CIRJE-F-938

Matrix Exponential Stochastic Volatility with Cross Leverage

Tsunehiro Ishihara
Hitotsubashi University

Yasuhiro Omori
The University of Tokyo

Manabu Asai
Soka University

August 2014

Matrix exponential stochastic volatility with cross leverage

Tsunehiro Ishihara* Yasuhiro Omori † and Manabu Asai ‡

August 26, 2014

Abstract

A multivariate stochastic volatility model with the dynamic correlation and the cross leverage effect is described and its efficient estimation method using Markov chain Monte Carlo is proposed. The time-varying covariance matrices are guaranteed to be positive definite by using a matrix exponential transformation. Of particular interest is our approach for sampling a set of latent matrix logarithm variables from their conditional posterior distribution, where we construct the proposal density based on an approximating linear Gaussian state space model. The proposed model and its extended models with fat-tailed error distribution are applied to trivariate returns data (daily stocks, bonds, and exchange rates) of Japan. Further, a model comparison is conducted including constant correlation multivariate stochastic volatility models with leverage and diagonal multivariate GARCH models.

Key words: Dynamic correlation, Leverage effect, Matrix exponential, Markov chain Monte Carlo, Multi-move sampler, Multivariate stochastic volatility

*Department of Economics, Hitotsubashi University, 2-1, Naka, Kunitachi, Tokyo 186-8601, Japan. Email: T.Ishihara@r.hit-u.ac.jp

†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

‡Faculty of Economics, Soka University, Hachioji-shi, Tokyo 192-8577, Japan. E-mail: m-asai@soka.ac.jp

1 Introduction

Over the last several decades, there has been a great deal of interest in modeling volatilities of multivariate stock market returns. The examples are multivariate generalized autoregressive conditional heteroskedasticity (GARCH) models (see the review of Bauwens, Laurent, and Rombouts (2006)), multivariate stochastic volatility (SV) models (see the review of Asai, McAleer, and Yu (2006), Chib, Omori, and Asai (2009)) and realized covariance models (see e.g. Golosnoy, Gribisch, and Liesenfeld (2012)). The realized covariance model uses the high-frequency data to estimate covariance matrices and regard them as observed covariance matrices, while they are latent variables in GARCH and SV models.

Various multivariate volatility models have been proposed in the literature to describe the dynamic properties of the covariance matrices such as the volatility clustering, the dynamic correlations, and the leverage effects. The DCC models, (Engle (2002)) and BEKK model (Engle and Kroner (1995)) are such multivariate GARCH models, and autoregressive Wishart models (Philipov and Glickman (2006), Gourieroux, Jasiak, and Sufana (2009), Golosnoy, Gribisch, and Liesenfeld (2012)) are examples in multivariate SV models. The common difficulty in these models is to keep the covariance matrices positive definite. To overcome this difficulty, reparameterization methods are considered in Yu and Meyer (2006), Tsay (2005), and Jungbacker and Koopman (2006). The Choleski decomposition of the covariance matrix is also considered in Lopes, McCulloch, and Tsay (2012) and Loddo, Ni, and Sun (2011).

However, there have been still few previous works on the multivariate volatility models with both dynamic correlations and cross leverage effects. Cross leverage refers to the correlation between the i-th asset return at time t and the function of j-th asset volatility at time $t+1$ (when $i = j$, we simply call it a leverage effect). Thus, to model these properties of covariance matrices, this paper considers the matrix logarithm transformation which is known useful to model positive definite matrices in a flexible way. Since the seminal work of Chiu, Leonard, and Tsui (1996), the matrix exponential model for the covariance matrix has been applied to the spatial model to simplify the calculation of log-likelihood functions (LeSage and Pace (2007)), and is extended to the GARCH model (Kawakatsu (2006)), the SV model (Asai, McAleer, and Yu (2006)) and the realized covariation model (Bauer and Vorkink (2010) and Sheppard (2007)) for multivariate financial time series.

We consider the general multivariate volatility model using the matrix exponential SV model with cross leverage effects and propose an efficient computational algorithm. This is a generalization of Ishihara and Omori (2012) who propose the following multivariate stochastic volatility (MSV) model with cross-asset leverage effect of the form

$$\boldsymbol{y}_t = \mathrm{diag}\left(\exp(\alpha_{1t}/2), \ldots, \exp(\alpha_{pt}/2)\right) \boldsymbol{\varepsilon}_t, \tag{1}$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \tag{2}$$

$$\left(\boldsymbol{\varepsilon}_t', \boldsymbol{\eta}_t'\right)' \sim \mathcal{N}_{2p}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{3}$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{pt})'$, $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{pt})'$, $\boldsymbol{\Phi} = \mathrm{diag}(\phi_1, \ldots, \phi_p)$ and $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the p-dimensional normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$. This is fairly general in the sense that there is no restriction imposed on the covariance matrix $\boldsymbol{\Sigma}$, while, in the previous literature, various parameter restrictions are imposed (e.g. Asai and McAleer (2006), Asai and McAleer (2009), Chan, Kohn, and Kirby (2006), and Daníelsson (1998)) to estimate parameters based on the Monte Carlo likelihood. We, further, model the dynamic covariance matrices (dynamic variances and correlations) using a matrix logarithm transformation. Since it is difficult to implement a maximum likelihood estimation for our proposed model without imposing restrictions on parameters, we take Bayesian approach and estimate posterior distributions of model parameters using Markov chain Monte Carlo (MCMC) method. The simple sampling algorithm for the latent covariance matrices is known to be inefficient as discussed in Ishihara and Omori (2012). They showed that the single-move sampler which samples one volatility variable given others is highly inefficient and proposed the efficient multi-move sampler (block sampler) which divides the vector of all latent variables into blocks and samples one block given other blocks based on Omori and Watanabe (2008). Thus we construct the multi-move sampler for our matrix exponential model and show that it is efficient in comparison with the alternative simple sampling algorithm.

The rest of the paper is organized as follows. In Section 2, we introduce a matrix exponential stochastic volatility model with cross leverage effects. Bayesian estimation method and the associated particle filter for calculating likelihood functions are described in Section 3. Section 4 shows the efficiency of our proposed algorithm using the simulated data, and, in Section 5, the empirical studies are given using the trivariate asset returns data (stock indices, bond indices and foreign exchange rates). We conduct a model selection among the proposed model, extended models with fat-tailed error distribution and some constant correlation multivariate SV models. Section 6 concludes the paper.

2 Matrix exponential stochastic volatility

This section proposes the matrix exponential stochastic volatility (MESV) model with cross leverage effects. The MESV model is based on the matrix exponential transformation as below. A matrix exponential is widely studied in the context of multidimensional differential equations and Lie algebra. The statistical applications of the matrix exponential transformation are given, for example, in Chiu, Leonard, and Tsui (1996), and Kawakatsu (2006). For any $p \times p$ matrix $\mathbf{A}$, the matrix exponential is defined by the following power series expansion

$$\exp(\mathbf{A}) \equiv \sum_{s=0}^{\infty} \frac{1}{s!}\mathbf{A}^s,$$

where the series converges absolutely if all eigenvalues of $\mathbf{A}$ are finite. (see e.g. Abadir and Magnus (2005) for various properties of the matrix exponential transformation). For any real symmetric positive definite matrix $\mathbf{C}$, there exists a real symmetric $p \times p$ matrix $\mathbf{A}$ such that $\mathbf{C} = \exp(\mathbf{A})$, and the matrix $\mathbf{A}$ is obtained by the matrix logarithm transformation. Conversely, for any real symmetric matrix $\mathbf{A}$, $\mathbf{C} = \exp(\mathbf{A})$ is a symmetric positive definite matrix (Chiu, Leonard, and Tsui (1996)). If $\mathbf{A}$ is a $p \times p$ real symmetric matrix, there exists a $p \times p$ orthogonal matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{\Lambda}$ such that $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$ and

$$\exp(\mathbf{A}) = \mathbf{U}\left(\sum_{s=0}^{\infty} \frac{1}{s!}\mathbf{\Lambda}^s\right)\mathbf{U}' = \mathbf{U}\exp(\mathbf{\Lambda})\mathbf{U}'.$$

Now let $\boldsymbol{y}_t = (y_{1t}, \dots, y_{pt})'$ denote the p-dimensional asset return vector at time t, and let $\mathbf{H}_t$ denote the matrix logarithm of the variance-covariance matrix of $\boldsymbol{y}_t$. The MESV model with leverage effects is given by

$$\begin{aligned}
\boldsymbol{y}_t &= \exp(\mathbf{H}_t/2)\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \text{i.i.d. } \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p), \quad t = 1, \dots, n. && (4)\\
\mathbf{H}_{t+1} &= \mathbf{M} + \tilde{\mathbf{\Phi}} \odot (\mathbf{H}_t - \mathbf{M}) + \mathbf{E}_t, && (5)\\
\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} &\sim \text{i.i.d. } \mathcal{N}_{p+q}(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{I}_p & \mathbf{\Sigma}_{\varepsilon\eta} \\ \mathbf{\Sigma}_{\eta\varepsilon} & \mathbf{\Sigma}_{\eta\eta} \end{pmatrix}, \quad t = 1, \dots, n-1, && (6)\\
\boldsymbol{h}_1 &\sim \mathcal{N}_q(\boldsymbol{\mu}, \mathbf{\Sigma}_0), && (7)
\end{aligned}$$

where $\boldsymbol{\eta}_t = \text{vech}(\mathbf{E}_t)$, $q = p(p+1)/2$, $\mathbf{M} = \{\mu_{ij}\}$, and $\tilde{\mathbf{\Phi}} = \{\phi_{ij}\}$ are $p \times p$ symmetric matrices of parameters, and $\odot$ denotes the Hadamard product. For the identifiability, we set the covariance matrix of $\boldsymbol{\varepsilon}_t$ equal to $\mathbf{I}_p$.

If we let $\boldsymbol{h}_t = \text{vech}(\mathbf{H}_t) = (h_{11,t}, h_{21,t}, \ldots, h_{p1,t}, h_{22,t} \ldots, h_{pp,t})'$ denote the stacked column vector of the lower triangle elements of the $\mathbf{H}_t$, then the "vech form" of (5) is given by

$$\boldsymbol{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \tag{8}$$

with $\boldsymbol{\mu} = \text{vech}(\mathbf{M}) = (\mu_{11}, \mu_{21}, \ldots \mu_{p1}, \mu_{22}, \ldots, \mu_{pp})'$, $\boldsymbol{\Phi} = \text{diag}(\boldsymbol{\phi})$ (a diagonal matrix whose diagonal elements are equal to $\boldsymbol{\phi}$) and $\boldsymbol{\phi} = \text{vech}(\tilde{\boldsymbol{\Phi}}) = (\phi_{11}, \phi_{21}, \ldots \phi_{p1}, \phi_{22}, \ldots, \phi_{pp})'$. The number of parameters in the MESV model is $q(q+2p+5)/2$. The covariance matrix of the initial latent variable, $\boldsymbol{\Sigma}_0$, is assumed to satisfy a stationary condition such that

$$\text{vec}(\boldsymbol{\Sigma}_0) = (\mathbf{I}_{q^2} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi})^{-1}\text{vec}(\boldsymbol{\Sigma}_{\eta\eta}),$$

where $\otimes$ denotes the Kronecker product.

We let $\boldsymbol{\Sigma}_{\eta\eta} = \{\rho_{ij,\eta\eta}\sigma_{i,\eta\eta}\sigma_{j,\eta\eta}\}$, and $\boldsymbol{\Sigma}_{\varepsilon\eta} = \{\rho_{ij,\varepsilon\eta}\sigma_{j,\eta\eta}\}$ where $\sigma_{i,\eta\eta}$ is the standard deviation of η_{it} and $\rho_{ij,xy}$ is the correlation coefficient between x_{it} and y_{jt}. Further, for convenience, we use the notation $E(i,j) = k$ based on the relationship $\boldsymbol{\eta}_t = \text{vech}(\mathbf{E}_t)$ such that the (i,j)-th element of $\mathbf{E}_t$, $\mathbf{E}_t(i,j)$, is equal to the k-th element of $\boldsymbol{\eta}_t$, η_{kt} (*i.e.*, $E(1,1) = 1$, $E(2,1) = 2$, $\ldots$, $E(p,1) = p$, $E(2,2) = p+1, \ldots$, $E(p,p) = p(p+1)/2$). Thus, $Cov(\varepsilon_{lt}, \eta_{kt}) = \rho_{lk,\varepsilon\eta}\sigma_{k,\eta\eta}$ is equal to $Cov(\varepsilon_{lt}, \mathbf{E}_t(i,j)) = \rho_{lE(i,j),\varepsilon\eta}\sigma_{E(i,j),\eta\eta}$.

Remark 1. Due to the nonlinearity of the matrix exponential transformation, the interpretation of the (untransformed) parameters will depend on the dimension of $\boldsymbol{y}_t$. Thus, we consider estimates of transformed parameters to investigate the properties of interest, such as volatilities and correlations.

Remark 2. We can also interpret it terms of the principal component analysis. Using the diagonalization, $\mathbf{H}_t = \mathbf{U}_t\boldsymbol{\Lambda}_t\mathbf{U}_t'$, the element of $\mathbf{U}_t'\boldsymbol{y}_t$ is the principal component of the returns at time t, and the row vector of the loading matrix $\mathbf{U}_t'$ is a weight vector for the portfolio which represents the stock market principal component. We note that it is also possible to model $\mathbf{U}_t$ and $\boldsymbol{\Lambda}_t$ as in Plataniotis (2011) using Givens rotation matrix for the eigenvectors matrix.

3 Bayesian estimation and associated particle filter

In this section, we describe an efficient Bayesian estimation method and an associated particle filter to compute the likelihood for the MESV model. Let $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{h} = (\boldsymbol{h}_1', \ldots, \boldsymbol{h}_n')'$ and $Y_n = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$. Then the joint probability density function of Y_n and $\boldsymbol{h}$ given $\boldsymbol{\theta}$ for (4) and (8) is given by

$$\begin{aligned} f(Y_n, \boldsymbol{h}|\boldsymbol{\theta}) &= f(\boldsymbol{h}_1|\boldsymbol{\theta}) \prod_{t=1}^{n-1} f(\boldsymbol{y}_t, \boldsymbol{h}_{t+1}|\boldsymbol{h}_t, \boldsymbol{\theta}) f(\boldsymbol{y}_n|\boldsymbol{h}_n, \boldsymbol{\theta}) \qquad (9) \\ &\propto |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}} |\boldsymbol{\Sigma}|^{-\frac{n-1}{2}} |\boldsymbol{\Sigma}_{\varepsilon\varepsilon}|^{-\frac{1}{2}} \exp\left\{ \sum_{t=1}^{n} l_t - \frac{1}{2}(\boldsymbol{h}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{h}_1 - \boldsymbol{\mu}) \right\} \\ &\quad \times \exp\left[-\frac{1}{2} \sum_{t=1}^{n-1} \{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu})\}' \boldsymbol{\Sigma}_{\eta\eta}^{-1} \{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu})\} \right], \end{aligned}$$

where $l_t = -\frac{1}{2}\left\{\mathrm{tr}(\mathbf{H}_t) + (\boldsymbol{y}_t - \boldsymbol{\mu}_t)' \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{y}_t - \boldsymbol{\mu}_t)\right\}$ and

$$\begin{aligned} \boldsymbol{\mu}_t &= \exp(\mathbf{H}_t/2) \boldsymbol{m}_t, \qquad (10) \\ \boldsymbol{\Sigma}_t &= \exp(\mathbf{H}_t/2) \mathbf{S}_t \exp(\mathbf{H}_t/2), \qquad (11) \\ \boldsymbol{m}_t &= \boldsymbol{\Sigma}_{\varepsilon\eta} \boldsymbol{\Sigma}_{\eta\eta}^{-1} (\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu})) I(t < n), \qquad (12) \\ \mathbf{S}_t &= \mathbf{I}_p - \boldsymbol{\Sigma}_{\varepsilon\eta} \boldsymbol{\Sigma}_{\eta\eta}^{-1} \boldsymbol{\Sigma}_{\eta\varepsilon} I(t < n), \qquad (13) \end{aligned}$$

where $I(t < n)$ is an indication function which is equal to 1 if $t < n$ and 0 otherwise.

3.1 Prior distributions

For prior distributions of $(\boldsymbol{\phi}, \boldsymbol{\mu})$, we assume

$$\begin{aligned} \frac{\phi_{ij} + 1}{2} &\sim \mathcal{B}(a_{ij}, b_{ij}), \quad i = 1, \ldots, p, \quad j = 1, \ldots, i, \qquad (14) \\ \boldsymbol{\mu} &\sim \mathcal{N}_q(\boldsymbol{m}_0^*, \mathbf{V}_0^*), \qquad (15) \end{aligned}$$

where $\mathcal{B}(a, b)$ denotes a beta distribution with parameters a and b. To define a prior distribution of $\boldsymbol{\Sigma}$, we first denote

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} \end{bmatrix},$$

where $\mathbf{\Sigma}^{11}$, $\mathbf{\Sigma}^{12}$ and $\mathbf{\Sigma}^{22}$ are $p \times p$, $p \times q$ and $q \times q$ matrices. Noting that $\mathbf{\Sigma}^{11} = \mathbf{I}_p + \mathbf{\Sigma}^{12}\mathbf{\Sigma}^{22-1}\mathbf{\Sigma}^{21}$, we assume the prior distributions such that

$$\mathrm{vec}(\mathbf{\Sigma}^{21})|\mathbf{\Sigma}^{22} \sim \mathcal{N}_{pq}(\mathrm{vec}(\mathbf{\Sigma}^{22}\mathbf{\Delta}_0), \mathbf{\Omega}_0 \otimes \mathbf{\Sigma}^{22}), \quad \mathbf{\Sigma}^{22} \sim \mathcal{W}(n_0, \mathbf{R}_0), \tag{16}$$

where $\mathcal{W}(n, \mathbf{R})$ denotes Wishart distribution with parameters n and $\mathbf{R}$. We construct the natural conjugate prior of $(\mathbf{\Sigma}^{21}, \mathbf{\Sigma}^{22})$ for the likelihood function of the multivariate normal distribution with $\Sigma_{\varepsilon\varepsilon} = I_p$ (in our case, excluding some modification). In our empirical study, we assumed a fairly at prior by setting $\mathbf{\Delta}_0 = 0$ and $\mathbf{\Omega}_0 = 5I_p$ to reflect that we do not have sufficient information. For ϕ_{ij}, we assume $|\phi_{ij}| < 1$ $(0 < (\phi_{ij} + 1)/2 < 1)$. Thus assuming Beta prior distribution for such a parameter is quite popular in empirical studies. Also, it is stated that $vec(\mathbf{\Sigma}_0)$ is based on a stationary condition.

