Multivariate stochastic volatility

Siddhartha Chib, Washington University in St. Louis, USA
Yasuhiro Omori, University of Tokyo, Japan
Manabu Asai, Soka University, Japan

April 1, 2007

Contents

1 Introduction

The success of univariate stochastic volatility (SV) models in relation to univariate GARCH models has spurred an enormous interest in generalizations of SV models to a multivariate setting. A large number of multivariate SV (MSV) models are now available along with clearly articulated estimation recipes. Our goal in this paper is to provide the first detailed summary of the various model formulations, along with connections and differences, and discuss how the models are estimated. We aim to show that the developments and achievements in this area represent one of the great success stories of financial econometrics.

As is to be expected, MSV models are generalizations of the univariate SV model. To fix notation and set the stage for our discussion, the canonical version of the univariate SV model is given by (Ghysels, Harvey, and Renault (1996), Broto and Ruiz (2004) and Shephard (2004))

$$
\begin{aligned}
y_t &= \exp(h_t/2)\varepsilon_t, \quad t = 1, \ldots, n, &(1)\\
h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, \quad t = 1, \ldots, n-1, &(2)\\
& h_1 \sim \mathcal{N}\left(\mu, \sigma_\eta^2/(1-\phi^2)\right), &(3)\\
& \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} | h_t \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}, &(4)
\end{aligned}
$$

where y_t is a univariate outcome, h_t is a univariate latent variable and $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a univariate normal distribution with mean μ and variance σ^2, and an m-variate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. In this model, conditioned on the parameters $(\mu, \phi, \sigma_\eta^2)$, the first generating equation represents the distribution of y_t conditioned on h_t, and the second generating equation represents the Markov evolution of h_{t+1} given h_t. The conditional mean of y_t is assumed to be zero because that is a reasonable assumption in the setting of high frequency financial data. The SV model is thus a state-space model, with a linear evolution of the state variable h_t but with a non-linear measurement equation (because h_t enters the outcome model non-linearly). Furthermore, from the measurement equation we see that $\mathrm{Var}(y_t|h_t) = \exp(h_t)$, which implies that h_t may be understood as the log of the conditional variance of the outcome. To ensure that the evolution of these log-volatilities is stationarity, one generally assumes that $|\phi| < 1$. Many other versions of the univariate SV model are possible. For example, it is possible let the model errors have a non-Gaussian fat-tailed distribution, to permit jumps, and incorporate the leverage effect (through a non-zero off-diagonal element in $\boldsymbol{\Sigma}$). The estimation of the canonical SV model and its various extensions was at one time considered difficult since the likelihood function of these models is not easily calculable. This problem has fully resolved by the creative use of Monte Carlo methods, primarily Bayesian Markov chain Monte Carlo (MCMC) methods (for example, Jacquier, Polson, and Rossi (1994), Kim, Shephard, and Chib (1998), Chib, Nardari, and Shephard (2002) and Omori, Chib, Shephard, and Nakajima (2007)).

In the multivariate context, when one is dealing with a collection of financial time series denoted by $\mathbf{y}_t = (y_{1t}, \ldots, y_{pt})'$, the main goal is to model the time-varying conditional covariance matrix of $\mathbf{y}_t$. There are several ways in which this can be done within the SV context (see Asai, McAleer, and Yu (2006) for a recent survey). A typical starting point is the assumption of series-specific log-volatilites h_{tj} $(j \leq p)$ whose joint evolution is governed by a first-order stationary vector autoregressive process

$$\begin{aligned}\mathbf{h}_{t+1} &= \boldsymbol{\mu}+\boldsymbol{\Phi}(\mathbf{h}_t-\boldsymbol{\mu})+\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t|\mathbf{h}_t\sim\mathcal{N}_p(0,\boldsymbol{\Sigma}_{\eta\eta}),\ t=1,\dots,n-1\\ \mathbf{h}_1 &\sim \mathcal{N}_p\left(\boldsymbol{\mu},\boldsymbol{\Sigma}_0\right),\end{aligned}$$

