\varepsilon}$, they are about 20 to 30 times larger for the single-move sampler. This implies that our multi-move sampler is highly efficient, as expected.

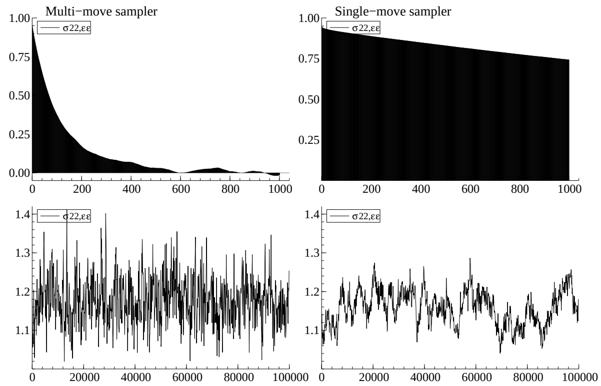

Figure 1: Sample autocorrelation functions and paths of $\sigma_{2,\varepsilon\varepsilon}$:
Single-move sampler versus Multi-move sampler.

Furthermore, for $\sigma_{2,\varepsilon\varepsilon}$, Figure 1 shows the sample autocorrelation functions and the sample paths for the single-move and the multi-move samplers. In contrast with the very slow decay of the sample autocorrelations in the single-move sampler, these autocorrelations vanish fairly rapidly under the multi-move sampler. Further, the sample path of the single-move sampler shows relatively slow movement in state space. This also indicates that the MCMC mixes well under the multi-move sampler.

Table 3: Maximum of inefficiency factors for $K = 50, 100, 140, 200$, and 270

K	$n/(K+1)$	ϕ_j	$\sigma_{i,\varepsilon\varepsilon}$	$\sigma_{i,\eta\eta}$	$\rho_{i,\varepsilon\eta}$	$\rho_{ij,\varepsilon\varepsilon}$	$\rho_{ij,\eta\eta}$	$\rho_{ij,\varepsilon\eta}$	ν
50	79	221	407	388	168	8	323	163	152
100	40	153	160	256	84	4	219	101	66
140	29	151	242	214	86	4	185	100	29
200	20	111	236	181	81	3	160	90	48
270	15	92	257	154	69	3	155	73	46
single-move	1	786	6732	901	297	27	937	509	151

Selection of a tuning parameter K. We set $K = 200$ in this example as follows. Table 3 shows the maxima of inefficiency factors for ϕ_js, $\sigma_{i,\varepsilon\varepsilon}$s, $\sigma_{i,\eta\eta}$s, $\rho_{i,\varepsilon\eta}$s, $\rho_{ij,\varepsilon\varepsilon}$s, $\rho_{ij,\eta\eta}$s, $\rho_{ij,\varepsilon\eta}$s, and ν using $K = 50, 100, 140, 200$, and 270. Since the maxima for $K = 200$ are overall smaller than those for other Ks, this is selected as an optimal value in our MCMC simulation. If this value is greater than 200 (*i.e.*, the number of elements in one block becomes small on an average), the Markov chain would not move quickly in state space due to high correlations among adjacent $\boldsymbol{\alpha}_t$s. However, if K is less than 200, the proposed states, namely, $\boldsymbol{\alpha}_t$s, would be rejected too often in the MH algorithm, which also results in the slowed mixing of the chain.