3.2 MCMC algorithm

Using Equations (9), (14), (15) and (16), we obtain the joint posterior density function of $(\boldsymbol{\theta}, \boldsymbol{h})$ given by

$$\begin{aligned}
&\pi(\boldsymbol{\theta}, \boldsymbol{h}|Y_n) \\
&\propto \; f(Y_n, \boldsymbol{h}|\boldsymbol{\theta}) \times \prod_{i=1}^{p}\prod_{j=1}^{i}(1+\phi_{ij})^{a_{ij}-1}(1-\phi_{ij})^{b_{ij}-1} \times f_N(\boldsymbol{\mu}|\boldsymbol{m}_0^*, \mathbf{V}_0^*) \\
&\quad \times f_N\left(\mathrm{vec}(\mathbf{\Sigma}^{21})|\mathrm{vec}(\mathbf{\Sigma}^{22}\mathbf{\Delta}_0), \mathbf{\Omega}_0 \otimes \mathbf{\Sigma}^{22}\right) \times |\mathbf{\Sigma}^{22}|^{\frac{n-q-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_0^{-1}\mathbf{\Sigma}^{22}\right)\right\}, \quad (17)
\end{aligned}$$

where $f_N(\cdot|\boldsymbol{\mu}, \mathbf{\Sigma})$ denotes a normal density with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. To obtain the posterior quantities of the parameters $\boldsymbol{\theta}$ and volatility variables $\{\boldsymbol{h}_t\}_{t=1}^n$ from the posterior distribution, we implement the MCMC algorithm in six blocks:

1. Initialize $\boldsymbol{h}, \boldsymbol{\phi}, \boldsymbol{\mu}, \mathbf{\Sigma}$.
2. Generate $\boldsymbol{h}|\boldsymbol{\phi}, \boldsymbol{\mu}, \mathbf{\Sigma}, Y_n$.
3. Generate $\boldsymbol{\mu}|\boldsymbol{\phi}, \mathbf{\Sigma}, \boldsymbol{h}, Y_n$.
4. Generate $\mathbf{\Sigma}|\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{h}, Y_n$.
5. Generate $\boldsymbol{\phi}|\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{h}, Y_n$.
6. Go to Step 2.

3.2.1 Generation of $\boldsymbol{h}$

As is often pointed out in the literature, it is important to sample the latent volatility variables $\{\boldsymbol{h}_t\}_{t=1}^n$ in an efficient way. The simple sample method, which samples one $\boldsymbol{h}_t$ at a time given the other $\boldsymbol{h}_s$'s and parameters, is known to be inefficient, often producing highly autocorrelated MCMC samples. This is because the estimates of autoregressive parameters ϕ_i are often found to be very close to one in empirical studies. Thus, we propose the sampling method based on a multi-move sampler which samples a set of $\boldsymbol{h}_t$'s as one block at a time (see e.g. Shephard and Pitt (1997), Watanabe and Omori (2004), Omori and Watanabe (2008), Ishihara and Omori (2012)). We first describe a simple algorithm which we call a single-move sampler as we use it as a benchmark to evaluate the estimation efficiencies of the multi-move algorithm.

Single-move sampler. Let $\boldsymbol{\Sigma_h} = \boldsymbol{\Sigma}_{\eta\eta} - \boldsymbol{\Sigma}_{\eta\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\eta}$ and $\boldsymbol{\mu}_{h,t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu}) + \boldsymbol{\Sigma}_{\eta\varepsilon}\exp(-\mathbf{H}_t/2)\boldsymbol{y}_t$. Then, the conditional posterior density of $\boldsymbol{h}_t$ given $\{\boldsymbol{h}_s\}_{s\neq t}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ is

$$\pi(\boldsymbol{h}_t|\{\boldsymbol{h}_s\}_{s\neq t}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{Y}_n) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{h}_t - \boldsymbol{\gamma}_t)'\boldsymbol{\Gamma}_t^{-1}(\boldsymbol{h}_t - \boldsymbol{\gamma}_t) + g(\boldsymbol{h}_t)\right\},$$

where

$$\begin{aligned} g(\boldsymbol{h}_t) &= -\frac{1}{2}\mathrm{tr}(\mathbf{H}_t) - \frac{1}{2}\boldsymbol{y}_t'\exp(-\mathbf{H}_t/2)\mathbf{S}_t^{-1}\exp(-\mathbf{H}_t/2)\boldsymbol{y}_t \\ &\quad + (\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu}))'\boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\exp(-\mathbf{H}_t/2)\boldsymbol{y}_t I(t<n), \end{aligned}$$

and

$$\boldsymbol{\Gamma}_t = \begin{cases} (\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1}\boldsymbol{\Phi} + \boldsymbol{\Sigma}_0^{-1})^{-1}, & t=1, \\ (\boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1} + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1}\boldsymbol{\Phi})^{-1}, & 1<t<n, \\ \boldsymbol{\Sigma}_{\boldsymbol{h}}, & t=n, \end{cases}$$

$$\boldsymbol{\gamma}_t = \begin{cases} \boldsymbol{\Gamma}_1\Phi\boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1}(\boldsymbol{h}_2 - (\mathbf{I}_p - \boldsymbol{\Phi})\boldsymbol{\mu}), & t=1, \\ \boldsymbol{\Gamma}_t(\Phi\boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1}(\boldsymbol{h}_{t+1} - (\mathbf{I}_p - \boldsymbol{\Phi})\boldsymbol{\mu}) + \boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1}\boldsymbol{\mu}_{h,t}), & 1<t<n, \\ \boldsymbol{\mu}_{h,t}, & t=n. \end{cases}$$

We generate a candidate $\boldsymbol{h}_t^\dagger$ from $\boldsymbol{h}_t^\dagger \sim \mathcal{N}_q(\boldsymbol{\gamma}_t, \boldsymbol{\Gamma}_t)$ and accept it with probability

$$\min\left\{\exp\{g(\boldsymbol{h}_t^\dagger) - g(\boldsymbol{h}_t)\}, 1\right\}, \quad t = 1, \ldots, n.$$

Multi-move sampler. In this algorithm, we first divide $\boldsymbol{h}$ into several blocks, and sample one block at a time from its conditional posterior distribution given other blocks. Using the Taylor expansion of the logarithm of the conditional posterior density around the conditional posterior mode, we derive a candidate distribution as a posterior distribution for some linear Gaussian state space model to exploit various smoothing and simulation algorithms as in Omori and Watanabe (2008). Details are given in Appendix A.

Mixture of the single-move and multi-move samplers. Although the single-move sampler is easier to implement, it is inefficient in the sense that it produces highly autocorrelated MCMC samples as shown in Section 4. On the other hand, the multi-move sampler is efficient, but it is more complicated than the single-move sampler, which may need more computational cost and time. Thus, we could propose a mixture of the single-move sampler and the multi-move sampler. That is, we implement the single-move sampler with probability p (say 0.9) and the multi-move sampler with probability $1-p$ (0.1).

3.2.2 Generation of $\boldsymbol{\Sigma}$

Since the $p \times p$ leading principal submatrix of $\boldsymbol{\Sigma}$ is an identity matrix, we first generate $\boldsymbol{\Sigma}^{22}$ and then sample $\mathrm{vec}(\boldsymbol{\Sigma}^{21})$ conditional on $\boldsymbol{\Sigma}^{22}$, to conduct MH algorithm using the property of Wishart distribution (see, e.g., Theorem 3.3.9 of Gupta and Nagar (2000)).
Let

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{R}^{11} & \mathbf{R}^{12} \\ \mathbf{R}^{21} & \mathbf{R}^{22} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{n-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' & \sum_{t=1}^{n-1} \boldsymbol{\varepsilon}_t \boldsymbol{\eta}_t' \\ \sum_{t=1}^{n-1} \boldsymbol{\eta}_t \boldsymbol{\varepsilon}_t' & \sum_{t=1}^{n-1} \boldsymbol{\eta}_t \boldsymbol{\eta}_t' \end{bmatrix},$$

where $\mathbf{R}^{11}$, $\mathbf{R}^{12} = \mathbf{R}^{21\prime}$, $\mathbf{R}^{22}$ are $p \times p$, $p \times q$, $q \times q$ matrices, $\boldsymbol{\varepsilon}_t = \exp(-\mathbf{H}_t/2)\boldsymbol{y}_t$ and $\boldsymbol{\eta}_t = \boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu})$. Using $\mathrm{tr}(\mathbf{AB}) = \mathrm{vec}(\mathbf{A}')'\mathrm{vec}(\mathbf{B})$ and $\mathrm{vec}(\mathbf{AXB}) = (\mathbf{B}' \otimes \mathbf{A})\mathrm{vec}(\mathbf{X})$, for $\mathbf{X}(n \times n)$, $\mathbf{A}(m \times n)$ and $\mathbf{B}(n \times m)$, the joint conditional posterior probability density of $\mathbf{\Sigma}^{12}$ and $\mathbf{\Sigma}^{22}$ is obtained as follows.

$$
\begin{aligned}
&\pi(\mathbf{\Sigma}^{12},\mathbf{\Sigma}^{22}|\mathbf{\Phi},\boldsymbol{h},Y_n) = \pi(\mathbf{\Sigma}^{22}|\mathbf{\Phi},\boldsymbol{h},Y_n)\pi(\mathrm{vec}(\mathbf{\Sigma}^{21})|\mathbf{\Sigma}^{22},\mathbf{\Phi},\boldsymbol{h},Y_n)\\
&\propto \; h(\mathbf{\Sigma})\times|\mathbf{\Sigma}|^{-\frac{n-1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\right)\right\}\times|\mathbf{\Sigma}^{22}|^{\frac{n_0-q-1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_0^{-1}\mathbf{\Sigma}^{22}\right)\right\}\\
&\quad\times f_N(\mathrm{vec}(\mathbf{\Sigma}^{21})|\mathrm{vec}(\mathbf{\Sigma}^{22}\mathbf{\Delta}_0),\mathbf{\Omega}_0\otimes\mathbf{\Sigma}^{22})\\
&\propto \; h(\mathbf{\Sigma})\times|\mathbf{\Sigma}^{22}|^{\frac{n_0+n-1-q-1}{2}}\exp\left[-\frac{1}{2}\mathrm{tr}\left\{(\mathbf{R}^{22}-\mathbf{R}^{21}\mathbf{R}^{11\,-1}\mathbf{R}^{12}+\mathbf{R}_0^{-1})\mathbf{\Sigma}^{22}\right\}\right]\\
&\quad\times\exp\left\{-\frac{1}{2}\mathrm{vec}(\mathbf{\Sigma}^{21}+\mathbf{\Sigma}^{22}\mathbf{R}^{21}\mathbf{R}^{11\,-1})'(\mathbf{R}^{11}\otimes\mathbf{\Sigma}^{22\,-1})\mathrm{vec}(\mathbf{\Sigma}^{21}+\mathbf{\Sigma}^{22}\mathbf{R}^{21}\mathbf{R}^{11\,-1})\right\}\\
&\quad\times f_N(\mathrm{vec}(\mathbf{\Sigma}^{21})|\mathrm{vec}(\mathbf{\Sigma}^{22}\mathbf{\Delta}_0),\mathbf{\Omega}_0\otimes\mathbf{\Sigma}^{22})\\
&\propto \; h(\mathbf{\Sigma})\times|\mathbf{\Sigma}^{22}|^{\frac{n_1-q-1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_1^{-1}\mathbf{\Sigma}^{22}\right)\right\}\times f_N(\mathrm{vec}(\mathbf{\Sigma}^{21})|\mathrm{vec}(\mathbf{\Sigma}^{22}\mathbf{\Delta}_1),\mathbf{\Omega}_1\otimes\mathbf{\Sigma}^{22})
\end{aligned}
$$

where $n_1 = n_0 + n - 1$, $\mathbf{R}_1^{-1} = \mathbf{R}^{22} - \mathbf{\Delta}_1\mathbf{\Omega}_1^{-1}\mathbf{\Delta}_1' + \mathbf{R}_0^{-1} + \mathbf{\Delta}_0\mathbf{\Omega}_0^{-1}\mathbf{\Delta}_0'$ and

$$
\begin{aligned}
\mathbf{\Omega}_1 &= \left(\mathbf{R}^{11}+\mathbf{\Omega}_0^{-1}\right)^{-1},\\
\mathbf{\Delta}_1 &= \left(-\mathbf{R}^{21}+\mathbf{\Delta}_0\mathbf{\Omega}_0^{-1}\right)\mathbf{\Omega}_1\\
h(\mathbf{\Sigma}) &= |\mathbf{\Sigma}_0|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{h}_1-\boldsymbol{\mu})'\mathbf{\Sigma}_0^{-1}(\boldsymbol{h}_1-\boldsymbol{\mu})\right\}.
\end{aligned}
$$

Thus, we generate a candidate $\mathbf{\Sigma}^\dagger$ in three steps.

1. Draw $\mathbf{\Sigma}^{22\dagger} \sim \mathcal{W}(n_1, \mathbf{R}_1)$.
2. Draw $\mathrm{vec}(\mathbf{\Sigma}^{21\dagger})|\mathbf{\Sigma}^{22\dagger} \sim \mathcal{N}_{pq}(\mathrm{vec}(\mathbf{\Sigma}^{22\dagger}\mathbf{\Delta}_1), \mathbf{\Omega}_1 \otimes \mathbf{\Sigma}^{22\dagger})$.
3. Compute $\mathbf{\Sigma}^\dagger_{\varepsilon\eta} = -\mathbf{\Sigma}^{12\dagger}\mathbf{\Sigma}^{22\dagger-1}$ and $\mathbf{\Sigma}^\dagger_{\eta\eta} = \mathbf{\Sigma}^{22\dagger-1} + \mathbf{\Sigma}^\dagger_{\eta\varepsilon}\mathbf{\Sigma}^\dagger_{\varepsilon\eta}$.

and accept it with probability

$$\min\left\{\frac{h(\mathbf{\Sigma}^\dagger)}{h(\mathbf{\Sigma})}, 1\right\}.$$

3.2.3 Generation of ($\boldsymbol{\mu}$, ϕ)

Generation of $\boldsymbol{\mu}$. The conditional posterior distribution of $\boldsymbol{\mu}$ is

$$\boldsymbol{\mu}|\mathbf{\Sigma},\mathbf{\Phi},\boldsymbol{h},Y_n \sim \mathcal{N}_q(\boldsymbol{m}_1^*, \mathbf{V}_1^*),$$

where

$$\mathbf{V}_1^* = \left(\mathbf{\Sigma}_0^{-1} + \sum_{t=1}^{n-1}(\mathbf{I}_p - \mathbf{\Phi})'\mathbf{\Sigma}_{\boldsymbol{h}}^{-1}(\mathbf{I}_p - \mathbf{\Phi}) + \mathbf{V}_0^{*\ -1}\right)^{-1},$$

$$\boldsymbol{m}_1^* = \mathbf{V}_1^*\left[\sum_{t=1}^{n-1}(\mathbf{I}_p - \mathbf{\Phi})'\mathbf{\Sigma}_{\boldsymbol{h}}^{-1}\left\{\boldsymbol{h}_{t+1} - \mathbf{\Phi}\boldsymbol{h}_t - \mathbf{\Sigma}_{\varepsilon\eta}\exp(-\mathbf{H}_t/2)\boldsymbol{y}_t\right\} + \mathbf{\Sigma}_0^{-1}\boldsymbol{h}_1 + \mathbf{V}_0^{*\ -1}\boldsymbol{m}_0^*\right].$$

Generation of $\boldsymbol{\phi}$. Let $\mathbf{A} = \sum_{t=1}^{n-1}(\boldsymbol{h}_t - \boldsymbol{\mu})(\boldsymbol{h}_t - \boldsymbol{\mu})'$, $\mathbf{B} = \sum_{t=1}^{n-1}\{(\boldsymbol{h}_t - \boldsymbol{\mu})\boldsymbol{y}_t'\exp(-\mathbf{H}_t/2)\mathbf{\Sigma}^{12} + (\boldsymbol{h}_t - \boldsymbol{\mu})(\boldsymbol{h}_{t+1} - \boldsymbol{\mu})'\mathbf{\Sigma}^{22}\}$ 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

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

where $\boldsymbol{\mu}_{\boldsymbol{\phi}} = \mathbf{\Sigma}_{\boldsymbol{\phi}}\boldsymbol{b}$, $\mathbf{\Sigma}_{\boldsymbol{\phi}}^{-1} = \mathbf{\Sigma}^{22} \odot \mathbf{A}$. 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, $\boldsymbol{\phi}^\dagger \sim \mathcal{TN}_R(\boldsymbol{\mu}_{\boldsymbol{\phi}},\mathbf{\Sigma}_{\boldsymbol{\phi}})$, $R = \{\boldsymbol{\phi} : |\phi_j| < 1, j = 1,\ldots,p\}$ and accept it with probability $\min\{k(\boldsymbol{\phi}^\dagger)/k(\boldsymbol{\phi}), 1\}$.

Remark 3. The MH acceptance ratio may become low, for example, when the elements of $\mathbf{\Phi}$ are nearly one. To improve the MH algorithm by reducing the number of rejected proposals, Mira (2001) proposes the delaying rejection algorithm. We note that we can apply the delaying rejection algorithm to our independent MH algorithm for $\mathbf{\Phi}, \mathbf{\Sigma}$, and $\boldsymbol{h}$. The performance of the delaying rejection algorithm is discussed in Section 4. See Appendix B for the details of the algorithm.

