CARF Working Paper

CARF-F-198

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

Tsunehiro Ishihara
Yasuhiro Omori
The University of Tokyo

December 2009

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

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

Tsunehiro Ishihara* and Yasuhiro Omori †

December, 2009

Abstract

The efficient Bayesian estimation method using Markov chain Monte Carlo is proposed for a multivariate stochastic volatility model that is a natural extension of the univariate stochastic volatility model with leverage and heavy-tailed errors, where we further incorporate cross leverage effects among stock returns. Our method is based on a multi-move sampler which samples a block of latent volatility vectors and is described first in the literature for 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 which samples one latent volatility vector at a time given other latent vectors and parameters. The empirical studies are given using five dimensional stock return indices in Tokyo Stock Exchange.

Key words: Asymmetry, Bayesian analysis, Heavy-tailed error, Leverage effect, Markov chain Monte Carlo, Multi-move sampler, Multivariate stochastic volatility, Stock returns.

*Graduate School of Economics, University of Tokyo. 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan.

†Faculty of Economics, University of Tokyo. 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan. Tel: +81-3-5841-5516. Email: omori@e.u-tokyo.ac.jp

1 Introduction

The univariate stochastic volatility (SV) models have been well known and successful to account for the time-varying variance in financial time series (e.g. Broto and Ruiz (2004)). Efficient Bayesian estimation methods are proposed using Markov chain Monte Carlo (MCMC) methods since the likelihood functions are difficult to evaluate in the implementation of the maximum likelihood estimation (e.g. Shephard and Pitt (1997), Omori, Chib, Shephard, and Nakajima (2007)).

Extending these models to the multivariate SV (MSV) model has become recently a major concern to investigate the correlation structure of multivariate financial time series for the purpose of the portfolio optimisation, the risk management, and the derivative pricing. Multivariate factor modelling of stochastic volatilities has been widely introduced to describe the complex dynamic structure of the high dimensional stock returns data (Jacquier, Polson, and Rossi (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, Nardari, and Shephard (2006)). On the other hand, efficient estimation methods for MSV models with cross leverage (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, j) or asymmetry have not been well investigated in the literature except for simple bivariate models (see surveys by Asai, McAleer, and Yu (2006) and Chib, Omori, and Asai (2009)). Chan, Kohn, and Kirby (2006) considered the Bayesian estimation of MSV models with correlations between measurement errors and state errors, but their setup did not exactly correspond to the leverage effects. Asai and McAleer (2006) simplified the MSV model with leverage by assuming no cross leverage effects (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 describe the 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 efficient MCMC algorithm using a multi-move sampler which samples a block of many 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 the MCMC implementation for the SV models, it is critical to sample the latent volatility (or state) variables from their full conditional posterior distributions efficiently. The single-move sampler that draws a single volatility variable at a time given the rest of the volatility variables and other parameters is easy to implement, but obtained MCMC samples are known to have high autocorrelations. This implies we need to iterate the MCMC algorithm a huge number of times to obtain accurate estimates when we use 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 the state variables, the mode can be obtained easily 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 by Metropolis-Hastings (MH) algorithm in which a simulation smoother for the linear and Gaussian state space model (de Jong and Shephard (1995), Durbin and Koopman (2002)) is used to generate a candidate.

The rest of the paper is organised as follows. Section 2 discusses a Bayesian estimation of the MSV model using a multi-move sampler for the latent state variables. Extension to the model 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 regarding the sampling efficiencies. Section 4 gives empirical studies using five dimensional stock return indices. Section 5 concludes the paper.

2 MSV model with cross leverage and heavy-tailed errors

2.1 Basic MSV Model

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

$$y_t = \exp(\alpha_t/2)\varepsilon_t, \quad t = 1, \ldots, n, \tag{1}$$

$$\alpha_{t+1} = \phi\alpha_t + \eta_t, \quad t = 1, \ldots, n-1, \tag{2}$$

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

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 it to the MSV model, we let $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{pt})$' denote a p dimensional stock returns vector and $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{pt})'$ denote their corresponding log volatility vectors, respectively. We consider the MSV model given by

$$\begin{aligned} \boldsymbol{y}_t &= \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, n, & (5)\\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \quad t = 1, \ldots, n-1, & (6)\\ & \boldsymbol{\alpha}_1 \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_0), & (7) \end{aligned}$$

where

$$\begin{aligned} \mathbf{V}_t &= \operatorname{diag}(\exp(\alpha_{1t}), \ldots, \exp(\alpha_{pt})), & (8)\\ \boldsymbol{\Phi} &= \operatorname{diag}(\phi_1, \ldots, \phi_p), & (9)\\ \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} &\sim \mathcal{N}_{2p}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & \boldsymbol{\Sigma}_{\varepsilon\eta} \\ \boldsymbol{\Sigma}_{\eta\varepsilon} & \boldsymbol{\Sigma}_{\eta\eta} \end{pmatrix}. & (10) \end{aligned}$$

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 a stationarity condition $\boldsymbol{\Sigma}_0 = \boldsymbol{\Phi}\boldsymbol{\Sigma}_0\boldsymbol{\Phi} + \boldsymbol{\Sigma}_{\eta\eta}$ such that

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

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

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

where

$$\begin{aligned} l_t &= \text{const} - \frac{1}{2}\mathbf{1}_p'\boldsymbol{\alpha}_t - \frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{\mu}_t)'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_t), & (12)\\ & \boldsymbol{\mu}_t = \mathbf{V}_t^{1/2}\boldsymbol{m}_t, \quad \boldsymbol{\Sigma}_t = \mathbf{V}_t^{1/2}\mathbf{S}_t\mathbf{V}_t^{1/2}, & (13) \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{m}_t &= \begin{cases} \boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\boldsymbol{\alpha}_{t+1} - \Phi\boldsymbol{\alpha}_t), & t < n, \\ \mathbf{0} & t = n, \end{cases} & (14)\\ \mathbf{S}_t &= \begin{cases} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} - \boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}, & t < n, \\ \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & t = n. \end{cases} & (15) \end{aligned}$$

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 to evaluate the likelihood function of $\boldsymbol{\theta}$ analytically or using a high dimensional numerical integration. In this paper, by taking a Bayesian approach, we employ a simulation method, the MCMC method, to generate samples from the posterior distribution to conduct a statistical inference regarding 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 $\mathcal{IW}(n_0, \mathbf{R}_0)$ 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

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

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 blocks:

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 to sample $\boldsymbol{\alpha}$ from its conditional posterior distribution in Step 1. One is a so-called single-move sampler which samples one $\boldsymbol{\alpha}_t$ at a time given other $\boldsymbol{\alpha}_j$'s, while the other method is a multi-move sampler which samples a block of state vectors, say, $(\boldsymbol{\alpha}_t, \ldots, \boldsymbol{\alpha}_{t+k})$ given the rest of state vectors.

Generation of $\boldsymbol{\alpha}$

Single-move sampler. A simple but inefficient method is to sample one $\boldsymbol{\alpha}_t$ at a time given other $\boldsymbol{\alpha}_j$'s and 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

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

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

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

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, to sample from the conditional posterior distribution using 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 $t=1,\ldots,n$.

Multi-move sampler. As an alternative method, we propose an efficient block sampler for $\boldsymbol{\alpha}$ to sample a block of $\boldsymbol{\alpha}_t$'s from the posterior distribution extending Omori and Watanabe (2008) who considered the univariate SV model with leverage (see also Takahashi, Omori, and Watanabe (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 $k_i-k_{i-1}\geq 2$. The K knots $(k_1,\ldots,k_K)$ are generated randomly using

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

where U_i's are independent uniform random variable on $(0,1)$ (see e.g., Shephard and Pitt (1997)). These stochastic knots have an advantage to allow the points of conditioning to change over the MCMC iterations where K is a tuning parameter to obtain less autocorrelated MCMC samples.