4. Empirical studies

4.1. Data

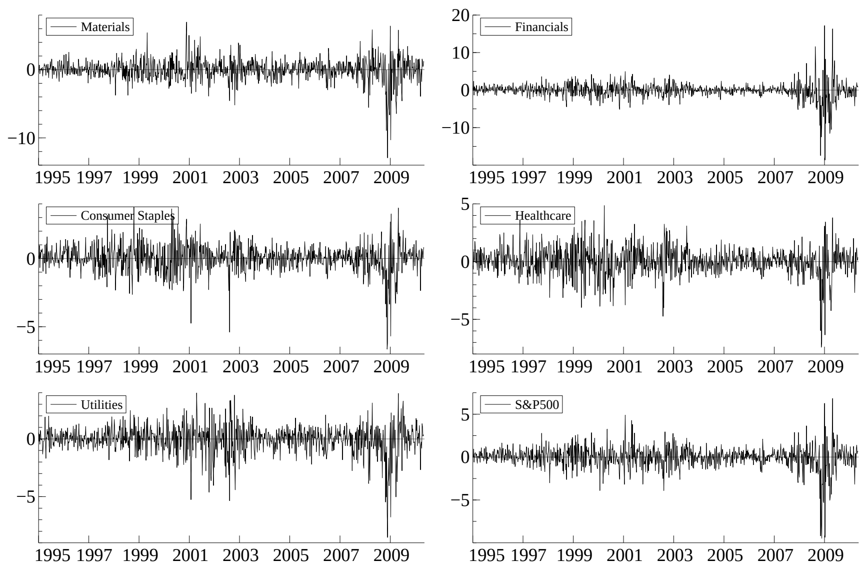

Figure 2: Returns of five S&P500 sector indices and the S&P500 index

This section applies our proposed MSVt model to S&P500 sector indices from January 2, 1995 to March 31, 2010. The dataset is obtained from Thomson Reuters Datastream. We excluded the annual market holidays and September 11-14, 2001, July 3, 2006, and January 2, 2007 since the same value is recorded on those days as the value from the previous day. The returns are defined by the log-difference of each sector index multiplied by 100 for five series, namely, 'Materials' (Series 1), 'Financials' (Series 2), 'Consumer Staples' (Series 3), 'Healthcare' (Series 4), and 'Utilities' (Series 5). There are 3,839 trading days, and the time series plot of the five return series with S&P500 index returns are shown in Figure 2, which indicates the co-movement of volatility among the five stock returns during this period.

Table 4: Univariate SV model with leverage and t-distributed errors.
Posterior means, standard deviations, and 95% credible intervals.

Param.	Series	Mean	Stdev	95% interval
	1	0.989	0.003	[0.983, 0.994]
	2	0.991	0.002	[0.987, 0.995]
ϕ_i	3	0.984	0.003	[0.977, 0.989]
	4	0.985	0.003	[0.978, 0.991]
	5	0.984	0.004	[0.977, 0.991]
	1	1.248	0.102	[1.067, 1.472]
	2	1.415	0.150	[1.142, 1.728]
$\sigma_{i,\varepsilon\varepsilon}$	3	0.871	0.055	[0.765, 0.983]
	4	1.051	0.074	[0.919, 1.210]
	5	0.976	0.077	[0.843, 1.146]
	1	0.124	0.012	[0.101, 0.150]
	2	0.143	0.012	[0.121, 0.167]
$\sigma_{i,\eta\eta}$	3	0.148	0.013	[0.124, 0.174]
	4	0.144	0.014	[0.118, 0.175]
	5	0.151	0.014	[0.126, 0.182]
	1	-0.527	0.065	[-0.644,-0.393]
	2	-0.684	0.049	[-0.771,-0.579]
$\rho_{i,\varepsilon\eta}$	3	-0.608	0.056	[-0.707,-0.489]
	4	-0.568	0.060	[-0.678,-0.443]
	5	-0.346	0.070	[-0.482,-0.208]
	1	22.7	11.3	[12.1, 55.9]
	2	25.0	15.4	[12.8, 69.8]
ν_i	3	25.5	12.6	[13.0, 59.0]
	4	16.5	5.2	[10.1, 28.7]
	5	46.5	27.6	[16.4,115.3]

4.2. The univariate SV model with leverage and t-distributed errors

First, we fit the univariate SV model with leverage and t-distributed errors to individual series as a benchmark for the MSVt model. The prior distributions for ϕ and $\mathbf{\Sigma}$ are assumed to be

$$\begin{aligned} \frac{\phi_i+1}{2} &\sim \mathcal{B}(20,1.5), \quad \nu_i \sim \mathcal{G}(0.01,0.01), \\ \mathbf{\Sigma}_i &\sim \mathcal{IW}(5,(5\mathbf{\Sigma}_i^*)^{-1}), \quad \mathbf{\Sigma}_i^* = \begin{pmatrix} 1 & -0.1 \\ -0.1 & 0.04 \end{pmatrix}, \end{aligned}$$

for $i = 1, \ldots, 5$, where ν_i is the degrees of freedom for the respective t-distribution. Since the posterior estimates of ν_i may be sensitive to the choice of the prior distribution (Nakajima and Omori (2009)), we use a relatively flat prior. To implement the MCMC algorithm, we draw 60,000 samples and discard 10,000 samples as a burn-in period. Posterior means, standard deviations, and 95% credible intervals are shown in Table 4.

The estimates of ϕ_i show the high persistence of log volatilities varying from 0.984 to 0.991; those negative values of ρ_i (ranging from -0.684 to -0.346) imply the credible existence of leverage effects for all series. Furthermore, the degrees of freedoms ν are relatively small (16.5 to 46.5) which indicate distributions with heavy-tails. These results are consistent with those found in previous empirical studies.