3.3 Associated particle filter

We describe the associated auxiliary particle filter introduced by Pitt and Shephard (1999) for the MESV model to compute the log likelihood function. In Section 5, we use this algorithm to calculate DIC (deviance information criterion proposed by Spiegelhalter, Best, Carlin, and van der Linde (2002)). Let

$$
\begin{aligned}
f(\boldsymbol{y}_t|\boldsymbol{h}_t) &= (2\pi)^{-\frac{p}{2}}\,|\exp(\mathbf{H}_t)\,|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\boldsymbol{y}_t'\exp(-\mathbf{H}_t)\boldsymbol{y}\right\},\\
f(\boldsymbol{h}_{t+1}|\boldsymbol{y}_t,\boldsymbol{h}_t,\boldsymbol{\theta}) &= (2\pi)^{-\frac{q}{2}}\,|\boldsymbol{\Sigma}_h|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{h}_{t+1}-\boldsymbol{\mu}_{h,t+1})'\boldsymbol{\Sigma}_h^{-1}(\boldsymbol{h}_{t+1}-\boldsymbol{\mu}_{h,t+1})\right\},
\end{aligned}
$$

and $f(\boldsymbol{h}_t|Y_t,\boldsymbol{\theta})$ denote a conditional density of $\boldsymbol{h}_t$ given $(Y_t,\boldsymbol{\theta})$. Then the conditional joint density function of $\boldsymbol{h}_{t+1},\boldsymbol{h}_t$, given $(Y_{t+1},\boldsymbol{\theta})$ is

$$
f(\boldsymbol{h}_{t+1},\boldsymbol{h}_t|Y_{t+1},\boldsymbol{\theta}) \;\propto\; f(\boldsymbol{y}_{t+1}|\boldsymbol{h}_{t+1})f(\boldsymbol{h}_{t+1}|\boldsymbol{y}_t,\boldsymbol{h}_t,\boldsymbol{\theta})f(\boldsymbol{h}_t|Y_t,\boldsymbol{\theta}).
$$

We first construct an importance function to sample from the conditional joint distribution. Let $\hat{f}(\boldsymbol{h}_t|Y_t,\boldsymbol{\theta})$ denote a discrete probability mass function approximating $f(\boldsymbol{h}_t|Y_t,\boldsymbol{\theta})$ and

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

where

$$
\begin{aligned}
g(\boldsymbol{h}_t^i|Y_{t+1},\boldsymbol{\theta}) &= \frac{f(\boldsymbol{y}_{t+1}|\boldsymbol{\mu}_{h,t+1}^i)\hat{f}(\boldsymbol{h}_t^i|Y_t,\boldsymbol{\theta})}{\sum_{j=1}^I f(\boldsymbol{y}_{t+1}|\boldsymbol{\mu}_{h,t+1}^j)\hat{f}(\boldsymbol{h}_t^j|Y_t,\boldsymbol{\theta})},\\
f(\boldsymbol{y}_{t+1}|\boldsymbol{\mu}_{h,t+1}^i) &= (2\pi)^{-\frac{p}{2}}\,|\exp(\mathbf{M}_{h,t+1}^i)|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\boldsymbol{y}_{t+1}'\exp(-\mathbf{M}_{h,t+1}^i)\boldsymbol{y}_{t+1}\right\},\\
\boldsymbol{\mu}_{h,t+1}^i &= \boldsymbol{\mu}+\boldsymbol{\Phi}(\boldsymbol{h}_t^i-\boldsymbol{\mu})+\boldsymbol{\Sigma}_{\eta\varepsilon}\exp(-\mathbf{H}_t^i/2)\boldsymbol{y}_t,
\end{aligned}
$$

and $\mathbf{M}_{h,t+1}^i$ is a symmetric matrix such that $\boldsymbol{\mu}_{h,t+1}^i = \mathrm{vech}(\mathbf{M}_{h,t+1}^i)$.

Using this importance function, we implement the auxiliary particle filter as follows.

1. Set $t = 1$.
   (a) Generate $\boldsymbol{h}_1^i \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}_0)$ $(i = 1,\ldots,I)$.
   (b) Compute $w_i = f(\boldsymbol{y}_1|\boldsymbol{h}_1^i,\boldsymbol{\theta})$ and save $\bar{w}_1 = \frac{1}{I}\sum_{i=1}^I w_i$.
   (c) Let $\hat{f}(\boldsymbol{h}_1^i|Y_1,\boldsymbol{\theta}) = \pi_1^i = w_i/\sum_{j=1}^I w_j$ $(i = 1,\ldots,I)$.
2. Generate $\left(\boldsymbol{h}_{t+1}^i,\boldsymbol{h}_t^i\right) \sim g(\boldsymbol{h}_{t+1},\boldsymbol{h}_t^i|Y_{t+1},\boldsymbol{\theta})$ $(i = 1,\ldots,I)$:
   (a) Compute $\boldsymbol{\mu}_{h,t+1}^i = \boldsymbol{\mu}+\boldsymbol{\Phi}(\boldsymbol{h}_t^i-\boldsymbol{\mu})+\boldsymbol{\Sigma}_{\eta\varepsilon}\exp\left(-\mathbf{H}_t^i/2\right)\boldsymbol{y}_t$.
   (b) Generate $\boldsymbol{h}_t^i \sim g(\boldsymbol{h}_t^i|Y_{t+1},\boldsymbol{\theta})$.

(c) Generate $\boldsymbol{h}_{t+1}^i \sim f(\boldsymbol{h}_{t+1}|\boldsymbol{y}_t, \boldsymbol{h}_t^i, \boldsymbol{\theta})$.

Then compute

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

and save

$$
\overline{w}_t = \frac{1}{I} \sum_{i=1}^{I} w_i.
$$

Further let $\hat{f}(\boldsymbol{h}_{t+1}^i|Y_{t+1}, \boldsymbol{\theta}) = \pi_{t+1}^i = w_i / \sum_{j=1}^{I} w_j \ (i = 1, \ldots, I)$.

3. Set $t \leftarrow t+1$. Go to Step 2.

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

Remark 4. We may implement MH algorithm based on the sequential Monte Carlo approach using the likelihood computed by the particle filter. However, since the dimension of the parameter vector is very high in our multivariate model, we do not pursue this alternative approach in this paper.

4 Illustrative example with simulated data

This section shows the efficiency of our proposed method using a simulated data. Two examples are given where we generate $n = 4,000$ observations with $p = 3$. Prior distributions are assumed to be as follows.

$$
\begin{aligned}
\boldsymbol{\mu} &\sim \mathcal{N}_q(\mathbf{0}, 5\mathbf{I}_q), \\
\frac{\phi_{ij}+1}{2} &\sim \begin{cases} \mathcal{B}(20, 3/2), & i = j, \\ \mathcal{B}(1,1), & i \neq j, \end{cases} \\
\boldsymbol{\Sigma}^{22} &\sim \mathcal{W}(6, (6\boldsymbol{\Sigma}^{22*})^{-1}), \\
\mathrm{vec}(\boldsymbol{\Sigma}^{21})|\boldsymbol{\Sigma}^{22} &\sim \mathcal{N}_{pq}(\mathbf{0}, (5\mathbf{I}_p) \otimes \boldsymbol{\Sigma}^{22}),
\end{aligned}
$$

where $\boldsymbol{\Sigma}^{22*}$ is a true covariance matrix satisfying $E(\boldsymbol{\Sigma}^{22}) = \boldsymbol{\Sigma}^{22*}$. The mean and the standard deviation of the prior distribution of ϕ_{ii}, $j = 1, 2, 3$ are set 0.86 and 0.11 respectively.

Using the multi-move (single-move) sampler, we draw 110,000 (550,000) posterior samples and discard the first 10,000 (50,000) samples as burn-in periods.

Example 1. First, we consider the following MESV model to replicate the dynamics of the stock return series:

$$\mu_{ij} = \begin{cases} 0.5, & i = j, \\ 0.2, & i \neq j, \end{cases}, \quad \phi_{ij} = \begin{cases} 0.97, & i = j, \\ 0.85, & i \neq j, \end{cases}$$

$$\rho_{iE(j,k),\varepsilon\eta} \equiv Corr(\varepsilon_{it}, \mathbf{E}_t(j,k)) = \begin{cases} -0.3, & i = j = k, \\ -0.1, & \text{otherwise}, \end{cases}$$

$$\rho_{E(i,j)E(k,l),\eta\eta} \equiv Corr(\mathbf{E}_t(i,j), \mathbf{E}_t(k,l)) = \begin{cases} 0.3, & i = j \neq k = l \\ 0.1, & \text{otherwise}, \end{cases}$$

$$\sigma_{E(i,j),\eta\eta} \equiv \sqrt{Var(\mathbf{E}_t(i,j))} = \begin{cases} 0.2, & i = j \\ 0.15, & \text{otherwise}, \end{cases}$$

which are based on typical values in our empirical studies where $E(1,1) = 1$, $E(2,1) = 2$, $E(3,1) = 3$, $E(2,2) = 4$, $E(3,2) = 5$ and $E(3,3) = 6$.

Tables 1, 2 and 3 show the estimation summaries for all parameters via 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[1]. The inefficiency factors for the single-move sampler are about three times larger than those for the multi-move sampler. Further, Table 7 shows inefficiency factors for $\mathbf{H}_{2000}$. For these latent variables, inefficiency factors forthe single-move sampler are about seventeen times larger than those for the multi-move sampler. This implies that our proposed multi-move sampler is highly efficient than the single-move sampler as we expected.

Tables 4, 5, 6 and 7 compare the inefficiency factors of single-move and multi-move sampler with and without delaying rejection algorithm (DR). For the mixture of the single-move and the multi-move samplers, we use multi-move (single-move) sampler with probability 0.1 (0.9) and generate 120,000 samples and use 100,000 samples. The inefficiency factors of the

[1]The inefficiency factors are also shown for the multi-move sampler and the single-move sampler. The inefficiency factor is the ratio of the numerical variance of the estimate from the MCMC samples relative to that from hypothetical uncorrelated samples, and is defined as $1 + 2\sum_{s=1}^{\infty} \rho_s$ where ρ_s is the sample autocorrelation at lag s. It suggests the relative number of correlated draws necessary to attain the same variance of the posterior sample mean from the uncorrelated draws (Chib (2001)).

single-move sampler are much larger than those of the multi-move sample with and without DR especially for latent variables[2]. We note that the DR algorithm reduces the inefficiency factors but that it requires additional computational time. The single-move sampler with DR is less efficient than the multi-move sampler without DR, while the mixture of the single-move and the multi-move samplers with DR shows similar inefficiency factors to those of the multi-move sampler without DR. Taking account of the computational time, the mixture sampler may be preferred.

Table 1: Posterior means, 95% credible intervals, and inefficiency factors.

Param.	True	ij	Mean	95% interval	Inefficiency multi	[single]
ϕ_{ij}	0.97	11	0.967	[0.955, 0.978]	149	[334]
		22	0.964	[0.951, 0.975]	119	[115]
		33	0.975	[0.965, 0.984]	105	[437]
	0.85	21	0.802	[0.701, 0.877]	288	[795]
		31	0.841	[0.734, 0.910]	511	[1622]
		32	0.837	[0.751, 0.900]	332	[435]
μ_{ij}	0.5	11	0.572	[0.390, 0.755]	6	[38]
		22	0.456	[0.266, 0.649]	6	[75]
		33	0.485	[0.225, 0.744]	3	[42]
	0.2	21	0.201	[0.159, 0.243]	20	[465]
		31	0.203	[0.158, 0.247]	17	[355]
		32	0.204	[0.156, 0.252]	18	[304]
$\sigma_{E(i,j),\eta\eta}$	0.2	11	0.189	[0.158, 0.223]	285	[572]
		22	0.216	[0.185, 0.251]	219	[306]
		33	0.211	[0.181, 0.242]	180	[657]
	0.15	21	0.158	[0.118, 0.204]	460	[1182]
		31	0.145	[0.106, 0.194]	646	[1648]
		32	0.175	[0.129, 0.230]	438	[761]

$$\sigma_{E(i,j),\eta\eta} = \sqrt{\mathrm{Var}(\mathbf{E}_t(i,j))}$$

[2]The acceptance rates for $\boldsymbol{\phi}$, $\boldsymbol{\Sigma}$ and $\mathbf{H}_{2000}$ in MH algorithms are 0.93, 0.97 and 0.55 (0.94, 0.98 and 0.70) for the multi-move sampler (single-move sampler). For MH algorithms with DR, they are 0.99, 0.99 and 0.92 (0.99, 0.99 and 0.93) for the multi-move sampler (single-move sampler), and 0.99, 0.99, and 0.93 for the mixture of the multi-move and single-move samplers. The elapsed times for the simulation using the multi-move, single-move, multi-move with DR, single-move with DR, mixture with DR samplers are 9.38, 0.56, 17.81, 1.13 and 2.96 hours per 10,000 iterations with Richland AMD A10-6800K Black Edition (4.1GHz).

Table 2: Posterior means, 95% credible intervals and inefficiency factors.

$$\rho_{iE(j,k),\varepsilon\eta} = \mathrm{Corr}(\varepsilon_{it}, \mathbf{E}_t(j,k))$$

True	i jk	Mean	95% interval	Inefficiency multi	[single]
-0.3	1 11	-0.277	[-0.388,-0.159]	61	[133]
	2 22	-0.255	[-0.372,-0.134]	84	[126]
	3 33	-0.374	[-0.479,-0.264]	81	[220]
-0.1	1 21	0.046	[-0.100, 0.188]	73	[164]
	1 31	-0.041	[-0.190, 0.107]	96	[218]
	1 22	-0.113	[-0.226, 0.002]	34	[143]
	1 32	-0.035	[-0.163, 0.092]	63	[206]
	1 33	0.005	[-0.114, 0.123]	66	[155]
	2 11	-0.018	[-0.147, 0.111]	60	[231]
	2 21	-0.017	[-0.172, 0.131]	94	[164]
	2 31	-0.172	[-0.328,-0.022]	86	[289]
	2 32	-0.072	[-0.207, 0.061]	65	[189]
	2 33	-0.067	[-0.192, 0.057]	71	[184]
	3 11	-0.092	[-0.217, 0.030]	61	[159]
	3 21	-0.194	[-0.341,-0.047]	87	[113]
	3 31	-0.118	[-0.263, 0.028]	91	[293]
	3 22	-0.080	[-0.192, 0.035]	51	[174]
	3 32	-0.027	[-0.155, 0.104]	72	[149]

Table 3: Posterior means, 95% credible intervals and inefficiency factors.

$$\rho_{E(i,j)E(k,l),\eta\eta} = \mathrm{Corr}(\mathbf{E}_t(i,j)), \mathbf{E}_t(k,l))$$

True	ij kl	Mean	95% interval	Inefficiency multi	[single]
0.3	11 22	0.248	[0.080, 0.411]	102	[340]
	11 33	0.332	[0.170, 0.481]	135	[382]
	22 33	0.266	[0.104, 0.420]	68	[186]
0.1	11 21	-0.055	[-0.266, 0.177]	184	[971]
	11 31	0.150	[-0.066, 0.355]	215	[662]
	11 32	0.064	[-0.135, 0.267]	138	[109]
	21 31	0.092	[-0.199, 0.377]	301	[110]
	21 22	-0.034	[-0.239, 0.188]	205	[133]
	21 32	-0.127	[-0.386, 0.142]	239	[100]
	21 33	0.203	[-0.028, 0.424]	177	[106]
	31 22	0.055	[-0.173, 0.269]	234	[111]
	31 32	0.041	[-0.192, 0.275]	213	[161]
	31 33	0.213	[0.010, 0.413]	224	[931]
	22 32	-0.006	[-0.203, 0.189]	137	[536]
	32 33	0.053	[-0.134, 0.239]	138	[830]

Table 4: Inefficiency factors.

Param.	ij	Inefficiency multi	single	multi DR	single DR	mixed DR
ϕ_{ij}	11	149	334	95	191	198
	22	119	115	114	178	57
	33	105	437	113	151	71
	21	288	795	458	598	353
	31	511	1622	332	1149	560
	32	332	435	290	506	247
μ_{ij}	11	6	38	6	28	22
	22	6	75	6	20	10
	33	3	42	4	15	6
	21	20	465	20	91	20
	31	17	355	18	125	43
	32	18	304	6	50	38
$\sigma_{E(i,j),\eta\eta}$	11	285	572	182	399	296
	22	219	306	199	321	146
	33	180	657	177	248	161
	21	460	1182	583	957	478
	31	646	1648	510	1099	819
	32	438	761	422	668	324

$$\sigma_{E(i,j),\eta\eta} = \sqrt{\mathrm{Var}(\mathbf{E}_t(i,j))}$$

Table 5: Inefficiency factors.

$$\rho_{iE(j,k),\varepsilon\eta} = \mathrm{Corr}(\varepsilon_{it}, \mathbf{E}_t(j,k))$$

$i\ jk$	Inefficiency multi	single	multi DR	single DR	mixed DR
1 11	61	133	37	137	68
2 22	84	126	41	121	47
3 33	81	220	58	105	48
1 21	73	164	55	163	111
1 31	96	218	101	268	105
1 22	34	143	23	96	71
1 32	63	206	33	124	64
1 33	66	155	54	84	90
2 11	60	231	42	256	70
2 21	94	164	84	169	98
2 31	86	289	76	212	150
2 32	65	189	46	90	52
2 33	71	184	47	110	52
3 11	61	159	39	146	66
3 21	87	113	80	270	91
3 31	91	293	85	290	108
3 22	51	174	50	118	29
3 32	72	149	46	106	30

Table 6: Inefficiency factors.

