CIRJE-F-953

Dynamic Equicorrelation Stochastic Volatility

Yuta Kurose
Kwansei Gakuin University

Yasuhiro Omori
The University of Tokyo

January 2015

Dynamic Equicorrelation Stochastic Volatility

Yuta Kurose*and Yasuhiro Omori†

January, 2015

Abstract

A multivariate stochastic volatility model with dynamic equicorrelation and cross leverage effect is proposed and estimated. Using a Bayesian approach, an efficient Markov chain Monte Carlo algorithm is described where we use the multi-move sampler, which generates multiple latent variables simultaneously. Numerical examples are provided to show its sampling efficiency in comparison with the simple algorithm that generates one latent variable at a time given other latent variables. Furthermore, the proposed model is applied to the multivariate daily stock price index data. The model comparisons based on the portfolio performances and DIC show that our model overall outperforms competing models.

Key words: Asymmetry, cross leverage effect, dynamic equicorrelation, Markov chain Monte Carlo, multi-move sampler, multivariate stochastic volatility.

*School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda-shi, Hyogo 669-1337, JAPAN.

†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN.

1 Introduction

Over the last several decades, various multivariate volatility models have been proposed to model asset returns with time-varying variances. Two popular examples are generalized autoregressive conditional heteroskedasticity (GARCH) models (Bauwens, Laurent, and Rombouts (2006)) and multivariate stochastic volatility (SV) models (Asai, McAleer, and Yu (2006), Chib, Omori, and Asai (2009)).

They are proposed to model the volatility clustering and the dynamic correlations, which are found to exist in empirical studies of financial time series (Bauwens, Hafner, and Laurent (2012)). Dynamic conditional correlation (DCC) models (Engle (2002)) and BEKK models (Engle and Kroner (1995)) are such widely used multivariate GARCH models. They simplify the multivariate covariance structure since there is an increasing difficulty in estimating too many parameters for dynamic correlations for high dimensional data. To overcome the difficulty, Engle and Kelly (2012), Vargas (2009), Jin and Tang (2009) and Clements, Coleman-Fenn, and Smith (2011) proposed the dynamic equicorrelation (DECO) model, which is based on a DCC model with all correlations equal but time-varying. Making reference to Elton and Gruber (1973), they argue that the dynamic equicorrelation assumption gives a superior portfolio allocation. In a Bayesian context, Ledoit and Wolf (2004) proposed the covariance matrix estimator obtained by shrinking the sample correlation matrix to an equicorrelated matrix for the purpose of the portfolio optimization. Hafner and Reznikova (2012) applied the shrinkage methods to the DCC models and improved the estimation results of the DCC model. Lucas, Schwaab, and Zhang (2012) proposed the dynamic generalized hyperbolic (GH) skew-t-error model with generalized autoregressive score (GAS) equicorrelation structure. For an asset allocation, an equicorrelated factor model is sometimes considered as a mean of dimension reduction (e.g., McNeil, Frey, and Embrechts (2005)).

In volatilities of stock returns, we often observe the asymmetry or the cross leverage effect, which implies a decrease in the i-th dependent variable at date t followed by an increase in the j-th latent stochastic variance at date $(t+1)$. The simple univariate SV model with leverage effect is given in a state space form as:

$$y_t = m + \exp(h_t/2)\epsilon_t, \quad t = 1, \ldots, n, \tag{1}$$

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

$$h_1 \sim \mathrm{N}\left(\mu, \frac{\sigma_\eta^2}{1-\phi^2}\right), \tag{3}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & q \\ q & \sigma_\eta^2 \end{pmatrix} \right), \tag{4}$$

$$|\phi| < 1, \tag{5}$$

where y_t denotes a (univariate) asset return, h_t is a log-variance of y_t. The negative value of q implies the existence of the leverage effect. It can be extended to the multivariate SV model with cross leverage effect (Daníelsson (1998), Asai and McAleer (2006), Asai and McAleer (2009), Chan, Kohn, and Kirby (2006), Ishihara, Omori, and Asai (2011), Ishihara and Omori (2012) and Nakajima (2012)). The major difficulty in constructing such multivariate models is to make the covariance matrices positive definite, especially when some dynamic correlation structure between the asset returns is incorporated. It is desirable to model the dynamic covariance structure as simply as possible since the parameter estimation becomes difficult in the sense that there are too many latent variables to be integrated out analytically to obtain the likelihood function.

In this article, we propose the multivariate SV model with dynamic equicorrelation and cross leverage effect (DESV model) and describe an efficient Bayesian estimation using the Markov chain Monte Carlo (MCMC) method to generate the latent stochastic volatilities and dynamic equicorrelation from the posterior distributions. As discussed in Shephard and Pitt (1997), Watanabe and Omori (2004) and Omori and Watanabe (2008), we divide all latent variables into several blocks and generate one block given other blocks (multi-move sampler or block-sampler). This is known to be more efficient than the simple single-move sampler, which draws the single latent stochastic volatility (the single dynamic equicorrelation factor) at a time given the other latent variables and the parameters. It means that we only need to generate a fewer number of MCMC samples to estimate the posterior distribution of the interested parameters.

The rest of this article is organized as follows. In Section 2, we propose the multivariate stochastic volatility model with dynamic equicorrelation and cross leverage effect. Section 3 describes an efficient Bayesian estimation method for the proposed model using a multi-move sampling method. A single-move sampling method that is simple but inefficient is also described as a benchmark. In Section 4, we illustrate our estimation method using simulated data and show that our MCMC algorithm is efficient. Section 5 applies our proposed DESV model to the trivariate asset return data based on industrial sector indices of TOPIX (Tokyo stock price index). Section 6 concludes this article.

2 Equicorrelation model

2.1 Equicorrelation matrix

Suppose that a p-dimensional random variable has an equicorrelation structure where the $p \times p$ equicorrelation matrix takes the form

$$R = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{pmatrix} \tag{6}$$

$$= (1-\rho)I_p + \rho J_p, \tag{7}$$

I_p is a unit matrix of size p, and $J_p = \mathbf{1}_p \mathbf{1}_p'$ ($\mathbf{1}_p$ denotes a p-dimensional vector with all elements equal to one). The matrix R is positive definite if and only if ρ satisfies the condition $-(p-1)^{-1} < \rho < 1$. It is noted that the lower bound of ρ is depending on p, that is, as p becomes larger the lower bound approaches to zero. The determinant and the inverse are given by

$$|R| = (1-\rho)^{p-1}\{1 + (p-1)\rho\}, \tag{8}$$

$$R^{-1} = \frac{1}{1-\rho}\left(I_p - \frac{\rho}{1-\rho+p\rho}J_p\right). \tag{9}$$

We note that the eigenvalues of R are $1-\rho$ (multiplicity $p-1$) and $1+(p-1)\rho$. The eigenvector $\boldsymbol{x} = (x_1, \dots, x_p)'$ associated with $1-\rho$ satisfies the condition $\sum_{i=1}^{p} x_i = 0$, while the eigenvector associated with $1+(p-1)\rho$ satisfies the condition $x_1 = \cdots = x_p$. Thus the spectral decomposition of R is given by

$$R = \{1+(p-1)\rho\}\boldsymbol{r}_1\boldsymbol{r}_1' + (1-\rho)\boldsymbol{r}_2\boldsymbol{r}_2' + \cdots + (1-\rho)\boldsymbol{r}_p\boldsymbol{r}_p',$$

where $\boldsymbol{r}_1, \dots, \boldsymbol{r}_p$ are the associated orthonormalized eigenvectors.

Let $R_D = \mathrm{diag}\{1+(p-1)\rho, 1-\rho, \dots, 1-\rho\}$ and $R_O = (\boldsymbol{r}_1, \dots, \boldsymbol{r}_p)$. Then, $R = R_O R_D R_O'$ and we denote $R^{1/2} = R_O R_D^{1/2}$ where $R_D^{1/2} = \mathrm{diag}\{(1+(p-1)\rho)^{1/2}, (1-\rho)^{1/2}, \dots, (1-\rho)^{1/2}\}$. Alternatively we could decompose as $R = R_c R_c'$ (e.g., Choleski decomposition), but the spectral decomposition is advantageous in that it does not depend on the order of the random variables in the vector.

2.2 Multivariate SV model with dynamic equicorrelation

Let $\boldsymbol{y}_t = (y_{1t}, \dots, y_{pt})'$ denote a p-dimensional asset return vector at time t $(t = 1, \dots, n)$. Let $\boldsymbol{m}_t = (m_{1t}, \dots, m_{pt})'$ and $\boldsymbol{h}_t = (h_{1t}, \dots, h_{pt})'$ denote p-dimensional vectors of unobserved variables and g_t an unobserved variable. We consider the multivariate SV model given by

$$\boldsymbol{y}_t = \boldsymbol{m}_t + V_t^{1/2}\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathrm{N}_p(\mathbf{0}_p, R_t), \quad t = 1, \dots, n, \tag{10}$$

$$\boldsymbol{h}_{t+1} = \boldsymbol{\mu} + \Phi(\boldsymbol{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathrm{N}_p(\mathbf{0}_p, \Omega), \quad t = 1, \dots, n-1, \tag{11}$$

$$\boldsymbol{h}_1 = \boldsymbol{\mu} + \boldsymbol{\eta}_0, \quad \boldsymbol{\eta}_0 \sim \mathrm{N}_p(\mathbf{0}_p, \Omega_0), \tag{12}$$

$$g_{t+1} = \gamma + \theta(g_t - \gamma) + \zeta_t, \quad \zeta_t \sim \mathrm{N}(0, \sigma^2), \tag{13}$$

$$g_1 = \gamma + \zeta_0, \quad \zeta_0 \sim \mathrm{N}(0, \sigma^2/(1-\theta^2)), \tag{14}$$

$$\boldsymbol{m}_{t+1} = \boldsymbol{m}_t + \boldsymbol{\eta}_{\boldsymbol{m}t}, \quad \boldsymbol{\eta}_{\boldsymbol{m}t} \sim \mathrm{N}_p(\mathbf{0}_p, \Omega_{\boldsymbol{m}}), \quad t = 1, \dots, n-1, \tag{15}$$

$$\boldsymbol{m}_1 = \boldsymbol{\eta}_{\boldsymbol{m}0}, \quad \boldsymbol{\eta}_{\boldsymbol{m}0} \sim \mathrm{N}_p(\mathbf{0}_p, \kappa I_p), \tag{16}$$

where $\mathbf{0}_p$ is a p-dimensional zero vector,

$$V_t = \mathrm{diag}\{\exp(h_{1t}), \dots, \exp(h_{pt})\}, \; t = 1, \dots, n, \tag{17}$$

$$R_t = (1 - \rho_t)I_p + \rho_t J_p, \; t = 1, \dots, n, \tag{18}$$

$$\rho_t = \frac{\exp(g_t)}{\exp(g_t) + 1}, \; t = 1, \dots, n, \tag{19}$$

$$\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)', \tag{20}$$

$$\Phi = \mathrm{diag}(\phi_1, \dots, \phi_p), \tag{21}$$

$$\Omega_{\boldsymbol{m}} = \mathrm{diag}(\omega_{m_1}^2, \dots, \omega_{m_p}^2), \tag{22}$$

$$\begin{pmatrix} R_t^{-\frac{1}{2}}\boldsymbol{\epsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} \sim \mathrm{N}_{2p}(\mathbf{0}_{2p}, \Psi), \; t = 1, \dots, n-1, \tag{23}$$

$$\Psi = \begin{pmatrix} I_p & Q' \\ Q & \Omega \end{pmatrix}, \tag{24}$$

and Ω_0, the covariance matrix of the initial latent variable $\boldsymbol{h}_1$, satisfies the stationary condition $\Omega_0 = \Phi\Omega_0\Phi + \Omega$ such that

$$\mathrm{vec}(\Omega_0) = (I_{p^2} - \Phi \otimes \Phi)^{-1}\mathrm{vec}(\Omega). \tag{25}$$

We also set $\mathrm{Var}(g_1) = \sigma^2/(1-\theta^2)$, the variance of the initial latent variable g_1, for assuming the stationary condition and set κ, the variance of the initial latent variable $m_{j,1}$ $(j =$

$1,\ldots,p$), equal to some large known constant. The latent vector, $\boldsymbol{h}_t$, is a vector of log-variances of the returns, and the latent variable, g_t, is the transformed equicorrelation of $\boldsymbol{y}_t$. For the identifiability, we set the diagonal elements of the covariance matrix of $\boldsymbol{\epsilon}_t$ equal to 1.

Notice that we define ρ_t, $t = 1,\ldots n$, so as to take values on the unit interval, $(0,1)$. As shown in the previous subsection, the equicorrelation matrix R_t is positive definite if and only if ρ_t is in $(-(p-1)^{-1}, 1)$. It means that as p becomes large, the negative region of the parameter space becomes smaller. Therefore it is reasonable to restrict the parameter space of ρ_t to the positive region so that the parameter space is independent of p.

We assume, for simplicity, that $\{\boldsymbol{m}_t\}_{t=1}^n$, the latent sequence of the expectations of $\boldsymbol{y}_t$, $t = 1,\ldots,n$, follows a simple random walk. Empirical studies suggest that the expectation of the asset return is nearly 0 in most cases (e.g., McNeil, Frey, and Embrechts (2005)), but it is practically important to include non-zero asset means especially for portfolio optimizations as we shall see in Section 5.3.

3 Bayesian estimation

3.1 Priors and posterior densities

For prior distributions of $\{\boldsymbol{\mu}, \gamma, \Phi, \theta, \sigma^2, \Omega_{\boldsymbol{m}}\}$, we assume that

$$\boldsymbol{\mu} \sim \mathrm{N}_p(\boldsymbol{m}_{\boldsymbol{\mu}0}, S_{\boldsymbol{\mu}0}), \tag{26}$$

$$\gamma \sim \mathrm{N}(m_{\gamma 0}, s_{\gamma 0}^2), \tag{27}$$

$$(\phi_j + 1)/2 \sim \mathrm{Be}(a_{\phi_j}, b_{\phi_j}), \quad j = 1,\ldots,p, \tag{28}$$

$$(\theta + 1)/2 \sim \mathrm{Be}(a_\theta, b_\theta), \tag{29}$$

$$\sigma^2 \sim \mathrm{IG}(\alpha_{\sigma^2 0}/2, \beta_{\sigma^2 0}/2), \tag{30}$$

$$\omega_{m_j}^2 \sim \mathrm{IG}(\alpha_{m_j 0}/2, \beta_{m_j 0}/2), \quad j = 1,\ldots,p, \tag{31}$$

where $\mathrm{Be}(a,b)$ denotes a beta distribution with parameters a, b and $\mathrm{IG}(\alpha,\beta)$ denotes an inverted gamma distribution with shape parameter α and scale parameter β. For a prior distribution of Ψ, we let

$$\Psi^{-1} = \begin{pmatrix} \Psi^{11} & \Psi^{12} \\ \Psi^{21} & \Psi^{22} \end{pmatrix},$$

where Ψ^{11}, Ψ^{12} and Ψ^{22} are $p \times p$ matrices, respectively. Noting that $\Psi^{11} = I_p + \Psi^{12}(\Psi^{22})^{-1}\Psi^{21}$, we assume

$$\Psi^{22} \sim \mathrm{W}_p(n_0, S_0), \tag{32}$$

$$\Psi^{21}|\Psi^{22} \sim \mathrm{N}_{p\times p}(\Psi^{22}\Delta_0, \Lambda_0 \otimes \Psi^{22}), \tag{33}$$

where $\mathrm{W}(n, S)$ denotes a Wishart distribution with parameters (n, S).

