Matrix exponential stochastic volatility with cross leverage

Tsunehiro Ishihara, Yasuhiro Omori and Manabu Asai

Matrix exponential stochastic volatility with cross leverage

Tsunehiro Ishihara[*]   Yasuhiro Omori [†] and   Manabu Asai [‡]

October 14, 2013

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.

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 (e.g. Andersen, Bollerslev, Diebold, and Ebens (2001), Barndorff-Nielsen and Shephard (2004), Chiriac and Voev (2010), Audrino and Corsi (2010), Voev (2007), Bauer and Vorkink (2010), Jin and Maheu (2010) and Sheppard (2007)). 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) and Gourieroux, Jasiak, and Sufana (2009)) 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).

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

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

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

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

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 an 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.

3

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 $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{pt})'$ 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

$$\boldsymbol{y}_t = \exp(\mathbf{H}_t/2)\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \text{i.i.d. } \mathcal{N}(\mathbf{0}, \mathbf{I}_p), \quad t = 1, \ldots, n. \tag{4}$$

$$\mathbf{H}_{t+1} = \mathbf{M} + \tilde{\mathbf{\Phi}} \odot (\mathbf{H}_t - \mathbf{M}) + \mathbf{E}_t, \tag{5}$$

$$\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} \sim \text{ i.i.d. } \mathcal{N}_{p+q}(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{I}_p & \mathbf{\Sigma}_{\varepsilon\eta} \\ \mathbf{\Sigma}_{\eta\varepsilon} & \mathbf{\Sigma}_{\eta\eta} \end{pmatrix}, \quad t = 1, \ldots, n-1, \tag{6}$$

$$\boldsymbol{h}_1 \sim \mathcal{N}_q(\boldsymbol{\mu}, \mathbf{\Sigma}_0), \tag{7}$$

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 $q(q + 2p + 3)/2$. The covariance matrix of the initial latent variable, $\boldsymbol{\Sigma}_0$, is assumed to satisfy a stationary condition such that

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

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. 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, correlations, principal components, and the news impact curve.

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

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

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

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

3.1 Prior distributions

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

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

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

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

where $\mathcal{W}(n, \mathbf{R})$ denotes Wishart distribution with parameters n and $\mathbf{R}$.

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

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

where $f_N(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a normal density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. 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 $\boldsymbol{h}, \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$.

2. Generate $\boldsymbol{h}|\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, Y_n$.

3. Generate $\boldsymbol{\mu}|\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{h}, Y_n$.

4. Generate $\boldsymbol{\Sigma}|\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{h}, Y_n$.

5. Generate $\boldsymbol{\phi}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{h}, Y_n$.

6. Go to Step 2.

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

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

where

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

and

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

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

We generate a candidate $\boldsymbol{h}_t^{\dagger}$ from $\boldsymbol{h}_t^{\dagger} \sim \mathcal{N}_q(\boldsymbol{\gamma}_t, \mathbf{\Gamma}_t)$ 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 samples 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).

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

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

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

Consider the following linear Gaussian state space model:

$$\hat{\boldsymbol{y}}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{G}_t\boldsymbol{u}_t, \quad t = s+1, \ldots, s+m, \tag{18}$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \mathbf{K}_t\boldsymbol{u}_t, \quad t = s+1, \ldots, s+m-1, \tag{19}$$

$$\boldsymbol{u}_t = (\boldsymbol{\xi}_t', \boldsymbol{x}_t')' \sim \mathcal{N}_{2p}(\mathbf{0}, \mathbf{I}),$$

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. Define $\boldsymbol{d}_t, \mathbf{A}_t$ and $\mathbf{B}_t$ as in Appendix A. Let $\hat{\boldsymbol{d}}_t, \hat{\mathbf{A}}_t$ and $\hat{\mathbf{B}}_t$ denote the vector and matrices evaluated at $\boldsymbol{x}_t = \hat{\boldsymbol{x}}_t$ $(t = s, \ldots, s+m-1)$. Further, 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 $\hat{\boldsymbol{\alpha}}_s = \boldsymbol{h}_s - \boldsymbol{\mu}$.

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

$$\hat{\boldsymbol{y}}_t = \hat{\boldsymbol{\alpha}}_t + \mathbf{D}_t^{-1}(\boldsymbol{b}_t + \hat{\mathbf{B}}_t'\hat{\boldsymbol{\alpha}}_{t+1}), \quad \mathbf{Z}_t = \mathbf{I}_p + \mathbf{D}_t^{-1}\hat{\mathbf{B}}_t'\boldsymbol{\Phi},$$

$$\mathbf{G}_t = [\mathbf{D}_t^{-1/2\prime}, \mathbf{D}_t^{-1}\hat{\mathbf{B}}_t'\boldsymbol{\Sigma}_{\eta\eta}], \quad t = s+2, \ldots, s+m,$$

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 (18) and (19). 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

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

where $f(\{\boldsymbol{x}_t\}_{t=s}^{s+m-1})$ and $f^*(\{\boldsymbol{x}_t\}_{t=s}^{s+m-1})$ denote conditional posterior densities of $\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}$ for MESV model and the linear Gaussian state space model (18)–(19).

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

$$
\mathbf{R}^{-1} \;=\; \left[\begin{array}{cc} \mathbf{R}^{11} & \mathbf{R}^{12} \\ \mathbf{R}^{21} & \mathbf{R}^{22} \end{array}\right] = \left[\begin{array}{cc} \sum_{t=1}^{n-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' & \sum_{t=1}^{n-1}\boldsymbol{\varepsilon}_t\boldsymbol{\eta}_t' \\ \sum_{t=1}^{n-1}\boldsymbol{\eta}_t\boldsymbol{\varepsilon}_t' & \sum_{t=1}^{n-1}\boldsymbol{\eta}_t\boldsymbol{\eta}_t' \end{array}\right],
$$

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 $\boldsymbol{\Sigma}^{12}$ and $\boldsymbol{\Sigma}^{22}$ is obtained as follows.

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

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

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

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

1. Draw $\boldsymbol{\Sigma}^{22\dagger} \sim \mathcal{W}(n_1, \mathbf{R}_1)$.

2. Draw $\mathrm{vec}(\boldsymbol{\Sigma}^{21\dagger})|\boldsymbol{\Sigma}^{22\dagger} \sim \mathcal{N}_{pq}(\mathrm{vec}(\boldsymbol{\Sigma}^{22\dagger}\boldsymbol{\Delta}_1), \boldsymbol{\Omega}_1 \otimes \boldsymbol{\Sigma}^{22\dagger})$.