Suppose that $k_{i-1}=s$ and $k_i=s+m$ for the i-th block and consider sampling this block from its conditional posterior distribution given other state vectors and parameters. Let $\boldsymbol{\xi}_t=\mathbf{R}_t^{-1}\boldsymbol{\eta}_t$, where the matrix $\mathbf{R}_t$ denotes a Choleski decomposition of $\boldsymbol{\Sigma}_{\eta\eta}=\mathbf{R}_t\mathbf{R}_t'$ for

$t = s, s+1, \ldots, s+m$, and $\boldsymbol{\Sigma}_0 = \mathbf{R}_0\mathbf{R}_0'$ for $t = s = 0$. To construct a proposal distribution for MH algorithm, we focus on the distribution of the disturbance $\underline{\boldsymbol{\xi}} \equiv (\boldsymbol{\xi}_s', \ldots, \boldsymbol{\xi}_{s+m-1}')'$ which is fundamental in the sense that it derives the distribution of $\underline{\boldsymbol{\alpha}} \equiv (\boldsymbol{\alpha}_{s+1}', \ldots, \boldsymbol{\alpha}_{s+m}')'$. Then, the logarithm of the full conditional joint density distribution of $\underline{\boldsymbol{\xi}}$ excluding constant terms is given by

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

$$\text{where} \quad L = \sum_{t=s}^{s+m} l_s - \frac{1}{2}(\boldsymbol{\alpha}_{s+m+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_{s+m})'\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\boldsymbol{\alpha}_{s+m+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_{s+m})I(s+m<n). \tag{20}$$

Then using the second order Taylor expansion of (19) around the mode $\hat{\underline{\boldsymbol{\xi}}}$, we obtain approximating normal density f^* to be used for the MH algorithm as follows

$$\begin{aligned}
& \log f(\underline{\boldsymbol{\xi}}|\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+m+1}, \boldsymbol{y}_s, \ldots, \boldsymbol{y}_{s+m}) \\
\approx\ & \text{const.} - \frac{1}{2}\sum_{t=s}^{s+m-1} \boldsymbol{\xi}_t'\boldsymbol{\xi}_t + \hat{L} + \left.\frac{\partial L}{\partial \underline{\boldsymbol{\xi}}'}\right|_{\underline{\boldsymbol{\xi}}=\hat{\underline{\boldsymbol{\xi}}}}(\underline{\boldsymbol{\xi}} - \hat{\underline{\boldsymbol{\xi}}}) + \frac{1}{2}(\underline{\boldsymbol{\xi}} - \hat{\underline{\boldsymbol{\xi}}})'\mathbb{E}\left(\frac{\partial^2 L}{\partial\underline{\boldsymbol{\xi}}\partial\underline{\boldsymbol{\xi}}'}\right)(\underline{\boldsymbol{\xi}} - \hat{\underline{\boldsymbol{\xi}}}) \\
=\ & \text{const.} - \frac{1}{2}\sum_{t=s}^{s+m-1} \boldsymbol{\xi}_t'\boldsymbol{\xi}_t + \hat{L} + \hat{\boldsymbol{d}}'(\underline{\boldsymbol{\alpha}} - \hat{\underline{\boldsymbol{\alpha}}}) - \frac{1}{2}(\underline{\boldsymbol{\alpha}} - \hat{\underline{\boldsymbol{\alpha}}})'\hat{\mathbf{Q}}(\underline{\boldsymbol{\alpha}} - \hat{\underline{\boldsymbol{\alpha}}}), \qquad (21) \\
=\ & \text{const.} + \log f^*(\underline{\boldsymbol{\xi}}|\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+m+1}, \boldsymbol{y}_s, \ldots, \boldsymbol{y}_{s+m}) \qquad (22)
\end{aligned}$$

where $\hat{\mathbf{Q}}$ and $\hat{\boldsymbol{d}}$ are $\mathbf{Q} = -E(\partial^2 L/\partial\underline{\boldsymbol{\alpha}}\partial\underline{\boldsymbol{\alpha}}')$ and $\boldsymbol{d} = \partial L/\partial\underline{\boldsymbol{\alpha}}$ evaluated at $\underline{\boldsymbol{\alpha}} = \hat{\underline{\boldsymbol{\alpha}}}$ (*i.e.*, $\underline{\boldsymbol{\xi}} = \hat{\underline{\boldsymbol{\xi}}}$). 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 the equation (22) 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 a posterior probability density function of $\underline{\boldsymbol{\xi}}$ obtained from the state space model:

$$\begin{aligned}
\hat{\boldsymbol{y}}_t &= \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{G}_t\boldsymbol{u}_t, \quad t = s+1, \ldots, s+m, &(23)\\
\boldsymbol{\alpha}_{t+1} &= \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \mathbf{H}_t\boldsymbol{u}_t, \quad t = s+1, \ldots, s+m-1, &(24)\\
\boldsymbol{u}_t &\sim \mathcal{N}_{2p}(\mathbf{0}, \mathbf{I}_{2p}),
\end{aligned}$$

where $\hat{\boldsymbol{y}}_t$, $\mathbf{Z}_t$, $\mathbf{G}_t$ are defined in Appendix A.1, and $\mathbf{H}_t = [\mathbf{O}, \mathbf{R}_t]$. To find a mode $\hat{\underline{\boldsymbol{\xi}}}$, we repeat following steps until it converges,

1. Compute $\hat{\underline{\boldsymbol{\alpha}}}$ at $\underline{\boldsymbol{\xi}} = \hat{\underline{\boldsymbol{\xi}}}$ using (6).

2. Obtain the approximating linear Gaussian state-space model given by (23) and (24).

3. Applying the disturbance smoother by Koopman (1993) to the approximating linear Gaussian state-space model in Step 2, compute the posterior mode $\hat{\underline{\boldsymbol{\xi}}}$.

since these steps are equivalent to the method of scoring to find a maximiser of the conditional posterior density. As an initial value of $\hat{\underline{\boldsymbol{\xi}}}$, the current sample of $\underline{\boldsymbol{\xi}}$ may be used in the MCMC implementation. If the approximate linear Gaussian state-space model is obtained using a mode $\hat{\underline{\boldsymbol{\xi}}}$, then we draw a sample $\underline{\boldsymbol{\xi}}$ from the conditional posterior distribution by MH algorithm as follows.

1. Propose a candidate $\underline{\boldsymbol{\xi}}^\dagger$ by sampling from $q(\underline{\boldsymbol{\xi}}^\dagger) \propto \min(f(\underline{\boldsymbol{\xi}}^\dagger), cf^*(\underline{\boldsymbol{\xi}}^\dagger))$ using the Acceptance-Rejection algorithm where c can be constructed from a constant term and $\hat{L}$ of (21):

   (a) Generate $\underline{\boldsymbol{\xi}}^\dagger \sim f^*$ using a simulation smoother (e.g. de Jong and Shephard (1995), Durbin and Koopman (2002)) based on the approximating linear Gaussian state-space model (23) - (24).

   (b) Accept $\underline{\boldsymbol{\xi}}^\dagger$ with probability $\min\{f(\underline{\boldsymbol{\xi}}^\dagger)/cf^*(\underline{\boldsymbol{\xi}}^\dagger), 1\}$. If it is rejected, go back to (a).

2. Given the current value $\underline{\boldsymbol{\xi}}$, accept $\underline{\boldsymbol{\xi}}^\dagger$ with probability

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

   if rejected, accept the current $\underline{\boldsymbol{\xi}}$ as a sample.

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

Generation of $\boldsymbol{\Sigma}$ and $\boldsymbol{\phi}$

The sampling method for $\boldsymbol{\Sigma}$ and $\boldsymbol{\phi}$ is rather straightforward as we discuss below.

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

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

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

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

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

Generation of $\boldsymbol{\phi}$. Let $\boldsymbol{\Sigma}^{ij}$ be a $p \times p$ matrix and denote the (i,j)-th block of $\boldsymbol{\Sigma}^{-1}$. Further, 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 whose 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 $\boldsymbol{\phi}$ from its conditional posterior distribution using MH algorithm, we generate a candidate from a truncated normal distribution over the region R, $\boldsymbol{\phi}^\dagger \sim \mathcal{TN}_R(\boldsymbol{\mu}_{\boldsymbol{\phi}}, \boldsymbol{\Sigma}_{\boldsymbol{\phi}})$, $R = \{\boldsymbol{\phi} : |\phi_j| < 1, j = 1, \ldots, p\}$ and accept it with probability $\min\{h(\boldsymbol{\phi}^\dagger)/h(\boldsymbol{\phi}), 1\}$.