Then, the joint posterior density function is

$$\begin{aligned}
&f(\boldsymbol{\vartheta}, \{\boldsymbol{h}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n | \{\boldsymbol{y}_t\}_{t=1}^n) \\
&\propto \pi(\boldsymbol{\vartheta}) \times \exp\left(\sum_{t=1}^n l_t\right) \times |\Omega_0|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{h}_1 - \boldsymbol{\mu})'\Omega_0^{-1}(\boldsymbol{h}_1 - \boldsymbol{\mu})\right\} \\
&\quad \times |\Omega|^{-\frac{n-1}{2}} \exp\left[-\frac{1}{2}\sum_{t=1}^{n-1}\{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_t - \boldsymbol{\mu})\}'\Omega^{-1}\{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_t - \boldsymbol{\mu})\}\right] \\
&\quad \times \left(\frac{\sigma^2}{1-\theta^2}\right)^{-1/2} \exp\left\{-\frac{(g_1 - \gamma)^2}{2\sigma^2/(1-\theta^2)}\right\} \times (\sigma^2)^{-\frac{n-1}{2}} \exp\left[-\frac{\sum_{t=1}^{n-1}\{g_{t+1} - (1-\theta)\gamma - \theta g_t\}^2}{2\sigma^2}\right] \\
&\quad \times \exp\left(-\frac{1}{2\kappa}\boldsymbol{m}_1'\boldsymbol{m}_1\right) \times |\Omega_{\boldsymbol{m}}|^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{n-1}(\boldsymbol{m}_{t+1} - \boldsymbol{m}_t)'\Omega_{\boldsymbol{m}}^{-1}(\boldsymbol{m}_{t+1} - \boldsymbol{m}_t)\right\},
\end{aligned} \tag{34}$$

where

$$\begin{aligned}
l_t = &-\frac{1}{2}\left[R_t^{-\frac{1}{2}}V_t^{-\frac{1}{2}}(\boldsymbol{y}_t - \boldsymbol{m}_t) - Q'\Omega^{-1}\{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_t - \boldsymbol{\mu})\}\right]' \\
&\times (I_p - Q'\Omega^{-1}Q)^{-1}\left[R_t^{-\frac{1}{2}}V_t^{-\frac{1}{2}}(\boldsymbol{y}_t - \boldsymbol{m}_t) - Q'\Omega^{-1}\{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_t - \boldsymbol{\mu})\}\right] \\
&-\frac{1}{2}\log\{(1-\rho_t)^{p-1}(1+(p-1)\rho_t)\} - \frac{1}{2}\textstyle\sum_{j=1}^p h_{jt},
\end{aligned} \tag{35}$$

and $\boldsymbol{\vartheta} = \{\boldsymbol{\mu}, \gamma, \Phi, \theta, \Omega, Q, \sigma^2, \Omega_{\boldsymbol{m}}\}$.

We implement the MCMC algorithm in twelve blocks:

1. Initialize $\{\boldsymbol{h}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\mu}, \gamma, \Phi, \theta, \Omega, Q, \sigma^2, \Omega_{\boldsymbol{m}}$.

2. Generate $\boldsymbol{\mu}|\{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{h}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \Phi, \Omega, Q$.

3. Generate $\gamma|\{\boldsymbol{y}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \theta, \sigma^2$.

4. Generate $\Phi|\{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{h}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\mu}, \Omega, Q$.

5. Generate $\theta|\{\boldsymbol{y}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \gamma, \sigma^2$.

6. Generate $\Omega, Q|\{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{h}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\mu}, \Phi$.

7. Generate $\sigma^2|\{\boldsymbol{y}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \gamma, \theta$.

8. Generate $\Omega_{\boldsymbol{m}}|\{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n$.

9. Generate $\{\boldsymbol{h}_t\}_{t=1}^n|\{\boldsymbol{y}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\mu}, \Phi, \Omega, Q$.

10. Generate $\{g_t\}_{t=1}^n|\{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{h}_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\mu}, \gamma, \Phi, \theta, \Omega, Q, \sigma^2$.

11. Generate $\{\boldsymbol{m}_t\}_{t=1}^n|\{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{h}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \boldsymbol{\mu}, \Phi, \Omega, Q, \Omega_{\boldsymbol{m}}$.

12. Go to 2.

3.2 Generation of latent variable $\{\boldsymbol{h}_t\}_{t=1}^n$

3.2.1 Single-move sampling method

A simple sampling method for $\{\boldsymbol{h}_t\}_{t=1}^n$, is a single-move sampler that draws a single latent variable $\boldsymbol{h}_t$ at a time given the other $\boldsymbol{h}_t$'s and the parameters. The other method is a multi-move sampler that draws multiple $\boldsymbol{h}_t$'s simultaneously. A single-move sampler is simpler than a multi-move sampler, but a multi-move sampler is known to be more efficient (Shephard and Pitt (1997), Watanabe and Omori (2004) and Omori and Watanabe (2008)). As a benchmark, we first describe the single-move sampling method. The conditional posterior density of $\boldsymbol{h}_t$ is

$$f(\boldsymbol{h}_t|\{\boldsymbol{h}_s\}_{-t}, \{\boldsymbol{y}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\vartheta}) \propto g(\boldsymbol{h}_t) \times \exp\left\{-\frac{1}{2}(\boldsymbol{h}_t - \boldsymbol{m}_{\boldsymbol{h}t})' S_{\boldsymbol{h}t}^{-1}(\boldsymbol{h}_t - \boldsymbol{m}_{\boldsymbol{h}t})\right\}, \tag{36}$$

where

$$\begin{aligned}
S_{\boldsymbol{h}t} &= \{\Phi\Omega^{-1}\Phi + (\Omega - QQ')^{-1}\}^{-1}, \\
\boldsymbol{m}_{\boldsymbol{h}t} &= S_{\boldsymbol{h}t}\left[-\frac{1}{2}\mathbf{1}_p + \Phi\Omega^{-1}\{\boldsymbol{h}_{t+1} - (I_p - \Phi)\boldsymbol{\mu}\} + (\Omega - QQ')^{-1}\{\boldsymbol{\mu} + \Phi(\boldsymbol{h}_{t-1} - \boldsymbol{\mu}) + Q\boldsymbol{z}_{t-1}\}\right], \\
g(\boldsymbol{h}_t) &= \exp\left[-\frac{1}{2}\left[\boldsymbol{z}_t - Q'\Omega^{-1}\{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_t - \boldsymbol{\mu})\}\right]'(I_p - Q'\Omega^{-1}Q)^{-1}\right. \\
&\qquad \left.\times \left[\boldsymbol{z}_t - Q'\Omega^{-1}\{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_t - \boldsymbol{\mu})\}\right]\right], \\
\boldsymbol{z}_t &= R_t^{-1/2} V_t^{-1/2}(\boldsymbol{y}_t - \boldsymbol{m}_t).
\end{aligned}$$

We propose a candidate $\boldsymbol{h}_t^\dagger|\{\boldsymbol{h}_s\}_{-t}, \{\boldsymbol{y}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\vartheta} \sim \mathrm{N}(\boldsymbol{m}_{\boldsymbol{h}t}, S_{\boldsymbol{h}t})$ and accept it with probability $\min[1, g(\boldsymbol{h}_t^\dagger)/g(\boldsymbol{h}_t)]$ using Metropolis-Hastings (MH) algorithm, where $\boldsymbol{h}_t$ is a current value.

3.2.2 Efficient multi-move sampling method

A simulation smoother, an efficient sampler for the state variables was proposed by de Jong and Shephard (1995) and by Durbin and Koopman (2002) for the linear Gaussian state space model. However, such a simulation smoother cannot be applied directly to our nonlinear model. As discussed in Shephard and Pitt (1997), Watanabe and Omori (2004) and Omori and Watanabe (2008), we approximate the nonlinear Gaussian likelihood function by the linear Gaussian likelihood function and implement the MH algorithm.

In this algorithm, we first divide $\{\boldsymbol{h}_t\}_{t=1}^n$ into $K+1$ blocks, $(\boldsymbol{h}_{k_{m-1}+1}, \dots, \boldsymbol{h}_{k_m})$, $m = 1, \dots, K$ with $k_0 = 0$, $k_{K+1} = n$, $k_i - k_{i-1} \geq 2$, using stochastic knots $k_m = \mathrm{int}[n(m + U_m)/(K+2)]$, where U_m's are independent uniform random variables on $(0,1)$. Next, we generate $(\boldsymbol{h}_{k_{m-1}+1}, \dots, \boldsymbol{h}_{k_m})$ given other blocks by generating $(\underline{\boldsymbol{\eta}}_{k_{m-1}}, \dots, \underline{\boldsymbol{\eta}}_{k_m - 1})$, where $\underline{\boldsymbol{\eta}}_t = \Omega^{-1/2}\boldsymbol{\eta}_t$.

The conditional posterior density of $\underline{\boldsymbol{\eta}} = (\underline{\boldsymbol{\eta}}_s', \dots, \underline{\boldsymbol{\eta}}_{s+r-1}')'$ is given by

$$
\begin{aligned}
& f(\underline{\boldsymbol{\eta}}_s, \dots, \underline{\boldsymbol{\eta}}_{s+r-1}|\{\boldsymbol{y}_t\}_{t=1}^n, \boldsymbol{h}_s, \boldsymbol{h}_{s+r+1}, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\vartheta}) \\
& \propto \prod_{t=s}^{s+r} f(\boldsymbol{y}_t|\boldsymbol{h}_t, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\vartheta}) \prod_{t=s}^{s+r-1} f(\boldsymbol{\eta}_t|\boldsymbol{\vartheta}) \times (f(\boldsymbol{h}_{s+r+1}|\boldsymbol{h}_{s+r}, \boldsymbol{\vartheta}))^{\mathrm{I}\{s+r<n\}} \\
& \propto \exp\left(L - \frac{1}{2}\sum_{t=s}^{s+r-1} \underline{\boldsymbol{\eta}}_t'\underline{\boldsymbol{\eta}}_t\right),
\end{aligned}
$$

where

$$
L = \begin{cases}
\displaystyle\sum_{t=s}^{s+r} l_t - \frac{1}{2}\{\boldsymbol{h}_{s+r+1} - (I_p - \Phi)\boldsymbol{\mu} - \Phi\boldsymbol{h}_{s+r}\}'\Omega^{-1}\{\boldsymbol{h}_{s+r+1} - (I_p - \Phi)\boldsymbol{\mu} - \Phi\boldsymbol{h}_{s+r}\} \\
\quad \text{if } s+r<n, \\
\displaystyle\sum_{t=s}^{n} l_t \quad \text{if } s+r=n.
\end{cases}
$$

Using Taylor expansion around the conditional posterior mode of $\underline{\boldsymbol{\eta}} = (\underline{\boldsymbol{\eta}}_s', \dots, \underline{\boldsymbol{\eta}}_{s+r-1}')'$, we approximate L and construct a proposal density (linear and Gaussian state-space model) as

$$
\begin{aligned}
&\log f(\underline{\boldsymbol{\eta}}_s, \ldots, \underline{\boldsymbol{\eta}}_{s+r-1} | \boldsymbol{h}_s, \boldsymbol{h}_{s+r+1}, \boldsymbol{y}_s, \ldots, \boldsymbol{y}_{s+r}) \\
&\approx \text{const.} - \frac{1}{2}\sum_{t=s}^{s+r-1} \underline{\boldsymbol{\eta}}_t' \underline{\boldsymbol{\eta}}_t + \hat{L} + \left.\frac{\partial L}{\partial \underline{\boldsymbol{\eta}}'}\right|_{\underline{\boldsymbol{\eta}}=\hat{\underline{\boldsymbol{\eta}}}} (\underline{\boldsymbol{\eta}} - \hat{\underline{\boldsymbol{\eta}}}) + \frac{1}{2}(\underline{\boldsymbol{\eta}} - \hat{\underline{\boldsymbol{\eta}}})' \mathrm{E}\left.\left(\frac{\partial^2 L}{\partial \underline{\boldsymbol{\eta}} \partial \underline{\boldsymbol{\eta}}'}\right)\right|_{\underline{\boldsymbol{\eta}}=\hat{\underline{\boldsymbol{\eta}}}} (\underline{\boldsymbol{\eta}} - \hat{\underline{\boldsymbol{\eta}}}) \\
&= \text{const.} - \frac{1}{2}\sum_{t=s}^{s+r-1} \underline{\boldsymbol{\eta}}_t' \underline{\boldsymbol{\eta}}_t + \hat{L} + \hat{\boldsymbol{d}}'(\boldsymbol{h} - \hat{\boldsymbol{h}}) + \frac{1}{2}(\boldsymbol{h} - \hat{\boldsymbol{h}})' \mathrm{E}\left.\left(\frac{\partial^2 L}{\partial \boldsymbol{h} \partial \boldsymbol{h}'}\right)\right|_{\boldsymbol{h}=\hat{\boldsymbol{h}}} (\boldsymbol{h} - \hat{\boldsymbol{h}}) \\
&= \log f^*(\underline{\boldsymbol{\eta}}_s, \ldots, \underline{\boldsymbol{\eta}}_{s+r-1} | \boldsymbol{h}_s, \boldsymbol{h}_{s+r+1}, \boldsymbol{y}_s, \ldots, \boldsymbol{y}_{s+r}),
\end{aligned}
$$

where $\boldsymbol{h} = (\boldsymbol{h}_{s+1}', \ldots, \boldsymbol{h}_{s+r}')'$ and

$$
\boldsymbol{d} = (\boldsymbol{d}_{s+1}', \ldots, \boldsymbol{d}_{s+r}')', \;\; \boldsymbol{d}_t = \partial L / \partial \boldsymbol{h}_t, \tag{37}
$$

$$
-\mathrm{E}\left(\frac{\partial^2 L}{\partial \boldsymbol{h} \partial \boldsymbol{h}'}\right) = \begin{pmatrix} A_{s+1} & B_{s+2}' & O & \cdots & O \\ B_{s+2} & A_{s+2} & B_{s+3}' & \cdots & O \\ O & B_{s+3} & A_{s+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & B_{s+r}' \\ O & \cdots & O & B_{s+r} & A_{s+r} \end{pmatrix}, \tag{38}
$$

$$
A_t = -\mathrm{E}\left(\frac{\partial^2 L}{\partial \boldsymbol{h}_t \partial \boldsymbol{h}_t'}\right), \; t = s+1, \ldots, s+r, \tag{39}
$$

$$
B_t = -\mathrm{E}\left(\frac{\partial^2 L}{\partial \boldsymbol{h}_t \partial \boldsymbol{h}_{t-1}'}\right), \; t = s+2, \ldots, s+r, \; B_{s+1} = O \tag{40}
$$

(see Appendix A.1 for the derivation of $\boldsymbol{d}_t, A_t, B_t$). We generate $\{\underline{\boldsymbol{\eta}}_t\}_{t=s}^{s+r-1}$ in two steps:

Step 1 (Disturbance smoother).