4.3. The MSVt model

For the MSVt model, the prior distributions are assumed to be

$$\begin{aligned} \frac{\phi_i+1}{2} &\sim \mathcal{B}(20,1.5), \quad i=1,\ldots,5, \\ \mathbf{\Sigma} &\sim \mathcal{IW}(10,(10\mathbf{\Sigma}^*)^{-1}), \quad \nu \sim \mathcal{G}(0.01,0.01), \end{aligned}$$

where

$$\begin{aligned} \mathbf{\Sigma}^* &= \begin{pmatrix} \mathbf{\Sigma}_{\varepsilon\varepsilon}^* & \mathbf{\Sigma}_{\varepsilon\eta}^* \\ \mathbf{\Sigma}_{\varepsilon\eta}^{*\prime} & \mathbf{\Sigma}_{\eta\eta}^* \end{pmatrix} \\ &= \begin{pmatrix} 1.2^2(0.5\mathbf{I}_5 + 0.5\mathbf{1}_5\mathbf{1}_5') & 1.2 \times 0.2 \times (-0.1)\mathbf{I}_5 \\ & 0.2^2(0.2\mathbf{I}_5 + 0.8\mathbf{1}_5\mathbf{1}_5') \end{pmatrix}, \end{aligned}$$

and $E(\mathbf{\Sigma}^{-1}) = \mathbf{\Sigma}^{*-1}$. Hyper-parameters of the prior distributions are chosen based on an analysis of univariate SV models. Using the MCMC algorithms described in Section 2, we draw 100,000 samples after discarding 20,000 samples as a burn-in period. The tuning parameter K is set to 100 based on an analysis similar to that presented in Table 3. This means that the average size of one block is about 40.

Table 5 shows posterior means, standard deviations, 95% credible intervals, and inefficiency factors for $\phi, \sigma_{i,\varepsilon\varepsilon}, \sigma_{i,\eta\eta}$, $\rho_{i,\varepsilon\eta}$, and ν. The estimated ϕ_is and $\sigma_{i,\varepsilon\varepsilon}$s are similar to those in the univariate model, while the estimated $\rho_{i,\varepsilon\eta}$ are smaller in absolute value. This may be because the variation of one series is partly explained by those of other series through the high correlations among ε_{it}s and η_{jt}s, as shown in Table 6.

The posterior mean of ν is similar to the least posterior mean (namely, 16.5 for ν_4) in the univariate model; this small degree of freedom indicates that the multivariate sector index returns follow a heavy-tailed error distribution. Furthermore, we note that posterior standard deviations are relatively smaller than those of univariate models.

Table 5: The MSVt model.
Posterior means, standard deviations, 95% credible intervals, and inefficiency factors for $\phi, \sigma_{i,\varepsilon\varepsilon}, \sigma_{i,\eta\eta}, \rho_{i,\varepsilon\eta}$, and ν

	i	Mean	Stdev	95% interval	Inefficiency
ϕ_i	1	0.987	0.002	[0.983, 0.991]	170
	2	0.990	0.002	[0.987, 0.993]	109
	3	0.985	0.002	[0.980, 0.989]	168
	4	0.985	0.003	[0.979, 0.989]	236
	5	0.987	0.002	[0.982, 0.991]	160
$\sigma_{i,\varepsilon\varepsilon}$	1	1.416	0.103	[1.239, 1.649]	450
	2	1.580	0.134	[1.351, 1.872]	555
	3	0.970	0.058	[0.871, 1.100]	404
	4	1.184	0.079	[1.045, 1.358]	402
	5	1.106	0.086	[0.962, 1.300]	439
$\sigma_{i,\eta\eta}$	1	0.138	0.010	[0.120, 0.158]	181
	2	0.151	0.009	[0.133, 0.170]	78
	3	0.147	0.009	[0.130, 0.166]	142
	4	0.155	0.011	[0.135, 0.177]	260
	5	0.149	0.010	[0.130, 0.169]	158
$\rho_{i,\varepsilon\eta}$	1	-0.400	0.058	[-0.509,-0.282]	90
	2	-0.507	0.051	[-0.601,-0.402]	83
	3	-0.475	0.050	[-0.569,-0.373]	60
	4	-0.425	0.054	[-0.526,-0.313]	90
	5	-0.298	0.061	[-0.414,-0.177]	78
ν		17.6	1.8	[14.5, 21.6]	75

Table 6: The MSVt model.