3. Compute $\boldsymbol{\Sigma}_{\varepsilon\eta}^\dagger = -\boldsymbol{\Sigma}^{12\dagger}\boldsymbol{\Sigma}^{22\dagger-1}$ and $\boldsymbol{\Sigma}_{\eta\eta}^\dagger = \boldsymbol{\Sigma}^{22\dagger-1} + \boldsymbol{\Sigma}_{\eta\varepsilon}^\dagger\boldsymbol{\Sigma}_{\varepsilon\eta}^\dagger$.

and accept it with probability

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

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

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

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

where

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

Generation of $\boldsymbol{\phi}$. Let $\mathbf{A} = \sum_{t=1}^{n-1}(\boldsymbol{h}_t - \boldsymbol{\mu})(\boldsymbol{h}_t - \boldsymbol{\mu})'$, $\mathbf{B} = \sum_{t=1}^{n-1}\{(\boldsymbol{h}_t - \boldsymbol{\mu})\boldsymbol{y}_t'\exp(-\mathbf{H}_t/2)\boldsymbol{\Sigma}^{12} + (\boldsymbol{h}_t - \boldsymbol{\mu})(\boldsymbol{h}_{t+1} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{22}\}$ and $\boldsymbol{b}$ denote a vector whose i-th element is equal to the (i,i)-th element of $\mathbf{B}$. Then the conditional posterior probability density function of $\boldsymbol{\phi}$ is

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

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

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

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

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

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

where

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

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 $(\boldsymbol{h}_{t+1}^i, \boldsymbol{h}_t^i) \sim g(\boldsymbol{h}_{t+1}, \boldsymbol{h}_t^i|Y_{t+1}, \boldsymbol{\theta})$ $(i = 1, \ldots, I)$:

(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

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

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 $\sum_{t=1}^n \log \bar{w}_t \xrightarrow{p} \sum_{t=1}^n \log f(y_t|Y_{t-1}, \boldsymbol{\theta})$.

13

4 Illustrative example with simulated data

This section shows the efficiency of our proposed method using a simulated data. To replicate the dynamics of the stock return series, we use the MESV model with

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

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

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

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

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 $E(3,3) = 6$.

We generate $n = 4,000$ observations with $p = 3$ and $q = 6$. Prior distributions are assumed to be as follows.

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

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.

Parameter estimation. 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. 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 (see e.g. Chib (2001)).

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]

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

The inefficiency factors for the single-move sampler are about three times larger than those for the multi-move sampler. This implies that our proposed multi-move sampler is more efficient than the single-move sampler as we expected.

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]
	1 11	-0.277	[-0.388,-0.159]	61	[133]
−0.3	2 22	-0.255	[-0.372,-0.134]	84	[126]
	3 33	-0.374	[-0.479,-0.264]	81	[220]
	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]
−0.1	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]
	11 22	0.248	[0.080, 0.411]	102	[340]
0.3	11 33	0.332	[0.170, 0.481]	135	[382]
	22 33	0.266	[0.104, 0.420]	68	[186]
	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]
0.1	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]

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.
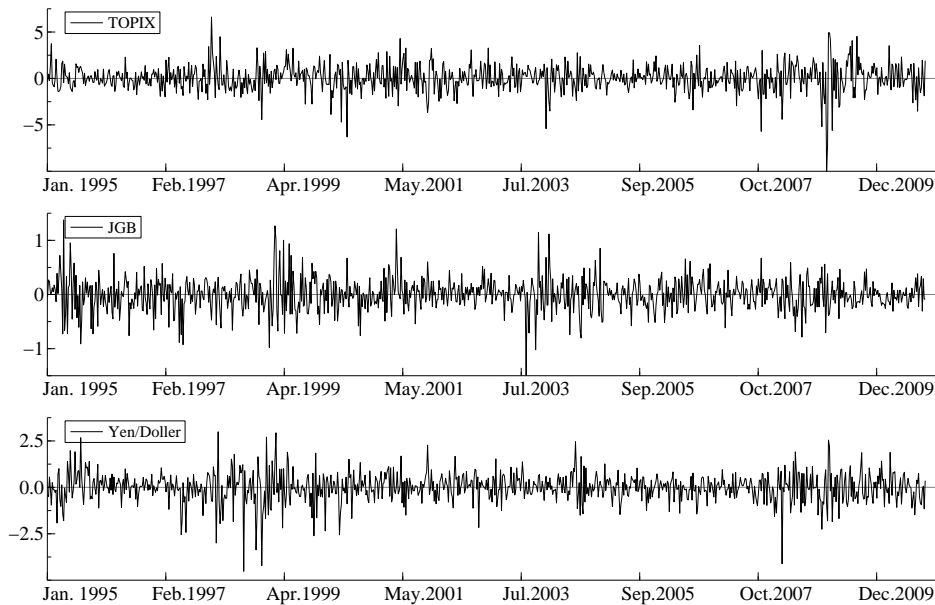
Figure 1: TOPIX, JGB and Yen/USD

5.2 Estimation results

5.2.1 Univariate SV and MSV models

We first estimate the (univariate) SV model with leverage effect defined by (1)–(3) with $p = 1$ for each series, and the constant correlation multivariate SV (MSV) model defined by (1)–(3) with $p = 3$. For the SV model, the prior distributions for ϕ_i and Σ_i are assumed

to be

$$\frac{\phi_i + 1}{2} \sim \mathcal{B}(20, 1.5), \quad \mathbf{\Sigma}_i \sim \mathcal{IW}(5, (5\mathbf{\Sigma}^*)^{-1}), \quad \mathbf{\Sigma}^* = \begin{pmatrix} 1 & 0 \\ 0 & 0.04 \end{pmatrix}, \quad i = 1, 2, 3.$$

For the MSV model, we assume prior distributions such that

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

where

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

Tables 4 and 5 show summary statistics of posterior distributions of the parameters for the SV and the MSV models. In both models, the posterior means of the autoregressive parameters (ϕ_j's) are very high (between 0.938 and 0.972 in the SV model and between 0.948 and 0.966 in the MSV model) showing that volatilities are highly persistent.

Table 4: Univariate SV model.
Posterior means and 95% credible intervals.