2.3 Associated Particle filter

This subsection describes the auxiliary particle filter (see Pitt and Shephard (1999)) to compute the likelihood function ordinate given the parameter $\boldsymbol{\theta}$.

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 $\hat{f}(\boldsymbol{\alpha}_t|Y_t, \boldsymbol{\theta})$ denote the corresponding discrete probability mass function which 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

$$f(\boldsymbol{\alpha}_{t+1}, \boldsymbol{\alpha}_t|Y_{t+1}, \boldsymbol{\theta}) \propto f(\boldsymbol{y}_{t+1}|\boldsymbol{\alpha}_{t+1}) f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{y}_t, \boldsymbol{\alpha}_t, \boldsymbol{\theta}) f(\boldsymbol{\alpha}_t|Y_t, \boldsymbol{\theta}), \qquad (25)$$

where

$$f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t) = (2\pi)^{-p/2}\,|\mathbf{V}_t^{1/2}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\mathbf{V}_t^{1/2}|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{y}_t'\mathbf{V}_t^{-1/2}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{V}_t^{-1/2}\boldsymbol{y}_t\right\}, \quad (26)$$

$$f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{y}_t,\boldsymbol{\alpha}_t,\boldsymbol{\theta}) = (2\pi)^{-p/2}\,|\boldsymbol{\Sigma}_\alpha|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\alpha}_{t+1}-\boldsymbol{\mu}_{\alpha,t+1})'\boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\alpha}_{t+1}-\boldsymbol{\mu}_{\alpha,t+1})\right\}, \quad (27)$$

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

Then using the importance probability density function given by

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

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 proceed the auxiliary particle filter as follows:

Step 1. Initialise $t = 1$ and generate $\boldsymbol{\alpha}_1^i \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_0)$, $(i = 1,\dots,I)$.

(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 $\hat{f}(\boldsymbol{\alpha}_1^i|Y_1,\boldsymbol{\theta}) = \pi_1^i = w_i/\sum_{j=1}^I w_j$, $(i = 1,\dots,I)$.

Step 2. For each i, generate $\left(\boldsymbol{\alpha}_{t+1}^i,\boldsymbol{\alpha}_t^i\right)$ from $g(\boldsymbol{\alpha}_{t+1},\boldsymbol{\alpha}_t|Y_{t+1},\boldsymbol{\theta})$, $(i = 1,\dots,I)$.

(a) Compute

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

for $i = 1,\dots,I$, and record $\overline{w}_t = \sum_{i=1}^I w_i/I$.

(b) Let $\hat{f}(\boldsymbol{\alpha}_{t+1}^i|Y_{t+1},\boldsymbol{\theta}) = \pi_{t+1}^i = w_i/\sum_{j=1}^I w_j$ $(i = 1,\dots,I)$.

Step 3 Increment t and go 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 Extension to MSV model with heavy-tailed errors

The basic MSV model can be extended to incorporate heavy-tailed errors in stock returns. Although the jump components can also be introduced, they are not considered here for simplicity and heavy-tailed errors would be sufficient to account for similar behaviours of stock returns as discussed in Nakajima and Omori (2009). To describe the fat-tailed distributions, two types of multivariate t distributions through a scale mixture distribution are considered. One of them has the common degrees of freedom for all stock returns, while the other has individual degrees of freedom for each stock return. For convenience, we refer the former to the type-1 multivariate t distribution and the latter to the type-2 multivariate t distribution.

Both of them can be expressed as a scale mixture of normal distributions as follows. Let $\mathcal{G}(a,b)$ denotes a gamma distribution with a probability density function $f(\lambda|a,b) \propto \lambda^{a-1}\exp(-b\lambda)$ (a,b are known positive constants). Using a common scalar gamma random variable λ_t, the type-1 multivariate t random variable with ν degrees of freedom is obtained as

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

On the other hand, the type-2 multivariate t random variable is obtained as (see e.g., Harvey, Ruiz, and Shephard (1994))

$$\text{diag}\left(\lambda_{1t}^{-1/2},\ldots,\lambda_{pt}^{-1/2}\right)\boldsymbol{\varepsilon}_t, \quad \text{where} \quad \lambda_{jt} \sim i.i.d.\ \mathcal{G}(\nu_j/2,\nu_j/2), \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0},\boldsymbol{\Sigma}_{\varepsilon\varepsilon}). \tag{29}$$

Thus we extend our basic model as

$$\boldsymbol{y}_t^* = \boldsymbol{\Lambda}_t^{-1/2}\mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\Lambda}_t = \begin{cases} \lambda_t\mathbf{I}_p, & \text{if type-1,} \\ \text{diag}(\lambda_{1t},\ldots,\lambda_{pt}), & \text{if type-2,} \end{cases} \tag{30}$$

for $t=1,\ldots,n$. The prior distributions for the parameters (ν,ν_j) are assumed to be

$$\nu \sim \mathcal{G}(m_0^\nu,S_0^\nu), \quad \nu_j \sim \mathcal{G}(m_{0j}^\nu,S_{0j}^\nu), \quad j=1,\ldots,p. \tag{31}$$

We denote the prior probability density functions of (ν,ν_j) by $\pi(\nu)$, $\pi(\nu_j)$, respectively.

To implement MCMC simulation, we sample $(\boldsymbol{\alpha},\boldsymbol{\phi},\boldsymbol{\Sigma})$ as in Section 2.2 by setting $\boldsymbol{y}_t = \boldsymbol{\Lambda}_t^{1/2}\boldsymbol{y}_t^*$. Thus we focus on sampling from the conditional posterior distributions for other parameters. To illustrate the algorithm, we consider the model with the type-1 multivariate t distribution error in this section since Yu and Meyer (2006) found that the type-1 formulation

(28) was empirically better supported than the type-2 formulation (29) in their bivariate models. (see Appendix B for the model of the type-2 multivariate t distribution error). The conditional posterior probability density function is given by

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

where $Y_n^* = \{\boldsymbol{y}_t^*\}_{t=1}^n$, $\boldsymbol{\lambda} = \{\lambda_t\}_{t=1}^n$,. 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 using $\boldsymbol{y}_t = \boldsymbol{\Lambda}_t^{1/2} \boldsymbol{y}_t^*$.

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 densities of ν is given by

$$
\pi(\nu | \boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, Y_n^*) \quad \propto \quad \pi(\nu) \left\{ \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \right\}^n \left(\prod_{t=1}^{n} \lambda_t \right)^{\frac{\nu}{2}} \times \exp\left\{ -\frac{\sum_{t=1}^n \lambda_t}{2} \nu \right\}. \tag{32}
$$

To sample from this conditional posterior distribution, we first find a conditional mode $\hat{\nu}$ of $\pi(\nu | \boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, Y_n^*)$ numerically. Using MH algorithm, we propose a candidate from a normal distribution truncated over the region $(0, \infty)$, $\nu^\dagger \sim \mathcal{TN}_{(0,\infty)} N(\mu_\nu, \sigma_\nu^2)$, where

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

We accept the candidate with probability

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

where $f_N(\nu | \mu, \sigma^2)$ denotes a probability density function of a normal distribution with mean μ and variance σ^2.