Posterior means, standard deviations, 95% credible intervals, and inefficiency factors for $\rho_{ij,\varepsilon\varepsilon}, \rho_{ij,\eta\eta}, \rho_{ij,\varepsilon\eta}$

	ij	Mean	Stdev	95% interval	Inefficiency
$\rho_{ij,\varepsilon\varepsilon}$	12	0.648	0.010	[0.628, 0.668]	3
	13	0.565	0.012	[0.542, 0.588]	2
	14	0.526	0.013	[0.500, 0.550]	3
	15	0.489	0.013	[0.463, 0.515]	3
	23	0.675	0.010	[0.657, 0.694]	3
	24	0.662	0.010	[0.642, 0.681]	2
	25	0.560	0.012	[0.537, 0.583]	3
	34	0.707	0.009	[0.689, 0.723]	2
	35	0.553	0.012	[0.529, 0.577]	3
	45	0.509	0.013	[0.483, 0.534]	2
$\rho_{ij,\eta\eta}$	12	0.829	0.034	[0.755, 0.886]	191
	13	0.802	0.036	[0.722, 0.862]	147
	14	0.773	0.041	[0.684, 0.845]	160
	15	0.792	0.041	[0.703, 0.861]	160
	23	0.814	0.033	[0.742, 0.872]	152
	24	0.781	0.040	[0.695, 0.851]	152
	25	0.717	0.050	[0.610, 0.806]	180
	34	0.878	0.024	[0.823, 0.919]	109
	35	0.785	0.041	[0.698, 0.855]	200
	45	0.742	0.048	[0.639, 0.823]	180
$\rho_{ij,\varepsilon\eta}$	12	-0.397	0.053	[-0.499,-0.289]	69
	13	-0.456	0.049	[-0.550,-0.358]	70
	14	-0.382	0.054	[-0.485,-0.275]	92
	15	-0.353	0.060	[-0.465,-0.229]	94
	21	-0.442	0.060	[-0.554,-0.316]	119
	23	-0.436	0.052	[-0.534,-0.332]	77
	24	-0.369	0.056	[-0.475,-0.256]	97
	25	-0.415	0.062	[-0.530,-0.287]	129
	31	-0.373	0.060	[-0.485,-0.250]	89
	32	-0.366	0.057	[-0.474,-0.250]	69
	34	-0.397	0.056	[-0.502,-0.284]	83
	35	-0.359	0.064	[-0.479,-0.228]	111
	41	-0.365	0.060	[-0.476,-0.240]	106
	42	-0.367	0.056	[-0.473,-0.252]	80
	43	-0.399	0.052	[-0.497,-0.294]	79
	45	-0.367	0.064	[-0.487,-0.238]	112
	51	-0.231	0.060	[-0.345,-0.111]	73
	52	-0.250	0.056	[-0.356,-0.137]	70
	53	-0.299	0.052	[-0.399,-0.195]	65
	54	-0.252	0.055	[-0.357,-0.139]	73

Tables 6 shows the posterior means, standard deviations, 95% credible intervals and inefficiency factors for $\rho_{ij,\varepsilon\varepsilon}$, $\rho_{ij,\eta\eta}$, and $\rho_{ij,\varepsilon\eta}$. Very high correlations are found among ε_{it}s (with $\rho_{ij,\varepsilon\varepsilon}$s ranging from 0.489 to 0.707), and among η_{it}s (with $\rho_{ij,\eta\eta}$s ranging from 0.717 to 0.878), which may cause the low absolute values of $\rho_{i,\epsilon\eta}$s as mentioned above.

The posterior means of $\rho_{ij,\varepsilon\eta}$ are all negative and vary from -0.456 to -0.231, and none of the 95% credible intervals for $\rho_{ij,\varepsilon\eta}$ include zero. This is strong evidence for the existence of cross asset leverage effects. Furthermore, we note that the cross leverage effects between series i and j are found to be asymmetric, *i.e.*, $\rho_{ij,\varepsilon\eta} \neq \rho_{ji,\varepsilon\eta}$ for $i \neq j$. For example, in Table 6, $\rho_{i5,\varepsilon\eta}$s vary from -0.415 to -0.353, while $\rho_{5j,\varepsilon\eta}$s vary from -0.299 to -0.231. This means the cross leverage effects from Series 1, 2, 3, and 4 on the volatility of Series 5 (i.e., 'Utilities') are relatively stronger than vice-versa. The volatility of the 'Utilities' series is more influenced by decreases in returns of four other series, while those of the other series are less subject to change in the returns of 'Utilities' series. This would suggest that market participants do not react so sharply when a decrease in the return is limited to 'Utilities' series, but they do react in a more sensitive way if decreases occur in other series.

4.4. Model comparison

Finally, we conduct a comparison of the MSV and MSVt models using the deviance information criterion (DIC) (Spiegelhalter et al. (2002)). The DIC is defined by

$$DIC = E_{\theta|y}[D(\theta)] + p_D,$$

where

$$p_D = E_{\theta|y}[D(\theta)] - D(E_{\theta|y}[\theta]), \quad D(\theta) = -2\log f(Y_n|\boldsymbol{\theta}).$$

To compute $E_{\theta|y}[D(\theta)]$, we use a sample analogue, $\frac{1}{M}\sum_{m=1}^{M} D(\theta^{(m)})$, where we set $M = 100$, and $\theta^{(m)}$s are resampled from the posterior distribution. The numerical standard error of the estimate is obtained by repeatedly estimating $E_{\theta|y}[D(\theta)]$ ten times. Regarding $D(E_{\theta|y}[\theta])$, which equals to $D(\theta)$ evaluated at the posterior mean, we implement an auxiliary particle filter to compute the likelihood ordinate $\log f(Y_n|\boldsymbol{\theta})$ as discussed in Section 2.3 and set the number of particles, $I = 10,000$. We repeat this ten times to obtain the numerical standard error.