Param.	i	Mean	95% interval
	1	0.971	[0.960, 0.981]
ϕ_i	2	0.972	[0.959, 0.983]
	3	0.938	[0.897, 0.966]
	1	1.177	[1.078, 1.285]
$\sigma_{i\varepsilon}$	2	0.267	[0.240, 0.298]
	3	0.652	[0.609, 0.699]
	1	0.177	[0.148, 0.207]
$\sigma_{i\eta}$	2	0.184	[0.150, 0.220]
	3	0.239	[0.175, 0.321]
	1	-0.511	[-0.604,-0.408]
ρ_i	2	-0.176	[-0.293,-0.052]
	3	-0.201	[-0.316,-0.079]

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

Also, the leverage effects are estimated to be negative as in the previous literature where that of the stock return (-0.511 in the SV model and -0.445 in the MSV model) is much stronger than those of the bond return and the foreign exchange return (-0.176 and -0.201 in the SV model, -0.177 and -0.167 in the MSV model). In the MSV model, 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 5: MSV model.
Posterior means and 95% credible intervals.

Param.	i	Mean	95% interval	Param.	ij	Mean	95% interval
	1	0.965	[0.953, 0.975]		12	-0.312	[-0.343,-0.281]
ϕ_i	2	0.966	[0.952, 0.978]	$\rho_{ij,\varepsilon\varepsilon}$	13	0.062	[0.028, 0.095]
	3	0.948	[0.918, 0.969]		23	-0.014	[-0.048, 0.020]
	1	1.212	[1.115, 1.322]		12	0.196	[0.024, 0.363]
$\sigma_{i,\varepsilon\varepsilon}$	2	0.275	[0.249, 0.305]	$\rho_{ij,\eta\eta}$	13	0.559	[0.402, 0.694]
	3	0.662	[0.615, 0.716]		23	0.390	[0.221, 0.552]
	1	0.191	[0.163, 0.224]		12	0.189	[0.073, 0.301]
$\sigma_{i,\eta\eta}$	2	0.210	[0.174, 0.249]		13	-0.003	[-0.127, 0.117]
	3	0.225	[0.173, 0.293]		21	0.091	[-0.021, 0.200]
	1	-0.445	[-0.543,-0.339]	$\rho_{ij,\varepsilon\eta}$	23	0.035	[-0.077, 0.145]
$\rho_{ii,\varepsilon\eta}$	2	-0.177	[-0.281,-0.067]		31	-0.036	[-0.152, 0.079]
	3	-0.167	[-0.282,-0.045]		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

Next, we estimate the MESV model assuming that prior distributions are

$$\frac{\phi_{ij} + 1}{2} \sim \begin{cases} \mathcal{B}(20, 3/2), & \text{if } i = j, \\ \mathcal{B}(33/4, 11/4), & \text{otherwise,} \end{cases}$$
$$\boldsymbol{\mu} \sim \mathcal{N}_q(\mathbf{0}, 5\mathbf{I}_q),$$
$$\boldsymbol{\Sigma}^{22} \sim \mathcal{W}(6, (6\boldsymbol{\Sigma}^{22*})^{-1}), \quad \mathrm{vec}(\boldsymbol{\Sigma}^{21})|\boldsymbol{\Sigma}^{22} \sim \mathcal{N}_{pq}(\mathbf{0}, (5\mathbf{I}_p) \otimes \boldsymbol{\Sigma}^{22}).$$

The (i, j)-th element of $\boldsymbol{\Sigma}^{22*-1}$ is $\rho^*_{E(k,l)E(m,n),\eta\eta}\sigma^*_{E(k,l),\eta\eta}\sigma^*_{E(m,n),\eta\eta}$ such that

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

19

The mean and the standard deviation of the prior distribution of ϕ_{ij}, $i \neq j$ are set 0.5 and 0.5 respectively.

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.

Table 6: MESV model.
Posterior means, standard deviations, 95% credible intervals, and inefficiency factors.

	ij	Mean	Stdev	95% interval	IF
	11	0.968	0.005	[0.958, 0.977]	141
	22	0.968	0.006	[0.955, 0.979]	94
ϕ_{ij}	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
	11	0.286	0.087	[0.114, 0.457]	6
	22	-2.923	0.113	[-3.143,-2.699]	11
μ_{ij}	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
	11	0.176	0.014	[0.150, 0.204]	253
	22	0.211	0.018	[0.176, 0.249]	277
$\sigma_{E(i,j),\eta\eta}$	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

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

The estimation results are summarized in the Tables 6, 7 and 8. 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 5 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 $(-0.445, -0.177, -0.167)$.

Table 7: 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 8: 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

Estimated volatilities and correlations. However, as mentioned in Section 2, the parameters such as $\boldsymbol{\Phi}$, $\mathbf{M}$ and $\boldsymbol{\Sigma}_{\varepsilon\eta}$ 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 of 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.

Next, we investigate the posterior means of 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.
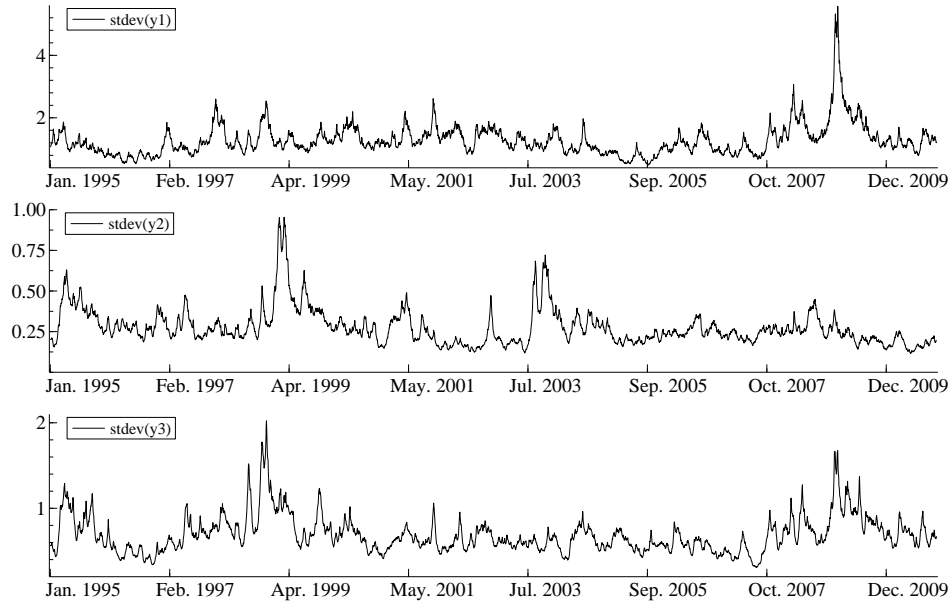