Generation of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$. The conditional posterior probability density function of λ_t is

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

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

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

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 example using simulated data

This section illustrates our proposed methods using simulated data, and show the efficiency of our proposed multi-move sampler in comparison with the single-move sampler. To simulate the data from the basic MSV model in Section 2, we set

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

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

$$\begin{aligned} &\rho_{ij,\varepsilon\varepsilon} \equiv Corr(\varepsilon_{it}, \varepsilon_{jt}) = 0.6, \quad \rho_{ij,\eta\eta} \equiv Corr(\eta_{it}, \eta_{jt}) = 0.7, \\ &\rho_{ij,\varepsilon\eta} \equiv Corr(\varepsilon_{it}, \eta_{jt}) = -0.1, \quad \text{for } i \neq j, \end{aligned}$$

where the negative value of $\rho_{ij,\varepsilon\eta}$ indicates the cross leverage effects. Using these parameters, we generated $n = 2,000$ observations of five stocks returns ($p = 5$). For prior distributions, we assume

$$\frac{\phi_j + 1}{2} \sim \mathcal{B}(20, 1.5), \quad j = 1, \ldots, 5, \quad \boldsymbol{\Sigma} \sim \mathcal{IW}(10, (10\boldsymbol{\Sigma}^*)^{-1}),$$

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.

Using MCMC algorithm described in Section 2.2, we generated 110,000 samples using the multi-move sampler and 550,000 samples using the single-move sampler, and discard first 10,000 and 50,000 samples respectively as burn-in periods.

Tables 1 and 2 show the estimation results using the multi-move sampler (with a tuning parameter $K = 100$) for $(\phi_i, \sigma_{i,\varepsilon\varepsilon}, \sigma_{i,\eta\eta}, \rho_{i,\varepsilon\eta})$ 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 estimates are sufficiently close to true values. The inefficiency factor is defined as $1 + 2\sum_{s=1}^{\infty} \rho_s$ where ρ_s is the sample autocorrelation at lag s, and are computed to measure how well the MCMC chain mixes (see e.g. Chib (2001)). It is the ratio of the numerical variance of the posterior sample mean to the variance of the sample mean from uncorrelated draws. When the inefficiency factor is equal to m, we need to draw MCMC samples m times as many as uncorrelated samples.

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

	True	i	Mean	95% interval	Inefficiency	
					multi	[single]
		1	0.972	[0.962, 0.981]	60	[486]
		2	0.976	[0.967, 0.983]	55	[493]
ϕ_i	0.97	3	0.963	[0.950, 0.974]	87	[451]
		4	0.967	[0.956, 0.977]	74	[713]
		5	0.967	[0.956, 0.977]	59	[506]
		1	1.204	[1.047, 1.375]	380	[6881]
		2	1.318	[1.113, 1.549]	378	[7254]
$\sigma_{i,\varepsilon\varepsilon}$	1.2	3	1.275	[1.135, 1.431]	301	[3369]
		4	1.233	[1.097, 1.378]	240	[5014]
		5	1.349	[1.187, 1.529]	274	[5003]
		1	0.203	[0.174, 0.234]	122	[384]
		2	0.215	[0.186, 0.248]	97	[433]
$\sigma_{i,\eta\eta}$	0.2	3	0.220	[0.187, 0.256]	142	[376]
		4	0.200	[0.171, 0.234]	143	[1080]
		5	0.205	[0.175, 0.238]	115	[750]
		1	-0.251	[-0.373,-0.120]	55	[283]
		2	-0.190	[-0.317,-0.058]	60	[388]
$\rho_{i,\varepsilon\eta}$	-0.2	3	-0.125	[-0.257, 0.013]	80	[528]
		4	-0.219	[-0.348,-0.088]	47	[307]
		5	-0.114	[-0.255, 0.026]	55	[232]

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	0.612	[0.583, 0.640]	2	[14]
		13	0.609	[0.580, 0.637]	2	[23]
		14	0.608	[0.579, 0.636]	3	[29]
		15	0.594	[0.564, 0.623]	1	[9]
		23	0.586	[0.555, 0.615]	3	[10]
		31	0.606	[0.577, 0.634]	2	[18]
		32	0.602	[0.572, 0.631]	2	[15]
		34	0.601	[0.572, 0.630]	3	[16]
		35	0.603	[0.573, 0.632]	2	[16]
		45	0.616	[0.587, 0.644]	3	[29]
$\rho_{ij,\eta\eta}$	0.7	12	0.773	[0.669, 0.856]	124	[667]
		13	0.741	[0.629, 0.832]	101	[1002]
		14	0.720	[0.601, 0.811]	138	[867]
		15	0.685	[0.556, 0.787]	104	[459]
		23	0.763	[0.645, 0.849]	159	[744]
		31	0.773	[0.668, 0.852]	129	[703]
		32	0.697	[0.566, 0.799]	159	[509]
		34	0.758	[0.641, 0.847]	165	[716]
		35	0.685	[0.548, 0.793]	137	[389]
		45	0.684	[0.551, 0.789]	151	[591]
$\rho_{ij,\varepsilon\eta}$	-0.1	12	-0.147	[-0.282,-0.011]	70	[281]
		13	-0.080	[-0.218, 0.060]	48	[286]
		14	-0.184	[-0.319,-0.043]	83	[493]
		15	-0.046	[-0.196, 0.106]	53	[279]
		21	-0.136	[-0.262,-0.003]	54	[429]
		23	-0.021	[-0.155, 0.114]	44	[286]
		24	-0.204	[-0.331,-0.067]	79	[453]
		25	-0.035	[-0.176, 0.107]	49	[210]
		31	-0.127	[-0.264, 0.015]	83	[503]
		32	-0.110	[-0.247, 0.029]	64	[157]
		34	-0.137	[-0.279, 0.007]	78	[500]
		35	-0.015	[-0.161, 0.132]	63	[292]
		41	-0.097	[-0.228, 0.040]	61	[592]
		42	-0.132	[-0.260, 0.003]	61	[115]
		43	-0.095	[-0.226, 0.041]	42	[205]
		45	-0.061	[-0.203, 0.081]	51	[251]
		51	-0.044	[-0.176, 0.091]	60	[532]
		52	-0.115	[-0.243, 0.017]	41	[261]
		53	-0.049	[-0.179, 0.081]	53	[335]
		54	-0.192	[-0.321,-0.056]	61	[354]

The inefficiency factors of parameters obtained from the single-move sampler are much larger (about 10 $\sim$ 20 times for $\sigma_{i,\varepsilon\varepsilon}$) larger than those from the block sampler, which implies our proposed algorithm using the multi-move sampler is highly efficient.

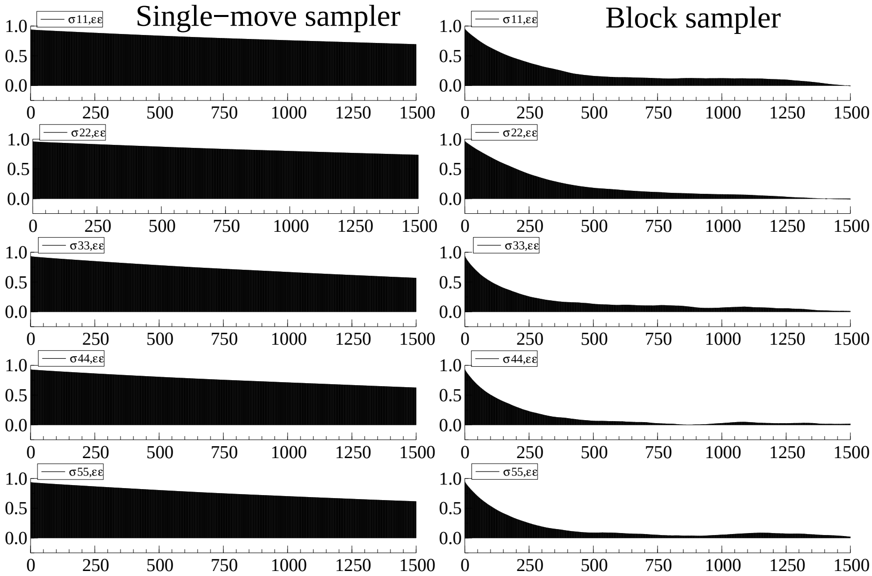

Figure 1: Sample autocorrelation functions of $\sigma_{i,\varepsilon\varepsilon}$:
the single-move sampler vs. the multi-move sampler.