Table 7 shows the sample means of DIC, the standard errors, and the smallest and the largest values among one hundred DIC values computed for two competing models. The DIC of the MSVt model much smaller than that of the MSV model, which indicates that the MSVt models outperforms the MSV model. This result also constitutes the evidence that the S&P500 sector index returns are heavy-tailed.

Table 7: Estimates, standard errors and the smallest and the largest values of DIC

Model	DIC	(s.e.)	$\text{DIC}_{\max}$	$\text{DIC}_{\min}$
MSV	48741.6	(1.4)	48749.9	48728.4
MSVt	48650.3	(2.6)	48664.8	48634.2

5. Conclusion

This paper proposes an efficient MCMC algorithm using a multi-move sampler for latent volatility vectors for MSV models with cross leverage and heavy-tailed errors. To sample a block of state vectors, we construct a proposal density function for the MH algorithm based on the approximate normal distribution using Taylor expansion of the logarithm of the target likelihood. We then exploit the sampling algorithms, which are developed for the linear and Gaussian state space models. We show that our proposed methods are easy to implement and that they are highly efficient. Extending to the model with respect to multivariate t-distributed errors is also discussed. Illustrative examples and empirical analyses are presented based on five sectors of S&P500 indices.

Acknowledgment

The authors are grateful to Erricos Kontoghiorghes, Guest editor, two anonymous referees, Siddhartha Chib, Mike K P So, and Boris Choy for helpful comments and discussions. This work is supported by the Research Fellowship (DC1) from the Japan Society for the Promotion of Science and 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)).

Appendix

A. Multi-move sampler for the MSV model

A.1. Derivation of the approximate state space model

First, noting that $E[\partial^2 L/\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}'_{t+k}] = \mathbf{0}$ $(k \geq 2)$, define $\mathbf{A}_t$ and $\mathbf{B}_t$ as

$$\begin{aligned}
\mathbf{A}_t &= -E\left[\frac{\partial^2 L}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}'_t}\right], \quad t = s+1, \ldots, s+m, & (32)\\
\mathbf{B}_t &= -E\left[\frac{\partial^2 L}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}'_{t-1}}\right], \quad t = s+2, \ldots, s+m, \quad \mathbf{B}_{s+1} = \mathbf{0}, & (33)
\end{aligned}$$

and let $\boldsymbol{d}_t = \partial L/\partial \boldsymbol{\alpha}_t$ for $t = s+1, \ldots, s+m$. $\boldsymbol{d}_t, \mathbf{A}_t$, and $\mathbf{B}_t$ are shown in Appendix A.2 for the MSV model. To obtain the approximating state space model, first evaluate $\boldsymbol{d}_t, \mathbf{A}_t$, and $\mathbf{B}_t$ at the current mode, with $\boldsymbol{\alpha}_t = \hat{\boldsymbol{\alpha}}_t$. Use $\hat{\boldsymbol{d}}_t, \hat{\mathbf{A}}_t$ and $\hat{\mathbf{B}}_t$ in the following computations.

1. Set $\boldsymbol{b}_s = \mathbf{0}$ and $\hat{\mathbf{B}}_{s+m+1} = \mathbf{0}$. Compute

   $$\mathbf{D}_t = \hat{\mathbf{A}}_t - \hat{\mathbf{B}}_t \mathbf{D}_{t-1}^{-1} \hat{\mathbf{B}}'_t, \quad \boldsymbol{b}_t = \hat{\boldsymbol{d}}_t - \hat{\mathbf{B}}_t \mathbf{D}_{t-1}^{-1} \boldsymbol{b}_{t-1}, \quad \hat{\boldsymbol{\gamma}}_t = \hat{\boldsymbol{\alpha}}_t + \mathbf{D}_t^{-1} \hat{\mathbf{B}}'_{t+1} \hat{\boldsymbol{\alpha}}_{t+1},$$

   for $t = s+1, \ldots, s+m$ recursively, where $\mathbf{K}_t$ denotes a Choleski decomposition of $\mathbf{D}_t$ such that $\mathbf{D}_t = \mathbf{K}_t \mathbf{K}'_t$.
2. Define auxiliary vectors and matrices

   $$\hat{\boldsymbol{y}}_t = \hat{\boldsymbol{\gamma}}_t + \mathbf{D}_t^{-1} \boldsymbol{b}_t, \quad \mathbf{Z}_t = \mathbf{I}_p + \mathbf{D}_t^{-1} \hat{\mathbf{B}}'_{t+1} \boldsymbol{\Phi}, \quad \mathbf{G}_t = [\mathbf{K}_t'^{-1}, \mathbf{D}_t^{-1} \hat{\mathbf{B}}'_{t+1} \mathbf{R}_t],$$

   for $t = s+1, \ldots, s+m$.