where $\mathbf{h}_t=(h_{1t},\dots,h_{pt})'$. To reduce the computational load, especially when p is large, the log volatilities can be assumed to be conditionally independent. In that case,

$$\begin{aligned}\boldsymbol{\Phi} &= \mathrm{diag}(\phi_{11},...,\phi_{pp}) \text{ and}\\ \boldsymbol{\Sigma}_{\eta\eta} &= \mathrm{diag}(\sigma_{1,\eta\eta},...,\sigma_{p,\eta\eta})\end{aligned}$$

are both diagonal matrices. We refer to the former specification as the *VAR(1)* model and the latter as the *IAR(1)* (for independent AR) model. Beyond these differences, the various models primarily differ in the way in which the outcomes y_t are modeled. In one formulation, the outcomes are assumed to be generated as

$$\mathbf{y}_t=\mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \mathbf{V}_t^{1/2}=\mathrm{diag}\left(\exp(h_{1t}/2),\dots,\exp(h_{pt}/2)\right), \quad t=1,\dots,n,$$

with the additional assumptions that

$$\begin{pmatrix}\boldsymbol{\varepsilon}_t\\ \boldsymbol{\eta}_t\end{pmatrix}|\mathbf{h}_t\sim\mathcal{N}_{2p}(\mathbf{0},\boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{\varepsilon\varepsilon} & \mathbf{O}\\ \mathbf{O} & \boldsymbol{\Sigma}_{\eta\eta}\end{pmatrix}$$

and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ is a matrix in correlation (with units on the main diagonal). Thus, conditioned on $\mathbf{h}_t$, $\mathrm{Var}(\mathbf{y}_t)=\mathbf{V}_t^{1/2}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\mathbf{V}_t^{1/2}$ is time-varying (as required), but the conditional correlation matrix is $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ which is not time-varying. In the sequel we refer to this model as the *basic MSV* model.

A second approach for modeling the outcome process is via a latent factor approach. In this case, the outcome model is specified as

$$\mathbf{y}_t=\mathbf{B}\mathbf{f}_t+\mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t,\ \mathbf{V}_t^{1/2}=\mathrm{diag}\left(\exp(h_{1t}/2),\dots,\exp(h_{pt}/2)\right)$$

where $\mathbf{B}$ is a $p\times q$ matrix ($q\le p$) called the loading matrix, and $\mathbf{f}_t=(f_{1t},...,f_{qt})$ is a $q\times 1$ latent factor at time t. For identification reasons, the loading matrix is subject to some restrictions (that we present later in the paper), and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ is the identity matrix. The model is closed by assuming that the latent variables are distributed independently across time as

$$\mathbf{f}_t|\mathbf{h}_t\sim\mathcal{N}_q(\mathbf{0},\mathbf{D}_t)$$

where

$$\mathbf{D}_t=\mathrm{diag}\left(\exp(h_{p+1,t}),\dots,\exp(h_{p+q,t})\right)$$

is a diagonal matrix that depends on additional latent variables $h_{p+k,t}$. The full set of log-volatilities, namely

$$\mathbf{h}_t = (h_{1t}, \ldots, h_{pt}, h_{p+1,t}, ..., h_{p+q,t}),$$

are assumed to follow a VAR(1) or IAR(1) process. In this model, the variance of $\mathbf{y}_t$ conditional on the parameters and $\mathbf{h}_t$ is

$$\mathrm{Var}(\mathbf{y}_t|\mathbf{h}_t) = \mathbf{V}_t + \mathbf{B}\mathbf{D}_t\mathbf{B}'$$

and therefore the conditional correlation matrix is time-varying.