Figure 1 shows the sample autocorrelation functions of $\sigma_{i,\varepsilon\varepsilon}$ $(i = 1, \dots, 5)$ for the single-move and the multi-move samplers. In contrast with the slow decay of the sample autocorrelations in the single-move sampler, they vanish fairly quickly in the multi-move sampler. This also indicates that the MCMC chain mixes well when the multi-move sampler is used.

Table 3: Maximum of inefficiency factors for $K = 25, 50, 60, 100, 150$
and for the single-move sampler.

K	ϕ_j	$\sigma_{i,\varepsilon\varepsilon}$	$\sigma_{i,\eta\eta}$	$\rho_{i,\varepsilon\eta}$	$\rho_{ij,\varepsilon\varepsilon}$	$\rho_{ij,\eta\eta}$	$\rho_{ij,\varepsilon\eta}$
25	322	1099	507	259	7	601	329
50	182	612	245	130	4	324	149
66	125	449	198	95	3	214	120
100	87	380	143	80	3	165	83
150	71	495	120	81	3	151	88
single-move	713	7254	1080	528	29	1002	592

Regarding a selection of a tuning parameter K, we set $K = 100$ in this example as follows. Table 3 shows the maxima of inefficiency factors for ϕ_j's, $\sigma_{i,\varepsilon\varepsilon}$'s, $\sigma_{i,\eta\eta}$'s, $\rho_{i,\varepsilon\eta}$'s, $\rho_{ij,\varepsilon\varepsilon}$'s, $\rho_{ij,\eta\eta}$'s and $\rho_{ij,\varepsilon\eta}$'s using $K = 25, 50, 66, 100$ and 150 (the average block sizes n/K are 80, 40, 30, 20 and 13). Since the maxima for $K = 100$ are overall smaller than those for other K's, it is selected as an optimal value in our MCMC simulation. If it is greater than 100 (*i.e.*, the number of elements in one block becomes small on an average), the Markov chain would not move fast around in its state space due to the high correlations among adjacent $\boldsymbol{\alpha}_t$'s. On the other hand, if it is less than 100, the proposed states, $\boldsymbol{\alpha}_t$'s, would be rejected too often in the MH algorithm, which also results in the slow mixing of the chain.

4 Empirical studies

4.1 Data

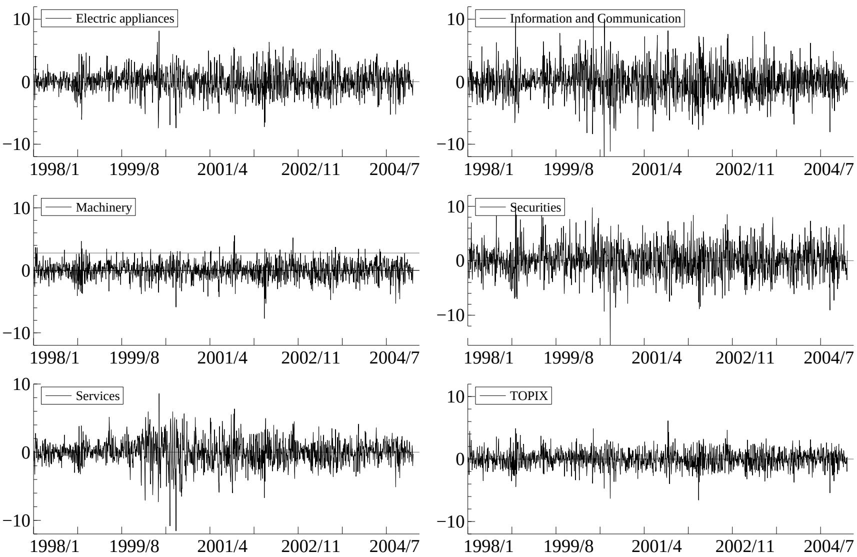

Figure 2: Returns of five sector indices and TOPIX

This section applies the MSV models to five stock returns data in Tokyo Stock Exchange. We consider stock returns for five industry sectors defined by the log-difference of the closing values of TOPIX (Tokyo Stock Exchange Stock Price Index) sector indices: 'Electric appliances' (Series 1), 'Information and Communication' (Series 2), 'Machinery' (Series 3), 'Securities' (Series 4) and 'Services' (Series 5). During the sample period from January 5, 1998 to December 30, 2004, there are 1,722 trading days. The time series plot of five return series and TOPIX are presented in Figure 2, which indicates the co-movement of the volatility among five stock returns during this period.

4.2 Univariate SV model

First we fit univariate SV models with leverage effects (1)-(2) to individual series. The SV models with leverage effects and t-distributed error (SVt) are also estimated as benchmarks for the multivariate models. The prior distributions for ϕ and $\mathbf{\Sigma}$ are assumed to be

$$\begin{aligned} \frac{\phi_i+1}{2} &\sim \mathcal{B}(20, 1.5), \quad \mathbf{\Sigma}_i \sim \mathcal{IW}(5, (5\mathbf{\Sigma}_i^*)^{-1}), \quad \mathbf{\Sigma}_i^* = \begin{pmatrix} 1 & -0.02 \\ -0.02 & 0.04 \end{pmatrix}, \\ \nu_i &\sim \mathcal{G}(1, 0.05), \quad i = 1, \ldots, 5, \end{aligned}$$

where ν_i's are the degrees of freedom for the t-distributed error models. Since the posterior estimates of ν_i may be sensitive to the choice of the prior distribution (see e.g. Nakajima and Omori (2009)), we use a flat prior with mean 20 and standard deviation 20.

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

		SV			SVt		
	i	Mean	Stdev	95% interval	Mean	Stdev	95% interval
ϕ_i	1	0.973	0.009	[0.953, 0.988]	0.974	0.009	[0.955, 0.988]
	2	0.969	0.010	[0.948, 0.986]	0.973	0.009	[0.954, 0.988]
	3	0.940	0.016	[0.904, 0.967]	0.944	0.015	[0.911, 0.969]
	4	0.955	0.014	[0.923, 0.978]	0.958	0.013	[0.927, 0.980]
	5	0.976	0.007	[0.960, 0.989]	0.976	0.007	[0.961, 0.989]
$\sigma_{i,\varepsilon\varepsilon}$	1	1.556	0.118	[1.320, 1.788]	1.505	0.117	[1.269, 1.732]
	2	2.004	0.182	[1.655, 2.366]	1.961	0.185	[1.589, 2.332]
	3	1.269	0.051	[1.171, 1.372]	1.237	0.050	[1.140, 1.338]
	4	2.351	0.144	[2.066, 2.632]	2.298	0.138	[2.038, 2.578]
	5	1.383	0.143	[1.102, 1.676]	1.358	0.129	[1.101, 1.627]
$\sigma_{i,\eta\eta}$	1	0.165	0.022	[0.125, 0.210]	0.159	0.021	[0.123, 0.203]
	2	0.204	0.028	[0.154, 0.263]	0.188	0.026	[0.143, 0.242]
	3	0.186	0.025	[0.142, 0.239]	0.173	0.023	[0.132, 0.223]
	4	0.200	0.028	[0.147, 0.260]	0.189	0.028	[0.142, 0.249]
	5	0.200	0.023	[0.158, 0.248]	0.195	0.023	[0.154, 0.244]
$\rho_{i,\varepsilon\eta}$	1	-0.290	0.087	[-0.453,-0.114]	-0.311	0.092	[-0.483,-0.124]
	2	-0.260	0.084	[-0.416,-0.088]	-0.262	0.087	[-0.423,-0.084]
	3	-0.422	0.081	[-0.572,-0.255]	-0.445	0.082	[-0.599,-0.276]
	4	-0.271	0.083	[-0.425,-0.103]	-0.270	0.084	[-0.426,-0.095]
	5	-0.281	0.073	[-0.418,-0.131]	-0.288	0.075	[-0.430,-0.136]
ν_i	1				41.44	21.33	[16.39, 98.93]
	2				45.99	23.88	[15.70,108.69]
	3				39.67	21.41	[14.25, 93.05]
	4				41.66	22.54	[15.07, 98.83]
	5				58.65	25.30	[22.08,119.77]

To implement the MCMC algorithm, we draw 60,000 samples and discard 10,000 samples as burn-in periods. Posterior means, standard deviations and 95% credible intervals are shown in Table 4.