Then, derive the approximate linear Gaussian state space model given using (22) and (23).

A.2. $\boldsymbol{d}_t$, $\mathbf{A}_t$ and $\mathbf{B}_t$

Matrix differentiation

We first summarize definitions for the first and second derivatives of a matrix and some results (Magnus and Neudecker (1999), and Magnus and Abadir (2007)). Let F be a twice differentiable $m \times p$ matrix function of an $n \times q$ matrix $\boldsymbol{X}$. Then the first derivative (Jacobian matrix) of F at $\boldsymbol{X}$ is defined by the $mp \times nq$ matrix

$$\mathbf{D}F(\boldsymbol{X}) = \frac{\partial F(\boldsymbol{X})}{\partial \boldsymbol{X}} = \frac{\partial \mathrm{vec}(F(\boldsymbol{X}))}{\partial \mathrm{vec}(\boldsymbol{X})'},$$

and the second derivative (Hessian matrix) of F at $\boldsymbol{X}$ is defined by the $mnpq \times nq$ matrix

$$\mathbf{H}F(\boldsymbol{X}) = \mathbf{D}\left((\mathbf{D}F(X))'\right) = \frac{\partial}{\partial(\mathrm{vec}(\boldsymbol{X}))'}\mathrm{vec}\left(\left(\frac{\partial \mathrm{vec}(F(\boldsymbol{X}))}{\partial(\mathrm{vec}(\boldsymbol{X}))'}\right)'\right).$$

Chain rule: Let S a subset of $\mathbb{R}^{n\times q}$, and assume that $F : S \to \mathbb{R}^{m\times p}$ is differentiable at an interior point C of S. Let T be a subset of $\mathbb{R}^{m\times p}$ such that $F(X) \in T$ for all $X \in S$, and assume that $G : T \to \mathbb{R}^{r\times s}$ is differentiable at an interior point $B = F(C)$ of T. Then the composite function $H : S \to \mathbb{R}^{r\times s}$ defined by $H(X) = G(F(X))$ is differentiable at C, and

$$\mathbf{D}H(X) = (\mathbf{D}G(F(X)))(\mathbf{D}F(X)) = \frac{\partial \mathrm{vec}(G(F(\boldsymbol{X})))}{\partial(\mathrm{vec}(F(\boldsymbol{X})))'}\frac{\partial \mathrm{vec}(F(\boldsymbol{X}))}{\partial(\mathrm{vec}(\boldsymbol{X}))'}. \quad (34)$$

When $q = 1$, $x \in \mathbb{R}^{n\times 1}$, $f : \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times p}$, $g : \mathbb{R}^{m\times p} \to \mathbb{R}^{r\times s}$,

$$\frac{\partial g(f(\mathbf{x}))}{\partial \mathbf{x}'} = \frac{\partial \mathrm{vec}(g(f(\mathbf{x})))}{\partial \mathrm{vec}(f(\mathbf{x}))'}\frac{\partial \mathrm{vec}(f(\mathbf{x}))}{\partial \mathrm{vec}(\mathbf{x})'}. \quad (35)$$

Product rule: Let S a subset of $\mathbb{R}^{n\times q}$, and assume that $F : S \to \mathbb{R}^{m\times p}$ and $G : S \to \mathbb{R}^{p\times r}$ are differentiable at an interior point C of S. Then

$$\frac{\partial \mathrm{vec}(FG)}{\partial(\mathrm{vec}(X))'} = (G' \otimes \mathbf{I}_m)\frac{\partial \mathrm{vec}(F)}{\partial(\mathrm{vec}(X))'} + (\mathbf{I}_r \otimes F)\frac{\partial \mathrm{vec}(G)}{\partial(\mathrm{vec}(X))'}. \quad (36)$$

$\boldsymbol{d_t}$*,* $\mathbf{A}_t$*, and* $\mathbf{B}_t$

Let $\boldsymbol{z}_t = \mathbf{V}_t^{-1/2}\boldsymbol{y}_t$. Then, the logarithm of the conditional posterior probability density is given by

$$l_t = \mathrm{const} - \frac{1}{2}\mathbf{1}_p'\boldsymbol{\alpha}_t - \frac{1}{2}(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1}(\boldsymbol{z}_t - \boldsymbol{m}_t).$$