(a) Initialize $\hat{\underline{\boldsymbol{\eta}}}_t, \; t = s, \ldots, s+r-1$.

(b) Compute $\hat{\boldsymbol{d}}_t, \hat{A}_t, \hat{B}_t, \; t = s+1, \ldots, s+r$. ($\boldsymbol{d}_t, A_t, B_t$ evaluated at $\hat{\underline{\boldsymbol{\eta}}}_t$)

(c) For $t = s+2, \ldots, s+r$, compute

$$
\begin{aligned}
&C_t = \hat{A}_t - \hat{B}_t C_{t-1}^{-1} \hat{B}'t, \; C_{s+1} = \hat{A}_{s+1}, \; C_t = F_t F_t', \\
&M_t = \hat{B}_t F'^{-1}_{t-1}, \; M_{s+1} = O, \; M_{s+r+1} = O, \\
&\boldsymbol{b}_t = \hat{\boldsymbol{d}}_t - M_t F_{t-1}^{-1} \boldsymbol{b}_{t-1}, \; \boldsymbol{b}_{s+1} = \hat{\boldsymbol{d}}_{s+1}.
\end{aligned}
$$

(d) For $t = s+1, \ldots, s+r$, define

$\hat{\boldsymbol{y}}_t = \hat{\gamma}_t + C_t^{-1} \boldsymbol{b}_t, \;\; \hat{\gamma}_t = \hat{\boldsymbol{h}}_t + F_t'^{-1} M_{t+1}' \hat{\boldsymbol{h}}_{t+1}.$

(e) Consider the linear and Gaussian state-space model:

$$\hat{\boldsymbol{y}}_t = Z_t \boldsymbol{h}_t + G_t \boldsymbol{u}_t, \tag{41}$$

$$\boldsymbol{h}_{t+1} = (I_p - \Phi)\boldsymbol{\mu} + \Phi \boldsymbol{h}_t + H_t \boldsymbol{u}_t, \tag{42}$$

$$Z_t = I_p + F_t'^{-1} M_{t+1}' \Phi, \; G_t = F_t'^{-1}[I_p, M_{t+1}' \mathrm{chol}(\Omega)], \tag{43}$$

$$H_t = [O, \mathrm{chol}(\Omega)]. \tag{44}$$

(chol(Ω) denotes Choleski decomposition of Ω.)

(f) Apply Kalman filter and disturbance smoother (Koopman (1993)) and update $\underline{\hat{\boldsymbol{\eta}}}$.

(g) Go to (b) until $\underline{\hat{\boldsymbol{\eta}}}$ converges to the mode.

Step 2. (a) Update $\underline{\hat{\boldsymbol{\eta}}}$ using the disturbance smoother and find a linear and Gaussian state-space model (41)-(44).

(b) Generate $\underline{\boldsymbol{\eta}}^\dagger \sim f^*$ (the linear and Gaussian state-space model) using Kalman filter and simulation smoother (de Jong and Shephard (1995) or Durbin and Koopman (2002)). Conduct Accept-Reject MH (AR-MH) algorithm (Chib and Greenberg (1995)).

3.3 Generation of latent variables $\{g_t\}_{t=1}^n$ and $\{\boldsymbol{m}_t\}_{t=1}^n$

We consider the two following sampling methods for $\{g_t\}_{t=1}^n$ corresponding to the last subsection.

Single-move sampling method. Generate g_t given $\{g_s\}_{-t}, \{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{h}_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\vartheta}$ using a random walk MH algorithm for $t = 1, \ldots, n$.

Multi-move sampling method. We divide $\{g_t\}_{t=1}^n$ into $K+1$ blocks using stochastic knots and generate one block given other blocks using the block sampler as mentioned in the last subsection. See Appendix A.1 for the derivation of $\boldsymbol{d}_t, A_t$ and B_t.

We generate all of $\boldsymbol{m}_t$'s simultaneously using a simulation smoother (de Jong and Shephard (1995) or Durbin and Koopman (2002)).

3.4 Generation of $\boldsymbol{\mu}, \gamma, \Phi, \theta, \sigma^2, \Omega_{\boldsymbol{m}}$

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

$$\boldsymbol{\mu}|\{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{h}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \Phi, \Omega, Q \sim \mathrm{N}(\boldsymbol{m}_{\boldsymbol{\mu}}, S_{\boldsymbol{\mu}}), \tag{45}$$

where

$$S_{\boldsymbol{\mu}} = \{S_{\boldsymbol{\mu}0}^{-1} + \Omega_0^{-1} + (n-1)(I_p - \Phi)(\Omega - QQ')^{-1}(I_p - \Phi)\}^{-1}, \tag{46}$$

$$\boldsymbol{m}_{\boldsymbol{\mu}} = S_{\boldsymbol{\mu}}\left\{S_{\boldsymbol{\mu}0}^{-1}\boldsymbol{m}_{\boldsymbol{\mu}0} + \Omega_0^{-1}\boldsymbol{h}_1 + (I_p - \Phi)(\Omega - QQ')^{-1}\sum_{t=1}^{n-1}(\boldsymbol{h}_{t+1} - \Phi\boldsymbol{h}_t - Q\boldsymbol{z}_t)\right\}.$$

Generation of γ*.* The conditional distribution of γ is

$$\gamma|\{\boldsymbol{y}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \theta, \sigma^2 \sim \mathrm{N}(m_\gamma, s_\gamma^2), \tag{47}$$

where

$$s_\gamma^2 = \{s_{\gamma 0}^{-2} + (1-\theta^2)\sigma^{-2} + (n-1)(1-\theta)^2\sigma^{-2}\}^{-1}, \tag{48}$$

$$m_\gamma = s_\gamma^2\left[s_{\gamma 0}^{-2}m_{\gamma 0} + (1-\theta^2)\sigma^{-2}g_1 + (1-\theta)\sigma^{-2}\left\{\sum_{t=2}^{n} g_t - \theta\sum_{t=1}^{n-1} g_t\right\}\right]. \tag{49}$$

Generation of Φ*.* The conditional posterior density of $\boldsymbol{\phi} = \Phi\boldsymbol{1}_p$ is given by

$$f(\boldsymbol{\phi}|\{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{h}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\mu}, \Omega, Q) \propto g(\boldsymbol{\phi}) \times \exp\left\{-\frac{1}{2}(\boldsymbol{\phi} - \boldsymbol{m}_{\boldsymbol{\phi}})' S_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\phi} - \boldsymbol{m}_{\boldsymbol{\phi}})\right\}, \tag{50}$$

where

$$S_{\boldsymbol{\phi}} = \left\{\sum_{t=1}^{n-1}((\boldsymbol{h}_t - \boldsymbol{\mu})(\boldsymbol{h}_t - \boldsymbol{\mu})') \odot (\Omega - QQ')^{-1}\right\}^{-1}, \tag{51}$$

$$(\boldsymbol{m}_{t,\boldsymbol{\phi}})_i = \left\{(\Omega - QQ')^{-1}(\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - Q\boldsymbol{z}_t)(\boldsymbol{h}_t - \boldsymbol{\mu})'\right\}_{i,i}, \quad \boldsymbol{m}_{\boldsymbol{\phi}} = S_{\boldsymbol{\phi}}\sum_{t=1}^{n-1}\boldsymbol{m}_{t,\boldsymbol{\phi}}, \tag{52}$$

$$g(\boldsymbol{\phi}) = \prod_{j=1}^{p}\pi(\phi_j) \times |\Omega_0|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{h}_1 - \boldsymbol{\mu})'\Omega_0^{-1}(\boldsymbol{h}_1 - \boldsymbol{\mu})\right\}, \tag{53}$$

and $\odot$ denotes Hadamard product. Generate a candidate from a truncated normal distribution over the region $\{\boldsymbol{\phi}; |\phi_i| < 1,\ i - 1, \ldots, p\}$, $\boldsymbol{\phi}^\dagger \sim \mathrm{TN}_{(-1,1)}(\boldsymbol{m}_{\boldsymbol{\phi}}, S_{\boldsymbol{\phi}})$, and accept it with probability $\min[1, g(\boldsymbol{\phi}^\dagger)/g(\boldsymbol{\phi})]$.

Generation of θ. The conditional posterior density of θ is given by

$$f(\theta|\{\boldsymbol{y}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \gamma, \sigma^2) \propto g(\theta) \times \exp\left\{ -\frac{1}{2s_\theta^2}(\theta - m_\theta)^2 \right\}, \tag{54}$$

where

$$s_\theta^2 = \sigma^2 \left\{ \sum_{t=1}^{n-1} (g_t - \gamma)^2 \right\}^{-1}, \quad m_\theta = s_\theta^2 \sum_{t=1}^{n-1} \sigma^{-2}(g_{t+1} - \gamma)(g_t - \gamma), \tag{55}$$

$$g(\theta) = \pi(\theta) \times (1-\theta^2)^{1/2} \exp\left\{ -\frac{(g_1 - \gamma)^2}{2\sigma^2/(1-\theta^2)} \right\}. \tag{56}$$

Generate a candidate $\theta^\dagger \sim \mathrm{TN}_{(-1,1)}(m_\theta, s_\theta^2)$ and accept it with probability $\min[1, g(\theta^\dagger)/g(\theta)]$.

Generation of σ^2. The conditional distribution of σ^2 is

$$\sigma^2|\{\boldsymbol{y}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \gamma, \theta \sim \mathrm{IG}(\alpha_{\sigma^2 1}/2, \beta_{\sigma^2 1}/2), \tag{57}$$

where $\alpha_{\sigma^2 1} = \alpha_{\sigma^2 0} + n$ and

$$\beta_{\sigma^2 1} = \beta_{\sigma^2 0} + (g_1 - \gamma)^2(1-\theta^2) + \sum_{t=1}^{n-1} \{g_{t+1} - \gamma - \theta(g_t - \gamma)\}^2. \tag{58}$$

Generation of $\Omega_{\boldsymbol{m}}$. The conditional distribution of $\omega_{m_j}^2$, $j = 1, \ldots, p$, is

$$\omega_{m_j}^2|\{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n \sim \mathrm{IG}(\alpha_{m_j 1}/2, \beta_{m_j 1}/2), \tag{59}$$

where $\alpha_{m_j 1} = \alpha_{m_j 0} + n - 1$ and

$$\beta_{m_j 1} = \beta_{m_j 0} + \sum_{t=1}^{n-1} (m_{j,t+1} - m_{jt})^2. \tag{60}$$

3.5 Generation of Ω, Q

The conditional posterior density of Ψ^{12} and Ψ^{22} is given by

$$\begin{aligned}
& f(\Psi^{12}, \Psi^{22}|\{\boldsymbol{y}_t\}_{t=1}^n, \{\boldsymbol{h}_t\}_{t=1}^n, \{g_t\}_{t=1}^n, \{\boldsymbol{m}_t\}_{t=1}^n, \boldsymbol{\mu}, \Phi) \\
& \propto \prod_{t=1}^{n-1} f(\boldsymbol{z}_t, \boldsymbol{\eta}_t|\boldsymbol{\vartheta}) \times \pi(\Psi^{21}, \Psi^{22}) \\
& \propto |\Omega_0|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\eta}_0'\Omega_0^{-1}\boldsymbol{\eta}_0 \right) \times |\Psi^{22}|^{(n_1-p-1)/2} \exp\left\{ -\frac{1}{2}\mathrm{tr}(S_1^{-1}\Psi^{22}) \right\} \\
& \quad \times |\Psi^{22}|^{-p/2} \exp\left[-\frac{1}{2}\{\mathrm{vec}(\Psi^{21} - \Psi^{22}\Delta_1)\}'(\Lambda_1 \otimes \Psi^{22})^{-1}\mathrm{vec}(\Psi_{21} - \Psi_{22}\Delta_1) \right]
\end{aligned} \tag{61}$$

(see, e.g., Gupta and Nagar (2000) and Ishihara, Omori, and Asai (2011)), where

$$n_1 = n_0 + n - 1,\ S_1 = (S_0^{-1} + \Xi_{22} + \Delta_0\Lambda_0^{-1}\Delta_0' - \Delta_1\Lambda_1^{-1}\Delta_1')^{-1}, \tag{62}$$

$$\Lambda_1 = (\Lambda_0^{-1} + \Xi_{11})^{-1},\ \Delta_1 = (-\Xi_{21} + \Delta_0\Lambda_0^{-1})\Lambda_1, \tag{63}$$

$$\Xi = \begin{pmatrix} \Xi_{11} & \Xi_{12} \\ \Xi_{21} & \Xi_{22} \end{pmatrix} = \sum_{t=1}^{n-1} \begin{pmatrix} \boldsymbol{z}_t \\ \boldsymbol{\eta}_t \end{pmatrix} \begin{pmatrix} \boldsymbol{z}_t \\ \boldsymbol{\eta}_t \end{pmatrix}'. \tag{64}$$

We generate a candidate $\Psi^\dagger$ in three steps:

1. Generate $(\Psi^{22})^\dagger \sim \mathrm{W}(n_1, S_1)$.
2. Generate $(\Psi^{21})^\dagger|(\Psi^{22})^\dagger \sim \mathrm{N}_{p\times p}((\Psi^{22})^\dagger\Delta_1, \Lambda_1 \otimes (\Psi^{22})^\dagger)$.
3. Compute $Q^\dagger = -((\Psi^{22})^\dagger)^{-1}(\Psi^{21})^\dagger$, $\Omega^\dagger = ((\Psi^{22})^\dagger)^{-1} + Q^\dagger(Q^\dagger)'$ and accept $\Psi^\dagger$ with probability

$$\min\left[1, \exp\left\{ -\frac{1}{2}\log|\Omega_0^\dagger| - \frac{1}{2}\boldsymbol{\eta}_0'(\Omega_0^\dagger)^{-1}\boldsymbol{\eta}_0 + \frac{1}{2}\log|\Omega_0| + \frac{1}{2}\boldsymbol{\eta}_0'\Omega_0^{-1}\boldsymbol{\eta}_0 \right\}\right].$$