Another way to model time-varying correlations is by direct modeling of the variance matrix $\mathbf{\Sigma}_t = \mathrm{Var}(\mathbf{y}_t)$. One such model is the Wishart process model proposed by Philipov and Glickman (2006b) who assume that

$$\begin{aligned} \mathbf{y}_t|\mathbf{\Sigma}_t &\sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}_t), \\ \mathbf{\Sigma}_t|\nu, \mathbf{S}_{t-1} &\sim \mathcal{IW}_p(\nu, \mathbf{S}_{t-1}), \end{aligned}$$

where $\mathcal{IW}_p(\nu_0, \mathbf{Q}_0)$ denotes a p-dimensional inverted Wishart distribution with parameters $(\nu_0, \mathbf{Q}_0)$, and $\mathbf{S}_{t-1}$ is a function of $\mathbf{\Sigma}_{t-1}$. Several models along these lines have been proposed as we discuss in Section 4.

The rest of the article is organized as follows. In Section 2, we first discuss the basic MSV model along with some of its extensions. Section 3 is devoted to the class of factor MSV models while Section 4 deals with models in which the dynamics of the covariance matrix are modeled directly.

2 Basic MSV model

2.1 No Leverage model

As in the preceding section, let $\mathbf{y}_t = (y_{1t}, \ldots, y_{pt})'$ denote a set of observations at time t on p financial variables and let $\mathbf{h}_t = (h_{1t}, \ldots, h_{pt})'$ be the corresponding vector of log volatilities. Then the basic MSV model is defined in terms of the generating processes

$$\mathbf{y}_t = \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, n, \tag{5}$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \mathbf{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad t = 1, \ldots, n-1, \tag{6}$$

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

where

$$\mathbf{V}_t^{1/2} = \mathrm{diag}(\exp(h_{1t}/2), \ldots, \exp(h_{pt}/2)), \tag{8}$$

$$\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} |\mathbf{h}_t \sim \mathcal{N}_{2p}(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{\varepsilon\varepsilon} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\eta\eta} \end{pmatrix}, \tag{9}$$

and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$. Notice that in this version of the model, $\text{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\eta}_t | \mathbf{h}_t) = 0$, which rules out the leverage effect. This assumption can be relaxed as we discuss shortly. For identification purposes, the diagonal elements of $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ in (8) must be one which means that the matrix $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ is a correlation matrix.

Analyzes of this model are given by Harvey, Ruiz, and Shephard (1994), Daníelsson (1998), Smith and Pitts (2006) and Chan, Kohn, and Kirby (2006). Actually, Harvey, Ruiz, and Shephard (1994) dealt with a special case of this model in which $\boldsymbol{\Phi} = \text{diag}(\phi_1, \ldots, \phi_p)$. To estimate the resulting parameters, namely $(\phi_1, \ldots, \phi_p)$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, Harvey, Ruiz, and Shephard (1994) linearized the measurement equation (5) by squaring both sides of the equation and then taking its logarithm. Let $w_{it} = \log y_{it}^2$ and noting that

$$\text{E}(\log \varepsilon_{it}^2) = -1.27, \quad \text{Var}(\log \varepsilon_{it}^2) = \pi^2/2, \tag{10}$$

they obtained a linear state space model in which they assumed a random walk process for $\mathbf{h}_t$,

$$\begin{aligned} \mathbf{w}_t &= (-1.27)\mathbf{1} + \mathbf{h}_t + \boldsymbol{\xi}_t, & (11) \\ \mathbf{h}_{t+1} &= \mathbf{h}_t + \boldsymbol{\eta}_t, & (12) \end{aligned}$$

where $\mathbf{w}_t = (w_{1t}, \ldots, w_{pt})'$, $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{pt})'$, $\xi_{it} = \log \varepsilon_{it}^2 + 1.27$ and $\mathbf{1} = (1, \ldots, 1)'$. Although the new state error $\boldsymbol{\xi}_t$ does not follow a normal distribution, they regarded (11) and (12) as a linear Gaussian state-space model and obtained the corresponding maximum likelihood estimators using the Kalman filter algorithm. Since the likelihood function is misspecified, their method delivers the quasi-maximum likelihood (QML) estimates. Implementation also requires the covariance matrix of $\boldsymbol{\xi}_t$. Harvey, Ruiz, and Shephard (1994) showed that the (i, j)-th element of the covariance matrix of $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{pt})'$ is given by $(\pi^2/2)\rho_{ij}^*$ where $\rho_{ii}^* = 1$ and

$$\rho_{ij}^* = \frac{2}{\pi^2} \sum_{n=1}^{\infty} \frac{(n-1)!}{\{\prod_{k=1}^{n}(x+k-1)\}n} \rho_{ij}^{2n} \tag{13}$$