The gradient vector $\boldsymbol{d_t} = \partial L/\partial\boldsymbol{\alpha}_t$ is given by

$$\boldsymbol{d_t} = \frac{\partial l_t}{\partial\boldsymbol{\alpha}_t} + \frac{\partial l_{t-1}}{\partial\boldsymbol{\alpha}_t} + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t)I(t = s+m < n), \quad t = s+1,\ldots,s+m,$$

where

$$
\begin{aligned}
\frac{\partial l_t}{\partial \boldsymbol{\alpha}_t'} &= -\frac{1}{2}\mathbf{1}_p' - (\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1}\frac{\partial(\boldsymbol{z}_t - \boldsymbol{m}_t)}{\partial \boldsymbol{\alpha}_t'} \\
&= -\frac{1}{2}\mathbf{1}_p' + \frac{1}{2}(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1}\left\{\mathrm{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi} I(t<n)\right\}, \quad (37) \\
\frac{\partial l_{t-1}}{\partial \boldsymbol{\alpha}_t'} &= (\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1})'\mathbf{S}_{t-1}^{-1}\frac{\partial \boldsymbol{m}_{t-1}}{\partial \boldsymbol{\alpha}_t'} \\
&= (\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1})'\mathbf{S}_{t-1}^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1} I(t>1), \quad (38)
\end{aligned}
$$

using the chain rule (35). Thus

$$
\begin{aligned}
\boldsymbol{d_t} &= -\frac{1}{2}\mathbf{1}_p + \frac{1}{2}\left\{\mathrm{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon} I(t<n)\right\}\mathbf{S}_t^{-1}(\boldsymbol{z}_t - \boldsymbol{m}_t) \\
&\quad + \boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_{t-1}^{-1}(\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1}) I(t>1) \\
&\quad + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t) I(t=s+m<n),
\end{aligned}
$$

for $t = s+1, \ldots, s+m$.

To compute $\mathbf{A}_t$ and $\mathbf{B}_t$ using the product rule (36), we first obtain the Hessian matrix

$$
\frac{\partial^2 L}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}_t'} = \frac{\partial^2 l_t}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}_t'} + \frac{\partial^2 l_{t-1}}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}_t'} - \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi} I(t=s+m<n),
$$

for $t = s+1, \ldots, s+m$ where

$$
\begin{aligned}
&\frac{\partial^2 l_t}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}_t'} = \\
&\frac{1}{2}\left\{(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1} \otimes \mathbf{I}_p\right\}\frac{\partial \mathrm{vec}(\mathrm{diag}(\boldsymbol{z}_t))}{\partial \boldsymbol{\alpha}_t'} \\
&-\frac{1}{4}\left(\mathrm{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon} I(t<n)\right)\mathbf{S}_t^{-1}\left(\mathrm{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi} I(t<n)\right), \quad (39)
\end{aligned}
$$

and

$$
\frac{\partial^2 l_{t-1}}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}_t'} = -\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_{t-1}^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1} I(t>1). \quad (40)
$$

Noting that

$$\frac{\partial \text{vec}(\text{diag}(\boldsymbol{z}_t))}{\partial \boldsymbol{\alpha}_t'} = -\frac{1}{2}\begin{bmatrix} z_{1t}\boldsymbol{e}_1\boldsymbol{e}_1' \\ \vdots \\ z_{pt}\boldsymbol{e}_p\boldsymbol{e}_p' \end{bmatrix}, \quad E\left[z_{jt}(\boldsymbol{z}_t - \boldsymbol{m}_t)\right] = \mathbf{S}_t\boldsymbol{e}_j, \tag{41}$$

where $\boldsymbol{e}_j$ is a $p \times 1$ unit vector with j-th component equal to 1, the expected value of the first term in (39) is

$$\begin{aligned} E\left[\left\{(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1} \otimes \mathbf{I}_p\right\} \frac{\partial \text{vec}(\text{diag}(\boldsymbol{z}_t))}{\partial \boldsymbol{\alpha}_t'}\right] &= -\frac{1}{2}\sum_{j=1}^{p}(\boldsymbol{e}_j'\mathbf{S}_t^{-1})(\mathbf{S}_t\boldsymbol{e}_j)\boldsymbol{e}_j\boldsymbol{e}_j' \\ &= -\frac{1}{2}\mathbf{I}_p, \end{aligned} \tag{42}$$