4 Illustrative example using simulated data

This section illustrates our proposed DESV model using simulated data. We consider a trivariate case ($p = 3$) and investigate the efficiency of our multi-move sampling method in comparison with the single-move sampling method. Using the following parameters based on our empirical studies in Section 5,

$$\begin{aligned} &\boldsymbol{\mu}_* = \mathbf{0}_3,\ \gamma_* = 1.7,\ \Phi_* = 0.97I_3,\ \theta_* = 0.97, \Omega_* = 0.015I_3 + 0.015J_3, \\ &Q_* = (-0.1 \times \mathbf{1}_3, \mathbf{0}_3, \mathbf{0}_3),\ \sigma_*^2 = 0.05,\ \Omega_{\boldsymbol{m}*} = 0.001I_3, \end{aligned}$$

we generate 2,000 observations ($n = 2000$). For prior distributions, we assume

$$\begin{aligned} &\boldsymbol{\mu} \sim \mathrm{N}(\boldsymbol{\mu}_*, 100I_3),\ \gamma \sim \mathrm{N}(\gamma_*, 100),\ \frac{\phi_i + 1}{2} \sim \mathrm{Be}(20, 1.5),\ i = 1, 2, 3, \\ &\frac{\theta + 1}{2} \sim \mathrm{Be}(20, 1.5),\ \sigma^2 \sim \mathrm{IG}\left(\frac{5}{2}, \frac{3\sigma_*^2}{2}\right),\ \omega_{m_j}^2 \sim \mathrm{IG}\left(\frac{5}{2}, \frac{3\omega_{m_j*}^2}{2}\right),\ j = 1, 2, 3, \\ &\Psi^{22} \sim \mathrm{W}(6, 6^{-1}\Omega_*),\ \Psi^{21}|\Psi^{22} \sim \mathrm{N}_{3\times 3}(O, 10I_3 \otimes \Psi^{22}). \end{aligned}$$

We set $\kappa = 10$. Using the single-move sampler, we generate 200,000 MCMC samples after discarding the first 10,000 samples as the burn-in period. Also, the multi-move sampler is

Table 1: Posterior means, 95% credible intervals, p-values of convergence diagnostic test and inefficiency factors.

	True	Mean	95% interval	CD	IF (m-move)	IF (s-move)
μ_1	0	-0.062	(-0.433, 0.322)	0.065	3.9	31.2
μ_2	0	-0.055	(-0.372, 0.277)	0.393	5.9	56.2
μ_3	0	0.094	(-0.134, 0.331)	0.460	11.2	137.2
γ	1.7	1.597	(1.319, 1.873)	0.226	8.4	173.5
ϕ_1	0.97	0.979	(0.966, 0.989)	0.280	46.7	145.0
ϕ_2	0.97	0.976	(0.962, 0.988)	0.212	50.7	140.8
ϕ_3	0.97	0.966	(0.948, 0.981)	0.616	98.8	467.5
θ	0.97	0.959	(0.935, 0.978)	0.130	73.6	787.7
Ω_{11}	0.03	0.028	(0.018, 0.042)	0.264	136.2	627.2
Ω_{21}	0.015	0.011	(0.004, 0.019)	0.509	128.1	843.8
Ω_{31}	0.015	0.015	(0.006, 0.024)	0.320	147.2	1006.0
Ω_{22}	0.03	0.025	(0.016, 0.037)	0.560	137.3	412.8
Ω_{32}	0.015	0.011	(0.004, 0.019)	0.980	135.9	496.4
Ω_{33}	0.03	0.027	(0.016, 0.041)	0.431	**172.4**	771.2
Q_{11}	-0.1	-0.065	(-0.094, -0.035)	0.333	69.0	308.7
Q_{21}	-0.1	-0.062	(-0.092, -0.034)	0.953	89.4	365.9
Q_{31}	-0.1	-0.081	(-0.111, -0.051)	0.555	87.2	367.4
Q_{12}	0	-0.008	(-0.042, 0.026)	0.215	86.3	237.2
Q_{22}	0	-0.007	(-0.041, 0.030)	0.016	105.2	554.2
Q_{32}	0	-0.016	(-0.050, 0.019)	0.380	96.4	625.0
Q_{13}	0	-0.006	(-0.044, 0.031)	0.038	99.4	108.0
Q_{23}	0	-0.030	(-0.065, 0.004)	0.360	106.6	177.6
Q_{33}	0	-0.008	(-0.041, 0.028)	0.135	91.6	319.2
σ^2	0.05	0.051	(0.029, 0.078)	0.213	133.7	**1098.0**
$\omega_{m_1}^2 \times 10^3$	1	0.897	(0.502, 1.419)	0.881	112.5	97.8
$\omega_{m_2}^2 \times 10^3$	1	0.695	(0.391, 1.110)	0.776	110.9	218.5
$\omega_{m_3}^2 \times 10^3$	1	0.838	(0.497, 1.341)	0.758	115.4	176.6

The maximum IF's are indicated in bold type.

Table 2: Inefficiency factors.

	(m-move)	(s-move)		(m-move)	(s-move)		(m-move)	(s-move)
$h_{1,500}$	35.8	331.2	$h_{1,1000}$	8.2	313.9	$h_{1,1500}$	43.9	966.5
$h_{2,500}$	37.5	453.5	$h_{2,1000}$	15.4	376.0	$h_{2,1500}$	46.4	697.7
$h_{3,500}$	31.5	424.9	$h_{3,1000}$	15.2	488.3	$h_{3,1500}$	73.5	1138.7
g_{500}	18.7	811.0	g_{1000}	8.7	330.4	g_{1500}	33.0	1396.9

used to generate 40,000 MCMC samples after discarding the first 5,000 samples as the burn-in period. We set the number of the blocks to 301 ($K = 300$) based on several trials.

Table 1 reports the true values, posterior means, 95% credible intervals, p-values of convergence diagnostic test (CD) by Geweke (1992) and estimated inefficiency factors (IF). The inefficiency factor is defined as $1 + 2\sum_{g=1}^{\infty} \rho(g)$, where $\rho(g)$ is the sample autocorrelation at lag g. This is interpreted as the ratio of the numerical variance of the posterior mean from the chain to the variance of the posterior mean from hypothetical uncorrelated draws. The smaller the inefficiency factor becomes, the closer the MCMC sampling is to the uncorrelated sampling.

The posterior means are all close to the true values, which suggests that our proposed algorithms work well. All p-values of convergence diagnostic (CD) tests are greater than 0.01, suggesting that there is no significant evidence against the convergence of the distribution of MCMC samples to the posterior distribution. The inefficiency factors for the single-move sampler (the maximum is 1098.0) are larger than those for the multi-move sampler (the maximum is 172.4). Further, Table 2 shows inefficiency factors for the latent variables $\boldsymbol{h}_t$ and g_t, $t = 500, 1000, 1500$. For $\boldsymbol{h}_t$ (g_t), $t = 500, 1000, 1500$, the inefficiency factors for the single-move sampler are about ten (forty) larger times than those for the multi-move sampler, which suggests that our proposed multi-move sampling method is efficient compared with the single-move sampling method[1].

5 Empirical study

5.1 Data

This section applies our proposed model to returns of subindices of Tokyo stock price index (TOPIX) — three industrial sector indices: (1) Machinery, (2) Electric Appliances and (3) Precision Instruments. The sample period is from January 4, 2005 to December 28, 2012 (1964 observations in total). The asset return is calculated as $y_t = (\log p_t - \log p_{t-1}) \times 100$, where p_t is the asset price at time t. Figure 1 shows the time series plot of the three returns. The trajectories are relatively similar to each other, so it is expected that the equicorrelation

[1]The acceptance rates for $\{\boldsymbol{h}_t\}_{t=1}^n$ and $\{g_t\}_{t=1}^n$ are 0.643 and 0.803 using the multi-move sampler and 0.748 and 0.615 using the single-move sampler. The elapsed time for the simulation using the multi-move (single-move) sampler is 3.00 (0.23) hours per 10,000 iterations with Intel Core i7 3970X Extreme Edition (3.5GHz).

parameters are estimated to be positive.

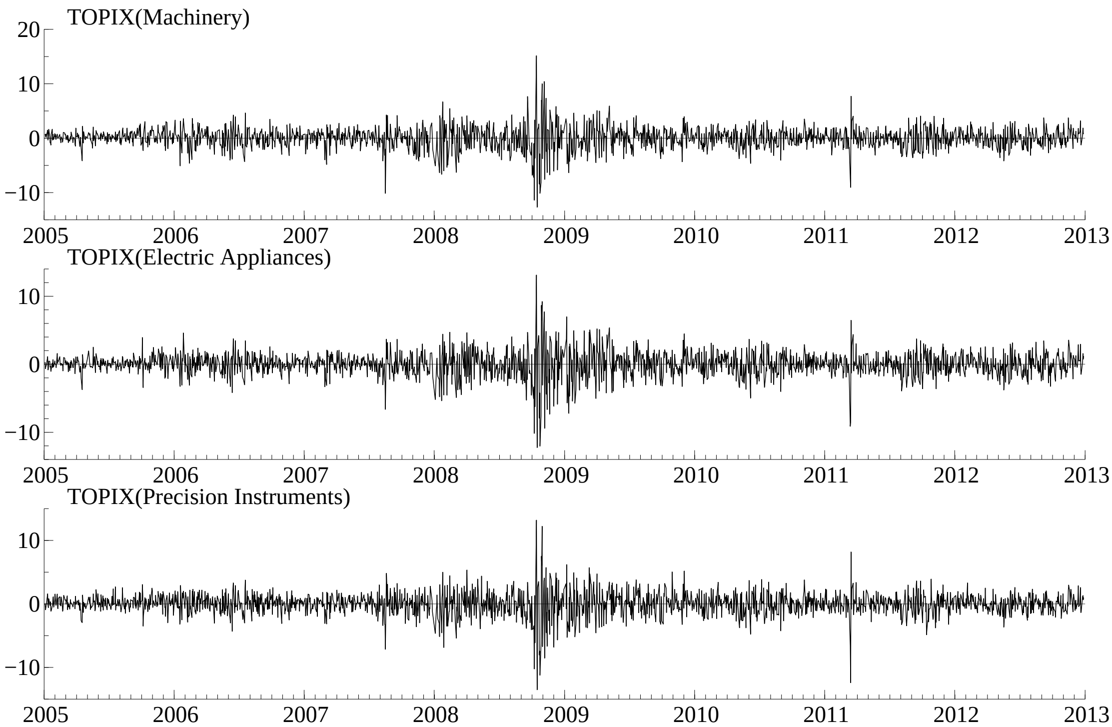

Figure 1: Time series plot of TOPIX.

5.2 Estimation results

Using the same prior distributions for the parameters as in Section 4, we implement the MCMC algorithm to conduct a Bayesian inference on the parameters of interest. We generate 100,000 MCMC samples from the posterior distributions of the parameters in the model after discarding the first 10,000 samples as the burn-in period.

Table 3 reports the summary of the estimation results. Figure 2 shows the posterior means of $\exp(h_{j,t}/2)$, square root of the estimated time-varying variances and ρ_t, the dynamic equicorrelation of $\boldsymbol{y}_t$.

The posterior means of μ_j's are similar and the levels of the volatilities are not different from one another. We also find that the posterior means of the diagonal elements of Ω are similar and it is consistent with the observed fluctuations shown in Figures 2. The credible intervals of off-diagonal elements of Ω are positive and do not include zero, which means that

Table 3: Posterior means, 95% credible intervals, p-values of convergence diagnostic tests and inefficiency factors.

	Mean	95% interval	CD	IF
μ_1	0.767	(0.478, 1.040)	0.002	28.6
μ_2	0.614	(0.268, 0.939)	0.796	17.3
μ_3	0.683	(0.375, 0.973)	0.219	33.1
γ	1.724	(1.491, 1.946)	0.016	19.4
ϕ_1	0.971	(0.959, 0.982)	0.618	450.1
ϕ_2	0.978	(0.968, 0.987)	0.104	476.0
ϕ_3	0.976	(0.965, 0.985)	0.970	360.3
θ	0.943	(0.897, 0.975)	0.013	315.7
Ω_{11}	0.030	(0.020, 0.043)	0.502	259.7
Ω_{21}	0.027	(0.018, 0.038)	0.252	269.9
Ω_{31}	0.026	(0.017, 0.037)	0.794	239.9
Ω_{22}	0.025	(0.016, 0.037)	0.121	393.0
Ω_{32}	0.024	(0.015, 0.035)	0.394	351.9
Ω_{33}	0.024	(0.015, 0.036)	0.890	422.6
Q_{11}	-0.108	(-0.138, -0.080)	0.839	262.0
Q_{21}	-0.086	(-0.113, -0.060)	0.045	240.7
Q_{31}	-0.084	(-0.110, -0.059)	0.095	193.0
Q_{12}	-0.015	(-0.039, 0.012)	0.396	186.7
Q_{22}	-0.005	(-0.032, 0.021)	0.209	368.1
Q_{32}	-0.002	(-0.029, 0.024)	0.895	372.8
Q_{13}	0.008	(-0.026, 0.043)	0.366	178.3
Q_{23}	0.004	(-0.031, 0.039)	0.838	582.5
Q_{33}	-0.014	(-0.047, 0.019)	0.496	255.9
σ^2	0.061	(0.028, 0.111)	0.011	361.1
$\omega^2_{m_1} \times 10^3$	0.261	(0.131, 0.486)	0.856	166.0
$\omega^2_{m_2} \times 10^3$	0.205	(0.110, 0.364)	0.102	130.0
$\omega^2_{m_3} \times 10^3$	0.224	(0.118, 0.408)	0.049	250.8

unobserved volatilities are correlated positively each other.

The posterior means of autoregressive coefficients (ϕ_j's) are very high (over 0.97), which shows that the log volatilities follow highly persistent processes. In addition, the top three panels of Figure 2 indicate the comovement of the volatilities. The trajectories sharply increased in September 2008, corresponding to the financial crisis during which Lehman Brothers filed for Chapter 11 bankruptcy protection (September 15, 2008). We also observe the increase in March 2011, resulted from Tohoku Region Pacific Coast Earthquake.