$\rho_{E(i,j)E(k,l),\eta\eta} = \mathrm{Corr}(\mathbf{E}_t(i,j)), \mathbf{E}_t(k,l))$

ij kl	Inefficiency				
	multi	single	multi DR	single DR	mixed DR
11 22	102	340	64	198	136
11 33	135	382	160	325	156
22 33	68	186	78	232	124
11 21	184	971	130	403	203
11 31	215	662	122	435	162
11 32	138	109	151	340	370
21 31	301	110	263	405	457
21 22	205	133	192	330	218
21 32	239	100	215	595	324
21 33	177	106	201	353	272
31 22	234	111	110	422	201
31 32	213	161	144	390	161
31 33	224	931	130	366	268
22 32	137	536	69	299	131
32 33	138	830	103	245	81

Table 7: Inefficiency factors of $\mathbf{H}_{2000}$

Parameter	multi	single	multi DR	single DR	mixed DR
$H_{11,2000}$	8	121	4	99	32
$H_{22,2000}$	8	141	4	88	28
$H_{33,2000}$	10	171	11	168	47
$H_{12,2000}$	6	36	9	45	16
$H_{13,2000}$	6	62	5	60	25
$H_{23,2000}$	5	40	3	36	20

Example 2. As another example, we consider the model with highly persistent ϕ_{ii}'s such that

$$\mu_{ij} = \begin{cases} 0.5, & i = j, \\ 0.2, & i \neq j, \end{cases} \quad , \quad \phi_{ij} = \begin{cases} 0.99, & i = j, \\ 0.90, & i \neq j, \end{cases}$$

$$\rho_{iE(j,k),\varepsilon\eta} \equiv Corr(\varepsilon_{it}, \mathbf{E}_t(j,k)) = \begin{cases} -0.3, & i = j = k, \\ -0.1, & \text{otherwise}, \end{cases}$$

$$\rho_{E(i,j)E(k,l),\eta\eta} \equiv Corr(\mathbf{E}_t(i,j), \mathbf{E}_t(k,l)) = \begin{cases} 0.6, & i = j \neq k = l \\ 0.2, & \text{otherwise,} \end{cases}$$

$$\sigma_{E(i,j),\eta\eta} \equiv \sqrt{Var(\mathbf{E}_t(i,j))} = \begin{cases} 0.2, & i = j \\ 0.2, & \text{otherwise.} \end{cases}$$

Tables 8, 9, and 10 show the estimation summaries for all parameters via the multi-move sampler. As in the previous example, 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 factors for the single-move sampler are about twice larger than those for the multi-move sampler, and as shown in Table 14, these inefficiency factors of $\mathbf{H}_{2000}$ for the single-move sampler are about twenty times larger than those for the multi-move sampler, implying that our proposed multi-move sampler is efficient. Further, Tables 11, 12, 13 and 14 compare inefficiency factors for the single-move and the multi-move samplers with and without DR. [3] The DR algorithm reduces the inefficiency factors, but, overall, its performance is similar to those in the previous example.

Table 8: Posterior means, 95% credible intervals, and inefficiency factors.

Param.	True	ij	Mean	95% interval	Inefficiency multi	[single]
ϕ_{ij}	0.99	11	0.992	[0.989, 0.996]	60	[297]
		22	0.992	[0.989, 0.995]	44	[222]
		33	0.992	[0.989, 0.996]	57	[151]
	0.90	21	0.925	[0.897, 0.947]	367	[478]
		31	0.879	[0.843, 0.909]	168	[239]
		32	0.901	[0.868, 0.929]	178	[304]
μ_{ij}	0.5	11	0.350	[-0.406, 1.102]	2	[5]
		22	0.551	[-0.149, 1.236]	1	[7]
		33	0.354	[-0.394, 1.080]	2	[6]
	0.2	21	0.240	[0.160 , 0.320]	6	[25]
		31	0.264	[0.202 , 0.327]	9	[47]
		32	0.209	[0.140 , 0.278]	6	[24]
$\sigma_{E(i,j),\eta\eta}$	0.2	11	0.189	[0.167, 0.214]	206	[620]
		22	0.187	[0.165, 0.212]	192	[547]
		33	0.187	[0.165, 0.210]	197	[346]
	0.2	21	0.174	[0.145, 0.208]	514	[632]
		31	0.215	[0.181, 0.253]	279	[409]
		32	0.196	[0.163, 0.233]	296	[442]

$\sigma_{E(i,j),\eta\eta} = \sqrt{\text{Var}(\mathbf{E}_t(i,j))}$

[3]The acceptance rates for $\boldsymbol{\phi}$, $\mathbf{\Sigma}$ and $\mathbf{H}_{2000}$ in MH algorithms are 0.82, 0.98 and 0.31 for the multi-move sampler, 0.83, 0.98 and 0.70 for the single-move sampler, 0.97, 1.99 and 0.81 for the multi-move sampler with DR, 0.97, 1.00 and 0.91 for the single-move sampler with DR, and 0.97, 1.00, and 0. 90 for the mixed-move sampler with DR.

Table 9: Posterior means, 95% credible intervals and inefficiency factors. $\rho_{iE(j,k),\varepsilon\eta} = \text{Corr}(\varepsilon_{it}, \mathbf{E}_t(j,k))$

True	i jk	Mean	95% interval	Inefficiency multi	[single]
-0.3	1 11	-0.252	[-0.351,-0.147]	80	[173]
	2 22	-0.296	[-0.396,-0.189]	80	[234]
	3 33	-0.334	[-0.436,-0.226]	89	[151]
-0.1	1 21	-0.060	[-0.161, 0.044]	63	[76]
	1 31	-0.215	[-0.311,-0.116]	60	[98]
	1 22	-0.122	[-0.223,-0.013]	76	[150]
	1 32	-0.184	[-0.282,-0.086]	50	[99]
	1 33	-0.101	[-0.207, 0.008]	58	[202]
	2 11	-0.140	[-0.248,-0.028]	80	[271]
	2 21	0.007	[-0.098, 0.114]	60	[151]
	2 31	-0.091	[-0.191, 0.007]	70	[74]
	2 32	-0.116	[-0.217,-0.009]	58	[117]
	2 33	-0.103	[-0.217, 0.010]	110	[297]
	3 11	-0.143	[-0.255,-0.026]	111	[203]
	3 21	-0.038	[-0.147, 0.073]	83	[82]
	3 31	-0.121	[-0.221,-0.018]	58	[99]
	3 22	-0.167	[-0.278,-0.049]	111	[236]
	3 32	-0.130	[-0.236,-0.025]	66	[204]

Table 10: Posterior means, 95% credible intervals and inefficiency factors. $\rho_{E(i,j)E(k,l),\eta\eta} = \text{Corr}(\mathbf{E}_t(i,j)), \mathbf{E}_t(k,l))$

True	ij kl	Mean	95% interval	Inefficiency multi	[single]
0.6	11 22	0.670	[0.563, 0.762]	167	[521]
	11 33	0.585	[0.451, 0.697]	203	[796]
	22 33	0.651	[0.534, 0.751]	265	[1177]
0.2	11 21	0.166	[-0.005, 0.333]	241	[407]
	11 31	0.168	[0.003, 0.321]	192	[318]
	11 32	0.231	[0.070, 0.388]	178	[318]
	21 31	0.261	[0.107, 0.413]	169	[385]
	21 22	0.201	[0.048, 0.355]	203	[340]
	21 32	0.044	[-0.112, 0.198]	142	[374]
	21 33	0.212	[0.045, 0.378]	169	[489]
	31 22	0.209	[0.046, 0.364]	190	[467]
	31 32	0.034	[-0.126, 0.192]	141	[201]
	31 33	0.165	[0.003, 0.319]	213	[404]
	22 32	0.097	[-0.059, 0.252]	142	[271]
	32 33	0.122	[-0.038, 0.275]	131	[238]

Table 11: Inefficiency factors.

Param.	ij	Inefficiency				
		multi	single	multi DR	single DR	mixed DR
ϕ_{ij}	11	60	297	32	56	98
	22	44	222	12	37	76
	33	57	151	18	73	73
	21	367	478	99	185	279
	31	168	239	23	75	104
	32	178	304	58	110	49
μ_{ij}	11	2	5	0.2	5	1
	22	1	7	0.5	4	1
	33	2	6	0.1	6	2
	21	6	25	4	18	9
	31	9	47	1	15	14
	32	6	24	1	14	12
$\sigma_{E(i,j),\eta\eta}$	11	206	620	13	143	211
	22	192	547	21	223	300
	33	197	346	95	311	131
	21	514	632	108	272	292
	31	279	409	15	125	247
	32	296	442	113	174	102

$$\sigma_{E(i,j),\eta\eta} = \sqrt{\mathrm{Var}(\mathbf{E}_t(i,j))}$$

Table 12: Inefficiency factors.

$$\rho_{iE(j,k),\varepsilon\eta} = \mathrm{Corr}(\varepsilon_{it}, \mathbf{E}_t(j,k))$$

i jk	Inefficiency				
	multi	single	multi DR	single DR	mixed DR
1 11	80	173	9	150	146
2 22	80	234	1	198	87
3 33	89	151	8	227	65
1 21	63	76	35	61	30
1 31	60	98	3	70	59
1 22	76	150	9	98	26
1 32	50	99	3	105	39
1 33	58	202	3	58	76
2 11	80	271	4	59	79
2 21	60	151	7	177	114
2 31	70	74	2	66	14
2 32	58	117	42	218	37
2 33	110	297	30	86	66
3 11	111	203	10	43	63
3 21	83	82	1	97	52
3 31	58	99	54	204	62
3 22	111	236	7	72	80
3 32	66	204	30	117	112

Table 13: Inefficiency factors.

$\rho_{E(i,j)E(k,l),\eta\eta} = \text{Corr}(\mathbf{E}_t(i,j)), \mathbf{E}_t(k,l))$

ij kl	Inefficiency				
	multi	single	multi DR	single DR	mixed DR
11 22	167	521	15	510	356
11 33	203	796	83	151	259
22 33	265	1177	44	344	114
11 21	241	407	1	306	259
11 31	192	318	25	228	88
11 32	178	318	2	207	257
21 31	169	385	26	127	41
21 22	203	340	32	251	218
21 32	142	374	9	171	140
21 33	169	489	37	171	177
31 22	190	467	16	160	115
31 32	141	201	4	107	89
31 33	213	404	40	83	45
22 32	142	271	74	84	131
32 33	131	238	70	115	89

Table 14: Inefficiency factors of $\mathbf{H}_{2000}$

Parameter	multi	single	multi DR	single DR	mixed DR
$H_{11,2000}$	12	162	3	122	32
$H_{22,2000}$	9	65	6	39	22
$H_{33,2000}$	6	30	5	18	10
$H_{12,2000}$	14	236	5	142	50
$H_{13,2000}$	6	45	4	26	18
$H_{23,2000}$	13	277	6	156	47

5 Application to trivariate asset returns data

5.1 Data

This section applies our proposed MESV model to returns of three assets: (1) Tokyo stock price index (TOPIX), (2) the Japan government bond clean price index (JGB) provided by Thomson Reuters Datastream), and (3) the currency exchange rate of Japanese Yen to U.S. Dollar (Yen/USD) announced by the Federal Reserve Bank at noon in New York. We excluded those days when at least one of three observations is not reported. The sample period is from January 4, 1995 to July 30, 2010 for a total of 3710 observations. Figure

1 shows the time series plot of three returns which are 100 times the differences of the logarithm of the asset values.

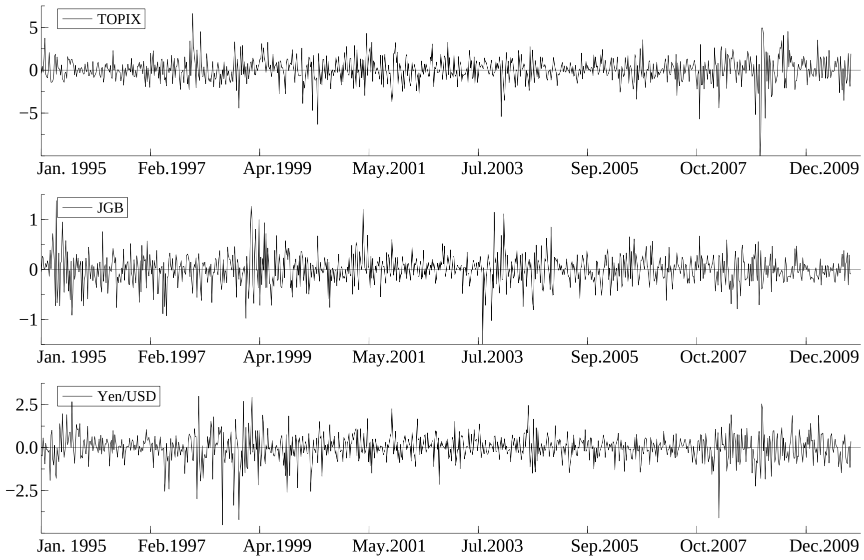

Figure 1: TOPIX, JGB and Yen/USD

5.2 Estimation results

5.2.1 MSV models

For the MSV model, we assume prior distributions such that

$$\frac{\phi_i + 1}{2} \sim \mathcal{B}(20, 1.5), \; i = 1, \ldots, 5, \quad \boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(6, (6\boldsymbol{\Sigma}^*)^{-1}),$$

where

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 1.5^2 & \mathbf{0} & \mathbf{0} \\ & 0.5\mathbf{I}_2 & \mathbf{0} \\ & & 0.2^2(0.3\mathbf{I}_3 + 0.7\mathbf{1}_3\mathbf{1}_3') \end{pmatrix}.$$

The prior mean and standard deviation of ϕ_i are 0.86 and 0.11 respectively reflecting the high persistence of log volatilities in past empirical studies. The prior mean of $\boldsymbol{\Sigma}^{-1}$ is equal to $\boldsymbol{\Sigma}^{*-1}$ where we choose $\boldsymbol{\Sigma}^*$ based on past empirical studies, that is volatilities is based on the univariate result and set the correlation of log-volatilities are high (0.7). Note that we take the degrees-of-freedom small but the prior distribution is still proper.

Tables 15 shows summary statistics of posterior distributions of the parameters for the MSV models. The posterior means of the autoregressive parameters (ϕ_j's) are very high (between 0.948 and 0.966) showing that volatilities are highly persistent. The leverage effects are estimated to be negative as in the previous literature where that of the stock return (-0.445) is much stronger than those of the bond return and the foreign exchange return (-0.177 and -0.167). The cross leverage effect, $\rho_{12,\varepsilon\eta}$, (from the stock return to the bond return volatility) is estimated to be positive (0.189), while the opposite effect, $\rho_{21,\varepsilon\eta}$, (from the bond return to the stock return volatility) is not credible in the sense that its 95% credible interval doesn't include zero. This implies the increase in the stock return at time t causes the high volatility in the bond return at time $t+1$, but the fall of the bond return seems to have a limited impact on the stock return volatility.

Table 15: MSV model.
Posterior means and 95% credible intervals.

Param.	i	Mean	95% interval
ϕ_i	1	0.965	[0.953, 0.975]
	2	0.966	[0.952, 0.978]
	3	0.948	[0.918, 0.969]
$\sigma_{i,\varepsilon\varepsilon}$	1	1.212	[1.115, 1.322]
	2	0.275	[0.249, 0.305]
	3	0.662	[0.615, 0.716]
$\sigma_{i,\eta\eta}$	1	0.191	[0.163, 0.224]
	2	0.210	[0.174, 0.249]
	3	0.225	[0.173, 0.293]
$\rho_{ii,\varepsilon\eta}$	1	-0.445	[-0.543,-0.339]
	2	-0.177	[-0.281,-0.067]
	3	-0.167	[-0.282,-0.045]

Param.	ij	Mean	95% interval
$\rho_{ij,\varepsilon\varepsilon}$	12	-0.312	[-0.343,-0.281]
	13	0.062	[0.028, 0.095]
	23	-0.014	[-0.048, 0.020]
$\rho_{ij,\eta\eta}$	12	0.196	[0.024, 0.363]
	13	0.559	[0.402, 0.694]
	23	0.390	[0.221, 0.552]
$\rho_{ij,\varepsilon\eta}$	12	0.189	[0.073, 0.301]
	13	-0.003	[-0.127, 0.117]
	21	0.091	[-0.021, 0.200]
	23	0.035	[-0.077, 0.145]
	31	-0.036	[-0.152, 0.079]
	32	-0.050	[-0.167, 0.067]

1: TOPIX, 2: JGB, 3: Yen/USD.

$\sigma_{i,\varepsilon\varepsilon} = \sqrt{\mathrm{Var}(\varepsilon_{it})}$, $\sigma_{i,\eta\eta} = \sqrt{\mathrm{Var}(\eta_{it})}$,

$\rho_{ij,\varepsilon\varepsilon} = \mathrm{Corr}(\varepsilon_{it}, \varepsilon_{jt})$, $\rho_{ij,\varepsilon\eta} = \mathrm{Corr}(\varepsilon_{it}, \eta_{jt})$, $\rho_{ij,\eta\eta} = \mathrm{Corr}(\eta_{it}, \eta_{jt})$.