The estimates of ϕ_i's vary from 0.940 to 0.976, and those of ρ_i's vary from -0.422 to -0.260, suggesting the strong persistences in the log volatilities and highly credible leverage effects. These results are consistent with those in the previous empirical studies, and the estimates of ν_i's are about the same as those of TOPIX returns during this period ($\nu = 43.75$, more detail results for TOPIX returns using the SVt model are omitted).

4.3 MSV model

The following three MSV models are considered.

(i) **MSV** model: the basic MSV model.

(ii) **MSVt1** model: the MSV model with type-1 multivariate t measurement errors.

(iii) **MSVt2** model: the MSV model with type-2 multivariate t measurement errors.

The prior distributions are assumed to be

$$\begin{aligned} \frac{\phi_i+1}{2} &\sim \mathcal{B}(20, 1.5), \quad \boldsymbol{\Sigma} \sim \mathcal{IW}(10, (10\boldsymbol{\Sigma}^*)^{-1}), \\ \nu &\sim \mathcal{G}(1, 0.05), \quad \nu_i \sim \mathcal{G}(1, 0.05), \quad i = 1, \ldots, 5, \end{aligned}$$

where

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} \boldsymbol{\Sigma}^*_{\varepsilon\varepsilon} & \boldsymbol{\Sigma}^*_{\varepsilon\eta} \\ \boldsymbol{\Sigma}^{*\prime}_{\varepsilon\eta} & \boldsymbol{\Sigma}^*_{\eta\eta} \end{pmatrix} = \begin{pmatrix} 1.5^2(0.5\mathbf{I}_5 + 0.5\mathbf{1}_5\mathbf{1}_5') & 1.5 \times 0.2 \times (-0.1)\mathbf{I}_5 \\ & 0.2^2(0.5\mathbf{I}_5 + 0.5\mathbf{1}_5\mathbf{1}_5') \end{pmatrix},$$

and $E(\boldsymbol{\Sigma}^{-1}) = \boldsymbol{\Sigma}^{*-1}$. Hyper-parameters of the prior distributions are chosen based on the analysis of the univariate SV models. Using the MCMC algorithms described in Section 2, we draw 100,000 samples after discarding 20,000 samples as the burn-in period for the MSV model, 150,000 samples after discarding 20,000 samples for the MSVt1 and the MSVt2 model. The tuning parameter K is set equal to 90 such that the average number of state vectors in one block is about 20.

Table 5 shows posterior means, standard deviations, 95% credible intervals and inefficiency factors of $\phi, \sigma_{i,\varepsilon\varepsilon}, \sigma_{i,\eta\eta}$ and $\rho_{i,\varepsilon\eta}$ for the MSV model (results for MSVt1 and MSVt2 models are similar and hence omitted). The estimated ϕ_i's and $\sigma_{i,\varepsilon\varepsilon}$'s are slightly less than those of univariate SV or SVt models, while those of $\rho_{i,\varepsilon\eta}$ are found to be larger. This may be because a part of the variation in one stock return is explained by that of other stock returns since we found high correlations among ε_{it}'s and η_{jt}'s as shown later. Further, we note that posterior standard deviations are mostly smaller than those of univariate models. On the other hand, the magnitude of the dispersion, $\sigma_{i,\eta\eta}$'s, in the state equations remains about the same.

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

	i	Mean	Stdev	95% interval	Inefficiency
ϕ_i	1	0.956	0.008	[0.940, 0.970]	114
	2	0.965	0.007	[0.951, 0.978]	177
	3	0.945	0.009	[0.926, 0.961]	78
	4	0.949	0.009	[0.929, 0.966]	119
	5	0.967	0.005	[0.956, 0.977]	131
$\sigma_{i,\varepsilon\varepsilon}$	1	1.497	0.073	[1.350, 1.640]	182
	2	1.844	0.128	[1.585, 2.082]	328
	3	1.225	0.046	[1.136, 1.316]	133
	4	2.212	0.109	[1.991, 2.420]	176
	5	1.298	0.084	[1.134, 1.463]	249
$\sigma_{i,\eta\eta}$	1	0.191	0.018	[0.158, 0.227]	147
	2	0.203	0.018	[0.169, 0.240]	138
	3	0.181	0.016	[0.153, 0.215]	96
	4	0.211	0.020	[0.174, 0.253]	161
	5	0.207	0.017	[0.174, 0.242]	118
$\rho_{i,\varepsilon\eta}$	1	-0.146	0.068	[-0.277,-0.011]	38
	2	-0.167	0.070	[-0.300,-0.028]	60
	3	-0.230	0.066	[-0.356,-0.099]	45
	4	-0.158	0.068	[-0.289,-0.023]	44
	5	-0.204	0.061	[-0.321,-0.081]	32

Table 6: MSVt1 and MSVt2 models. Posterior means, standard deviations, 95% credible intervals and inefficiency factors for ν and ν_i $(i = 1, \ldots, 5)$.

	i	Mean	Stdev	95% interval	Inefficiency
ν		42.43	15.21	[23.69,81.55]	298
ν_i	1	38.75	15.05	[20.37, 76.20]	714
	2	48.40	23.25	[20.25,108.03]	613
	3	76.94	27.26	[36.72,143.04]	326
	4	40.13	18.37	[18.20, 88.26]	372
	5	88.53	30.04	[43.51,159.11]	403

Table 7: MSV 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	0.688	0.013	[0.661, 0.713]	2
	13	0.788	0.010	[0.768, 0.806]	4
	14	0.730	0.012	[0.705, 0.752]	4
	15	0.766	0.011	[0.745, 0.787]	4
	23	0.633	0.015	[0.602, 0.662]	2
	24	0.631	0.015	[0.601, 0.660]	2
	25	0.693	0.013	[0.666, 0.718]	2
	34	0.737	0.012	[0.714, 0.759]	3
	35	0.748	0.011	[0.725, 0.770]	4
	45	0.698	0.013	[0.672, 0.723]	3
$\rho_{ij,\eta\eta}$	12	0.802	0.045	[0.702, 0.876]	126
	13	0.811	0.042	[0.717, 0.880]	93
	14	0.820	0.042	[0.727, 0.889]	87
	15	0.783	0.047	[0.680, 0.862]	67
	23	0.803	0.044	[0.707, 0.876]	125
	24	0.800	0.048	[0.691, 0.879]	98
	25	0.816	0.040	[0.728, 0.884]	62
	34	0.803	0.045	[0.705, 0.877]	109
	35	0.781	0.044	[0.683, 0.856]	75
	45	0.793	0.046	[0.689, 0.867]	126
$\rho_{ij,\varepsilon\eta}$	12	-0.203	0.067	[-0.332,-0.068]	37
	13	-0.151	0.071	[-0.289,-0.009]	49
	14	-0.123	0.069	[-0.255, 0.013]	51
	15	-0.195	0.067	[-0.323,-0.061]	47
	21	-0.106	0.075	[-0.251, 0.043]	67
	23	-0.116	0.075	[-0.263, 0.034]	63
	24	-0.111	0.073	[-0.252, 0.036]	56
	25	-0.084	0.074	[-0.227, 0.061]	53
	31	-0.159	0.066	[-0.288,-0.028]	38
	32	-0.209	0.064	[-0.332,-0.080]	38
	34	-0.134	0.065	[-0.260,-0.005]	35
	35	-0.193	0.065	[-0.318,-0.064]	43
	41	-0.127	0.071	[-0.267, 0.014]	55
	42	-0.180	0.070	[-0.313,-0.040]	55
	43	-0.169	0.073	[-0.309,-0.023]	52
	45	-0.112	0.071	[-0.249, 0.028]	56
	51	-0.133	0.065	[-0.258,-0.004]	28
	52	-0.199	0.063	[-0.319,-0.073]	34
	53	-0.204	0.065	[-0.330,-0.074]	55
	54	-0.183	0.063	[-0.303,-0.058]	33

Regarding the degrees of freedom for the heavy-tailed errors shown in Table 6, the posterior means of ν for the MSVt1 model is found to be about the average of ν_j's obtained for the univariate SVt models. However, a couple of posterior means of those ν_j's for the MSVt2 model are found to be much larger than those for the univariate models. These relatively large degrees of freedom for the SVt, the MSVt1 and the MSVt2 models indicate that the MSV model with normal errors would be appropriate to describe the dynamics of the stock returns during this sample period.