The posterior means of autoregressive coefficient (θ) is very high and the equicorrelation parameter is highly persistent, too. The bottom panel of Figure 2 shows that the equicor-

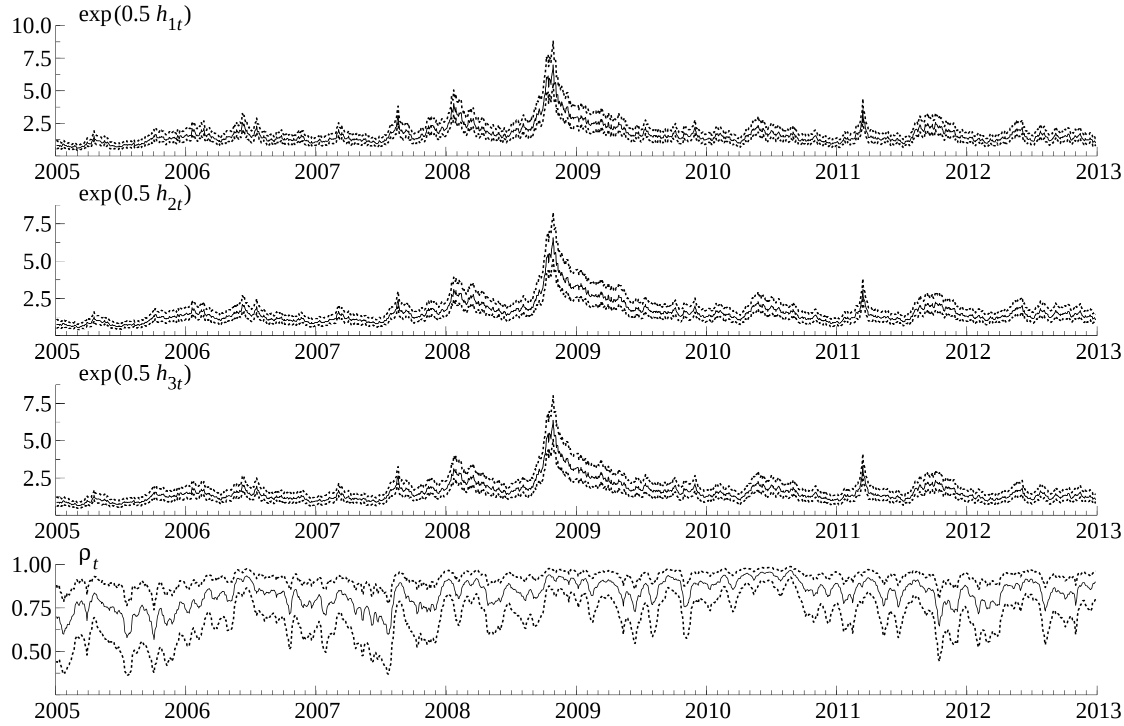

Figure 2: Posterior means (solid lines) and 95% credible intervals (between the two outmost dotted lines) of square root of the variances and the dynamic equicorrelation.

relation parameter varies at a high level (far from zero) and greatly, which means that it is time-varying and far from constant.

The posterior probability with which Q_{11} is negative is over 0.975 and this is similar for Q_{21} and Q_{31}. The 95% credible intervals of the other elements of Q include zeros. As stated in Section 2, Q is the covariance between $\boldsymbol{z}_t$, the transformed error term in the observed equation of our model and $\boldsymbol{\eta}_{t+1}$, the error term in the state equation for $t = 1, \ldots, n-1$. To verify the existence of the cross leverage effects, we should transform Q using R_t and Ω to obtain the posterior means of the (time-varying) correlations between $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_{t+1}$.

Figure 3 shows the posterior means of dynamic correlations between the return of i-th asset at time t ($y_{i,t}$) and the j-th log variance at time $t+1$ ($h_{j,t+1}$). Their trajectories are negative and far from zero, which indicates the existence of the cross leverage effect.

We note that the leverage effect of asset 1 is the strongest (the mean is -0.617) and the leverage effect of asset 3 is the weakest (the mean is -0.492) among the cross leverage

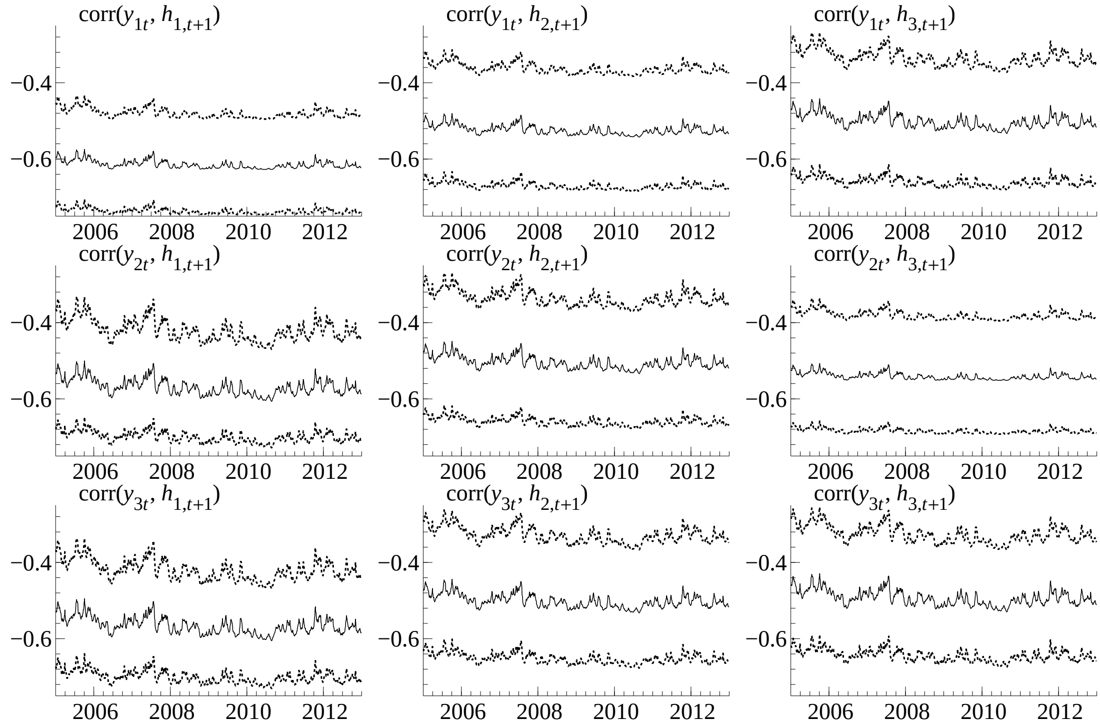

Figure 3: Posterior means (solid lines) and 95% credible intervals (between the two outmost dotted lines) of correlation of $(y_{it}, h_{j,t+1})$.

effects. In addition, the correlations between y_{1t} and $h_{2,t+1}$ is apparently weaker than the correlations between y_{2t} and $h_{1,t+1}$. It indicates that cross leverage effects from asset i to j, $i \neq j$, are not symmetric.

In conclusion, using our multivariate DESV model, we can therefore detect the volatility clustering, the dynamic equicorrelation and the cross leverage effects of the three subindices.

Figure 4 shows the posterior means of the expectations of the asset returns. We find that the 95% credible intervals include zero at almost all of the time points as expected. It seems to fluctuate slowly, which is consistent with the small values of the variances ($\omega_{m_j}^2$'s) reported in Table 3[2].

Note that the acceptance rates for $\{\boldsymbol{h}_t\}_{t=1}^n$ and $\{g_t\}_{t=1}^n$ in the independent MH algorithms are 0.645 and 0.791, respectively. It indicates that the generated candidates are accepted with relatively high probability and our sampling algorithm works well.

[2]Although the estimation results of $\omega_{m_j}^2$'s may be affected by the selection of hyperparameters of the corresponding prior distributions, the posterior estimates of other parameters are not affected.

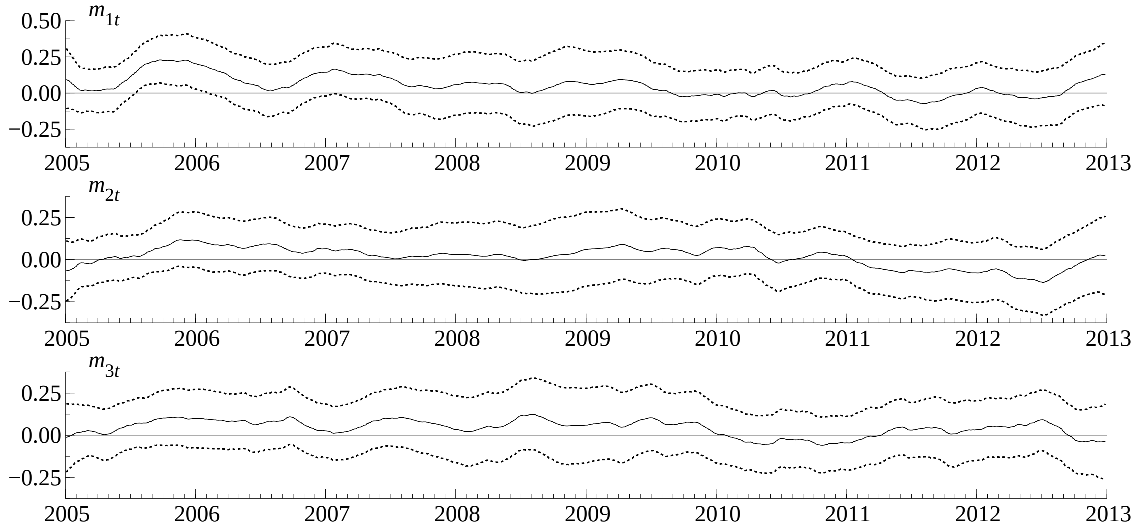

Figure 4: Posterior means (solid lines) and 95% credible intervals (between the two outmost dotted lines) of the expectations.

5.3 Model comparison

This subsection conducts a model comparison of our proposed model and the competing models based on forecasting and the DIC (Spiegelhalter, Best, Carlin, and van der Linde (2002)).

Forecasting. In modeling time-varying variances of asset returns, it is important to forecast the future covariance matrices of the time series for the financial risk management. To evaluate such a forecasting performance, we conduct out-of-sample covariance forecasts and give the minimum variance portfolios. It has often been implemented to investigate such a forecasting performance by the well-known mean-variance optimization (e.g., Luenberger (1997)).

Suppose that $\mathrm{E}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})$ and $\mathrm{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})$ denote, respectively, the conditional mean and covariance of a p-dimensional vector $\boldsymbol{y}_{t+1}$, the asset return at time $t+1$, given $\mathcal{F}_t$, the information at time t, and $\boldsymbol{\vartheta}$. In this study, we make two hedge portfolios: a global minimum variance (GMV) portfolio and a minimum variance (MV) portfolio. The GMV portfolio weights ($\boldsymbol{w}$) are obtained as the solution to the problem:

$$\min_{\boldsymbol{w}} \boldsymbol{w}'\mathrm{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})\boldsymbol{w} \text{ s.t. } \boldsymbol{w}'\mathbf{1}_p = 1. \tag{65}$$

We set the MV portfolio weights ($\boldsymbol{w}$) as the solution to the problem:

$$\min_{\boldsymbol{w}} \boldsymbol{w}'\text{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})\boldsymbol{w} \text{ s.t. } \boldsymbol{w}'\mathbf{1}_p = 1 \text{ and } \boldsymbol{w}'\text{E}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta}) \geq q_0, \tag{66}$$

where q_0 is the target value. It indicates that we make the expected returns exceed q_0 for this case. The optimal weights are given by

$$\boldsymbol{w}_{\text{GMV}} = \frac{1}{a}\text{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})^{-1}\mathbf{1}_p, \tag{67}$$

$$\boldsymbol{w}_{\text{MV}} = \frac{c - q_0 b}{ac - b^2}\text{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})^{-1}\mathbf{1}_p + \frac{q_0 a - b}{ac - b^2}\text{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})^{-1}\text{E}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta}), \tag{68}$$

where

$$a = \mathbf{1}_p'\text{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})^{-1}\mathbf{1}_p, \tag{69}$$

$$b = \mathbf{1}_p'\text{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})^{-1}\text{E}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta}), \tag{70}$$

$$c = \text{E}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})'\text{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta})^{-1}\text{E}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta}). \tag{71}$$

We implement the rolling forecast as follows:

1. Estimate the parameters of interest using the data from January 2005 to December 2010. (We set the data as $\{\boldsymbol{y}_t\}_{t=1}^n$ and the posterior mean of the parameter as $\overline{\boldsymbol{\vartheta}}$.)

2. For the next 3 months including n_1 trading days, i.e., $t = n+1, \ldots, n+n_1-1$,

    (a) use the particle filter (e.g., Doucet, de Freitas, and Gordon (2001)) to compute $\text{E}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \overline{\boldsymbol{\vartheta}})$ and $\text{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \overline{\boldsymbol{\vartheta}})$ numerically (Note that they cannot be obtained analytically). See Appendix A.2 for details.

    (b) compute the two hedge portfolio weights described above and the "realized" returns, $\boldsymbol{w}_{\text{GMV}}'\boldsymbol{y}_{t+1}$, $\boldsymbol{w}_{\text{MV}}'\boldsymbol{y}_{t+1}$.

3. Include the new observations of the next three months to our estimation period and remove the old observations of the first three months. Re-estimate the parameters of interest using the six-year-data (relabeled as $\{\boldsymbol{y}_t\}_{t=1}^n$).

4. Go to 2.

This is repeated until all one-step-ahead forecasts and portfolio choices are conducted through December 2012. In the end, we calculate the standard deviations of the "realized" returns

(493 in total). The numerical standard error of the estimate is obtained by repeating the particle filter forty times.

As a benchmark for the model comparison, we also estimate the univariate SV model with leverage effect introduced in Section 1 and the two multivariate GARCH-type models, the dynamic equicorrelation (DECO) model proposed by Engle and Kelly (2012) and the asymmetric diagonal BEKK model (see Appendix A.3 for details). We set the number of particles $M = 100,000$ and the target value $q_0 = -10, 0, 10, 20$ annually.

Table 4: Out-of-sample portfolio standard deviations (standard errors in parentheses).

	GMV	MV(-10)	MV(0)	MV(10)	MV(20)
DESV	1.477	1.745	1.835	2.094	2.446
	(0.001)	(0.053)	(0.072)	(0.115)	(0.154)
univariate SV	1.484	2.581	3.158	3.935	4.756
	(0.000)	(0.062)	(0.109)	(0.153)	(0.194)
DECO	1.576	2.606	2.194	1.973	1.942
	(0.000)	(0.034)	(0.022)	(0.013)	(0.011)
BEKK	1.543	2.920	2.349	2.368	2.665
	(0.000)	(0.246)	(0.176)	(0.160)	(0.248)

Table 4 reports the out-of-sample portfolio standard deviations using the six-year rolling estimation window. The prior distributions for each estimation are the same as those of the previous subsection. For each of the hedging strategies, the standard deviation based on our multivariate model is smaller than that of the univariate model. We note that MV portfolio strategy with $q_0 = 20$ (annually) makes the biggest difference between the two and GMV portfolio strategy makes the smallest difference between the two. For GMV strategy, the standard deviation based on our multivariate model is also smaller than those of the DECO model and the BEKK model. However, with respect to MV strategy, our model may not necessarily outperform the DECO model and the BEKK model for very large q_0, taking account of standard errors.