Figure 2: Posterior means of the square root of time-varying variances. Top: TOPIX, middle: JGB, bottom: Yen/USD.
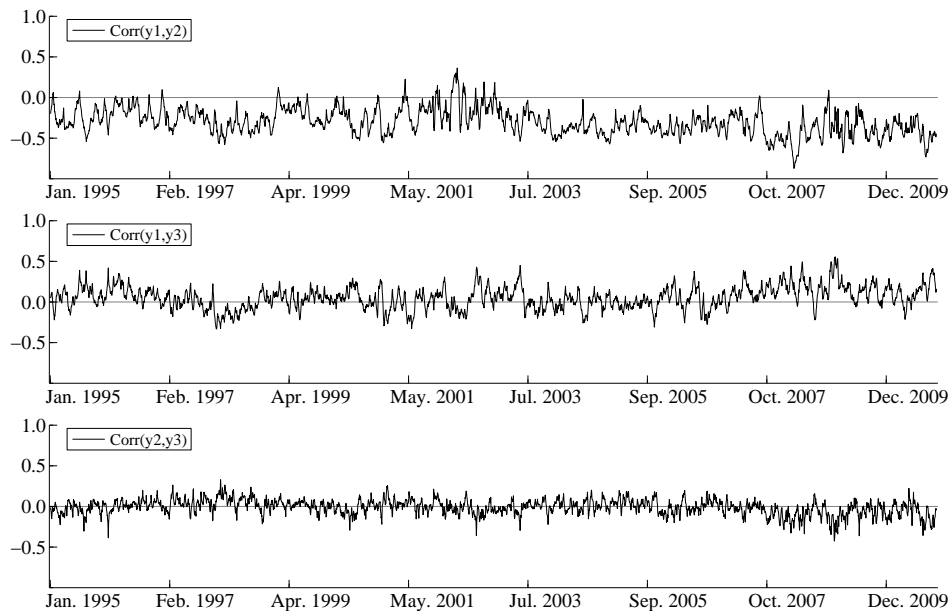

Figure 3: Posterior means of time-varying correlations. Top: (TOPIX, JGB), middle: (TOPIX, Yen/USD), bottom: (JGB, Yen/USD).

23

Dynamic principal components implied by the MESV model. Furthermore, we can interpret our estimation results for MESV model in terms of the principal component analysis. Using the diagonalization to calculate the matrix exponential, we calculate the principal components of $\boldsymbol{y}_t$. Since $\mathbf{H}_t = \mathbf{U}_t \boldsymbol{\Lambda}_t \mathbf{U}_t'$ where $\boldsymbol{\Lambda}_t = \text{diag}(\lambda_{1t}, \lambda_{2t}, \lambda_{3t})$ and $\mathbf{U}_t \mathbf{U}_t' = \mathbf{I}_3$, we obtain

$$\mathbf{U}_t' \boldsymbol{y}_t \quad = \quad \text{diag}(\exp(\lambda_{1t}/2), \exp(\lambda_{2t}/2), \exp(\lambda_{3t}/2)) \boldsymbol{\varepsilon}_t^*, \tag{20}$$

where $\boldsymbol{\varepsilon}_t^* := \mathbf{U}_t' \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$, and we assume $\lambda_{1t} \geq \lambda_{2t} \geq \lambda_{3t}$. The i-th element of $\mathbf{U}_t' \boldsymbol{y}_t$ in (20) can be interpreted as the i-th principal component of the returns at time t. Moreover, the row vector of the loading matrix $\mathbf{U}_t'$ can be a interpreted as a weight vector for the portfolio which represents the stock market principal component. Top and middle panels of Figure 4 show time series plots of the principal components, $\mathbf{U}_t' \boldsymbol{y}_t$, $t = 1, \ldots, n$ and their standard deviations.
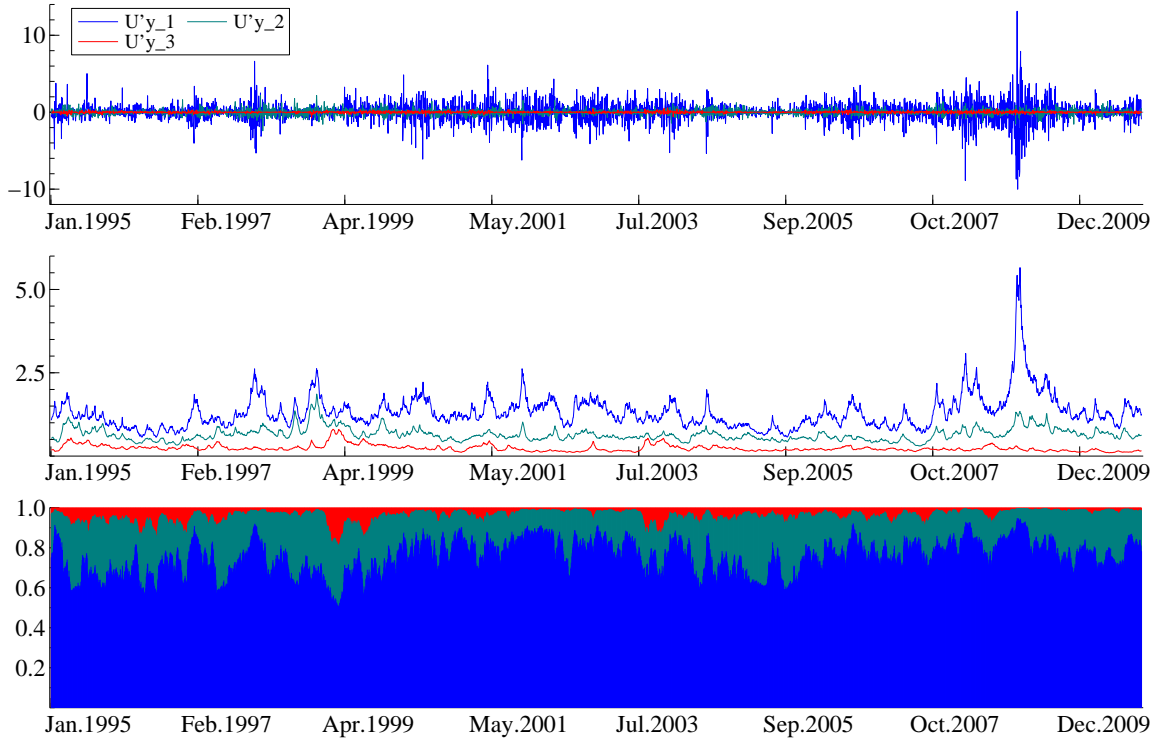
Figure 4: Posterior means of principal components series (top) and their standard deviations (middle), and contribution ratio series (bottom).