5.2.2 MESV model

For the MESV model, we assume that prior distributions are

$$
\begin{aligned}
\frac{\phi_{ij}+1}{2} &\sim \begin{cases} \mathcal{B}(20, 1.5), & \text{if } i = j, \\ \mathcal{B}(8.25, 2.75), & \text{otherwise,} \end{cases} \\
\boldsymbol{\mu} &\sim \mathcal{N}_q(\mathbf{0}, 5\mathbf{I}_q), \\
\boldsymbol{\Sigma}^{22} &\sim \mathcal{W}(6, 6^{-1}\boldsymbol{\Sigma}^{22*}), \quad \mathrm{vec}(\boldsymbol{\Sigma}^{21})|\boldsymbol{\Sigma}^{22} \sim \mathcal{N}_{pq}(\mathbf{0}, (5\mathbf{I}_p) \otimes \boldsymbol{\Sigma}^{22}).
\end{aligned}
$$

The mean and the standard deviation of the prior distribution of ϕ_{ij}, $i \neq j$ are set 0.5 and 0.25 respectively, which is fairly flat as we shall see in our posterior estimation results. [4] We take $\boldsymbol{\Sigma}^{22*} = \boldsymbol{\Sigma}_{22}^{*-1}$ assuming $\boldsymbol{\Sigma}^{*} = \mathrm{diag}(\mathbf{I}_3, \boldsymbol{\Sigma}_{22}^{*})$ and the (i,j)-th element of $\boldsymbol{\Sigma}_{22}^{*}$ is $\rho_{ij,\eta\eta}^{*}\sigma_{i,\eta\eta}^{*}\sigma_{j,\eta\eta}^{*}$ such that

$$\begin{aligned}
\rho^{*}_{E(k,l)E(m,n),\eta\eta} &= \begin{cases} 0.6, & \text{if } k = l \neq m = n, \\ 0.2, & \text{otherwise,} \end{cases} \\
\sigma^{*}_{E(k,l),\eta\eta} &= 0.2, \quad 1 \leq l \leq k \leq 3.
\end{aligned}$$

which is based on the MSV model result and we set the correlations between non diagonal elements of $\boldsymbol{H}_t$ is smaller than that of diagonal. We draw 220,000 samples for the multi-move sampler discarding the first 20,000 samples as a burn-in-period. The number of blocks is set to 185 based on several trials[5]. The acceptance rates for $\boldsymbol{\phi}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{h}$ in MH algorithms are on average 0.89, 0.98 and 0.81 respectively[6].

[4]We also tried a uniform prior for ϕ_{ij} with $i \neq j$. Although its sample path becomes a bit unstable in the sense that it sometimes takes low values around zero, but the parameter estimates are basically the same.

[5]The number of blocks is selected by trials and errors to minimize the maximum of inefficiency factors. The average number of variables in one block is chosen to be about 20.

[6]The elapsed time for the simulation using the multi-move (single-move) sampler is 8.95 (0.43) hours per 10,000 iterations with Richland AMD A10-6800K Black Edition (4.1GHz). The inefficiency factor for $H_{11,1855}$, for example, is 9.0 (250.2). Taking account of both elapsed time and inefficiency factors, the multi-move sampler is highly efficient.

Table 16: MESV model.
Posterior means, standard deviations, 95% credible intervals, and inefficiency factors.

	ij	Mean	Stdev	95% interval	IF
ϕ_{ij}	11	0.968	0.005	[0.958, 0.977]	141
	22	0.968	0.006	[0.955, 0.979]	94
	33	0.953	0.012	[0.927, 0.972]	648
	21	0.860	0.050	[0.738, 0.923]	693
	31	0.845	0.060	[0.699, 0.924]	560
	32	0.590	0.199	[0.051, 0.826]	741
μ_{ij}	11	0.286	0.087	[0.114, 0.457]	6
	22	-2.923	0.113	[-3.143,-2.699]	11
	33	-0.929	0.079	[-1.084,-0.773]	16
	21	-0.222	0.018	[-0.258,-0.186]	24
	31	0.054	0.023	[0.009, 0.099]	25
	32	0.007	0.017	[-0.027, 0.041]	75
$\sigma_{E(i,j),\eta\eta}$	11	0.176	0.014	[0.150, 0.204]	253
	22	0.211	0.018	[0.176, 0.249]	277
	33	0.204	0.027	[0.156, 0.263]	645
	21	0.106	0.016	[0.080, 0.144]	742
	31	0.133	0.022	[0.097, 0.185]	828
	32	0.144	0.022	[0.107, 0.193]	801

1:TOPIX, 2:JGB, 3:Yen/USD

$$\sigma_{E(i,j),\eta\eta} = \sqrt{Var(\mathbf{E}_t(i,j))}$$

Table 17: MESV model. Posterior means, standard deviations, 95% credible intervals and inefficiency factors.

$$\rho_{iE(j,k),\varepsilon\eta} = \mathrm{Corr}(\varepsilon_{it}, \mathbf{E}_t(j,k))$$

i jk	Mean	Stdev	95% interval	IF
1 11	-0.467	0.054	[-0.567,-0.358]	90
2 22	-0.109	0.059	[-0.224, 0.008]	64
3 33	-0.166	0.063	[-0.287,-0.041]	87
1 21	-0.100	0.065	[-0.230, 0.027]	71
1 31	-0.050	0.073	[-0.191, 0.094]	89
1 22	0.218	0.060	[0.097, 0.332]	65
1 32	0.058	0.085	[-0.108, 0.225]	162
1 33	0.026	0.065	[-0.107, 0.150]	99
2 11	-0.005	0.061	[-0.125, 0.114]	70
2 21	0.023	0.068	[-0.108, 0.160]	74
2 31	0.012	0.072	[-0.130, 0.152]	74
2 32	-0.046	0.080	[-0.204, 0.109]	120
2 33	0.055	0.063	[-0.069, 0.177]	76
3 11	-0.024	0.061	[-0.143, 0.095]	101
3 21	0.078	0.065	[-0.051, 0.206]	69
3 31	-0.128	0.074	[-0.273, 0.016]	109
3 22	-0.061	0.062	[-0.183, 0.061]	67
3 32	0.020	0.085	[-0.148, 0.188]	215

1:TOPIX, 2:JGB, 3:Yen/USD

Table 18: MESV model. Posterior means, standard deviations, 95% credible intervals and inefficiency factors.

$$\rho_{E(i,j)E(k,l),\eta\eta} = \mathrm{Corr}(\mathbf{E}_t(i,j), \mathbf{E}_t(k,l))$$

ij kl	Mean	Stdev	95% interval	IF
11 22	0.009	0.091	[-0.168, 0.187]	155
11 33	0.466	0.083	[0.293, 0.616]	221
22 33	0.263	0.097	[0.064, 0.443]	237
11 21	0.151	0.088	[-0.025, 0.320]	151
11 31	0.155	0.099	[-0.041, 0.350]	181
11 32	-0.233	0.131	[-0.459, 0.056]	362
21 31	0.020	0.108	[-0.185, 0.233]	230
21 22	0.081	0.096	[-0.108, 0.267]	154
21 32	-0.023	0.137	[-0.296, 0.240]	369
21 33	-0.010	0.104	[-0.211, 0.197]	202
31 22	0.045	0.109	[-0.175, 0.252]	193
31 32	-0.151	0.140	[-0.419, 0.135]	395
31 33	0.134	0.109	[-0.079, 0.348]	225
22 32	0.062	0.144	[-0.217, 0.346]	386
32 33	-0.145	0.136	[-0.399, 0.139]	472

1:TOPIX, 2:JGB, 3:Yen/USD

The estimation results are summarized in the Tables 16, 17 and 18. We notice that the parameters of the diagonal elements of $\mathbf{H}_t$ (the 1st, 4th, and 6th elements of $\boldsymbol{h}_t$) are similar to those of MSV models. The autoregressive parameters of log volatilities, $(\phi_{11}, \phi_{22}, \phi_{33})$, are (0.968, 0.968, 0.953) while (ϕ_1, ϕ_2, ϕ_3) for MSV models are (0.965, 0.966, 0.948). The posterior means of $(\mu_{11}, \mu_{22}, \mu_{33})$ are (0.286, -2.923, -0.929), while $(\log \sigma^2_{1,\varepsilon\varepsilon}, \log \sigma^2_{2,\varepsilon\varepsilon}, \log \sigma^2_{3,\varepsilon\varepsilon})$ evaluated at the posterior means in Table 15 are $(0.385, -2.582,\ -0.825)$. Further, the estimates of standard deviations of the η_{it}, $(\sigma_{E(1,1),\eta\eta}, \sigma_{E(2,2),\eta\eta}, \sigma_{E(3,3),\eta\eta})$, are (0.176, 0.211, 0.204), while those of $(\sigma_{1,\eta\eta}, \sigma_{2,\eta\eta}, \sigma_{3,\eta\eta})$ are (0.191, 0.210, 0.225). Regarding the leverage effects, the estimates of $(\rho_{1E(1,1),\varepsilon\eta},\ \rho_{2E(2,2),\varepsilon\eta},\ \rho_{3E(3,3),\varepsilon\eta})$ are $(-0.467, -0.109, -0.166)$, while $(\rho_{11,\varepsilon\eta},\ \rho_{22,\varepsilon\eta},\ \rho_{33,\varepsilon\eta})$ for the MSV models are $(-0.445, -0.177, -0.167)$.

Estimated volatilities and correlations. However, as mentioned in Section 2, the parameters such as $\mathbf{\Phi}$, $\mathbf{M}$ and $\mathbf{\Sigma}_{\varepsilon\eta}$ in the MESV model do not always correspond to those of the stock, the bond, and the exchange rate as they are in the MSV model. Thus, to interpret the estimation results of the MESV model in more intuitive way, we consider the posterior means of time-varying volatilities of each series, dynamic correlations among three returns and news impact curves using MCMC simulation technique.

First, we consider the posterior means with 95% credible intervals for the square root of the time-varying variances as shown in Figure 2. The estimated volatility series of the TOPIX returns sharply increased in September 2008, corresponding to the financial crisis during which Lehman Brothers filed for Chapter 11 bankruptcy protection. The volatilities of the JGB index returns increased in December 1998 and in September 2003 when the index dropped (or equivalently, the JGB interest rate ran up) in both periods. In December 1998, the JGB is supplied excessively because the Japanese government issued a large amount of JGB for the economic-stimulus measure, and the Ministry of Finance Japan announced to stop buying the new bonds in this month. Moreover, the Moody's downgraded the JGB rating in November 1998. In this month, the Nippon Credit Bank was brought under government control because of a large amount of the bad debt. In the mid 2003s, following the increase of the US bond interest rate and the economic boom, the deflationary concerns of Japan toned down. The expectation for the lifting of the zero-interest-rate policy of the Bank of Japan, the JGB interest rate shot up in June and September of 2003. The volatilities of the Yen/USD increase after the August 1998 when the ruble devaluation and the Long Term Capital Management report a large loss. Especially, the USD fell from 135.6 yen to 117 yen in five days of early October 1998. This is just after the decision of the monetary relaxation policy in the USA on September 29th and issuing the G-7 communique which urged the injection of taxpayers' money to financial institutions in Japan on October 5th.

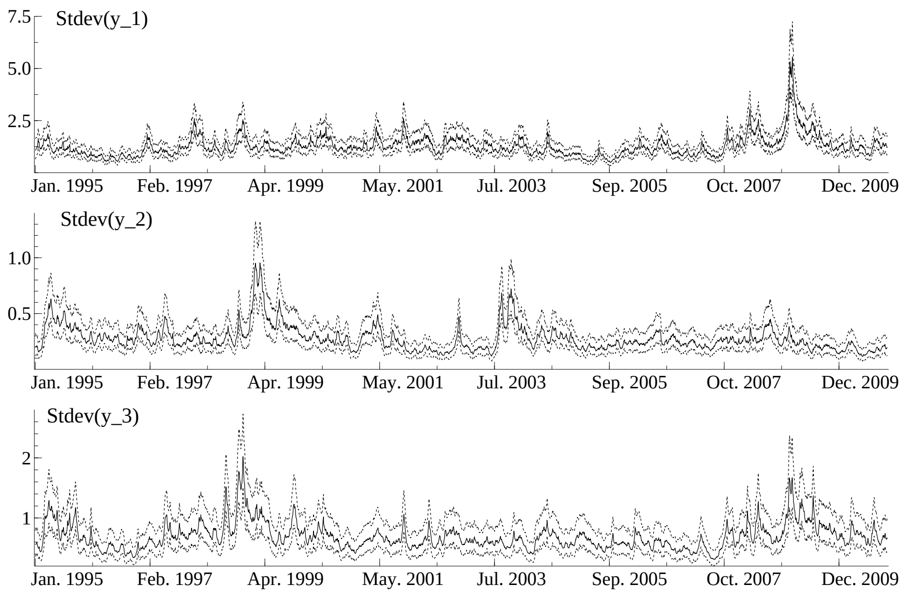

Figure 2: Posterior means with 95% credible intervals for the square root of time-varying variances. Top: TOPIX, middle: JGB, bottom: Yen/USD.

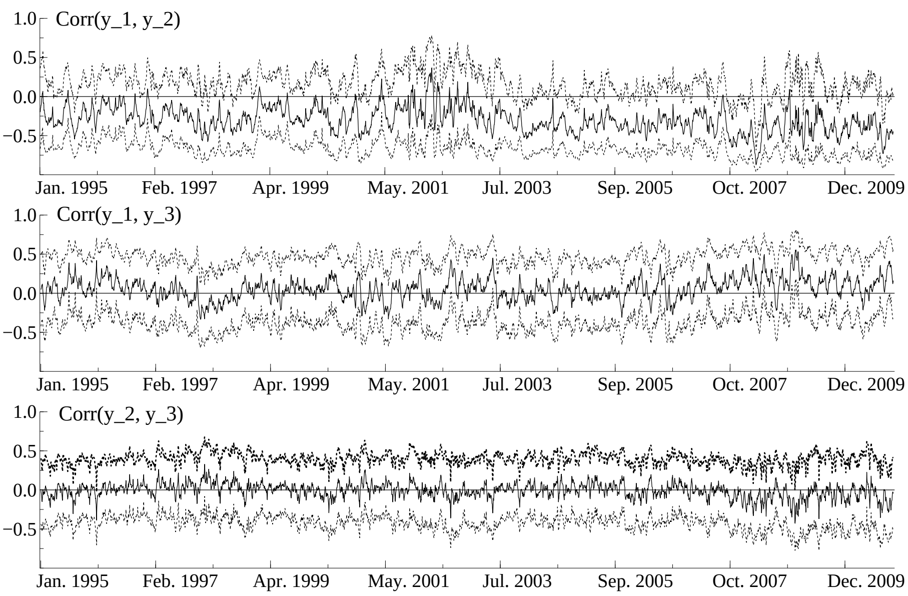

Figure 3: Posterior means with 95% credible intervals for time-varying correlations. Top: (TOPIX, JGB), middle: (TOPIX, Yen/USD), bottom: (JGB, Yen/USD).

Next, we investigate the posterior means with 95% credible intervals for dynamic correlations among three returns as shown in Figure 3. These correlations are computed using the MCMC samples of the covariance and the variances that are elements of $\exp(\mathbf{H}_t)$ which is the matrix exponential transformation of the log volatility matrix $\mathbf{H}_t$. The correlations between the stock and the JGB returns largely fluctuate taking negative values where they drop to less than -0.86 in January 2008 during the downturn of the stock market. The correlations between the stock market and the exchange rate returns fluctuates around zero. It is noted that this takes negative values during Asian crisis period from July 1997 to August 1998. In this period, the yen kept weakening and Japanese stock prices dropped. The JGB returns and the exchange rate returns seem to have no correlation throughout the sample period.

News impact curves. Finally, to show how the shocks in the returns at time t affect the volatilities at time $t+1$, we describe the news impact curve following Engle and Ng (1993). Similar ideas for stochastic volatility models are discussed by Yu (2005) and Asai and McAleer (2009). Let $\mathbf{H}_t = \mathbf{M}$ and $\mathbf{E}_t = \mathbf{O}$ and consider the case where $\boldsymbol{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Sigma}_{\eta\varepsilon} \exp(-\mathbf{M}/2)\boldsymbol{y}_t$.

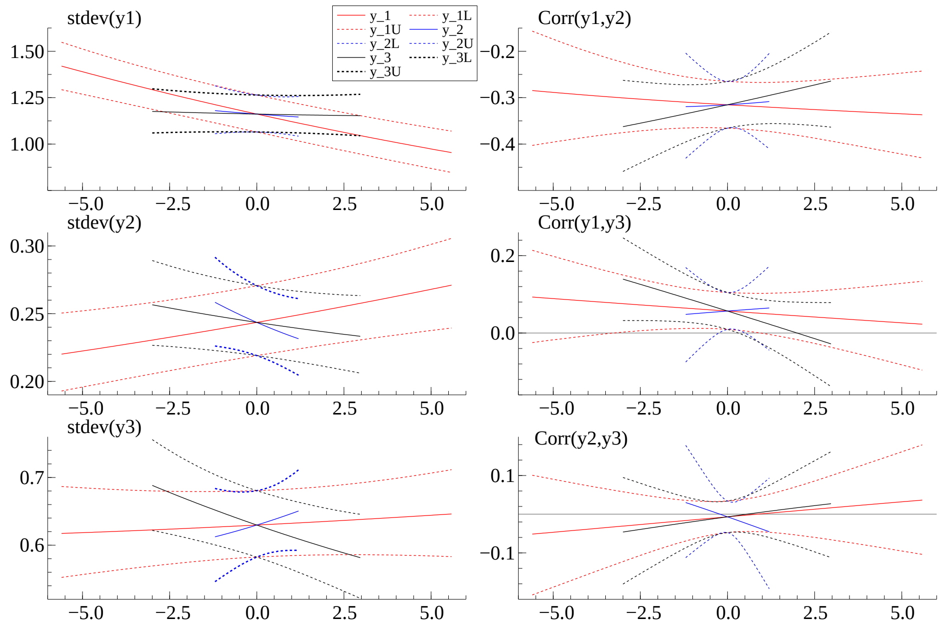

Figure 4: Posterior mean (solid line) and 95% interval (dotted lines) of the news impact curve* for the one-step-ahead conditional covariances of $\boldsymbol{y}_{t+1}$ when $\mathbf{H}_t = \mathbf{M}$ and $\mathbf{E}_t = \mathbf{O}$. (* The domain of the estimated curve is restricted to the range within ± 4 sample standard deviation of actual returns for each y_{it}.)

Figure 4 shows the posterior news impact curves on standard deviations and correlations of $\boldsymbol{y}_{t+1}$ from $\boldsymbol{y}_t$, obtained by calculating the posterior means and the 95% intervals of the normalized $\exp(\mathbf{H}_{t+1})$ under the various shocks of $\boldsymbol{y}_t$. The horizontal and vertical axes show the values of $\boldsymbol{y}_t$ and the standard deviations and correlations calculated from $\exp(\mathbf{H}_{t+1})$ respectively.