They applied the model to four daily foreign exchange rates (Pound/Dollar, Deutschemark/Dollar, Yen/Dollar and Swiss Franc/Dollar), and further considered a factor model and t distributions for the measurement error to account for heavy-tailed distributions of foreign exchange rates. As mentioned in Harvey, Ruiz, and Shephard (1994), this method cannot be extended to the leverage model.

So, Li, and Lam (1997) considered a similar transformation to that of Harvey, Ruiz, and Shephard (1994) (using $\tilde{w}_{it} = \log |y_{it} - \overline{y}_i|$ where $\overline{y}_i = \sum_{t=1}^{n} y_{it}/n$), but considered a vector AR(1) process for the latent volatility vector $\mathbf{h}_t$ (*i.e.* the non-diagonal element of $\boldsymbol{\Phi}$ are not set equal to zero). The model is given by

$$\tilde{\mathbf{w}}_t = \mathbf{h}_t + \tilde{\boldsymbol{\xi}}_t, \tag{14}$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \tag{15}$$

where $\tilde{\mathbf{w}}_t = (\tilde{w}_{1t}, \ldots, \tilde{w}_{pt})'$, $\tilde{\boldsymbol{\xi}}_t = (\tilde{\xi}_{1t}, \ldots, \tilde{\xi}_{pt})'$, $\mathrm{E}(\tilde{\xi}_t) = (-0.635)\mathbf{1}$, $\mathbf{1} = (1, \ldots, 1)'$ and $\mathrm{Var}(\tilde{\xi}_t) = (\pi^2/8)\mathbf{I}$. They obtained the QML estimates by a computationally efficient and numerically well-behaved EM algorithm. To describe this method, let $\boldsymbol{\theta} = (\boldsymbol{\mu}', \mathrm{vech}(\boldsymbol{\Sigma}_{\eta\eta})', \mathrm{vec}(\boldsymbol{\Phi})')'$ where vec and vech are respectively the vectorization operator for a matrix $\mathbf{A} = \{a_{ij}\}$ and the half-vectorization operator for a symmetric matrix $\mathbf{B} = \{b_{ij}\}$ such that

$$\begin{aligned} \mathrm{vec}(\mathbf{A}) &= (a_{11}, a_{21}, \ldots, a_{p1}, a_{12}, \ldots, a_{p2}, a_{13}, \ldots, a_{pp})', \\ \mathrm{vech}(\mathbf{B}) &= (b_{11}, b_{21}, \ldots, b_{p1}, b_{22}, \ldots, b_{p2}, b_{33}, \ldots, b_{pp})'. \end{aligned}$$

Let $l(\boldsymbol{\theta})$ denote the logarithm of the complete data likelihood given by

$$\begin{aligned} l(\boldsymbol{\theta}) &= -\frac{1}{2}\sum_{t=1}^{n-1} \{\mathbf{h}_{t+1} - \boldsymbol{\Phi}\mathbf{h}_t - (\mathbf{I} - \boldsymbol{\Phi})\boldsymbol{\mu}\}' \boldsymbol{\Sigma}_{\eta\eta}^{-1} \{\mathbf{h}_{t+1} - \boldsymbol{\Phi}\mathbf{h}_t - (\mathbf{I} - \boldsymbol{\Phi})\boldsymbol{\mu}\} \\ &\quad -\frac{n-1}{2}\log|\boldsymbol{\Sigma}_{\eta\eta}| + c, \end{aligned}$$

where c is a constant which does not depend on $\boldsymbol{\theta}$. Then the conditional expectation of $l(\boldsymbol{\theta})$ given $\mathbf{W}_n = (\tilde{\mathbf{w}}_1, \ldots, \tilde{\mathbf{w}}_n)$ is