Model selection based on DIC. The Deviance Information Criterion (DIC) is used as a Bayesian measure of fit or adequacy and is defined as

$$\mathrm{DIC} = \mathrm{E}_{\boldsymbol{\vartheta}|Y_n}[D(\boldsymbol{\vartheta})] + p_D, \tag{72}$$

where $D(\boldsymbol{\vartheta}) = -2 \log f(Y_n|\boldsymbol{\vartheta}),\ p_D = \mathrm{E}_{\boldsymbol{\vartheta}|Y_n}[D(\boldsymbol{\vartheta})] - D(\mathrm{E}_{\boldsymbol{\vartheta}|Y_n}[\boldsymbol{\vartheta}])$ represents model complexity as a penalty. We estimate $\mathrm{E}_{\boldsymbol{\vartheta}|Y_n}[D(\boldsymbol{\vartheta})]$ using the sample analogue $\overline{D(\boldsymbol{\vartheta}_{(d)})} = \frac{1}{d^*}\sum_{d=1}^{d^*} D(\boldsymbol{\vartheta}_{(d)})$, where $\boldsymbol{\vartheta}_{(d)}$'s are resampled from the posterior distribution. We set d^* equal to 100. Because we need to compute $D(\boldsymbol{\vartheta})$ numerically, we use the particle filter, where we set the number of particles $M = 10,000$ (see Appendix A.2 for details). The numerical standard error of the estimate is obtained by repeating the particle filter ten times.

Table 5: The means of DIC estimates, their standard errors, the maximum and minimum of DIC values.

	DIC	(s.e.)	DICmax	DICmin
DESV	16311.8	(1.3)	16320.7	16305.9
univariate SV	21774.4	(1.3)	21781.4	21769.6
DECO	16598.8	(0.8)	16604.9	16595.5
BEKK	16393.0	(1.0)	16398.8	16389.1

Table 5 shows the sample means of ten DICs, their standard errors, the maximum and minimum of ten DICs for each model. The DIC of the DESV model is the smallest and our model outperforms other competing models. We also note that the DICs of the DECO model and the BEKK model are smaller than that of the univariate SV model.

In summary, our proposed model with the time-varying covariance structure shows good out-of-sample forecasting performance with respect to dynamic GMV portfolio, and our multivariate model, the DECO model and the BEKK model show good out-of-sample forecasting performances with respect to dynamic MV portfolios. In addition, the DESV model attains the smallest DIC. It suggests that, for asset returns with time-varying variances, we should model the covariance structure among the asset returns and the one-step-ahead variances including the dynamic correlations between the asset returns. Furthermore, in comparison with the univariate SV and the two multivariate GARCH-type models, our method is shown to perform well regarding both one-ahead predictions and goodness-of-fit in the analysis of multivariate stock returns.

6 Conclusion

This article proposed the novel multivariate stochastic volatility model with dynamic equicorrelation and cross leverage effect. We took a Bayesian approach and described the efficient MCMC algorithm by dividing the latent variables of our nonlinear model into several blocks and approximating them to those of linear Gaussian state-space models. Its sampling efficiency is illustrated using simulated data in comparison with the single-move sampler.

An empirical study is provided using industrial sector index data of TOPIX. We find the persistence in the volatilities and equicorrelations and the existence of strong cross leverage effects. Model comparisons are conducted and our DESV model is found to outperform the univariate SV model and the multivariate GARCH-type models.

Our proposed model may be extended to the block-equicorrelation model. As shown in this article, the equicorrelation assumption is very simple but useful. Meanwhile, the assumption seems to be too strong and runs counter to the intuition, especially when the number of dependent variables is very large. However, this is beyond the scope of this paper and is left for our future work.

Acknowledgements

The authors would like to thank the associate editor, two anonymous referees, Tsunehiro Ishihara, Kazuhiko Kakamu, Shuichi Nagata, Teruo Nakatsuma, Kazumitsu Nawata, Kosuke Oya, Dale J. Poirier, Akimichi Takemura, Yoshihiro Yajima, Toshiaki Watanabe and participants in Japanese Economic Association 2012 Spring meeting and in International Society for Bayesian Analysis 2012 world meeting for their valuable comments. This work is supported by the Research Fellowship (DC1) from the Japan Society for the Promotion of Science, by the research fellowship from Ishii Memorial Securities Research Promotion Foundation and by the Grants-in-Aid for Scientific Research (A) 21243018 from the Japanese Ministry of Education, Science, Sports, Culture and Technology.

The computational results are obtained using Ox version 5.10 (see Doornik (2007)).

Appendix

A.1 Computations for the block-sampler

Noting that

$$\frac{\partial R_t^{-1/2}}{\partial g_t} = -\frac{1}{2}\mathrm{diag}\begin{bmatrix}(p-1)\{1+(p-1)\rho_t\}^{-3/2}(1-\rho_t)\rho_t \\ -(1-\rho_t)^{-1/2}\rho_t\cdot\mathbf{1}_{p-1}\end{bmatrix}\cdot R_O', \tag{73}$$

$$\frac{\partial^2 R_t^{-1/2}}{\partial g_t^2}$$

$$= -\frac{1}{2}\mathrm{diag}\left[\begin{array}{c} -\frac{3}{2}(p-1)^2\{1+(p-1)\rho_t\}^{-5/2}(1-\rho_t)\rho_t + (p-1)\{1+(p-1)\rho_t\}^{-3/2}(1-2\rho_t) \\ \{-\frac{1}{2}(1-\rho_t)^{-3/2}\rho_t - (1-\rho_t)^{-1/2}\}\cdot \mathbf{1}_{p-1} \end{array}\right]\cdot R_O', \tag{74}$$

$$\frac{\partial V_t^{-1/2}}{\partial h_{j_1 t}} = \mathrm{diag}\left\{\mathbf{0}', -\frac{1}{2}\exp\left(-\frac{1}{2}h_{j_1 t}\right), \mathbf{0}'\right\}, \quad j_1 = 1,\ldots,p, \tag{75}$$

$$\frac{\partial^2 V_t^{-1/2}}{\partial h_{j_1 t}^2} = \mathrm{diag}\left\{\mathbf{0}', \frac{1}{4}\exp\left(-\frac{1}{2}h_{j_1 t}\right), \mathbf{0}'\right\}, \quad j_1 = 1,\ldots,p, \tag{76}$$

and

$$\frac{\partial^2 V_t^{-1/2}}{\partial h_{j_1 t}\partial h_{j_2 t}} = O, \quad j_1 = 1,\ldots,p, \ j_2 = 1,\ldots,p, \ j_1 \neq j_2, \tag{77}$$

we compute $\boldsymbol{d}_t, A_t, B_t$ as below.

A.1.1 Computations for $\{h_t\}_{t=1}^n$

Derivation of $\boldsymbol{d}_t,\ t = s+1,\ldots,s+r.$ We can obtain

$$(\boldsymbol{d}_t)_{j_1} = \frac{\partial l_t}{\partial h_{j_1 t}} + \frac{\partial l_{t-1}}{\partial h_{j_1 t}}, \quad j_1 = 1,\ldots,p, \tag{78}$$

where

$$\frac{\partial l_t}{\partial h_{j_1 t}} = -\left\{R_t^{-\frac{1}{2}}\frac{\partial V_t^{-\frac{1}{2}}}{\partial h_{j_1 t}}(\boldsymbol{y}_t - \boldsymbol{m}_t) + Q'\Omega^{-1}\Phi\frac{\partial \boldsymbol{h}_t}{\partial h_{j_1 t}}\right\}'(I_p - Q'\Omega^{-1}Q)^{-1}$$
$$\cdot\left[R_t^{-\frac{1}{2}}V_t^{-\frac{1}{2}}(\boldsymbol{y}_t - \boldsymbol{m}_t) - Q'\Omega^{-1}\{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_t - \boldsymbol{\mu})\}\right] - \frac{1}{2}, \tag{79}$$

$$\frac{\partial l_{t-1}}{\partial h_{j_1 t}} = \left(Q'\Omega^{-1}\frac{\partial \boldsymbol{h}_t}{\partial h_{j_1 t}}\right)'(I_p - Q'\Omega^{-1}Q)^{-1}$$
$$\cdot\left[R_{t-1}^{-\frac{1}{2}}V_{t-1}^{-\frac{1}{2}}(\boldsymbol{y}_{t-1} - \boldsymbol{m}_{t-1}) - Q'\Omega^{-1}\{\boldsymbol{h}_t - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_{t-1} - \boldsymbol{\mu})\}\right]. \tag{80}$$

Derivation of $A_t,\ t = s+1,\ldots,s+r.$ We can obtain

$$(A_t)_{[j_1,j_1]} = -\mathrm{E}\left(\frac{\partial^2 l_t}{\partial h_{j_1 t}^2}\right) - \frac{\partial^2 l_{t-1}}{\partial h_{j_1 t}^2}, \quad j_1 = 1,\ldots,p, \tag{81}$$

$$(A_t)_{[j_1,j_2]} = -\mathrm{E}\left(\frac{\partial^2 l_t}{\partial h_{j_1 t}\partial h_{j_2 t}}\right) - \frac{\partial^2 l_{t-1}}{\partial h_{j_1 t}\partial h_{j_2 t}}, \quad j_1 = 1,\ldots,p, \ , j_2 = 1,\ldots,p, \ j_1 \neq j_2, \tag{82}$$

where

$$
\begin{aligned}
&\mathrm{E}(\partial^2 l_t/\partial h_{j_1t}^2) \\
&= -\mathrm{tr}\Bigg[\Bigg\{\frac{\partial^2 V_t^{-\frac{1}{2}}}{\partial h_{j_1t}^2} R_t^{-\frac{1}{2}'}(I_p - Q'\Omega^{-1}Q)^{-1} R_t^{-\frac{1}{2}} V_t^{-\frac{1}{2}} \\
&\qquad + \frac{\partial V_t^{-\frac{1}{2}}}{\partial h_{j_1t}} R_t^{-\frac{1}{2}'}(I_p - Q'\Omega^{-1}Q)^{-1} R_t^{-\frac{1}{2}} \frac{\partial V_t^{-\frac{1}{2}}}{\partial h_{j_1t}}\Bigg\} \mathrm{E}\{(\boldsymbol{y}_t - \boldsymbol{m}_t)(\boldsymbol{y}_t - \boldsymbol{m}_t)'\}\Bigg] \\
&\quad - \mathrm{E}(\boldsymbol{y}_t - \boldsymbol{m}_t)'\Bigg[-\frac{\partial^2 V_t^{-\frac{1}{2}}}{\partial h_{j_1t}^2} R_t^{-\frac{1}{2}'}(I_p - Q'\Omega^{-1}Q)^{-1} Q'\Omega^{-1}\{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_t - \boldsymbol{\mu})\} \\
&\qquad + 2\frac{\partial V_t^{-\frac{1}{2}}}{\partial h_{j_1t}} R_t^{-\frac{1}{2}'}(I_p - Q'\Omega^{-1}Q)^{-1} Q'\Omega^{-1}\Phi \frac{\partial \boldsymbol{h}_t}{\partial h_{j_1t}}\Bigg] \\
&\quad - \left(\frac{\partial \boldsymbol{h}_t}{\partial h_{j_1t}}\right)' \Phi\Omega^{-1}Q(I_p - Q'\Omega^{-1}Q)^{-1} Q'\Omega^{-1}\Phi \frac{\partial \boldsymbol{h}_t}{\partial h_{j_1t}},
\end{aligned} \tag{83}
$$

$$
\begin{aligned}
&\mathrm{E}(\partial^2 l_t/\partial h_{j_1t}\partial h_{j_2t}) \\
&= -\mathrm{tr}\Bigg\{\frac{\partial V_t^{-\frac{1}{2}}}{\partial h_{j_1t}} R_t^{-\frac{1}{2}'}(I_p - Q'\Omega^{-1}Q)^{-1} R_t^{-\frac{1}{2}} \frac{\partial V_t^{-\frac{1}{2}}}{\partial h_{j_2t}} \cdot \mathrm{E}\{(\boldsymbol{y}_t - \boldsymbol{m}_t)(\boldsymbol{y}_t - \boldsymbol{m}_t)'\}\Bigg\} \\
&\quad - \mathrm{E}(\boldsymbol{y}_t - \boldsymbol{m}_t)'\Bigg\{\frac{\partial V_t^{-\frac{1}{2}}}{\partial h_{j_1t}} R_t^{-\frac{1}{2}'}(I_p - Q'\Omega^{-1}Q)^{-1} Q'\Omega^{-1}\Phi \frac{\partial \boldsymbol{h}_t}{\partial h_{j_2jt}} \\
&\qquad + \frac{\partial V_t^{-\frac{1}{2}}}{\partial h_{j_2t}} R_t^{-\frac{1}{2}'}(I_p - Q'\Omega^{-1}Q)^{-1} Q'\Omega^{-1}\Phi \frac{\partial \boldsymbol{h}_t}{\partial h_{j_1t}}\Bigg\} \\
&\quad - \left(\frac{\partial \boldsymbol{h}_t}{\partial h_{j_1t}}\right)' \Phi\Omega^{-1}Q(I_p - Q'\Omega^{-1}Q)^{-1} Q'\Omega^{-1}\Phi \frac{\partial \boldsymbol{h}_t}{\partial h_{j_2t}},
\end{aligned} \tag{84}
$$

$$
\frac{\partial^2 l_{t-1}}{\partial h_{j_1t}\partial h_{j_2t}} = -\left(Q'\Omega^{-1}\frac{\partial \boldsymbol{h}_t}{\partial h_{j_1t}}\right)'(I_p - Q'\Omega^{-1}Q)^{-1}\left(Q'\Omega^{-1}\frac{\partial \boldsymbol{h}_t}{\partial h_{j_2t}}\right). \tag{85}
$$

We note that

$$
\boldsymbol{z}_t = R_t^{-1/2} V_t^{-1/2}(\boldsymbol{y}_t - \boldsymbol{m}_t) \sim \mathrm{N}(Q'\Omega^{-1}\{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_t - \boldsymbol{\mu})\}, I_p - Q'\Omega^{-1}Q), \tag{86}
$$

and hence that

$$
\mathrm{E}(\boldsymbol{y}_t - \boldsymbol{m}_t) = V_t^{1/2} R_t^{1/2} Q'\Omega^{-1}\{\boldsymbol{h}_{t+1} - \boldsymbol{\mu} - \Phi(\boldsymbol{h}_t - \boldsymbol{\mu})\}, \tag{87}
$$

$$
\mathrm{Var}(\boldsymbol{y}_t - \boldsymbol{m}_t) = V_t^{1/2} R_t^{1/2}(I_p - Q'\Omega^{-1}Q)(R_t^{1/2})' V_t^{1/2}, \tag{88}
$$

$$
\mathrm{E}\{(\boldsymbol{y}_t - \boldsymbol{m}_t)(\boldsymbol{y}_t - \boldsymbol{m}_t)'\} = \mathrm{Var}(\boldsymbol{y}_t - \boldsymbol{m}_t) + \mathrm{E}(\boldsymbol{y}_t - \boldsymbol{m}_t)\mathrm{E}(\boldsymbol{y}_t - \boldsymbol{m}_t)'. \tag{89}
$$