The left three panels in Figure 4 show the news impacts on the standard deviations of $\boldsymbol{y}_{t+1}$ caused by the elements, y_{1t} (red lines), y_{2t} (blue lines) and y_{3t} (black lines). The negative return on the i-th asset increases its own (i-th) future volatility, indicating the existence of leverage effects. The red lines in the middle and bottom left panels imply that the positive TOPIX return increases the future volatility of the JGB and the exchange rate returns. However, the impact of the TOPIX return on the exchange rate return volatility is smaller than that on the volatility of the JGB return. The positive shocks of the JGB return increase the future volatility of the exchange rate, while the negative shocks of the exchange rate return cause the higher future volatility of the JGB return. These results are generally consistent with those obtained with MSV model, taking account of the 95% credible intervals.

The right three panels in Figure 4 show the news impacts on the correlations among $y_{1,t+1}$, $y_{2,t+1}$ and $y_{3,t+1}$. The top right panel shows the impact on the correlation between the TOPIX and JGB returns. It is noted that the correlations are strongly affected by the exchange rate return but hardly affected by the TOPIX and the JGB returns. The black line in the middle right panel shows the impact on the correlation between the TOPIX and exchange rate returns caused by the shock of the exchange rate return. Interestingly, the sign of the correlation between the TOPIX and exchange rate returns strongly depends on the impact of the shock. More precisely, the large positive shock (greater than one) on the exchange rate return tends to produce the negative correlation between the TOPIX and exchange rate return, while the small or negative shock tends to produce the positive correlation. The sign of the correlation between the JGB and exchange rate returns also depends on the impacts of the shocks. However, we note that the impacts by shocks of the TOPIX and exchange rate returns are very small and the impacts by the JGB return shock have wide 95% intervals.

5.3 Extensions to fat-tailed error distributions and model comparison

Finally, we conduct a model comparison of the proposed MESV model with MSV models. In addition to the MESV and MSV models with normal errors, we consider extended models with fat-tailed error distribution given by

$$\boldsymbol{\varepsilon}_t = \xi_t^{1/2}\boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p),$$

where ξ_t is a random variable which takes positive values and independent of $\boldsymbol{e}_t$. We consider the multivariate Student-t error with $\xi_t^{-1} \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$. The extension is straightforward and hence we omit details of MCMC algorithms (similar specifications and MCMC sampling are also discussed in the Omori, Chib, Shephard, and Nakajima (2007) for univariate SV models, and Ishihara and Omori (2012) for MSV models). Thus, we consider the following six models:

- MSV-n model: MSV model with normal error distribution.
- MSV-t model: MSV model with multivariate Student-t error distribution.
- MESV-n model: MESV model with normal error distribution.
- MESV-t model: MESV model with multivariate Student-t error distribution.
- BEKK(Asymmetric) model: Multivariate asymmetric GARCH model (diagonal BEKK model, see e.g. Kroner and Ng (1998)) defined by

$$\boldsymbol{y}_t = \boldsymbol{\varepsilon}_t, \tag{18}$$

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2}\boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathcal{N}_3(\mathbf{0}, \mathbf{I}), \tag{19}$$

$$\mathbf{H}_t = \mathbf{W} + \mathbf{A}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'\mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1}\mathbf{B}' + \mathbf{C}\boldsymbol{\varepsilon}_{t-1}^{*}\boldsymbol{\varepsilon}_{t-1}^{*\prime}\mathbf{C}', \tag{20}$$

$$\mathbf{W} = \boldsymbol{\Omega} - \mathbf{A}\boldsymbol{\Omega}\mathbf{A}' - \mathbf{B}\boldsymbol{\Omega}\mathbf{B}' - \mathbf{C}\mathbf{N}\mathbf{C}',$$

where $\varepsilon_{it}^{*} = \epsilon_{it}I(\epsilon_{it} < 0)$, $\boldsymbol{\Omega} = E(\mathbf{H}_t) = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$, $\mathbf{N} = E(\boldsymbol{\varepsilon}_t^{*}\boldsymbol{\varepsilon}_t^{*\prime})$ and $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are assumed to be diagonal. The matrix $\mathbf{N}$ is given by

$$\mathbf{N} = (\boldsymbol{\Omega} \odot \mathbf{I})^{1/2}\{0.5\mathbf{I} + \mathbf{R} \odot (\mathbf{1}_3\mathbf{1}_3' - \mathbf{I})\}(\boldsymbol{\Omega} \odot \mathbf{I})^{1/2}, \tag{21}$$

where $\mathbf{1}_3$ is a 3×1 vector with all elements equal to one, and the (i, j)-th element of

R is

$$\begin{aligned} r_{ij} &= L(\rho_{ij}) \times \rho_{ij} + \frac{1}{2\pi}\sqrt{1-\rho_{ij}^2}, \\ & L(\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{0}\int_{-\infty}^{0} \exp\left\{-\frac{1}{2}(x^2 - 2\rho xy + y^2)/(1-\rho^2)\right\} dxdy, \\ & \rho_{ij} = \mathrm{Corr}(\varepsilon_{it}, \varepsilon_{jt}), \end{aligned}$$

(see e.g. Rosenbaum (1961)).

- BEKK(Symmetric) model: Multivariate symmetric GARCH model (obtained by setting $\mathbf{C} = \mathbf{O}$ in the asymmetric model).

We assume the prior distributions $\nu \sim \mathcal{G}(0.001, 0.001)$. The estimation results for ν are summarized in Table 19. The posterior means of ν for the MSV and MESV models are small, suggesting fat-tailed error distributions. The estimate of ν for the MSV model is smaller than that of the MESV model, probably because the MSV model fails to capture the dynamics of time-varying correlations. Other parameter estimates are similar to those of models with normal error and hence are omitted[7].

Table 19: The estimation results of ν

Model	Param.	Mean	Stdev	95% interval
MESV	ν	15.8	2.4	[11.8, 21.5]
MSV		11.8	1.2	[9.8, 14.5]

Comparison based on DIC. We compute the DIC (deviance information criterion) for the model comparison defined by

$$\begin{aligned} DIC &= E_{\boldsymbol{\theta}|Y_n}[D(\boldsymbol{\theta})] + p_D, \\ & p_D = E_{\boldsymbol{\theta}|Y_n}[D(\boldsymbol{\theta})] - D(E_{\boldsymbol{\theta}|Y_n}[\boldsymbol{\theta}]), \quad D(\boldsymbol{\theta}) = -2\log f(Y_n|\boldsymbol{\theta}) + C_y, \end{aligned}$$

where C_y is a constant term which depends only on the dataset Y_n. Since it cancels out in all calculations that compare different models, we set $C_y = 0$ for convenience. To estimate $E_{\theta|Y_n}[D(\boldsymbol{\theta})]$, we use a sample analogue $\frac{1}{M}\sum_{m=1}^{M} D(\boldsymbol{\theta}^{(m)})$, where we set $M = 100$, and $\boldsymbol{\theta}^{(m)}$s are resampled from the posterior samples generated by the MCMC method. To calculate $D(E_{\boldsymbol{\theta}|Y_n}[\boldsymbol{\theta}])$, which equals to $D(\boldsymbol{\theta})$ evaluated at the posterior mean, we implement

[7]Estimation results for the BEKK models are omitted to save space.

an auxiliary particle filter to compute the log-likelihood ordinate $\log f(Y_n|\boldsymbol{\theta})$, where we set the number of particles $I = 10,000$ for the MSV and the MESV models. We repeat this procedure ten times to obtain the numerical standard error.

Table 20: The averages of DIC estimates, their standard errors, the maximum and the minimum of DIC values.

Model	ranking	DIC	(s.e.)	DIC_{max}	DIC_{min}
MESV-t	1	20066.7	(1.7)	20074.2	20058.3
MESV-n	2	20090.1	(1.5)	20098.4	20084.1
MSV-t	3	20118.4	(1.3)	20124.3	20112.4
MSV-n	4	20201.2	(1.0)	20206.3	20196.0
BEKK (Asymmetric)	5	20564.8	(0.7)	20569.7	20562.8
BEKK (Symmetric)	6	20685.4	(0.8)	20688.8	20680.9

Table 20 shows the averages of DIC, their standard errors, and the maximum and the minimum of DIC values computed for six competing models. The DIC values for the MESV models are much smaller than those for the MSV models and the BEKK models, and hence the MESV models outperform other models. Among MESV models, models with fat-tailed error outperform the model with normal error, and the model with multivariate-t error distribution has the smallest DIC. This empirical study shows that our proposed model with multivariate-t error distribution performs quite well to describe the multivariate asset returns data escribe the multivariate asset returns data.

Comparison based on the prediction of the realized covariances. Further, we compare the volatility predictive performances of proposed models using both daily stock returns and daily realized covariances discussed in Noureldin, Shephard, and Sheppard (2012). We use 1511 daily return series (IBM, Alcoa and General Electric) from January 2nd, 2004 to December 31st, 2009. Model parameters are estimated using first 1411 observations, and T days ahead predictions ($T = 1, 5, 10$) are considered with a rolling window estimation method (the numbers of predictions are 100, 96 and 91 respectively). We predict the future realized covariance matrix using the posterior predictive means of the covariance matrix of $\boldsymbol{y}_t$, which we denote by $\hat{\mathbf{V}}_t$. The quasi-likelihood loss function (described in Noureldin, Shephard, and Sheppard (2012)) is used for the prediction comparison,

$$L(\Omega_t, \hat{V}_t) = \log|\hat{V}_t| + \text{tr}(\Omega_t \hat{V}_t^{-1}) - K_t,$$

where Ω_t is a realized covariance matrix, and K_t is constant which depends only on the data. The realized covariance matrix is used as a proxy of the true covariance matrix. Under this loss function, the ranking based on the conditionally unbiased estimator of the covariance matrices is consistent with the ranking based on the true covariance matrices.

Table 21: Average of loss differences from MESV-t model.

	$T = 1$		$T = 5$		$T = 10$	
Model	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
MESV-n	0.883	0.099	0.965	0.088	0.960	0.084
MSV-t	-0.308	0.155	1.753	0.167	3.116	0.208
MSV-n	0.934	0.126	1.164	0.121	1.289	0.115
BEKK (Asymmetric)	1.462	0.298	1.006	0.233	0.543	0.173
BEKK (Symmetric)	1.949	0.317	1.652	0.271	1.413	0.233

Table 21 shows the average of loss differences from MESV-t model for each model (the average loss of each model minus the average loss of the MESV-t model). Since the negative value implies that the model performs better than the MESV-t model, the MESV-t model performs better than other models for all periods except the MSV-t model with $T = 1$. However, taking account of standard errors, the MESV-t model and the MSV-t model have similar performances in one day ahead prediction. For $T = 5$ and 10, dynamic correlation models such as MESV models and asymmetric BEKK models outperform constant correlation MSV models. This result implies the evidence of time-varying correlations among the multivariate stock returns data.

6 Conclusion

In this paper, we extend the MSV model to allow the time-varying correlations and propose an efficient MCMC algorithm using a multi-move sampler. To sample a block of state vectors, we construct a proposal density using the normal approximation via a Taylor expansion of the logarithm of the target posterior density for the MH algorithm where the expectations of Hessian matrices are derived analytically. Moreover, to calculate the log-likelihood, we describe an auxiliary particle filter. An empirical analysis is presented using three returns of the TOPIX, the Japanese bond price index and the Yen/USD exchange rate. The correlation between returns of the TOPIX and the Japanese bond index is found to be time-varying. In contrast, the correlation between returns of the Japanese bond price index and the Yen/USD exchange rate is shown to be stable and less volatile. The positive cross leverage effects from the TOPIX on the Japanese bond price index is also found. The news impact curves for the MESV model are presented and investigated in detail. A model comparison between the MESV model with constant correlation MSV models including heavy-tailed error models is conducted. The MESV model with multivariate Student-t distributed error is found to outperform other models based on DIC and forecast performance of the future realized covariance matrix.

Acknowledgment

The authors are grateful to the Associate Editor, two anonymous referees, Siddhartha Chib, Toshiaki Watanabe, Teruo Nakatsuma, and Yoshi Baba 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 and (A) 24243031 from the Japanese Ministry of Education, Science, Sports, Culture and Technology, and Global COE Program "Research Unit for Statistical and Empirical Analysis in Social Sciences" from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. The computational results are generated using Ox (Doornik (2006)).

Appendix

A Multi-move sampler

To generate $\{\boldsymbol{h}_t\}_{t=s+1}^{s+m}$ given other $\boldsymbol{h}_t$'s, for example, we sample the normalized disturbances $\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}$ instead of $\{\boldsymbol{h}_t\}_{t=s+1}^{s+m}$ since such a sampling method is known to reduce the MCMC sample autocorrelations where

$$\boldsymbol{x}_t = \boldsymbol{\Sigma}_{\eta\eta}^{-1/2}\boldsymbol{\eta}_t, \quad t = 1, \dots, n-1, \quad \boldsymbol{x}_0 = \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\eta}_0,$$

and $\boldsymbol{\Sigma}_{\eta\eta}^{1/2}$, $\boldsymbol{\Sigma}_0^{1/2}$ denote Choleski decompositions such that $\boldsymbol{\Sigma}_{\eta\eta} = \boldsymbol{\Sigma}_{\eta\eta}^{1/2}\boldsymbol{\Sigma}_{\eta\eta}^{1/2'}$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_0^{1/2'}$.

The logarithm of the full conditional joint density of $\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}$ excluding constant terms is given by

$$\log f(\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}|\boldsymbol{h}_s, \boldsymbol{h}_{s+m+1}, \boldsymbol{y}_s, \dots, \boldsymbol{y}_{s+m}) = -\frac{1}{2}\sum_{t=s}^{s+m-1}\boldsymbol{x}_t'\boldsymbol{x}_t + L, \qquad (22)$$

where

$$
\begin{aligned}
L &= \sum_{t=s}^{s+m} l_t - \frac{1}{2}\boldsymbol{\eta}'_{s+m}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\eta}_{s+m} I(s+m<n), \qquad (23)\\
&\quad \boldsymbol{\eta}_{s+m} = \boldsymbol{h}_{s+m+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{h}_{s+m} - \boldsymbol{\mu}), \qquad (24)\\
&\quad \boldsymbol{x}_t = \boldsymbol{\Sigma}_{\eta\eta}^{-1/2}(\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu})).
\end{aligned}
$$

Let $\boldsymbol{\alpha}_t = \boldsymbol{h}_t - \boldsymbol{\mu}$ and $\hat{\boldsymbol{\alpha}}_t = \hat{\boldsymbol{h}}_t - \boldsymbol{\mu}$ where $\hat{\boldsymbol{h}}_t$ is $\boldsymbol{h}_t$ evaluated at $\boldsymbol{x}_t = \hat{\boldsymbol{x}}_t$.

Consider the approximation via the second order Taylor expansion of L around the $\boldsymbol{x}_t = \hat{\boldsymbol{x}}_t$ $(t = s, \ldots, s+m-1)$, and replace the Hessian matrices by the negative of information matrices to obtain

$$
\begin{aligned}
& \log f(\{\boldsymbol{x}_t\}_{t=s}^{s+m-1} | \boldsymbol{h}_s, \boldsymbol{h}_{s+m+1}, \boldsymbol{y}_s, \ldots, \boldsymbol{y}_{s+m})\\
\approx\ & \text{const.} - \frac{1}{2}\sum_{t=s}^{s+m-1} \boldsymbol{x}'_t\boldsymbol{x}_t + \hat{L}\\
& + \sum_{t=s+1}^{s+m}\left[\hat{\boldsymbol{d}}'_t - \frac{1}{2}(\boldsymbol{\alpha}_t - \hat{\boldsymbol{\alpha}}_t)'\hat{\mathbf{A}}_t + (\boldsymbol{\alpha}_{t-1} - \hat{\boldsymbol{\alpha}}_{t-1})'\hat{\mathbf{B}}_t\right](\boldsymbol{\alpha}_t - \hat{\boldsymbol{\alpha}}_t)\\
=\ & \text{const.} + \log f^*(\{\boldsymbol{x}_t\}_{t=s}^{s+m-1} | \boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+m+1}, \boldsymbol{y}_s, \ldots, \boldsymbol{y}_{s+m}), \qquad (25)
\end{aligned}
$$

where $\boldsymbol{d}'_t = \partial L / \partial \boldsymbol{\alpha}'_t$,

$$
\begin{aligned}
\mathbf{A}_t &= -E\left[\frac{\partial^2 L}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}'_t}\right], \quad t = s+1, \ldots, s+m, \qquad (26)\\
\mathbf{B}_t &= -E\left[\frac{\partial^2 L}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}'_{t-1}}\right], \quad t = s+2, \ldots, s+m, \quad \mathbf{B}_{s+1} = \mathbf{O}, \qquad (27)
\end{aligned}
$$

and $\hat{\boldsymbol{d}}_t$, $\hat{\mathbf{A}}_t$, $\hat{\mathbf{B}}_t$ are those evaluated at $\hat{\boldsymbol{\alpha}}_t$. The expectations are taken with respect to $\boldsymbol{y}_t$ given parameters and other latent variables.

Then $f^*(\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}|\cdot)$ is a conditional posterior density of $\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}$ for the following linear Gaussian state space model (28)–(29):

$$
\begin{aligned}
\hat{\boldsymbol{y}}_t &= \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{G}_t\boldsymbol{u}_t, \quad t = s+1, \ldots, s+m, \qquad (28)\\
\boldsymbol{\alpha}_{t+1} &= \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \mathbf{K}_t\boldsymbol{u}_t, \quad t = s+1, \ldots, s+m-1, \qquad (29)\\
\boldsymbol{u}_t &= (\boldsymbol{\xi}'_t, \boldsymbol{x}'_t)' \sim \mathcal{N}_{p+q}(\mathbf{0}, \mathbf{I}),
\end{aligned}
$$

where $\mathbf{K}_t = \left[\mathbf{O}, \boldsymbol{\Sigma}_{\eta\eta}^{1/2}\right]$ for $t = 1, \ldots, n-1$ and $\mathbf{K}_0 = \left[\mathbf{O}, \boldsymbol{\Sigma}_0^{1/2}\right]$, and $\hat{\boldsymbol{y}}_t$, $\mathbf{Z}_t$, $\mathbf{G}_t$ are computed as follows:

1. First we set $\hat{\boldsymbol{x}}_t = \boldsymbol{x}_t$ $(t = s, \ldots, s+m-1)$ where $\boldsymbol{x}_t$ is a current sample.