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) &= \mathrm{E}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(k)}}(l(\boldsymbol{\theta})|\mathbf{W}_n) \\ &= -\frac{1}{2}\sum_{t=1}^{n-1}\left\{\mathrm{tr}\left(\boldsymbol{\Sigma}_{\eta\eta}^{-1}\mathbf{M}_{t+1}\right) + (\mathbf{m}_{t+1} - \mathbf{d})'\boldsymbol{\Sigma}_{\eta\eta}^{-1}(\mathbf{m}_{t+1} - \mathbf{d}) - \log|\boldsymbol{\Sigma}_{\eta\eta}^{-1}|\right\} + c \end{aligned}$$

where

$$\begin{aligned} \mathbf{d} &= (\mathbf{I} - \boldsymbol{\Phi})\boldsymbol{\mu}, \quad \mathbf{m}_t = \mathbf{a}_t^{(k)} - \boldsymbol{\Phi}\mathbf{a}_{t-1}^{(k)}, \\ \mathbf{M}_t &= \mathbf{B}_t^{(k)} - \mathbf{C}_t^{(k)}\boldsymbol{\Phi}' - \boldsymbol{\Phi}\mathbf{C}_t^{(k)\prime} + \boldsymbol{\Phi}\mathbf{B}_{t-1}^{(k)}\boldsymbol{\Phi}' - \mathbf{m}_t\mathbf{m}_t', \\ \mathbf{a}_t^{(k)} &= \mathrm{E}(\mathbf{h}_t|W_n, \hat{\boldsymbol{\theta}}^{(k)}), \quad \mathbf{B}_t^{(k)} = \mathrm{E}(\mathbf{h}_t\mathbf{h}_t'|W_n, \hat{\boldsymbol{\theta}}^{(k)}), \quad \mathbf{C}_t^{(k)} = \mathrm{E}(\mathbf{h}_t\mathbf{h}_{t-1}'|W_n, \hat{\boldsymbol{\theta}}^{(k)}). \end{aligned}$$

The EM algorithm now proceeds by the recursive implementation of the following two steps.

1. Maximization step. Given $\hat{\boldsymbol{\theta}}^{(k)}$ and $\mathbf{a}_t^{(k)}, \mathbf{B}_t^{(k)}, \mathbf{C}_t^{(k)}$, maximize $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$

and obtain $\hat{\boldsymbol{\theta}}^{(k+1)} = (\hat{\boldsymbol{\mu}}^{(k+1)\prime}, \mathrm{vech}(\hat{\boldsymbol{\Sigma}}_{\eta\eta}^{(k+1)})', \mathrm{vec}(\hat{\boldsymbol{\Phi}}^{(k+1)})')'$ where

$$\begin{aligned}
\hat{\boldsymbol{\mu}}^{(k+1)} &= \frac{1}{n-1}\left(\mathbf{I}-\hat{\boldsymbol{\Phi}}^{(k+1)}\right)^{-1}\sum_{t=1}^{n-1}\left(\mathbf{a}_{t+1}^{(k)}-\hat{\boldsymbol{\Phi}}^{(k+1)}\mathbf{a}_t^{(k)}\right), \\
\hat{\boldsymbol{\Sigma}}_{\eta\eta}^{(k+1)} &= \frac{1}{n-1}\sum_{t=1}^{n-1}\left\{\mathbf{M}_{t+1}+(\mathbf{m}_{t+1}-\hat{\mathbf{d}}^{(k+1)})(\mathbf{m}_{t+1}-\hat{\mathbf{d}}^{(k+1)})'\right\}, \\
\hat{\boldsymbol{\Phi}}^{(k+1)} &= \left\{\left(\sum_{t=1}^{n-1}\mathbf{a}_{t+1}^{(k)}\right)\left(\sum_{t=1}^{n-1}\mathbf{a}_t^{(k)}\right)'-(n-1)\sum_{t=1}^{n-1}\mathbf{C}_{t+1}^{(k)}\right\} \\
&\quad\times\left\{\left(\sum_{t=1}^{n-1}\mathbf{a}_t^{(k)}\right)\left(\sum_{t=1}^{n-1}\mathbf{a}_t^{(k)}\right)'-(n-1)\sum_{t=1}^{n-1}\mathbf{B}_t^{(k)}\right\}^{-1}
\end{aligned}$$