24

The first principal component (blue lines) fluctuates sharply and its variance shot up when at least one of the volatilities of the original series increase around 1998, 2003 and 2008. The bottom panel of Figure 4 shows the contribution ratio of each principal component, defined by $\frac{\exp(\lambda_{1t})}{\sum_{i=1}^{3}\exp(\lambda_{it})}, \frac{\exp(\lambda_{2t})}{\sum_{i=1}^{3}\exp(\lambda_{it})}, \frac{\exp(\lambda_{3t})}{\sum_{i=1}^{3}\exp(\lambda_{it})}$ for $t = 1, \ldots, n$. The ratio of the first principal component (blue part) increases when one of three return series became more volatile. On the other hand, the contribution ratio of the second series (green part) decreases in those volatile periods. The third principal component (red part) has the ratio less than 10 % over almost all the sample period. This implies that most of the variation of the original three series can be explained by two principal components.

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 $\boldsymbol{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Sigma}_{\eta\varepsilon}\exp(-\mathbf{M}/2)\boldsymbol{y}_t$.
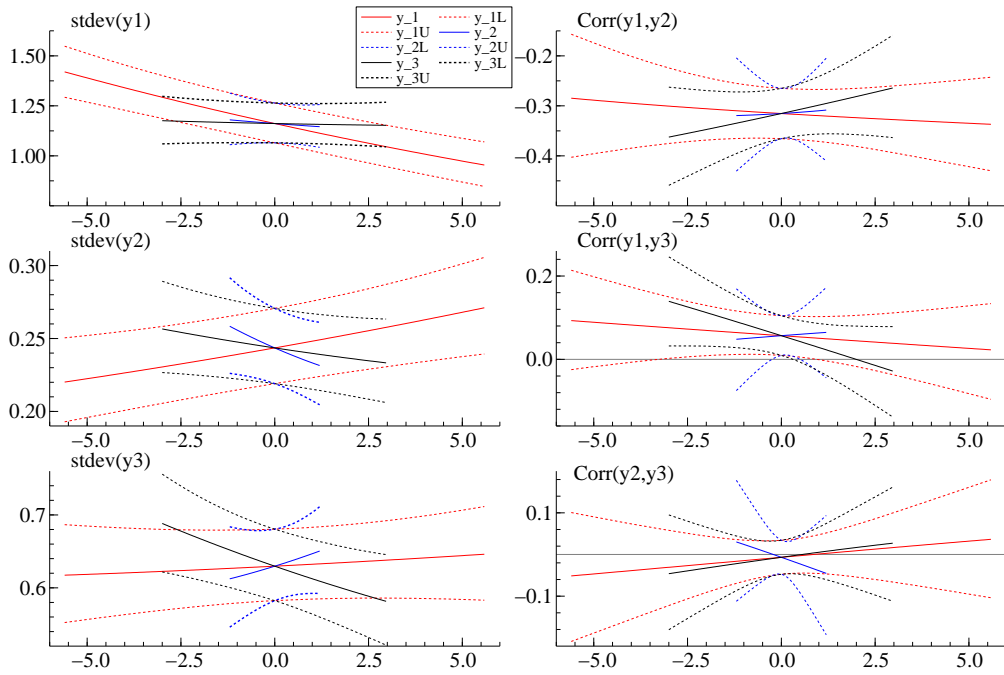

Figure 5: Posterior mean (solid line) and 95% interval (dotted lines) of the news impact curve* for the one-step-ahead conditional covariances of $\boldsymbol{y}_{t+1}$ when $\mathbf{H}_t = \mathbf{M}$ and $\mathbf{E}_t = \mathbf{O}$. (* The domain of the estimated curve is restricted to the range within ± 4 sample standard deviation of actual returns for each y_{it}.)

Figure 5 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 5 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 5 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.

It is noted that the 95% intervals of the impact on the standard deviations and correlations looks very large and that some of them include the horizontal line (*i.e.* no leverage case). Thus, we calculate the posterior probability that the standard deviations and the correlations calculated from $\exp(\mathbf{H}_{t+1})$ are larger than the posterior means of the standard

deviations and the correlations from $\exp(\mathbf{M})$. It means that the news impacts are larger than the posterior means of them at $\boldsymbol{y}_t = 0$. Figure 6 shows the posterior probabilities of such events. The left three panels show the results for standard deviations. The posterior probability for the standard deviation of the TOPIX (exchange rate) return approaches one as the TOPIX (exchange rate) return decreases in the top (middle) left panel. The probability for the standard deviation of the JGB return also approaches one as the TOPIX return increases in the middle left panel. The right three panels show the results for correlation coefficients. The posterior probability for the correlation between the TOPIX and exchange rate returns rapidly goes to one as the exchange rate return decreases in the middle right panel.
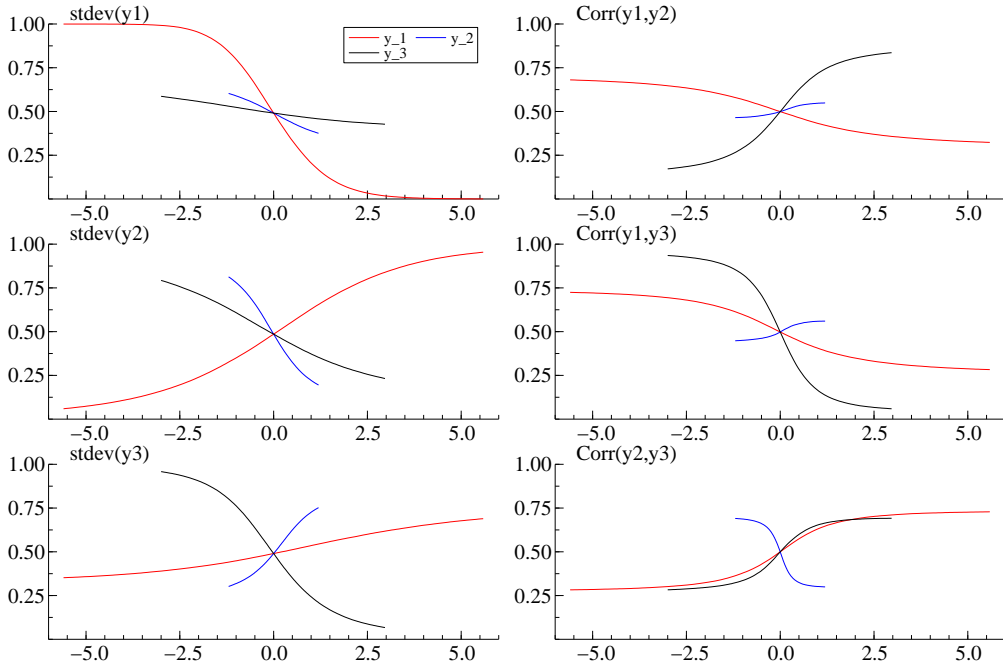
Figure 6: Posterior probability that news impact curves are larger than the posterior mean of $\exp(\mathbf{M})$.