2. Let $\hat{\boldsymbol{\alpha}}_{t+1} = \boldsymbol{\Phi}\hat{\boldsymbol{\alpha}}_t + \boldsymbol{\Sigma}_{\eta\eta}^{1/2}\hat{\boldsymbol{x}}_t$ $(t = s, \ldots, s+m-1)$ with $\hat{\boldsymbol{\alpha}}_s = \hat{\boldsymbol{h}}_s - \boldsymbol{\mu}$.

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

$$\mathbf{D}_t = \hat{\mathbf{A}}_t - \hat{\mathbf{B}}_t\mathbf{D}_{t-1}^{-1}\hat{\mathbf{B}}_t', \quad \boldsymbol{b}_t = \hat{\boldsymbol{d}}_t - \hat{\mathbf{B}}_t\mathbf{D}_{t-1}^{-1}\boldsymbol{b}_{t-1}, \quad t = s+1, \ldots, s+m.$$

4. Let

$$\begin{aligned}\hat{\boldsymbol{y}}_t &= \hat{\boldsymbol{\alpha}}_t + \mathbf{D}_t^{-1}(\boldsymbol{b}_t + \hat{\mathbf{B}}_t'\hat{\boldsymbol{\alpha}}_{t+1}), \quad \mathbf{Z}_t = \mathbf{I}_p + \mathbf{D}_t^{-1}\hat{\mathbf{B}}_t'\boldsymbol{\Phi}, \\ \mathbf{G}_t &= [\mathbf{D}_t^{-1/2\prime}, \mathbf{D}_t^{-1}\hat{\mathbf{B}}_t'\boldsymbol{\Sigma}_{\eta\eta}], \quad t = s+2, \ldots, s+m,\end{aligned}$$

where $\mathbf{D}_t^{1/2}$ denotes a Choleski decomposition such that $\mathbf{D}_t = \mathbf{D}_t^{1/2}\mathbf{D}_t^{1/2\prime}$.

5. Implement the disturbance smoother (Koopman (1993)) to obtain $\{\hat{\boldsymbol{x}}_t\}_{t=s}^{s+m-1}$, the mode of the conditional posterior density of $\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}$ for the model (28) and (29). If the the mode converges (however, usually several iterations will be sufficient to construct a proposal distribution), save $\hat{\boldsymbol{y}}_t$, $\mathbf{Z}_t$ and $\mathbf{G}_t$. Otherwise, go to Step 2.

Then we apply the simulation smoother (e.g. de Jong and Shephard (1995), Durbin and Koopman (2002)) to generate a candidate $\{\boldsymbol{x}_t^\dagger\}_{t=s}^{s+m-1}$ from this state space model for Metropolis-Hastings algorithm. We accept a candidate with probability

$$\min\left\{1, \frac{f(\{\boldsymbol{x}_t^\dagger\}_{t=s}^{s+m-1}|\cdot)f^*(\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}|\cdot)}{f(\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}|\cdot)f^*(\{\boldsymbol{x}_t^\dagger\}_{t=s}^{s+m-1}|\cdot)}\right\}.$$

A.1 Derivation of $\boldsymbol{d}_t$, $\mathbf{A}_t$ and $\mathbf{B}_t$

Summary of 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

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

where vec$(\cdot)$ is a vectorizing operator, 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

$$\mathbf{D}H(X) = (\mathbf{D}G(F(X)))(\mathbf{D}F(X)) = \frac{\partial \mathrm{vec}(G(F(\boldsymbol{X})))}{\partial(\mathrm{vec}(F(\boldsymbol{X})))'}\frac{\partial \mathrm{vec}(F(\boldsymbol{X}))}{\partial(\mathrm{vec}(\boldsymbol{X}))'}. \tag{30}$$

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

$$\frac{\partial g(f(\mathbf{x}))}{\partial \mathbf{x}'} = \frac{\partial \mathrm{vec}(g(f(\mathbf{x})))}{\partial \mathrm{vec}(f(\mathbf{x}))'}\frac{\partial \mathrm{vec}(f(\mathbf{x}))}{\partial \mathrm{vec}(\mathbf{x})'}. \tag{31}$$

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

$$\frac{\partial \mathrm{vec}(FG)}{\partial(\mathrm{vec}(X))'} = (G' \otimes \mathbf{I}_m)\frac{\partial \mathrm{vec}(F)}{\partial(\mathrm{vec}(X))'} + (\mathbf{I}_r \otimes F)\frac{\partial \mathrm{vec}(G)}{\partial(\mathrm{vec}(X))'}. \tag{32}$$

Derivation of $\boldsymbol{d}_t$

Let $\mathbf{F}_t = -\frac{1}{2}\mathbf{H}_t$ and $\boldsymbol{z}_t = \exp(\mathbf{F}_t)\boldsymbol{y}_t$. The logarithm of the conditional probability density of $\boldsymbol{y}_t$ given $\boldsymbol{h}_t$ excluding the constant term is

$$l_t = -\frac{1}{2}\mathrm{tr}(\mathbf{H}_t) - \frac{1}{2}(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1}(\boldsymbol{z}_t - \boldsymbol{m}_t),$$

where $E[\boldsymbol{z}_t] = \boldsymbol{m}_t = \boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t)I(t < n)$ and $E[\boldsymbol{z}_t\boldsymbol{z}_t'] = \mathbf{S}_t + \boldsymbol{m}_t\boldsymbol{m}_t'$ with $\mathbf{S}_t = \mathbf{I}_p - \boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}I(t < n)$. Further, let $\mathbf{D}_p$ denote a $p^2 \times q$ duplication matrix (whose elements are 0 or 1) such that $\mathrm{vec}(\mathbf{A}) = \mathbf{D}_p\mathrm{vech}(\mathbf{A})$ for a symmetric matrix $\mathbf{A}$. Then

$$\frac{\partial \mathrm{tr}(\mathbf{H}_t)}{\partial \boldsymbol{\alpha}_t'} = \frac{\partial \mathrm{tr}(\mathbf{H}_t)}{\partial \mathrm{vech}(\mathbf{H}_t)'} = \mathrm{vec}(\mathbf{I}_p)'\mathbf{D}_p = \mathrm{vech}(\mathbf{I}_p)', \tag{33}$$

$$\frac{\partial \boldsymbol{m}_t}{\partial \boldsymbol{\alpha}_t'} = -\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t < n), \tag{34}$$

$$\frac{\partial \boldsymbol{m}_{t-1}}{\partial \boldsymbol{\alpha}_t'} = \boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}I(t > 1), \tag{35}$$

where we used the chain rule and $\mathbf{D}_p'\text{vec}(\mathbf{A}) = \text{vech}(\mathbf{A} + \mathbf{A}' - (\mathbf{A} \odot \mathbf{I}_p))$ for a $p \times p$ matrix $\mathbf{A}$ in (33) (e.g. Magnus and Neudecker (1999), Magnus (1988)). Further, using the product rule for the power series expansion of $\mathbf{F}_t$ with

$$\frac{\partial \text{vec}(\mathbf{F}_t^i)}{\partial \text{vec}(\mathbf{F}_t)'} = \sum_{j=1}^{i} \left[\mathbf{F}_t^{i-j} \otimes \mathbf{F}_t^{j-1} \right], \tag{36}$$

we obtain

$$\begin{aligned} \frac{\partial \text{vec}(\exp(\mathbf{F}_t))}{\partial \text{vech}(\mathbf{F}_t)'} &= \sum_{i=1}^{\infty} \frac{1}{i!} \frac{\partial \text{vec}(\mathbf{F}_t^i)}{\partial \text{vech}(\mathbf{F}_t)'} \\ &= \mathbf{V}_t \mathbf{D}_p, \quad \mathbf{V}_t \equiv \frac{\partial \text{vec}(\exp(\mathbf{F}_t))}{\partial \text{vec}(\mathbf{F}_t)'} = \sum_{i=1}^{\infty} \frac{1}{i!} \sum_{j=1}^{i} \left[\mathbf{F}_t^{i-j} \otimes \mathbf{F}_t^{j-1} \right]. \end{aligned}$$

Let $\mathbf{Q}_t = \{\exp(-\mathbf{F}_t) \otimes \mathbf{I}_p\}\mathbf{V}_t$. Noting that $\boldsymbol{z}_t = (\boldsymbol{y}_t' \otimes \mathbf{I}_p)\text{vec}(\exp(\mathbf{F}_t))$,

$$\frac{\partial \boldsymbol{z}_t}{\partial \boldsymbol{\alpha}_t'} = -\frac{1}{2}(\boldsymbol{y}_t' \otimes \mathbf{I}_p)\mathbf{V}_t \mathbf{D}_p = -\frac{1}{2}(\boldsymbol{z}_t' \otimes \mathbf{I}_p)\mathbf{Q}_t \mathbf{D}_p. \tag{37}$$

Using (33) – (37) and $\partial \boldsymbol{x}'\mathbf{A}\boldsymbol{x}/\partial \boldsymbol{x}' = 2\boldsymbol{x}'\mathbf{A}$ for a $p \times p$ symmetric matrix $\mathbf{A}$ and a $p \times 1$ vector $\boldsymbol{x}$, we obtain

$$\begin{aligned} \boldsymbol{d}_t &= \left[\frac{\partial l_t}{\partial \boldsymbol{\alpha}_t'}\right]' + \left[\frac{\partial l_{t-1}}{\partial \boldsymbol{\alpha}_t'}\right]' + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t)I(t = s + m < n) \\ &= -\frac{1}{2}\text{vech}(\mathbf{I}_p) + \left\{ \frac{1}{2}\mathbf{D}_p'\mathbf{Q}_t'(\boldsymbol{z}_t \otimes \mathbf{I}_p) - \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}I(t < n) \right\} \mathbf{S}_t^{-1}(\boldsymbol{z}_t - \boldsymbol{m}_t) \\ &\quad + \boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_{t-1}^{-1}(\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1})I(t > 1) + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t)I(t = s + m < n). \end{aligned} \tag{38}$$

Although $\mathbf{Q}_t$ involves an infinite series of matrices, its computation is easy as shown in Appendix A.2.

Derivation of $\mathbf{A}_t$

By (34)–(38) and $\mathbf{Q}_t'(\boldsymbol{z}_t \otimes \mathbf{I}_p) = \mathbf{V}_t(\boldsymbol{y}_t \otimes \mathbf{I}_p)$,

$$\begin{aligned} \mathbf{A}_t &= -\frac{1}{2}\mathbf{D}_p' E\left[\frac{\partial \mathbf{V}_t(\boldsymbol{y}_t \otimes \mathbf{I}_p)\mathbf{S}_t^{-1}(\boldsymbol{z}_t - \boldsymbol{m}_t)}{\partial \boldsymbol{\alpha}_t'}\right] - \frac{1}{2}\boldsymbol{\Phi}\mathbf{N}_t' \\ &\quad + \boldsymbol{\Phi}\mathbf{M}_t\boldsymbol{\Phi} + \mathbf{M}_{t-1} + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t = s + m < n), \end{aligned} \tag{39}$$

where

$$\mathbf{M}_t = \boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}I(1 \le t < n), \tag{40}$$

$$\mathbf{N}_t = \mathbf{D}_p'\mathbf{Q}_t'(\boldsymbol{m}_t \otimes \mathbf{I}_p)\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}I(t < n). \tag{41}$$

Using a product rule and $\partial\mathrm{vec}(\mathbf{F}_t)/\partial\boldsymbol{\alpha}_t' = -\frac{1}{2}\mathbf{D}_p$, we obtain

$$\begin{aligned}
&\frac{\partial\mathbf{V}_t(\boldsymbol{y}_t \otimes \mathbf{I}_p)\mathbf{S}_t^{-1}(\boldsymbol{z}_t - \boldsymbol{m}_t)}{\partial\boldsymbol{\alpha}_t'} \\
&= -\frac{1}{2}\left\{(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1} \otimes \mathbf{I}_{p^2}\right\}\left\{(\boldsymbol{y}_t' \otimes \mathbf{I}_p) \otimes \mathbf{I}_{p^2}\right\}\mathbf{W}_t\mathbf{D}_p + \mathbf{V}_t(\boldsymbol{y}_t \otimes \mathbf{I}_p)\mathbf{S}_t^{-1}\frac{\partial(\boldsymbol{z}_t - \boldsymbol{m}_t)}{\partial\boldsymbol{\alpha}_t'},
\end{aligned}$$

where $\mathbf{W}_t \equiv \partial\mathrm{vec}(\mathbf{V}_t)/\partial\mathrm{vec}(\mathbf{F}_t)'$. Since

$$\begin{aligned}
&E\left[\{(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1} \otimes \mathbf{I}_{p^2}\}\{(\boldsymbol{y}_t' \otimes \mathbf{I}_p) \otimes \mathbf{I}_{p^2}\}\right] \\
&= E\left[(\boldsymbol{z}_t - \boldsymbol{m}_t)'\mathbf{S}_t^{-1}(\boldsymbol{y}_t' \otimes \mathbf{I}_p)\right] \otimes \mathbf{I}_{p^2} \\
&= E\left[\mathrm{vec}(\mathbf{S}_t^{-1}(\boldsymbol{z}_t - \boldsymbol{m}_t)\boldsymbol{y}_t')'\right] \otimes \mathbf{I}_{p^2} = \mathrm{vec}(\exp(-\mathbf{F}_t))' \otimes \mathbf{I}_{p^2},
\end{aligned}$$

and

$$\begin{aligned}
&E\left[\mathbf{V}_t(\boldsymbol{y}_t \otimes \mathbf{I}_p)\mathbf{S}_t^{-1}\frac{\partial(\boldsymbol{z}_t - \boldsymbol{m}_t)}{\partial\boldsymbol{\alpha}_t'}\right] \\
&= -\frac{1}{2}\mathbf{Q}_t'E\left[(\boldsymbol{z}_t \otimes \mathbf{I}_p)\mathbf{S}_t^{-1}(\boldsymbol{z}_t' \otimes \mathbf{I}_p)\right]\mathbf{Q}_t\mathbf{D}_p + \mathbf{Q}_t'(\boldsymbol{m}_t \otimes \mathbf{I}_p)\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi} \\
&= -\frac{1}{2}\mathbf{Q}_t'E\left[(\boldsymbol{z}_t\boldsymbol{z}_t') \otimes \mathbf{S}_t^{-1}\right]\mathbf{Q}_t\mathbf{D}_p + \mathbf{Q}_t'(\boldsymbol{m}_t \otimes \mathbf{I}_p)\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi} \\
&= -\frac{1}{2}\mathbf{Q}_t'\left\{(\mathbf{S}_t + \boldsymbol{m}_t\boldsymbol{m}_t') \otimes \mathbf{S}_t^{-1}\right\}\mathbf{Q}_t\mathbf{D}_p + \mathbf{Q}_t'(\boldsymbol{m}_t \otimes \mathbf{I}_p)\mathbf{S}_t^{-1}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi},
\end{aligned}$$

Equation (39) reduces to

$$\begin{aligned}
\mathbf{A}_t &= \frac{1}{4}\mathbf{D}_p'\left[\mathbf{P}_t + \mathbf{Q}_t'\left\{(\mathbf{S}_t + \boldsymbol{m}_t\boldsymbol{m}_t') \otimes \mathbf{S}_t^{-1}\right\}\mathbf{Q}_t\right]\mathbf{D}_p - \frac{1}{2}(\mathbf{N}_t\boldsymbol{\Phi} + \boldsymbol{\Phi}\mathbf{N}_t') \\
&\quad + \boldsymbol{\Phi}\mathbf{M}_t\boldsymbol{\Phi} + \mathbf{M}_{t-1} + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t = s + m < n),
\end{aligned} \tag{42}$$

where $\mathbf{D}_p'\mathbf{P}_t = \mathbf{D}_p'\left\{\mathrm{vec}(\exp(-\mathbf{F}_t))' \otimes \mathbf{I}_{p^2}\right\}\mathbf{W}_t$. The computation of $\mathbf{P}_t$ as well as $\mathbf{Q}_t$ is discussed in Appendix A.2.