and $\mathbf{M}_{t+1}, \mathbf{m}_{t+1}$ are evaluated at $\boldsymbol{\Phi} = \hat{\boldsymbol{\Phi}}^{(k+1)}$.

2. Expectation step. Compute $\mathbf{a}_t^{(k+1)}, \mathbf{B}_t^{(k+1)}, \mathbf{C}_t^{(k+1)}$ using the augmented state space model

$$\begin{aligned}
\mathbf{y}_t^* &= \boldsymbol{\alpha}_t + \mathbf{u}_t, \quad \mathbf{u}_t = \tilde{\boldsymbol{\xi}}_t + (0.635)\mathbf{1}, \\
\boldsymbol{\alpha}_{t+1} &= \hat{\boldsymbol{\Phi}}^{(k+1)}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t,
\end{aligned}$$

where $\mathbf{y}_t^* = \tilde{\mathbf{w}}_t - \hat{\boldsymbol{\mu}}^{(k+1)} + (0.635)\mathbf{1}$ and $\boldsymbol{\alpha}_t = \mathbf{h}_t - \hat{\boldsymbol{\mu}}^{(k+1)}$. By applying the Kalman filter and smoother algorithm, we obtain the smoothed state $\boldsymbol{\alpha}_{t|n} = \mathrm{E}(\boldsymbol{\alpha}_t|Y_n^*)$, the variance $\mathbf{P}_{t|n} = \mathrm{Var}(\boldsymbol{\alpha}_t|Y_n^*)$, $\mathbf{P}_{t|t} = \mathrm{Var}(\boldsymbol{\alpha}_t|Y_t^*)$, and $\mathbf{P}_{t+1|t} = \mathrm{Var}(\boldsymbol{\alpha}_{t+1}|Y_t^*)$ where $Y_t^* = \{\mathbf{y}_1^*, \ldots, \mathbf{y}_t^*\}$.

Then So, Li, and Lam (1997) showed that

$$\begin{aligned}
\mathbf{a}_t^{(k+1)} &= \boldsymbol{\alpha}_{t|n} + \hat{\boldsymbol{\mu}}^{(k+1)}, \\
\mathbf{B}_t^{(k+1)} &= \mathbf{P}_{t|n} + \mathbf{a}_t^{(k+1)}\mathbf{a}_t^{(k+1)\prime}, \\
\mathbf{C}_t^{(k+1)} &= \mathbf{P}_{t+1|n}\left(\mathbf{P}_{t|t}\hat{\boldsymbol{\Phi}}^{(k+1)\prime}\mathbf{P}_{t+1|t}^{-1}\right)' + \mathbf{a}_t^{(k+1)}\mathbf{a}_{t-1}^{(k+1)\prime}.
\end{aligned}$$

They also derived an asymptotic variance-covariance matrix for the EM estimates based on the information matrix.

Another related contribution is that of Daníelsson (1998) where the model

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon}), \\
\mathbf{h}_{t+1} &= \boldsymbol{\mu} + \mathrm{diag}(\phi_1, \ldots, \phi_p)(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\eta}),
\end{aligned}$$

is analyzed. The parameters of this model are estimated by the simulated maximum likelihood (SML) method. The SML procedure in this work is a multivariate extension of the accelerated importance sampling method proposed by Daníelsson and Richard (1993). The model and fitting method is applied in the estimation of a bivariate model for foreign exchange rates (Deutschemark/Dollar, Yen/Dollar) and stock indices (S&P500 and Tokyo stock exchange). Based on the log-likelihood values they concluded that the MSV model is superior to alternative GARCH models such as the vector GARCH, diagonal vector GARCH (Bollerslev, Engle, and Woodridge (1988)), Baba-Engle-Kraft-Kroner (BEKK) model (Engle and Kroner (1995)) and the constant conditional correlation (CCC) model (Bollerslev (1990)).