Tables 7 shows the posterior means, standard deviations, 95% credible intervals and inefficiency factors for $\rho_{ij,\varepsilon\varepsilon}, \rho_{ij,\eta\eta}$ and $\rho_{ij,\varepsilon\eta}$ for the MSV model. The correlations among ε_{it}'s and η_{it}'s, ($\rho_{ij,\varepsilon\varepsilon}$'s and $\rho_{ij,\eta\eta}$'s) are found to be very high, suggesting the possible co-movement among volatilities. The posterior means of $\rho_{ij,\varepsilon\eta}$ are all negative varying from -0.209 to -0.084 in the MSV model. The cross leverage effects from Series 1, 3, 4 & 5 to the volatility of Series 2, and from Series 5 to the volatilities of Series 1, 2, 3 & 4 seem to be relatively stronger and the posterior probability of negative $\rho_{i2,\varepsilon\eta}$ (or $\rho_{5j,\varepsilon\eta}$) is greater than 0.95 for the MSV model. We note that the cross leverages effects between series i and j are asymmetric, *i.e.*, $\rho_{ij,\varepsilon\eta} \neq \rho_{ji,\varepsilon\eta}$ for $i \neq j$. For example, in Table 7, the $\rho_{i2,\varepsilon\eta}$'s are from -0.209 to -0.180, while $\rho_{2j,\varepsilon\eta}$'s are from -0.116 to -0.084. This implies the volatility of Series 2 (Information and Communication) is influenced by other series, but the volatilities of other series are less subject to the return of Series 2.

5 Conclusion

This paper proposes efficient MCMC algorithms using a multi-move sampler for the latent volatility vectors for MSV models with cross leverage and heavy-tailed errors. To sample a block of such state vectors, we construct a proposal density for MH algorithm based on the normal approximation using Taylor expansion of the logarithm of the target likelihood and exploit the sampling algorithms which are developed for the linear and Gaussian state space model. We show that our proposed methods are easy to implement and that they are highly efficient. Extensions to the models with two types of multivariate t-distributed errors are also discussed. Illustrative examples and empirical studies are given using five stock market indices in Tokyo Stock Exchange.

Acknowledgement

The authors are grateful to 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 approximating state space model

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

$$\mathbf{A}_t = -E\left[\frac{\partial^2 L}{\partial\boldsymbol{\alpha}_t\partial\boldsymbol{\alpha}'_t}\right], \quad t = s+1,\ldots,s+m, \tag{33}$$

$$\mathbf{B}_t = -E\left[\frac{\partial^2 L}{\partial\boldsymbol{\alpha}_t\partial\boldsymbol{\alpha}'_{t-1}}\right], \quad t = s+2,\ldots,s+m, \quad \mathbf{B}_{s+1} = \mathbf{O}, \tag{34}$$

and let $\boldsymbol{d}_t = \partial L/\partial\boldsymbol{\alpha}_t$ for $t = s+1,\ldots,s+m$. The $\boldsymbol{d}_t, \mathbf{A}_t$ 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, $\underline{\boldsymbol{\alpha}} = \underline{\hat{\boldsymbol{\alpha}}}$. Using $\hat{\boldsymbol{d}}_t, \hat{\mathbf{A}}_t$ and $\hat{\mathbf{B}}_t$,

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

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

   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

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

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

Then, we obtain the approximating linear Gaussian state-space model given by (23) and (24).

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

Matrix differentiation

We first summarise definitions of first and second derivatives of a matrix and some results (see, 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

$$\mathbf{D}H(X) = (\mathbf{D}G(F(X)))(\mathbf{D}F(X)) = \frac{\partial \mathrm{vec}(G(F(\boldsymbol{X})))}{\partial(\mathrm{vec}(F(\boldsymbol{X})))'}\frac{\partial \mathrm{vec}(F(\boldsymbol{X}))}{\partial(\mathrm{vec}(\boldsymbol{X}))'}. \tag{35}$$

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

$$\frac{\partial g(f(\mathbf{x}))}{\partial \mathbf{x}'} = \frac{\partial \mathrm{vec}(g(f(\mathbf{x})))}{\partial \mathrm{vec}(f(\mathbf{x}))'}\frac{\partial \mathrm{vec}(f(\mathbf{x}))}{\partial \mathrm{vec}(\mathbf{x})'}. \tag{36}$$

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

$$\frac{\partial \mathrm{vec}(FG)}{\partial(\mathrm{vec}(X))'} = (G' \otimes \mathbf{I}_m)\frac{\partial \mathrm{vec}(F)}{\partial(\mathrm{vec}(X))'} + (\mathbf{I}_r \otimes F)\frac{\partial \mathrm{vec}(G)}{\partial(\mathrm{vec}(X))'}. \tag{37}$$

$\boldsymbol{d_t}$, $\mathbf{A}_t$ 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

$$\frac{\partial l_t}{\partial \boldsymbol{\alpha}_t'} = -\frac{1}{2}\mathbf{1}_p' - (\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1}\frac{\partial(\boldsymbol{z}_t - \boldsymbol{m}_t)}{\partial \boldsymbol{\alpha}_t'}$$
$$= -\frac{1}{2}\mathbf{1}_p' + \frac{1}{2}(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1}\left\{\text{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t<n)\right\}, \quad (38)$$
$$\frac{\partial l_{t-1}}{\partial \boldsymbol{\alpha}_t'} = (\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1})'\mathbf{S}_{t-1}^{-1}\frac{\partial \boldsymbol{m}_{t-1}}{\partial \boldsymbol{\alpha}_t'}$$
$$= (\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1})'\mathbf{S}_{t-1}^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}I(t>1), \quad (39)$$

using the chain rule (36). Thus

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

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

To compute $\mathbf{A}_t$ and $\mathbf{B}_t$ using the product rule (37), we first obtain the Hessian matrix

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

where

$$\frac{\partial^2 l_t}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}_t'} = \frac{1}{2}\left\{(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1} \otimes \mathbf{I}_p\right\}\frac{\partial \text{vec}(\text{diag}(\boldsymbol{z}_t))}{\partial \boldsymbol{\alpha}_t'}$$
$$-\frac{1}{4}\left(\text{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}I(t<n)\right)\mathbf{S}_t^{-1}\left(\text{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t<n)\right), \quad (40)$$
$$\frac{\partial^2 l_{t-1}}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}_t'} = -\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_{t-1}^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}I(t>1). \quad (41)$$

Noting that

$$\frac{\partial \text{vec}(\text{diag}(\boldsymbol{z}_t))}{\partial \boldsymbol{\alpha}_t'} = -\frac{1}{2}\begin{bmatrix} z_{1t}\boldsymbol{e}_1\boldsymbol{e}_1' \\ \vdots \\ z_{pt}\boldsymbol{e}_p\boldsymbol{e}_p' \end{bmatrix}, \quad E\left[z_{jt}(\boldsymbol{z}_t - \boldsymbol{m}_t)\right] = \mathbf{S}_t\boldsymbol{e}_j, \quad (42)$$

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 (40) is

$$E\left[\left\{(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1} \otimes \mathbf{I}_p\right\}\frac{\partial \text{vec}(\text{diag}(\boldsymbol{z}_t))}{\partial \boldsymbol{\alpha}_t'}\right] = -\frac{1}{2}\sum_{j=1}^{p}(\boldsymbol{e}_j'\mathbf{S}_t^{-1})(\mathbf{S}_t\boldsymbol{e}_j)\boldsymbol{e}_j\boldsymbol{e}_j' = -\frac{1}{2}\mathbf{I}_p, \quad (43)$$