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

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

27

where ξ_t is a random variable which takes positive values and independent of e_t. We consider two cases: (i) multivariate Student-t error with $\xi_t^{-1} \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ and (ii) log-Gaussian scale mixture error with $\log \xi_t \sim \mathcal{N}\left(-\frac{\delta^2}{2}, \delta^2\right)$. The extensions to these models are 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.

- MSV-g model: MSV model with log-Gaussian scale mixture error distribution.

- MESV-n model: MESV model with normal error distribution.

- MESV-t model: MESV model with multivariate Student-t error distribution.

- MESV-g model: MESV model with log-Gaussian scale mixture error distribution.

We assume the prior distributions $\nu \sim \mathcal{G}(0.001, 0.001)$ and $\delta^2 \sim \mathcal{G}(1,1)$ respectively. The estimation results for ν and δ^2 are summarized in Table 9. The posterior means of ν for the MSV and MESV models are small, suggesting fat-tailed error distributions. The estimate of ν (δ^2) for the MSV model is smaller (larger) 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.

Table 9: The estimation results of ν and δ^2

Model	Param.	Mean	Stdev	95% interval	Param.	Mean	Stdev	95% interval
MESV	ν	15.8	2.4	[11.8, 21.5]	δ^2	0.17	0.02	[0.13, 0.22]
MSV		11.8	1.2	[9.8, 14.5]		0.34	0.02	[0.30, 0.38]

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 $E_{\theta|Y_n}[D(\theta)]$, we use a sample analogue $\frac{1}{M} \sum_{m=1}^{M} D(\theta^{(m)})$, where we set $M = 100$, and $\theta^{(m)}$s are resampled from the posterior samples generated by the MCMC method. To calculate $D(E_{\theta|Y_n}[\theta])$, which equals to $D(\theta)$ evaluated at the posterior mean, we implement an auxiliary particle filter to compute the log-likelihood ordinate $\log f(Y_n|\theta)$, 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 10: The averages of DIC estimates, their standard errors,
the maximum and the minimum of DIC values.

Model	ranking	DIC	(s.e.)	$\mathrm{DIC_{max}}$	$\mathrm{DIC_{min}}$
MESV-t	1	20066.7	(1.7)	20074.2	20058.3
MESV-g	2	20087.4	(1.8)	20099.6	20079.6
MESV-n	3	20090.1	(1.5)	20098.4	20084.1
MSV-t	4	20118.4	(1.3)	20124.3	20112.4
MSV-g	5	20197.3	(3.2)	20214.3	20180.5
MSV-n	6	20201.2	(1.0)	20206.3	20196.0

Table 10 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 hence the MESV models out-perform the MSV 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.

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.

Appendix

A   Multi-move sampler

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

$$\log f(\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}|\boldsymbol{h}_s, \boldsymbol{h}_{s+m+1}, \boldsymbol{y}_s, \ldots, \boldsymbol{y}_{s+m}) = -\frac{1}{2}\sum_{t=s}^{s+m-1} \boldsymbol{x}_t'\boldsymbol{x}_t + L, \qquad (21)$$

where

$$L = \sum_{t=s}^{s+m} l_t - \frac{1}{2}\boldsymbol{\eta}'_{s+m}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\eta}_{s+m}I(s+m<n), \tag{22}$$

$$\boldsymbol{\eta}_{s+m} = \boldsymbol{h}_{s+m+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{h}_{s+m} - \boldsymbol{\mu}), \tag{23}$$

$$\boldsymbol{x}_t = \boldsymbol{\Sigma}_{\eta\eta}^{-1/2}(\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu})).$$

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 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

$$\log f(\{\boldsymbol{x}_t\}_{t=s}^{s+m-1}|\boldsymbol{h}_s, \boldsymbol{h}_{s+m+1}, \boldsymbol{y}_s, \ldots, \boldsymbol{y}_{s+m})$$
$$\approx \text{const.} - \frac{1}{2}\sum_{t=s}^{s+m-1} \boldsymbol{x}'_t\boldsymbol{x}_t + \hat{L}$$
$$+ \sum_{t=s+1}^{s+m} \left[\hat{\boldsymbol{d}}'_t - \frac{1}{2}(\boldsymbol{\alpha}_t - \hat{\boldsymbol{\alpha}}_t)'\hat{\mathbf{A}}_t + (\boldsymbol{\alpha}_{t-1} - \hat{\boldsymbol{\alpha}}_{t-1})'\hat{\mathbf{B}}_t\right](\boldsymbol{\alpha}_t - \hat{\boldsymbol{\alpha}}_t)$$
$$= \text{const.} + \log f^*(\boldsymbol{x}|\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+m+1}, \boldsymbol{y}_s, \ldots, \boldsymbol{y}_{s+m}), \tag{24}$$

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

$$\mathbf{A}_t = -E\left[\frac{\partial^2 L}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}'_t}\right], \quad t = s+1, \ldots, s+m, \tag{25}$$

$$\mathbf{B}_t = -E\left[\frac{\partial^2 L}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}'_{t-1}}\right], \quad t = s+2, \ldots, s+m, \quad \mathbf{B}_{s+1} = \mathbf{O}, \tag{26}$$

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.