Derivation of $B_t,\ t=s+2,\dots,s+r.$ We can obtain

$$(B_t)_{[j_1,j_2]} = -\mathrm{E}\left(\frac{\partial^2 l_{t-1}}{\partial h_{j_1 t}\partial h_{j_2,t-1}}\right), \quad j_1=1,\dots,p,\ j_2=1,\dots,p, \tag{90}$$

where

$$\begin{aligned}\mathrm{E}\left(\frac{\partial^2 l_{t-1}}{\partial h_{j_1 t}\partial h_{j_2,t-1}}\right) &= \left(Q'\Omega^{-1}\frac{\partial \boldsymbol{h}_t}{\partial h_{j_1 t}}\right)'(I_p - Q'\Omega^{-1}Q)^{-1}R_{t-1}^{-\frac{1}{2}}\frac{\partial V_{t-1}^{-\frac{1}{2}}}{\partial h_{j_2,t-1}}\mathrm{E}(\boldsymbol{y}_{t-1}-\boldsymbol{m}_{t-1}) \\ &\quad + \left(Q'\Omega^{-1}\frac{\partial \boldsymbol{h}_t}{\partial h_{j_1 t}}\right)'(I_p - Q'\Omega^{-1}Q)^{-1}\left(Q'\Omega^{-1}\Phi\frac{\partial \boldsymbol{h}_{t-1}}{\partial h_{j_2,t-1}}\right).\end{aligned} \tag{91}$$

A.1.2 Computations for $\{g_t\}_{t=1}^n$ (block-sampler)

Derivation of $\boldsymbol{d}_t,\ t=s+1,\dots,s+r.$ We can obtain

$$\boldsymbol{d}_t = \frac{\partial l_t}{\partial g_t}, \tag{92}$$

where

$$\begin{aligned}\frac{\partial l_t}{\partial g_t} &= -\left\{\frac{\partial R_t^{-\frac{1}{2}}}{\partial g_t}\cdot V_t^{-\frac{1}{2}}(\boldsymbol{y}_t-\boldsymbol{m}_t)\right\}'(I_p - Q'\Omega^{-1}Q)^{-1} \\ &\cdot\left[R_t^{-\frac{1}{2}}V_t^{-\frac{1}{2}}(\boldsymbol{y}_t-\boldsymbol{m}_t) - Q'\Omega^{-1}\{\boldsymbol{h}_{t+1}-\boldsymbol{\mu}-\Phi(\boldsymbol{h}_t-\boldsymbol{\mu})\}\right] + \frac{1}{2}p(p-1)\rho_t^2\{1+(p-1)\rho_t\}^{-1}.\end{aligned} \tag{93}$$

Derivation of $A_t,\ t=s+1,\dots,s+r.$ We can obtain

$$A_t = -\mathrm{E}\left(\frac{\partial^2 l_t}{\partial g_t^2}\right), \tag{94}$$

where

$$\begin{aligned}&\mathrm{E}(\partial^2 l_t/\partial g_t^2) \\ &= -\mathrm{tr}\left[\left\{V_t^{-\frac{1}{2}}\left(\frac{\partial^2 R_t^{-\frac{1}{2}}}{\partial g_t^2}\right)'(I_p - Q'\Omega^{-1}Q)^{-1}R_t^{-\frac{1}{2}}V_t^{-\frac{1}{2}}\right.\right. \\ &\qquad \left.\left.+ V_t^{-\frac{1}{2}}\left(\frac{\partial R_t^{-\frac{1}{2}}}{\partial g_t}\right)'(I_p - Q'\Omega^{-1}Q)^{-1}\frac{\partial R_t^{-\frac{1}{2}}}{\partial g_t}V_t^{-\frac{1}{2}}\right\}\mathrm{E}\{(\boldsymbol{y}_t-\boldsymbol{m}_t)(\boldsymbol{y}_t-\boldsymbol{m}_t)'\}\right] \\ &\quad + \mathrm{E}(\boldsymbol{y}_t-\boldsymbol{m}_t)'V_t^{-\frac{1}{2}}\left(\frac{\partial^2 R_t^{-\frac{1}{2}}}{\partial g_t^2}\right)'(I_p - Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}\{\boldsymbol{h}_{t+1}-\boldsymbol{\mu}-\Phi(\boldsymbol{h}_t-\boldsymbol{\mu})\} \\ &\quad + \frac{1}{2}p(p-1)\rho_t^2(1-\rho_t)\{1+(p-1)\rho_t\}^{-2}\{2+(p-1)\rho_t\}.\end{aligned} \tag{95}$$

Derivation of $B_t,\ t=s+2,\dots,s+r.$ We can obtain $B_t=0$.

A.2 Particle filter

Let $f(\boldsymbol{h}_t|Y_t,\boldsymbol{\vartheta})$ denote the density function of $\boldsymbol{h}_t$ given $(Y_t,\boldsymbol{\vartheta})$ where $Y_t=\{\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t\}$, and let $\hat{f}(\boldsymbol{h}_t|Y_t,\boldsymbol{\vartheta})$ denote the discrete approximation to $f(\boldsymbol{h}_t|Y_t,\boldsymbol{\vartheta})$.

We draw M samples from the conditional joint distribution of $(\boldsymbol{h}_{t+1},\boldsymbol{h}_t,g_{t+1},g_t,\boldsymbol{m}_{t+1},\boldsymbol{m}_t)$ given $(Y_{t+1},\boldsymbol{\vartheta})$ with the density

$$\begin{aligned}
&f(\boldsymbol{h}_{t+1},\boldsymbol{h}_t,g_{t+1},g_t,\boldsymbol{m}_{t+1},\boldsymbol{m}_t|Y_{t+1},\boldsymbol{\vartheta})\\
&\propto f(\boldsymbol{y}_{t+1}|\boldsymbol{h}_{t+1},g_{t+1},\boldsymbol{m}_{t+1},Y_t,\boldsymbol{\vartheta})f(\boldsymbol{h}_{t+1}|\boldsymbol{h}_t,g_t,\boldsymbol{m}_t,Y_t,\boldsymbol{\vartheta})f(g_{t+1}|g_t,\boldsymbol{\vartheta})f(\boldsymbol{m}_{t+1}|\boldsymbol{m}_t,\boldsymbol{\vartheta})\\
&\quad\times f(\boldsymbol{h}_t,g_t,\boldsymbol{m}_t|Y_t,\boldsymbol{\vartheta}). \qquad (96)
\end{aligned}$$

We implement the particle filter:

1. (a) Generate

$$\boldsymbol{h}_1^{(i)}\sim \mathrm{N}_p(\boldsymbol{\mu}_{\boldsymbol{h}_1},\Sigma_{\boldsymbol{h}_1}),\quad g_1^{(i)}\sim \mathrm{N}(\mu_{g_1},\sigma^2_{g_1}),\quad \boldsymbol{m}_1^{(i)}\sim \mathrm{N}_p(\boldsymbol{\mu}_{\boldsymbol{m}_1},\Sigma_{\boldsymbol{m}_1}),\quad i=1,\ldots,M,$$

where $\boldsymbol{\mu}_{\boldsymbol{h}_1},\boldsymbol{\mu}_{\boldsymbol{m}_1}$ are some constant vectors, $\Sigma_{\boldsymbol{h}_1},\Sigma_{\boldsymbol{m}_1}$ are some constant positive-definite matrices, μ_{g_1} is some constant and $\sigma^2_{g_1}$ is some positive constant (we adopt the posterior mean vectors of $\boldsymbol{h}_1,\boldsymbol{m}_1$, the posterior covariance matrices of $\boldsymbol{h}_1,\boldsymbol{m}_1$, the posterior mean of g_1 and the posterior variance of g_1), respectively.

   (b) Compute

$$\pi_i=\frac{\tilde{\pi}_i}{\sum_{j=1}^M\tilde{\pi}_i},\quad \tilde{\pi}_i=\frac{f(\boldsymbol{y}_1|\boldsymbol{h}_1,g_1,\boldsymbol{m}_1,\boldsymbol{\vartheta})f(\boldsymbol{h}_1,g_1,\boldsymbol{m}_1|\boldsymbol{\vartheta})}{g(\boldsymbol{h}_1,g_1,\boldsymbol{m}_1|\boldsymbol{\vartheta})}, \qquad (97)$$

where $g(\cdot)$ is a density generating $\boldsymbol{h}_1^{(i)},g_t^{(i)}$ and $\boldsymbol{m}_1^{(i)}$.

   (c) Set $\hat{f}(\boldsymbol{h}_1^{(i)},g_1^{(i)},\boldsymbol{m}_1^{(i)}|Y_1,\boldsymbol{\vartheta})=\pi_i$.

2. For $t=1,\ldots,n+n_1-1$,

   (a) generate $\boldsymbol{h}_t^{(i)},g_t^{(i)},\boldsymbol{m}_t^{(i)}\sim\hat{f}(\boldsymbol{h}_t,g_t,\boldsymbol{m}_t|Y_t,\boldsymbol{\vartheta})$.

   (b) generate $\boldsymbol{h}_{t+1}^{(i)}\sim f(\boldsymbol{h}_{t+1}|\boldsymbol{h}_t^{(i)},g_t^{(i)},\boldsymbol{m}_t^{(i)},Y_t,\boldsymbol{\vartheta})$, $g_{t+1}^{(i)}\sim f(g_{t+1}|g_t^{(i)},\boldsymbol{\vartheta})$, $\boldsymbol{m}_{t+1}^{(i)}\sim f(\boldsymbol{m}_{t+1}|\boldsymbol{m}_t^{(i)},\boldsymbol{\vartheta})$.

   (c) compute

$$\pi_i=\frac{\tilde{\pi}_i}{\sum_{j=1}^M\tilde{\pi}_j},\quad \tilde{\pi}_i=f(\boldsymbol{y}_{t+1}|\boldsymbol{h}_{t+1}^{(i)},g_{t+1}^{(i)},\boldsymbol{m}_{t+1}^{(i)},Y_t,\boldsymbol{\vartheta}). \qquad (98)$$

(d) set $\hat{f}(\boldsymbol{h}_{t+1}^{(i)}, g_{t+1}^{(i)}, \boldsymbol{m}_{t+1}^{(i)}|Y_{t+1}, \boldsymbol{\vartheta}) = \pi_i$.

Note that $\boldsymbol{y}_{t+1}|\boldsymbol{h}_{t+1}, g_{t+1}, \boldsymbol{m}_{t+1}, \sim \mathrm{N}_p(\boldsymbol{m}_{t+1}, V_{t+1}^{1/2} R_{t+1} V_{t+1}^{1/2}),\ \ t = 1, \ldots, n + n_1 - 1$. For $t = n, \ldots, n + n_1 - 1$, we obtain

$$\hat{\boldsymbol{\mu}}_{t+1|t} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{m}_{t+1}^{(i)} \to \mathrm{E}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta}), \tag{99}$$

$$\hat{\Sigma}_{t+1|t} = \frac{1}{M}\sum_{i=1}^{M} V_{t+1}^{1/2(i)} R_{t+1}^{(i)} V_{t+1}^{1/2(i)} + \Omega_m \to \mathrm{Var}(\boldsymbol{y}_{t+1}|\mathcal{F}_t, \boldsymbol{\vartheta}). \tag{100}$$

A.3 Benchmarking multivariate GARCH models

A.3.1 Dynamic equicorrelation (DECO) model

As a benchmark for the model comparison, we consider the following dynamic equicorrelation (DECO) model discussed in Engle and Kelly (2012):

$$\boldsymbol{y}_t = \boldsymbol{m}_t + \boldsymbol{\epsilon}_t,\ \ \boldsymbol{\epsilon}_t \sim \mathrm{N}(\mathbf{0}_p, \Sigma_t),\ \ t = 1, ..., n, \tag{101}$$

$$\boldsymbol{m}_{t+1} = \boldsymbol{m}_t + \boldsymbol{\eta}_{\boldsymbol{m}t},\ \ \boldsymbol{\eta}_{\boldsymbol{m}t} \sim \mathrm{N}(\mathbf{0}_p, \Omega_{\boldsymbol{m}}),\ \ t = 1, ..., n-1, \tag{102}$$

$$\boldsymbol{m}_1 = \boldsymbol{\eta}_{\boldsymbol{m}0},\ \ \boldsymbol{\eta}_{\boldsymbol{m}0} \sim \mathrm{N}(\mathbf{0}_p, \kappa I_p), \tag{103}$$

where

$$\Sigma_t = D_t R_t D_t, \tag{104}$$

$$D_t = \mathrm{diag}(h_{1t}^{1/2}, \ldots, h_{pt}^{1/2}), \tag{105}$$

$$R_t = (1-\rho_t) I_p + \rho_t J_t, \tag{106}$$

$$\rho_t = \frac{1}{p(p-1)}(\mathbf{1}_p' R_t^{\mathrm{DCC}} \mathbf{1}_p - p), \tag{107}$$

$$R_t^{\mathrm{DCC}} = \widetilde{Q}_t^{-\frac{1}{2}} Q_t \widetilde{Q}_t^{-\frac{1}{2}}, \tag{108}$$

$$\begin{aligned} Q_t = {} & (1 - \alpha^{\mathrm{DCC}} - \beta^{\mathrm{DCC}})\overline{Q} \\ & + \alpha^{\mathrm{DCC}} \widetilde{Q}_{t-1}^{\frac{1}{2}} D_{t-1}^{-\frac{1}{2}} (\boldsymbol{y}_{t-1} - \boldsymbol{m}_{t-1})(\boldsymbol{y}_{t-1} - \boldsymbol{m}_{t-1})' D_{t-1}^{-\frac{1}{2}} \widetilde{Q}_{t-1}^{\frac{1}{2}} + \beta^{\mathrm{DCC}} Q_{t-1}, \end{aligned} \tag{109}$$

$$Q_1 = \overline{Q}, \tag{110}$$

$\widetilde{Q}_t$ replaces the off-diagonal elements of Q_t with zeros but remains its main diagonal, $\overline{Q}$ is a positive definite matrix,

$$0 < \alpha^{\mathrm{DCC}} < 1, \tag{111}$$

$$0 < \beta^{\mathrm{DCC}} < 1, \tag{112}$$

$$\alpha^{\mathrm{DCC}} + \beta^{\mathrm{DCC}} < 1, \tag{113}$$

$$\Omega_{\boldsymbol{m}} = \mathrm{diag}(\omega_{m_1}^2, \ldots, \omega_{m_p}^2). \tag{114}$$

For simplicity, we assume that all the diagonal elements of $\overline{Q}$ are 1 and all the off-diagonal elements are the same positive constant, say, $\overline{q}$. It is reasonable to make the assumption because $\overline{Q}$ is the unconditional covariance matrix of $D_t^{-1}(\boldsymbol{y}_t - \boldsymbol{m}_t)$ and we apply the data which is expected to be positively correlated.