Further, using $\text{diag}(\boldsymbol{z}_t)\mathbf{S}_t^{-1}\text{diag}(\boldsymbol{z}_t) = \mathbf{S}_t^{-1} \odot (\boldsymbol{z}_t\boldsymbol{z}_t')$, the expected value of the second term in (39) is

$$\begin{aligned} &E\left[\left(\text{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}I(t<n)\right)\mathbf{S}_t^{-1}\left(\text{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t<n)\right)\right] \\ &= \mathbf{S}_t^{-1} \odot (\mathbf{S}_t + \boldsymbol{m}_t\boldsymbol{m}_t') + 4\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t<n) \qquad (43) \\ &\quad -2\left(\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\text{diag}(\boldsymbol{m}_t) + \text{diag}(\boldsymbol{m}_t)\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}\right)I(t<n) \end{aligned}$$

Thus, we obtain

$$\begin{aligned} \mathbf{A}_t &= \frac{1}{4}\left\{\mathbf{I}_p + \mathbf{S}_t^{-1} \odot (\mathbf{S}_t + \boldsymbol{m}_t\boldsymbol{m}_t')\right\} + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t<n) \\ &\quad -\frac{1}{2}\left(\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\text{diag}(\boldsymbol{m}_t) + \text{diag}(\boldsymbol{m}_t)\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}\right)I(t<n) \\ &\quad +\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_{t-1}^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}I(t>1) + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t=s+m<n), \end{aligned} \tag{44}$$

for $t = s+1, \ldots, s+m$. Similarly, it is straightforward to show that

$$\mathbf{B}_t = -E\left[\frac{\partial^2 l_{t-1}}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}_{t-1}'}\right] = \frac{1}{2}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_{t-1}^{-1}\left\{\text{diag}(\boldsymbol{m}_{t-1}) - 2\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}\right\}, \tag{45}$$

for $t = s+2, \ldots, s+m$.

References

Asai, M., McAleer, M., 2006. Asymmetric multivariate stochastic volatility. Econometric Reviews 25, 453–473.

Asai, M., McAleer, M., Yu, J., 2006. Multivariate stochastic volatility: A review. Econometric Reviews 25, 145–175.

Broto, C., Ruiz, E., 2004. Estimation methods for stochastic volatility models: a survey. Journal of Economic Survey 18, 613–649.

Chan, D., Kohn, R., Kirby, C., 2006. Multivariate stochastic volatility models with correlated errors. Econometric Reviews 25, 245–274.

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

Chib, S., Nardari, F., Shephard, N., 2006. Analysis of high dimensional multivariate stochastic volatility models. Journal of Econometrics 134, 341–371.

Chib, S., Omori, Y., Asai, M., 2009. Multivariate stochastic volatility. In: Andersen, T. G., Davis, R. A., Kreiss, J. P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer-Verlag, New York, pp. 365–400.

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

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

Doucet, A., de Freitas, N., Gordon, N. J. (Eds.), 2001. Sequential Monte Carlo Methods in Practice. Springer-Verlag, New York.

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

Jacquier, E., Polson, N. G., Rossi, P. E., 1999. Stochastic volatility: Univariate and multivariate extensions, cIRANO Working paper 99s–26, Montreal.

Koopman, S. J., 1993. Disturbance smoother for state space models. Biometrika 80, 117–126.

Liesenfeld, R., Richard, J.-F., 2003. Univariate and multivariate stochastic volatility models: Estimation and diagnostics. Journal of Empirical Finance 10, 505–531.

Lopes, H. F., Carvalho, C. M., 2007. Factor stochastic volatility with time varying loadings and Markov switching regimes. Journal of Statistical Planning and Inference 137, 3082–3091.

Magnus, J. R., Abadir, K. M., 2007. On some definitions in matrix algebra, discussion paper, CIRJE-F-476, Faculty of Economics, University of Tokyo.

Magnus, J. R., Neudecker, H., 1999. Matrix differential culculus with applications in statistics and econometrics, revised edition. John Wiley, Chichester.

Nakajima, J., Omori, Y., 2009. Leverage, heavy-tails and correlated jumps in stochastic volatility models. Computational Statistics and Data Analysis 53-6, 2335–2353.

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

Omori, Y., Watanabe, T., 2008. Block sampler and posterior mode estimation for asymmetric stochastic volatility models. Computational Statistics and Data Analysis 52-6, 2892–2910.

Pitt, M. K., Shephard, N., 1999. Time varying covariances: a factor stochastic volatility approach. In: Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. M. (Eds.), Bayesian Statistics. Vol. 6. Oxford University Press, Oxford, pp. 547–570.

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

So, M. K. P., Choi, C. Y., 2009. A threshold factor multivariate stochastic volatility model. Journal of Forecasting 28-8, 712–735.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society, Series B 64, 583–639.

Takahashi, M., Omori, Y., Watanabe, T., 2009. Estimating stochastic volatility models using daily returns and realized volatility simultaneously. Computational Statistics and Data Analysis 53-6, 2404–2426.