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 (40) is

$$
\begin{aligned}
&E\left[\left(\text{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}I(t<n)\right)\mathbf{S}_t^{-1}\left(\text{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t<n)\right)\right] \\
&= \mathbf{S}_t^{-1} \odot (\mathbf{S}_t + \boldsymbol{m}_t\boldsymbol{m}_t') - 2\left(\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\text{diag}(\boldsymbol{m}_t) + \text{diag}(\boldsymbol{m}_t)\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}\right)I(t<n) \\
&\quad +4\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t<n). \qquad (44)
\end{aligned}
$$

Thus, we obtain

$$
\begin{aligned}
\mathbf{A}_t &= \frac{1}{4}\left\{\mathbf{I}_p + \mathbf{S}_t^{-1} \odot (\mathbf{S}_t + \boldsymbol{m}_t\boldsymbol{m}_t')\right\} + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t<n) \\
&\quad -\frac{1}{2}\left(\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\text{diag}(\boldsymbol{m}_t) + \text{diag}(\boldsymbol{m}_t)\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}\right)I(t<n) \\
&\quad +\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_{t-1}^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}I(t>1) + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t=s+m<n), \qquad (45)
\end{aligned}
$$

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

$$
\mathbf{B}_t = -E\left[\frac{\partial^2 l_{t-1}}{\partial\boldsymbol{\alpha}_t\partial\boldsymbol{\alpha}_{t-1}'}\right] = \frac{1}{2}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_{t-1}^{-1}\left\{\text{diag}(\boldsymbol{m}_{t-1}) - 2\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}\right\}, \qquad (46)
$$

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

B Alternative multivariate t-distributed errors

This subsection describes MCMC implementation for the MSV model with the alternative (type-2) multivariate t-distributed errors. Since $(\boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{\Sigma})$ can be sampled as in Section 2, we focus on sampling parameters $\boldsymbol{\nu} = \{\nu_j\}_{j=1}^p$ and $\boldsymbol{\lambda} = (\{\lambda_{j1}\}_{j=1}^p, \ldots, \{\lambda_{jn}\}_{j=1}^p)$. The conditional posterior probability density function is given by

$$
\begin{aligned}
&\pi(\boldsymbol{\nu}, \boldsymbol{\lambda}|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, Y_n^*) \\
&\propto \prod_{j=1}^p \pi(\nu_j)\left\{\frac{\left(\frac{\nu_j}{2}\right)^{\frac{\nu_j}{2}}}{\Gamma\left(\frac{\nu_j}{2}\right)}\right\}^n \prod_{t=1}^n \lambda_{jt}^{\frac{\nu_j+1}{2}-1}\exp\left[-\frac{1}{2}\sum_{t=1}^n\left\{\nu_j\lambda_{jt} + \left(\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{y}_t^* - \boldsymbol{\mu}_t\right)'\boldsymbol{\Sigma}_t^{-1}\left(\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{y}_t^* - \boldsymbol{\mu}_t\right)\right\}\right].
\end{aligned}
$$

We implement MCMC simulation as follows.

1. Generate $(\boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{\Sigma})$ as in Section 2.

2. Generate $\nu_j \sim \pi(\nu_j|\boldsymbol{\lambda}_j)$.

3. Generate $\lambda_{jt} \sim \pi(\lambda_{jt}|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}_t, \nu_j, \boldsymbol{y}_t^*)$ for $t = 1, \ldots, n$ and $j = 1, \ldots, p$.

Generation of $\boldsymbol{\nu}$. The conditional posterior probability densities of $\boldsymbol{\nu}$ is given by

$$\pi(\boldsymbol{\nu}|\boldsymbol{\phi},\boldsymbol{\Sigma},\nu,\boldsymbol{\lambda},\boldsymbol{\alpha},Y_n^*) \quad \propto \quad \prod_{j=1}^{p}\pi(\nu_j)\left\{\frac{\left(\frac{\nu_j}{2}\right)^{\frac{\nu_j}{2}}}{\Gamma\left(\frac{\nu_j}{2}\right)}\right\}^n \prod_{t=1}^{n}\lambda_{jt}^{\frac{\nu_j}{2}}\exp\left\{-\frac{\sum_{t=1}^{n}\lambda_{jt}}{2}\nu_j\right\}. \tag{47}$$

We first find a conditional mode $\hat{\boldsymbol{\nu}}$ of $\pi(\boldsymbol{\nu}|\boldsymbol{\phi},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\alpha},Y_n^*)$ numerically. Then we propose a candidate using a truncated normal distribution over the region $\mathbf{R}_+^p$, $\boldsymbol{\nu}^\dagger \sim \mathcal{TN}_{\mathbf{R}_+^p}(\boldsymbol{\mu}_{\boldsymbol{\nu}},\boldsymbol{\Sigma}_{\boldsymbol{\nu}})$

$$\boldsymbol{\mu}_{\boldsymbol{\nu}} \quad = \quad \hat{\boldsymbol{\nu}}+\boldsymbol{\Sigma}_{\boldsymbol{\nu}}\left[\left.\frac{\partial\log\pi(\boldsymbol{\nu}|\boldsymbol{\phi},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\alpha},Y_n^*)}{\partial\boldsymbol{\nu}}\right|_{\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}}\right],\boldsymbol{\Sigma}_{\boldsymbol{\nu}}=\left[-\left.\frac{\partial^2\log\pi(\boldsymbol{\nu}|\boldsymbol{\phi},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\alpha},Y_n^*)}{\partial\boldsymbol{\nu}\partial\boldsymbol{\nu}'}\right|_{\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}}\right]^{-1},$$

and accept it with probability

$$\min\left[\frac{\pi(\boldsymbol{\nu}^\dagger|\boldsymbol{\phi},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\alpha},Y_n^*)f_N(\boldsymbol{\nu}|\boldsymbol{\mu}_{\boldsymbol{\nu}},\boldsymbol{\Sigma}_{\boldsymbol{\nu}})}{\pi(\boldsymbol{\nu}|\boldsymbol{\phi},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\alpha},Y_n^*)f_N(\boldsymbol{\nu}^\dagger|\boldsymbol{\mu}_{\boldsymbol{\nu}},\boldsymbol{\Sigma}_{\boldsymbol{\nu}})},1\right].$$

Generation of λ_{jt}, $j=1,\ldots,p,t=1,\ldots,n$. The conditional posterior probability density function of λ_{jt} is

$$\pi(\lambda_{jt}|\boldsymbol{\phi},\boldsymbol{\Sigma},\nu,\boldsymbol{\alpha},\boldsymbol{\lambda}_{-\lambda_{jt}},Y_n^*) \quad \propto \quad \lambda_{jt}^{\frac{\nu_j+1}{2}-1}\exp\left[-\frac{1}{2}\left\{\nu_j\lambda_{jt}+\left(\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{y}_t^*-\boldsymbol{\mu}_t\right)'\boldsymbol{\Sigma}_t^{-1}\left(\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{y}_t^*-\boldsymbol{\mu}_t\right)\right\}\right].$$

The λ_{jt} can be sampled from its posterior distribution using MH-algorithm using a proposal distribution

$$\lambda_{jt}^\dagger \sim \mathcal{G}\left(\frac{\nu_j+1}{2},\frac{1}{2}\left[\nu_j+y_{jt}^2\exp(-\alpha_{jt})s_t^j\right]\right),$$

where s_t^j is the j-th diagonal element of $\mathbf{S}_t^{-1}$.

References

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

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

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

Chan, D., R. Kohn, and C. Kirby (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 J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, pp. 3569–3649. Amsterdam: North-Holland.

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

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

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

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

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

Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*, 247–264.

Jacquier, E., N. G. Polson, and P. E. Rossi (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. and J.-F. Richard (2003). Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *Journal of Empirical Finance 10*, 505–531.

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

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

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

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

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

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

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

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

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

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

Yu, J. and R. Meyer (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews 25*, 361–384.