Following Engle and Kelly (2012), we assume that h_{it}, $i = 1, \ldots, p$ have asymmetric GARCH or GJR structure (Glosten, Jagannathan, and Runkle (1993)):

$$\begin{aligned} h_{it} = {} & \omega_i^{\mathrm{GARCH}} + \alpha_i^{\mathrm{GARCH}}(y_{i,t-1} - m_{i,t-1})^2 \\ & + \gamma_i^{\mathrm{GARCH}}(y_{i,t-1} - m_{i,t-1})^2 \mathrm{I}_{\{y_{i,t-1} - m_{i,t-1} < 0\}} + \beta_i^{\mathrm{GARCH}} h_{i,t-1}, \end{aligned} \tag{115}$$

$$h_{i1} = \omega_i^{\mathrm{GARCH}} / (1 - \alpha_i^{\mathrm{GARCH}} - \beta_i^{\mathrm{GARCH}} - \gamma_i^{\mathrm{GARCH}}), \tag{116}$$

$$0 < \alpha^{\mathrm{GARCH}} < 1, \tag{117}$$

$$0 < \alpha^{\mathrm{GARCH}} + \gamma^{\mathrm{GARCH}} < 1, \tag{118}$$

$$0 < \beta_i^{\mathrm{GARCH}} < 1, \tag{119}$$

$$0 < \alpha_i^{\mathrm{GARCH}} + \beta_i^{\mathrm{GARCH}} + \gamma_i^{\mathrm{GARCH}} < 1, \tag{120}$$

$$0 < \omega_i^{\mathrm{GARCH}}. \tag{121}$$

Priors. The prior specifications are noted as follows: $\alpha_i^{\mathrm{GARCH}} \sim \mathrm{U}(0,1)$, $\gamma_i^{\mathrm{GARCH}} | \alpha_i^{\mathrm{GARCH}} \sim \mathrm{U}(-\alpha_i^{\mathrm{GARCH}}, 1 - \alpha_i^{\mathrm{GARCH}})$, $\beta_i^{\mathrm{GARCH}} | \alpha_i^{\mathrm{GARCH}}, \gamma_i^{\mathrm{GARCH}} \sim \mathrm{U}(0, 1 - \alpha_i^{\mathrm{GARCH}} - \gamma_i^{\mathrm{GARCH}})$, $\widetilde{\omega}_i^{\mathrm{GARCH}} = \log(\omega_i^{\mathrm{GARCH}}) \sim \mathrm{N}(m_{\widetilde{\omega}_i}, s_{\widetilde{\omega}_i}^2)$ for $i = 1, \ldots, p$, $\alpha^{\mathrm{DCC}} \sim \mathrm{U}(0,1)$, $\beta^{\mathrm{DCC}} | \alpha^{\mathrm{DCC}} \sim \mathrm{U}(0, 1 - \alpha^{\mathrm{DCC}})$, $\widetilde{q} = \log\{\overline{q}/(1 - \overline{q})\} \sim \mathrm{N}(m_{\tilde{q}}, s_{\tilde{q}}^2)$, and $\omega_{m_j}^2 \sim \mathrm{IG}(\alpha_{m_j 0}/2, \beta_{m_j 0}/2)$, $j = 1, \ldots, p$, where $\mathrm{U}(a,b)$ denotes a uniform distribution on (a, b).

Estimation. For the restriction of the parameters in the model, we consider the transformation $\widetilde{\alpha}_i^{\mathrm{GARCH}} = \log\{\alpha_i^{\mathrm{GARCH}}/(1 - \alpha_i^{\mathrm{GARCH}})\}$, $\widetilde{\gamma}_i^{\mathrm{GARCH}} = \log\{(\alpha_i^{\mathrm{GARCH}} + \gamma_i^{\mathrm{GARCH}})/(1 - \alpha_i^{\mathrm{GARCH}} - \gamma_i^{\mathrm{GARCH}})\}$, $\widetilde{\beta}_i^{\mathrm{GARCH}} = \log\{\beta_i^{\mathrm{GARCH}}/(1 - \alpha_i^{\mathrm{GARCH}} - \beta_i^{\mathrm{GARCH}} - \gamma_i^{\mathrm{GARCH}})\}$ for $i = 1, \ldots, p$, $\widetilde{\alpha}^{\mathrm{DCC}} = \log\{\alpha^{\mathrm{DCC}}/(1 - \alpha^{\mathrm{DCC}})\}$ and $\widetilde{\beta}^{\mathrm{DCC}} = \log\{\beta^{\mathrm{DCC}}/(1 - \alpha^{\mathrm{DCC}} - \beta^{\mathrm{DCC}})\}$.

We implement the MCMC algorithm described as follows:

1. Initialize $\{\boldsymbol{m}_t\}_{t=1}^n$, $(\widetilde{\alpha}_i^{\mathrm{GARCH}}, \widetilde{\beta}_i^{\mathrm{GARCH}}, \widetilde{\gamma}_i^{\mathrm{GARCH}}, \widetilde{\omega}_i^{\mathrm{GARCH}})$, for $i = 1, \ldots, p$, $(\widetilde{\alpha}^{\mathrm{DCC}}, \widetilde{\beta}^{\mathrm{DCC}})$, $\widetilde{q}$ and $\Omega_{\boldsymbol{m}}$.

2. Generate the model parameters given $\{\boldsymbol{m}_t\}_{t=1}^n$ using the random-walk MH algorithm except for $\Omega_{\boldsymbol{m}}$. Generation of $\Omega_{\boldsymbol{m}}$ given $\{\boldsymbol{m}_t\}_{t=1}^n$ is implemented similarly as DESV model.

3. For $\{\boldsymbol{m}_t\}_{t=1}^n$, we divide it into several blocks using stochastic knots, generate a candidate for one block given other blocks and the parameters with fixed Σ_t's using a simulation smoother (de Jong and Shephard (1995) or Durbin and Koopman (2002)), and accept it with the calculated acceptance rate.

4. Go to 2.

Estimation results for the DECO model are omitted to save space.

A.3.2 Asymmetric diagonal BEKK model

We also consider the following asymmetric diagonal BEKK model (see, e.g., Kroner and Ng (1998), Asai and McAleer (2011) and Ishihara, Omori, and Asai (2011)):

$$\boldsymbol{y}_t = \boldsymbol{m}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathrm{N}(\mathbf{0}_p, \Sigma_t), \quad t = 1, ..., n, \tag{122}$$

$$\boldsymbol{m}_{t+1} = \boldsymbol{m}_t + \boldsymbol{\eta}_{\boldsymbol{m}t}, \quad \boldsymbol{\eta}_{\boldsymbol{m}t} \sim \mathrm{N}(\mathbf{0}_p, \Omega_{\boldsymbol{m}}), \quad t = 1, ..., n-1, \tag{123}$$

$$\boldsymbol{m}_1 = \boldsymbol{\eta}_{\boldsymbol{m}0}, \quad \boldsymbol{\eta}_{\boldsymbol{m}0} \sim \mathrm{N}(\mathbf{0}_p, \kappa I_p), \tag{124}$$

where

$$\Sigma_t = W + A\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'A' + B\Sigma_t B' + C\boldsymbol{\epsilon}_{t-1}^*\boldsymbol{\epsilon}_{t-1}^{*'}C', \tag{125}$$

$$\epsilon_{it}^* = \epsilon_{it}\mathrm{I}_{\{\epsilon_{it}<0\}}, \tag{126}$$

$$W = \Omega - A\Omega A' - B\Omega B' - CQC' \tag{127}$$

$$Q = (\Omega \odot I_p)^{1/2}\{0.5I_p + R \odot (\mathbf{1}_p\mathbf{1}_p' - I_p)\}(\Omega \odot I_p)^{1/2} \tag{128}$$

$$(R)_{ij} = L(\rho_{ij}) \cdot \rho_{ij} + \frac{1}{2\pi}\sqrt{1-\rho_{ij}^2}, \tag{129}$$

$$L(\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^0\int_{-\infty}^0 \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\}dxdy, \tag{130}$$

$$\rho_{ij} = \mathrm{Corr}(\epsilon_{it}, \epsilon_{jt}), \tag{131}$$

$\Omega = \mathrm{E}(\Sigma_t) = \mathrm{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')$, $Q = \mathrm{E}(\boldsymbol{\epsilon}_t^*\boldsymbol{\epsilon}_t^{*'})$, $A = \mathrm{diag}(a_1,\ldots,a_p), B = \mathrm{diag}(b_1,\ldots,b_p)$, $C = \mathrm{diag}(c_1,\ldots,c_p)$ and $\Omega_{\boldsymbol{m}} = \mathrm{diag}(\omega_{m_1}^2,\ldots,\omega_{m_p}^2)$. We assume that

$$0 < a_i < 1,\ 0 < b_i < 1,\ 0 < c_i < 1,\ a_i^2 + b_i^2 + c_i^2 < 1,\ i = 1,\ldots,p, \tag{132}$$

and that, for simplicity,

$$\Omega = V^{\frac{1}{2}} R V^{\frac{1}{2}}, \tag{133}$$

$$V = \mathrm{diag}\{\exp(\tilde{\omega}_1),\ldots,\exp(\tilde{\omega}_p)\}, \tag{134}$$

$$R = (1-\rho_\omega)I_p + \rho_\omega J_p, \tag{135}$$

$$\rho_\omega = \frac{\exp(\tilde{\omega}_{p+1})}{1+\exp(\tilde{\omega}_{p+1})}, \tag{136}$$

(see Rosenbaum (1961) and Tallis (1961)).

Priors. $a_i \sim \mathrm{U}(0,1)$, $b_i|a_i \sim \mathrm{U}(0,(1-a_i^2)^{1/2})$, $c_i|a_i,b_i \sim \mathrm{U}(0,(1-a_i^2-b_i^2)^{1/2})$ for $i = 1,\ldots,p$, $\tilde{\boldsymbol{\omega}} = (\tilde{\omega}_1,\ldots,\tilde{\omega}_{p+1})' \sim \mathrm{N}_{p+1}(\boldsymbol{m}_{\tilde{\boldsymbol{\omega}}}, S_{\tilde{\boldsymbol{\omega}}})$ and $\omega_{m_j}^2 \sim \mathrm{IG}(\alpha_{m_j0}/2, \beta_{m_j0}/2)$, $j = 1,\ldots,p$.

Estimation. For the restriction of the parameters in the model, we consider the transformation $\widetilde{a}_i = \log\{a_i^2/(1-a_i^2)\}$, $\widetilde{b}_i = \log\{b_i^2/(1-a_i^2-b_i^2)\}$, $\widetilde{c}_i = \log\{c_i^2/(1-a_i^2-b_i^2-c_i^2)\}$.

We implement the MCMC algorithm described as follows:

1. Initialize $\{\boldsymbol{m}_t\}_{t=1}^n$, $(\widetilde{a}_i,\widetilde{b}_i,\widetilde{c}_i)$ for $i = 1,\ldots,p$, $\tilde{\boldsymbol{\omega}}$ and $\Omega_{\boldsymbol{m}}$.
2. Generate the model parameters given $\{\boldsymbol{m}_t\}_{t=1}^n$ using the random-walk MH algorithm except for $\Omega_{\boldsymbol{m}}$. Generation of $\Omega_{\boldsymbol{m}}$ given $\{\boldsymbol{m}_t\}_{t=1}^n$ is implemented similarly as DESV model.
3. Generate $\{\boldsymbol{m}_t\}_{t=1}^n$ similarly as noted in the last subsection.
4. Go to 2.

Estimation results for the BEKK model are omitted to save space.

References

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. and M. McAleer (2011). Dynamic conditional correlations for asymmetric processes. *Journal of the Japan Statistical Society 41*, 143–157.

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

Bauwens, L., C. M. Hafner, and S. Laurent (Eds.) (2012). *Handbook of Volatility Models and Their Applications*. Wiley.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate GARCH models: 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. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician 49*, 327–335.

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*, Volume 4, pp. 365–400. Springer-Verlag.

Clements, A. E., C. A. Coleman-Fenn, and D. R. Smith (2011). Forecasting equicorrelation. NCER Working Paper Series 72.

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. A. (2007). *An Object-oriented Martrix Programming Language: Ox 5*. Timberlake Consultants Press.

Doucet, A., N. de Freitas, and N. Gordon (Eds.) (2001). *Sequential Monte Carlo Methods in Practice*. Springer.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*, 603–615.

Elton, E. J. and M. Gruber (1973). Estimating the dependence structure of share prices–implications for portfolio selection. *The Journal of Finance 28*, 1203–1232.

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

Engle, R. and B. Kelly (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics 30*, 212–228.

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

Geweke, J. F. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In A. P. D. J. M. Bernardo, J. O. Berger and A. F. M. Smith (Eds.), *Bayesian Statistics*, Volume 4, pp. 169–188. Oxford University Press.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance 48*, 1779–1801.

Gupta, A. K. and D. K. Nagar (2000). *Matrix Variate Distributions*. Chapman and Hall/CRC.

Hafner, C. M. and O. Reznikova (2012). On the estimation of dynamic conditional correlation models. *Computational Statistics & Data Analysis 56*, 3533–3545.

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

Ishihara, T., Y. Omori, and M. Asai (2011). Matrix exponential stochastic volatility with cross leverage. Discussion Paper CIRJE-F-812.

Jin, X. and Y. Tang (2009). The conditional dynamic dependence between herding and return: Evidence from US equitty market. Working Paper.

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

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

Ledoit, O. and M. Wolf (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management 30*(4), 110–119.

Lucas, A., B. Schwaab, and X. Zhang (2012). Measuring credit risk in a large banking system: econometric modeling and empirics. Discussion Paper.

Luenberger, D. G. (1997). *Investment Science*. Oxford University Press.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques And Tools*. Princeton University Press.

Nakajima, J. (2012). Bayesian analysis of multivariate stochastic volatility with skew distribution. Working Paper.

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

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

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

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, Ser. B 64*, 583–639.

Tallis, G. M. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society, Ser. B 23*, 223–229.

Vargas, G. A. (2009). Dynamic equicorrelation with non-Gaussian innovations. Working Paper.

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