Smith and Pitts (2006) considered a bivariate model without leverage that is similar to the model of Daníelsson (1998). The model is given by

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \mathbf{V}_t^{1/2} = \mathrm{diag}(\exp(h_{1t}/2), \exp(h_{2t}/2)), \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon}),\\
\mathbf{h}_{t+1} &= \mathbf{Z}_t\boldsymbol{\alpha} + \mathrm{diag}(\phi_1, \phi_2)(\mathbf{h}_t - \mathbf{Z}_{t-1}\boldsymbol{\alpha}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\eta}),\\
\mathbf{h}_1 &\sim \mathcal{N}_2(\mathbf{Z}_1\boldsymbol{\alpha}_1, \boldsymbol{\Sigma}_0),
\end{aligned}$$

where the (i, j)-th element of $\boldsymbol{\Sigma}_0$ is the (i, j)-th element of $\boldsymbol{\Sigma}_{\eta\eta}$ divided by $1 - \phi_i\phi_j$ to enforce the stationarity of $\mathbf{h}_t - \mathbf{Z}_t\boldsymbol{\alpha}$. To measure the effect on daily returns in the Yen/Dollar foreign exchange of intervention by the Bank of Japan, they included in $\mathbf{Z}_t$ a variable that represents central bank intervention which they modeled by a threshold model. The resulting model was fit by Bayesian Markov chain Monte Carlo (MCMC) methods. To improve the efficiency of the MCMC algorithm, they sampled $\mathbf{h}_t$'s in blocks, as in Shephard and Pitt (1997) (see also Watanabe and Omori (2004)). For simplicity, we describe their algorithm without the threshold specification and without missing observations. Let $Y_t = \{\mathbf{y}_1, \ldots, \mathbf{y}_t\}$ denote the set of observations until time t. Then the Smith and Pitts (2006) MCMC algorithm is given by:

1. Sample $\{\mathbf{h}_t\}_{t=1}^n | \rho_{12}, \phi_1, \phi_2, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\eta\eta}, Y_n$. Divide $\{\mathbf{h}_t\}_{t=1}^n$ in to several blocks, and sample a block at a time given other blocks. Let $\mathbf{h}_{a:b} = (\mathbf{h}_a', \ldots, \mathbf{h}_b')'$ To sample a block $\mathbf{h}_{a:b}$ given other $\mathbf{h}_j$'s, we conduct a M-H algorithm using a proposal density of the type introduced by Chib and Greenberg (1994) and Chib and Greenberg (1998),

$$\mathbf{h}_{a:b} \sim \mathcal{N}_{2(b-a+1)}\left(\hat{\mathbf{h}}_{a:b}, \left[-\frac{\partial l(\mathbf{h}_{a:b})}{\partial \mathbf{h}_{a:b}\partial \mathbf{h}_{a:b}'}\right]^{-1}_{\mathbf{h}_{a:b}=\hat{\mathbf{h}}_{a:b}}\right)$$

where

$$\begin{aligned}
l(\mathbf{h}_{a:b}) &= \text{const} - \frac{1}{2}\sum_{t=a}^{b}\left(\mathbf{1}'\mathbf{h}_t + \mathbf{y}_t'\mathbf{V}_t^{-1/2}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{V}_t^{-1/2}\mathbf{y}_t\right)\\
&\quad -\frac{1}{2}\sum_{t=a}^{b+1}\left\{\mathbf{h}_t - \mathbf{Z}_t\boldsymbol{\alpha} - \boldsymbol{\Phi}(\mathbf{h}_{t-1} - \mathbf{Z}_{t-1}\boldsymbol{\alpha})\right\}'\boldsymbol{\Sigma}_{\eta\eta}^{-1}\left\{\mathbf{h}_t - \mathbf{Z}_t\boldsymbol{\alpha} - \boldsymbol{\Phi}(\mathbf{h}_{t-1} - \mathbf{Z}_{t-1}\boldsymbol{\alpha})\right\}.
\end{aligned}$$