A.1   Derivation of 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

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

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

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

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

$$\mathbf{D}H(X) = (\mathbf{D}G(F(X)))(\mathbf{D}F(X)) = \frac{\partial\text{vec}(G(F(\boldsymbol{X})))}{\partial(\text{vec}(F(\boldsymbol{X})))'}\frac{\partial\text{vec}(F(\boldsymbol{X}))}{\partial(\text{vec}(\boldsymbol{X}))'}. \tag{27}$$

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

$$\frac{\partial g(f(\mathbf{x}))}{\partial \mathbf{x}'} = \frac{\partial\text{vec}(g(f(\mathbf{x})))}{\partial\text{vec}(f(\mathbf{x}))'}\frac{\partial\text{vec}(f(\mathbf{x}))}{\partial\text{vec}(\mathbf{x})'}. \tag{28}$$

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

$$\frac{\partial\text{vec}(FG)}{\partial(\text{vec}(X))'} = (G' \otimes \mathbf{I}_m)\frac{\partial\text{vec}(F)}{\partial(\text{vec}(X))'} + (\mathbf{I}_r \otimes F)\frac{\partial\text{vec}(G)}{\partial(\text{vec}(X))'}. \tag{29}$$

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

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

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 $\text{vec}(\mathbf{A}) = \mathbf{D}_p\text{vech}(\mathbf{A})$ for a symmetric matrix $\mathbf{A}$. Then

$$\frac{\partial\text{tr}(\mathbf{H}_t)}{\partial\boldsymbol{\alpha}_t'} = \frac{\partial\text{tr}(\mathbf{H}_t)}{\partial\text{vec}(\mathbf{H}_t)'}\frac{\partial\text{vec}(\mathbf{H}_t)}{\partial\boldsymbol{h}_t'} = \text{vec}(\mathbf{I}_p)'\mathbf{D}_p = \text{vech}(\mathbf{I}_p)', \tag{30}$$

$$\frac{\partial\boldsymbol{m}_t}{\partial\boldsymbol{\alpha}_t'} = -\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t < n), \tag{31}$$

$$\frac{\partial\boldsymbol{m}_{t-1}}{\partial\boldsymbol{\alpha}_t'} = \boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\eta\eta}^{-1}I(t > 1), \tag{32}$$

32

where we used the chain rule and $\mathbf{D}'_p \mathrm{vec}(\mathbf{A}) = \mathrm{vech}(\mathbf{A} + \mathbf{A}' - (\mathbf{A} \odot \mathbf{I}_p))$ for a $p \times p$ matrix $\mathbf{A}$ in (30) (e.g. Magnus and Neudecker (1999), Magnus (1988)). Further, define

$$
\mathbf{V}_t \equiv \frac{\partial \mathrm{vec}(\exp(\mathbf{F}_t))}{\partial \mathrm{vec}(\mathbf{F}_t)'} = \sum_{i=1}^{\infty} \frac{1}{i!} \sum_{j=1}^{i} \left[\mathbf{F}_t^{i-j} \otimes \mathbf{F}_t^{j-1} \right],
$$

using the product rule and let $\mathbf{Q}_t = \{\exp(-\mathbf{F}_t) \otimes \mathbf{I}_p\} \mathbf{V}_t$. Noting that $\boldsymbol{z}_t = (\boldsymbol{y}'_t \otimes \mathbf{I}_p) \mathrm{vec}(\exp(\mathbf{F}_t))$,

$$
\frac{\partial \boldsymbol{z}_t}{\partial \boldsymbol{\alpha}'_t} = -\frac{1}{2}(\boldsymbol{y}'_t \otimes \mathbf{I}_p) \mathbf{V}_t \frac{\partial \mathrm{vec}(\mathbf{H}_t)}{\partial \boldsymbol{h}'_t} = -\frac{1}{2}(\boldsymbol{z}'_t \otimes \mathbf{I}_p) \mathbf{Q}_t \mathbf{D}_p. \tag{33}
$$

Using (30) – (33) 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

$$
\begin{aligned}
\boldsymbol{d}_t &= \left[\frac{\partial l_t}{\partial \boldsymbol{\alpha}'_t} \right]' + \left[\frac{\partial l_{t-1}}{\partial \boldsymbol{\alpha}'_t} \right]' + \boldsymbol{\Phi} \boldsymbol{\Sigma}_{\eta\eta}^{-1} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi} \boldsymbol{\alpha}_t) I(t = s + m < n) \\
&= -\frac{1}{2} \mathrm{vech}(\mathbf{I}_p) + \left\{ \frac{1}{2} \mathbf{D}'_p \mathbf{Q}'_t (\boldsymbol{z}_t \otimes \mathbf{I}_p) - \boldsymbol{\Phi} \boldsymbol{\Sigma}_{\eta\eta}^{-1} \boldsymbol{\Sigma}_{\eta\varepsilon} I(t < n) \right\} \mathbf{S}_t^{-1} (\boldsymbol{z}_t - \boldsymbol{m}_t) \\
&\quad + \boldsymbol{\Sigma}_{\eta\eta}^{-1} \boldsymbol{\Sigma}_{\eta\varepsilon} \mathbf{S}_{t-1}^{-1} (\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1}) I(t > 1) + \boldsymbol{\Phi} \boldsymbol{\Sigma}_{\eta\eta}^{-1} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi} \boldsymbol{\alpha}_t) I(t = s + m < n). \tag{34}
\end{aligned}
$$

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

$$
\begin{aligned}
\mathbf{A}_t &= -\frac{1}{2} \mathbf{D}'_p E \left[\frac{\partial \mathbf{V}_t (\boldsymbol{y}_t \otimes \mathbf{I}_p) \mathbf{S}_t^{-1} (\boldsymbol{z}_t - \boldsymbol{m}_t)}{\partial \boldsymbol{\alpha}'_t} \right] - \frac{1}{2} \boldsymbol{\Phi} \mathbf{N}'_t \\
&\quad + \boldsymbol{\Phi} \mathbf{M}_t \boldsymbol{\Phi} + \mathbf{M}_{t-1} + \boldsymbol{\Phi} \boldsymbol{\Sigma}_{\eta\eta}^{-1} \boldsymbol{\Phi} I(t = s + m < n), \tag{35}
\end{aligned}
$$

where

$$
\mathbf{M}_t = \boldsymbol{\Sigma}_{\eta\eta}^{-1} \boldsymbol{\Sigma}_{\eta\varepsilon} \mathbf{S}_t^{-1} \boldsymbol{\Sigma}_{\varepsilon\eta} \boldsymbol{\Sigma}_{\eta\eta}^{-1} I(1 \leq t < n), \tag{36}
$$

$$
\mathbf{N}_t = \mathbf{D}'_p \mathbf{Q}'_t (\boldsymbol{m}_t \otimes \mathbf{I}_p) \mathbf{S}_t^{-1} \boldsymbol{\Sigma}_{\varepsilon\eta} \boldsymbol{\Sigma}_{\eta\eta}^{-1}. \tag{37}
$$