Derivation of $\mathbf{B}_t$

By (34), (37) and (38),

$$\mathbf{B}_t = -\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_{t-1}^{-1}E\left[\frac{\partial(\boldsymbol{z}_{t-1}-\boldsymbol{m}_{t-1})}{\partial\boldsymbol{\alpha}_{t-1}'}\right] = \frac{1}{2}\mathbf{N}_{t-1}' - \mathbf{M}_{t-1}\boldsymbol{\Phi}, \quad t=2,\ldots,n. \tag{43}$$

A.2 Computation of $\mathbf{P}_t$ and $\mathbf{Q}_t$

(1) $\mathbf{Q}_t$

Since $\mathbf{F}_t$ is a symmetric matrix, there exists a $p \times p$ orthogonal matrix $\mathbf{U}_t$ such that

$$\mathbf{U}_t'\mathbf{F}_t\mathbf{U}_t = \boldsymbol{\Lambda}_t,$$

where $\boldsymbol{\Lambda}_t = \mathrm{diag}(\lambda_{1t}, \lambda_{2t}, \ldots, \lambda_{pt})$ and $\lambda_{1t} \geq \lambda_{2t} \geq \ldots \geq \lambda_{pt}$ are the ordered eigenvalues of $\mathbf{F}_t$. Then

$$\begin{aligned}
\mathbf{Q}_t &= \{\exp(-\mathbf{F}_t)\otimes\mathbf{I}_p\}\sum_{i=1}^{\infty}\frac{1}{i!}\sum_{j=1}^{i}\left[\mathbf{F}_t^{i-j}\otimes\mathbf{F}_t^{j-1}\right], \\
&= \sum_{i=1}^{\infty}\frac{1}{i!}\sum_{j=1}^{i}\left[\exp(-\mathbf{F}_t)\mathbf{F}_t^{i-j}\otimes\mathbf{F}_t^{j-1}\right], \\
&= (\mathbf{U}_t\otimes\mathbf{U}_t)\left[\sum_{i=1}^{\infty}\frac{1}{i!}\sum_{j=1}^{i}\left\{\exp(-\boldsymbol{\Lambda}_t)\boldsymbol{\Lambda}_t^{i-j}\otimes\boldsymbol{\Lambda}_t^{j-1}\right\}\right](\mathbf{U}_t'\otimes\mathbf{U}_t').
\end{aligned} \tag{44}$$

The second factor in (44) is a diagonal matrix with its (k,k)-th element given by

$$\sum_{i=1}^{\infty}\frac{1}{i!}\sum_{j=1}^{i}\exp(-\lambda_{at})\lambda_{at}^{i-j}\lambda_{bt}^{j-1} = \begin{cases} 1, & \text{if } \lambda_{at}=\lambda_{bt}, \\ \frac{\exp(\lambda_{bt}-\lambda_{at})-1}{\lambda_{bt}-\lambda_{at}}, & \text{if } \lambda_{at}\neq\lambda_{bt}, \end{cases} \tag{45}$$

where $a = \lfloor (k-1)/p \rfloor + 1$ and $b = k - p\lfloor (k-1)/p \rfloor$, and $\lfloor x \rfloor$ denotes the integer part of x. Note that $\lambda_{at} = \lambda_{bt}$ for $k = (i-1)p + i$ $(i = 1, \ldots, p)$.

(2) $\mathbf{P}_t$

Let $\mathbf{K}_{mn}$ denote a $mn \times mn$ 0-1 matrix called a commutation matrix such that $\mathrm{vec}(\mathbf{A}') = \mathbf{K}_{mn}\mathrm{vec}(\mathbf{A})$ holds for a $m \times n$ matrix $\mathbf{A}$ (see Chapter 3 of Magnus and Neudecker (1999)).

Using

$$\begin{aligned}
\frac{\partial \mathrm{vec}(\mathbf{A} \otimes \mathbf{B})}{\partial \mathrm{vec}(\mathbf{A})'} &= (\mathbf{I}_n \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_p)(\mathbf{I}_{mn} \otimes \mathrm{vec}(\mathbf{B})), \\
\frac{\partial \mathrm{vec}(\mathbf{A} \otimes \mathbf{B})}{\partial \mathrm{vec}(\mathbf{B})'} &= (\mathbf{I}_n \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_p)(\mathrm{vec}(\mathbf{A}) \otimes \mathbf{I}_{pq}),
\end{aligned}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{X}$ are $m \times n$, $p \times q$, $n \times p$ matrices respectively, we compute

$$\begin{aligned}
\mathbf{W}_t &= \frac{\partial \mathrm{vec}(\mathbf{V}_t)}{\partial \mathrm{vec}(\mathbf{F}_t)'} = \sum_{i=2}^{\infty} \frac{1}{i!} \sum_{j=1}^{i} \frac{\partial \mathrm{vec}(\mathbf{F}_t^{i-j} \otimes \mathbf{F}_t^{j-1})}{\partial \mathrm{vec}(\mathbf{F}_t)'} \\
&= \sum_{i=2}^{\infty} \frac{1}{i!} (\mathbf{I}_p \otimes \mathbf{K}_{pp} \otimes \mathbf{I}_p) \Bigg\{ \sum_{j=1}^{i-1} (\mathbf{I}_{p^2} \otimes \mathrm{vec}(\mathbf{F}_t^{j-1})) \frac{\partial \mathrm{vec}(\mathbf{F}_t^{i-j})}{\partial \mathrm{vec}(\mathbf{F}_t)'} \\
&\qquad + \sum_{j=2}^{i} (\mathrm{vec}(\mathbf{F}_t^{i-j}) \otimes \mathbf{I}_{p^2}) \frac{\partial \mathrm{vec}(\mathbf{F}_t^{j-1})}{\partial \mathrm{vec}(\mathbf{F}_t)'} \Bigg\}.
\end{aligned}$$

Using the product rule,

$$\frac{\partial \mathrm{vec}(\mathbf{F}_t^k)}{\partial \mathrm{vec}(\mathbf{F}_t)'} = \sum_{m=1}^{k} \mathbf{F}_t^{k-m} \otimes \mathbf{F}_t^{m-1},$$

we note that

$$\begin{aligned}
&\left\{\mathrm{vec}(\exp(-\mathbf{F}_t))' \otimes \mathbf{I}_{p^2}\right\} (\mathbf{I}_p \otimes \mathbf{K}_{pp} \otimes \mathbf{I}_p) \left\{\mathbf{I}_{p^2} \otimes \mathrm{vec}(\mathbf{F}_t^{j-1})\right\} \frac{\partial \mathrm{vec}(\mathbf{F}_t^{i-j})}{\partial \mathrm{vec}(\mathbf{F}_t)'} \\
= \; & \mathbf{K}_{pp} (\mathbf{F}_t^{j-1} \exp(-\mathbf{F}_t) \otimes \mathbf{I}_p) \frac{\partial \mathrm{vec}(\mathbf{F}_t^{i-j})}{\partial \mathrm{vec}(\mathbf{F}_t)'} = \mathbf{K}_{pp} \sum_{h=1}^{i-j} \mathbf{F}_t^{j-1} \exp(-\mathbf{F}_t) \mathbf{F}_t^{i-j-h} \otimes \mathbf{F}_t^{h-1},
\end{aligned}$$

and

$$\begin{aligned}
&\left\{\mathrm{vec}(\exp(-\mathbf{F}_t))' \otimes \mathbf{I}_{p^2}\right\} (\mathbf{I}_p \otimes \mathbf{K}_{pp} \otimes \mathbf{I}_p) \left\{\mathrm{vec}(\mathbf{F}_t^{i-j}) \otimes \mathbf{I}_{p^2}\right\} \frac{\partial \mathrm{vec}(\mathbf{F}_t^{j-1})}{\partial \mathrm{vec}(\mathbf{F}_t)'} \\
= \; & \left\{\mathbf{F}_t^{i-j} \exp(-\mathbf{F}_t) \otimes \mathbf{I}_p\right\} \frac{\partial \mathrm{vec}(\mathbf{F}_t^{j-1})}{\partial \mathrm{vec}(\mathbf{F}_t)'} = \sum_{h=1}^{j-1} \mathbf{F}_t^{i-j} \exp(-\mathbf{F}_t) \mathbf{F}_t^{j-1-h} \otimes \mathbf{F}_t^{h-1},
\end{aligned}$$

by using

$$\mathbf{K}_{nn}(\mathbf{B}\mathbf{A} \otimes \mathbf{I}_n) = (\mathrm{vec}(\mathbf{A})' \otimes \mathbf{I}_{n^2})(\mathbf{I}_n \otimes \mathbf{K}_{nn} \otimes \mathbf{I}_n)(\mathbf{I}_{n^2} \otimes \mathrm{vec}(\mathbf{B})), \tag{46}$$

$$(\mathbf{B}\mathbf{A}' \otimes \mathbf{I}_n) = (\mathrm{vec}(\mathbf{A})' \otimes \mathbf{I}_{n^2})(\mathbf{I}_n \otimes \mathbf{K}_{nn} \otimes \mathbf{I}_n)(\mathrm{vec}(\mathbf{B}) \otimes \mathbf{I}_{n^2}), \tag{47}$$

where $\mathbf{A}$, $\mathbf{B}$ are $m \times n$, $p \times q$ matrices (e.g. Theorem 4.4 of Rogers (1980) on p. 23). Thus, using $\mathbf{D}_p'\mathbf{K}_{pp} = \mathbf{D}_p'$,

$$
\begin{aligned}
\mathbf{D}_p'\mathbf{P}_t &= \mathbf{D}_p' \left\{ \mathrm{vec}(\exp(-\mathbf{F}_t))' \otimes \mathbf{I}_{p^2} \right\} \mathbf{W}_t \\
&= \mathbf{D}_p' \sum_{i=2}^{\infty} \frac{1}{i!} \left(\sum_{j=1}^{i-1} \sum_{h=1}^{i-j} \exp(-\mathbf{F}_t)\mathbf{F}_t^{i-h-1} \otimes \mathbf{F}_t^{h-1} + \sum_{j=2}^{i} \sum_{k=1}^{j-1} \exp(-\mathbf{F}_t)\mathbf{F}_t^{i-k-1} \otimes \mathbf{F}_t^{k-1} \right) \\
&= 2\mathbf{D}_p' \sum_{i=2}^{\infty} \frac{1}{i!} \sum_{j=1}^{i-1} \sum_{h=1}^{i-j} \exp(-\mathbf{F}_t)\mathbf{F}_t^{i-h-1} \otimes \mathbf{F}_t^{h-1} \\
&= 2\mathbf{D}_p' \sum_{i=2}^{\infty} \frac{1}{i!} \sum_{j=1}^{i-1} (i-j) \exp(-\mathbf{F}_t)\mathbf{F}_t^{i-j-1} \otimes \mathbf{F}_t^{j-1},
\end{aligned}
$$

and we obtain

$$
\begin{aligned}
\mathbf{P}_t &= 2 \sum_{i=2}^{\infty} \frac{1}{i!} \sum_{j=1}^{i-1} (i-j) \exp(-\mathbf{F}_t)\mathbf{F}_t^{i-j-1} \otimes \mathbf{F}_t^{j-1} \\
&= (\mathbf{U}_t \otimes \mathbf{U}_t) \left[2 \sum_{i=2}^{\infty} \frac{1}{i!} \sum_{j=1}^{i-1} (i-j) \left\{ \exp(-\mathbf{\Lambda}_t)\mathbf{\Lambda}_t^{i-j-1} \otimes \mathbf{\Lambda}_t^{j-1} \right\} \right] (\mathbf{U}_t' \otimes \mathbf{U}_t'). \quad (48)
\end{aligned}
$$

The second factor in (48) is a diagonal matrix with its (k,k)-th element given by

$$
2 \sum_{i=2}^{\infty} \frac{1}{i!} \sum_{j=1}^{i-1} (i-j) \exp(-\lambda_{at}) \lambda_{at}^{i-j-1} \lambda_{bt}^{j-1} = \begin{cases} 1, & \text{if } \lambda_{at} = \lambda_{bt}, \\ \frac{2\{\exp(\lambda_{bt}-\lambda_{at})-1-(\lambda_{bt}-\lambda_{at})\}}{(\lambda_{bt}-\lambda_{at})^2}, & \text{if } \lambda_{at} \neq \lambda_{bt}, \end{cases} \quad (49)
$$

where $a = \lfloor (k-1)/p \rfloor + 1$ and $b = k - p\lfloor (k-1)/p \rfloor$.

B Delaying rejection algorithm

To improve the MH algorithm by reducing the number of rejected proposals, Mira (2001) proposes the delaying rejection algorithm. Let $\boldsymbol{\theta}^0$ denote the current value, and let $\boldsymbol{\theta}^J$ $(J = 1, 2, \dots)$ denote the J-th stage proposal. Further, let $\pi(\cdot)$ and $q_J(\cdot|\cdots)$ be an invariant density and the J-th proposal density. The delaying rejection algorithm is described as follows:

1. For $J = 1$, we generate $\boldsymbol{\theta}^1$ from the proposal distribution with the density $q_1(\boldsymbol{\theta}^1|\boldsymbol{\theta}^0)$

and accept it with probability

$$\alpha_1(\boldsymbol{\theta}^0, \boldsymbol{\theta}^1) := \min\left\{1\frac{\pi(\boldsymbol{\theta}^1)q_1(\boldsymbol{\theta}^0|\boldsymbol{\theta}^1)}{\pi(\boldsymbol{\theta}^0)q_1(\boldsymbol{\theta}^1|\boldsymbol{\theta}^0)}\right\}, \tag{50}$$

If it is accepted, go to Step 3. If it is rejected, then increase J by one, and move to the second stage.

2. For the J-th stage, we generate $\boldsymbol{\theta}^J$ from the proposal distribution with the density $q_J(\boldsymbol{\theta}^J|\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^{J-1})$ and accept it with probability

$$\begin{aligned}&\alpha_J(\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^J)\\ &:= \min\left\{1, \frac{\pi(\boldsymbol{\theta}^J)\prod_{j=1}^J q_j(\boldsymbol{\theta}^{J-j}|\boldsymbol{\theta}^J, \boldsymbol{\theta}^{J-1}, \ldots, \boldsymbol{\theta}^{J-j+1})\prod_{j=1}^{J-1}\{1-\alpha_j(\boldsymbol{\theta}^J, \ldots, \boldsymbol{\theta}^{J-j})\}}{\pi(\boldsymbol{\theta}^0)\prod_{j=1}^J q_j(\boldsymbol{\theta}^j|\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^{j-1})\prod_{j=1}^{J-1}\{1-\alpha_j(\boldsymbol{\theta}^0, \ldots, \boldsymbol{\theta}^j)\}}\right\}.\end{aligned}$$

If it is accepted, go to Step 3. Otherwise, increase J by one, and repeat Step 2.

3. Accept $\boldsymbol{\theta}^J$ as a new MCMC sample, and go to Step 1.

In this paper, we only use the independence sampler for our MH algorithm, and hence $q_J(\boldsymbol{\theta}^J|\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^{J-1}) = q(\boldsymbol{\theta}^J)$. It makes the algorithm simple, but it may result in repeating Steps 2 and 3 with large J. Thus, sometimes the delaying rejection algorithm with independent proposal can not exit from Step 2. To reduce the computational time, we only implement the algorithm until $J \leq 5$ where

$$\alpha_J(\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^J) = \min\left\{1, \frac{\max\{0, g(\boldsymbol{\theta}^J) - \max_{j=1,\ldots,J-1} g(\boldsymbol{\theta}^j)\}}{g(\boldsymbol{\theta}^0) - \max_{j=1,\ldots,J-1} g(\boldsymbol{\theta}^j)}\right\}, \tag{51}$$

and $g(\theta^j) = \pi(\theta^j)/q(\theta^j)$. If the proposal is rejected at the fifth stage, we use the current value as an MCMC sample.

References

Abadir, K. M. and J. R. Magnus (2005). *Matrix algebra.* New York: Cambridge University Press.

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

Asai, M. and M. McAleer (2009). Multivariate stochastic volatility, leverage and news impact surfaces. *Econometrics Journal 12*, 292–309.

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

Bauer, G. H. and K. Vorkink (2010). Forecasting multivariate realized stock market volatility. *Journal of Econometrics 160*, 93–101.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate GARCH: a survey. *Journal of Applied Econometrics 21*, 79–109.

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., 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.

Chiu, T., T. Leonard, and K. Tsui (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association 91*, 198–210.

Daníelsson, J. (1998). Multivariate stochastic volatility models: Estimation and a comparison with VGARCH models. *Journal of Empirical Finance 5*, 155–173.

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.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics 20*, 339–350.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*, 122–150.

Engle, R. F. and V. Ng (1993). Measuring and testing the impact of news on volatility. *Journal of Finance 48*, 1749–1778.

Golosnoy, V., B. Gribisch, and R. Liesenfeld (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics 167*, 211–223.

Gourieroux, C., J. Jasiak, and R. Sufana (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics 150*, 167–181.

Gupta, A. K. and D. K. Nagar (2000). *Matrix variate Distributions*. Boca Raton: Chapman & Hall/CRC.

Ishihara, T. and Y. Omori (2012). Efficient Bayesian estimation of a multivariate stochastic volatility model with cross leverage and heavy-tailed errors. *Computational Statistics and Data Analysis 56*, 3674–3689.

Jungbacker, B. and S. J. Koopman (2006). Monte Carlo likelihood estimation for three multivariate stochastic volatility models. *Econometric Reviews 25*, 385–408.

Kawakatsu, H. (2006). Matrix exponential GARCH. *Journal of Econometrics 134*, 95–128.

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

Kroner, K. and V. Ng (1998). Modeling asymmetric comovements of asset returns. *Review Financial Studies 11*, 817–844.

LeSage, J. P. and R. K. Pace (2007). A matrix exponential spatial specification. *Journal of Econometrics 140*, 190–214.

Loddo, A., S. Ni, and D. Sun (2011). Selection of multivariate stochastic volatility models via bayesian stochastic search. *Journal of Business & Economic Statistics 29*(3), 342–355.

Lopes, H. F., R. E. McCulloch, and R. S. Tsay (2012). Cholesky stochastic volatility models for high-dimensional time series. Discussion papers.

Magnus, J. R. (1988). *Linear Structures.* Oxford: Oxford University press.

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.

Mira, A. (2001). On metropolis-hastings algorithms with delayed rejection. *Metron 59*(3-4), 231–241.

Noureldin, D., N. Shephard, and K. Sheppard (2012). Multivariate high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics 27*(6), 907–933.

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.

Philipov, A. and M. E. Glickman (2006). Multivariate stochastic volatility via Wishart processes. *Journal of Business and Economic Statistics 24*, 313–328.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association 94*(446), 590–599.

Plataniotis, A. (2011). High dimensional time-varying covariance matrices with applications in finance. PhD Thesis, Athens University of Economics and Busines.

Rogers, G. S. (1980). *Matrix Derivarives: Lecture Notes in Statistics, Vol. 2*. New York: Marcel Dekker.

Rosenbaum, S. (1961). Moments of a truncated bivariate normal distribution. *Journal of the Royal Statistical Society, Series B 23*, 405–408.

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

Sheppard, K. (2007). Positive semi-definite matrix multiplicative error models. University of Oxford.

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

Tsay, R. S. (2005). *Analysis of Financial Time Series: Financial Econometrics* (2 ed.). New York: Wiley.

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

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

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