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

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

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

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

and

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

Equation (35) reduces to

$$
\mathbf{A}_t = \frac{1}{4}\mathbf{D}_p'\left[\mathbf{P}_t + \mathbf{Q}_t'\left\{(\mathbf{S}_t + \boldsymbol{m}_t\boldsymbol{m}_t') \otimes \mathbf{S}_t^{-1}\right\}\mathbf{Q}_t\right]\mathbf{D}_p - \frac{1}{2}(\mathbf{N}_t\boldsymbol{\Phi} + \boldsymbol{\Phi}\mathbf{N}_t')
$$
$$
+ \boldsymbol{\Phi}\mathbf{M}_t\boldsymbol{\Phi} + \mathbf{M}_{t-1} + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Phi}I(t = s + m < n), \tag{38}
$$

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

Derivation of $\mathbf{B}_t$

By (31), (33) and (34),

$$
\mathbf{B}_t = -\boldsymbol{\Sigma}_{\eta\eta}^{-1}\boldsymbol{\Sigma}_{\eta\varepsilon}\mathbf{S}_{t-1}^{-1}E\left[\frac{\partial(\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1})}{\partial \boldsymbol{\alpha}_{t-1}'}\right] = \frac{1}{2}\mathbf{N}_{t-1}' - \mathbf{M}_{t-1}\boldsymbol{\Phi}, \quad t = 2, \ldots, n. \tag{39}
$$

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

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

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

$$
\begin{aligned}
\mathbf{Q}_t &= \{\exp(-\mathbf{F}_t) \otimes \mathbf{I}_p\} \sum_{i=1}^{\infty} \frac{1}{i!} \sum_{j=1}^{i} \left[\mathbf{F}_t^{i-j} \otimes \mathbf{F}_t^{j-1} \right], \\
&= \sum_{i=1}^{\infty} \frac{1}{i!} \sum_{j=1}^{i} \left[\exp(-\mathbf{F}_t)\mathbf{F}_t^{i-j} \otimes \mathbf{F}_t^{j-1} \right], \\
&= (\mathbf{U}_t \otimes \mathbf{U}_t) \left[\sum_{i=1}^{\infty} \frac{1}{i!} \sum_{j=1}^{i} \left\{ \exp(-\mathbf{\Lambda}_t)\mathbf{\Lambda}_t^{i-j} \otimes \mathbf{\Lambda}_t^{j-1} \right\} \right] (\mathbf{U}_t' \otimes \mathbf{U}_t'). \qquad (40)
\end{aligned}
$$

The second factor in (40) is a diagonal matrix with its (k, k)-th element given by

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

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

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

and, using the product rule,

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

Note that

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

and

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

by using

$$
\begin{aligned}
\mathbf{K}_{nn} (\mathbf{BA} \otimes \mathbf{I}_n) \quad &= \quad (\mathrm{vec}(\mathbf{A})' \otimes \mathbf{I}_{n^2})(\mathbf{I}_n \otimes \mathbf{K}_{nn} \otimes \mathbf{I}_n)(\mathbf{I}_{n^2} \otimes \mathrm{vec}(\mathbf{B})), \qquad (41) \\
(\mathbf{BA}' \otimes \mathbf{I}_n) \quad &= \quad (\mathrm{vec}(\mathbf{A})' \otimes \mathbf{I}_{n^2})(\mathbf{I}_n \otimes \mathbf{K}_{nn} \otimes \mathbf{I}_n)(\mathrm{vec}(\mathbf{B}) \otimes \mathbf{I}_{n^2}), \qquad (42)
\end{aligned}
$$

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 $\mathbf{D}'_p \mathbf{K}_{pp} = \mathbf{D}'_p$,

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

and we obtain

$$
\begin{aligned}
\mathbf{P}_t &= 2 \sum_{i=2}^{\infty} \frac{1}{i!} \sum_{j=1}^{i-1} \sum_{h=1}^{i-j} (i-j) \exp(-\mathbf{F}_t) \mathbf{F}_t^{i-j-1} \otimes \mathbf{F}_t^{j-1} \\
&= (\mathbf{U}_t \otimes \mathbf{U}_t) \left[2 \sum_{i=2}^{\infty} \frac{1}{i!} \sum_{j=1}^{i-1} (i-j) \left\{ \exp(-\mathbf{\Lambda}_t) \mathbf{\Lambda}_t^{i-j-1} \otimes \mathbf{\Lambda}_t^{j-1} \right\} \right] (\mathbf{U}'_t \otimes \mathbf{U}'_t). \quad (43)
\end{aligned}
$$

The second factor in (43) is a diagonal matrix with its (k, k)-th element given by

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

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

References

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

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). Thedistribution of stock return volatility. *Journal of Financial Economics 61*, 43–76.

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.

Audrino, F. and F. Corsi (2010). Modelling tick-by-tick realized correlations. *Computational Statistics and Data Analysis 54*, 2372–2382.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Econometric analysis of realized covariation: high frequency covariance, regression and correlation in financial economics. *Econometrica 72*, 885–925.

Bauer, G. H. and K. Vorkink (2010). Forecasting multivariate realized stock market volatility. *JoE*. doi:10.1016/j.jeconom.2010.03.021.

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.

Chiriac, R. and V. Voev (2010). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*. DOI: 10.1002/jae.1152.

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.

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

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.

Jin, X. and J. M. Maheu (2010). Modelling realized covariances and returns. Working Paper 408, University of Toronto.

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.

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

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.

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.

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

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.

Voev, V. (2007). Dynamic modelling of large-dimensional covariance matrices. In L. Bauwens, W. Pohlmeier, and D. Veredas (Eds.), *Recent Developments in High Frequency Financial Econometrics*, pp. 365–402. Berlin: Springer-Verlag.

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.