The proposal density is a Gaussian approximation of the conditional posterior density based on a Taylor expansion of the conditional posterior density around the mode $\hat{\mathbf{h}}_{a:b}$. The mode is found numerically by the Newton-Raphson method. The analytical derivatives are given in the Appendix B of Smith and Pitts (2006).

2. Sample $\rho_{12}|\{\mathbf{h}_t\}_{t=1}^n, \phi_1, \phi_2, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\eta\eta}, Y_n$ using the M-H algorithm.

3. Sample $\phi_1, \phi_2|\{\mathbf{h}_t\}_{t=1}^n, \rho_{12}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\eta\eta}, Y_n$ using the M-H algorithm.

4. Sample $\boldsymbol{\alpha}|\{\mathbf{h}_t\}_{t=1}^n, \rho_{12}, \phi_1, \phi_2, \boldsymbol{\Sigma}_{\eta\eta}, Y_n \sim \mathcal{N}_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ where

$$\begin{aligned}
\boldsymbol{\delta} &= \boldsymbol{\Sigma} \sum_{t=2}^{n} (\mathbf{Z}_t - \boldsymbol{\Phi}\mathbf{Z}_{t-1})' \boldsymbol{\Sigma}_{\eta\eta}^{-1} (\mathbf{h}_t - \boldsymbol{\Phi}\mathbf{h}_{t-1}) + \mathbf{Z}_1' \boldsymbol{\Sigma}_0^{-1} \mathbf{h}_1, \\
\boldsymbol{\Sigma}^{-1} &= \sum_{t=2}^{n} (\mathbf{Z}_t - \boldsymbol{\Phi}\mathbf{Z}_{t-1})' \boldsymbol{\Sigma}_{\eta\eta}^{-1} (\mathbf{Z}_t - \boldsymbol{\Phi}\mathbf{Z}_{t-1}) + \mathbf{Z}_1' \boldsymbol{\Sigma}_0^{-1} \mathbf{Z}_1,
\end{aligned}$$

5. Sample $\boldsymbol{\Sigma}_{\eta\eta}|\{\mathbf{h}_t\}_{t=1}^n, \rho_{12}, \phi_1, \phi_2, \boldsymbol{\alpha}, Y_n$ using the M-H algorithm.

Bos and Shephard (2006) considered a similar model but with the mean in the outcome specification driven by an $r \times 1$ latent process vector $\boldsymbol{\alpha}_t$

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{G}_t \mathbf{u}_t, \\
\boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \mathbf{u}_t, \\
\mathbf{u}_t &= \mathbf{V}_t^{1/2} \boldsymbol{\varepsilon}_t, \quad \mathbf{V}_t^{1/2} = \mathrm{diag}(\exp(h_{1t}/2), \ldots, \exp(h_{qt}/2)), \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}), \\
\mathbf{h}_{t+1} &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\eta}), \quad \mathbf{h}_t = (h_{1t}, \ldots, h_{qt})',
\end{aligned}$$

where $\mathbf{G}_t\mathbf{u}_t$ and $\mathbf{H}_t\mathbf{u}_t$ are independent and the off-diagonal element of $\boldsymbol{\Phi}$ may be non-zero. Given $\{\mathbf{h}_t\}_{t=1}^n$, this is a linear Gaussian state space model,

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{u}_t^*, \quad \mathbf{u}_t^* \sim \mathcal{N}_p(\mathbf{0}, \mathbf{G}_t \mathbf{V}_t \mathbf{G}_t'), \\
\boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{v}_t^*, \quad \mathbf{v}_t^* \sim \mathcal{N}_r(\mathbf{0}, \mathbf{H}_t \mathbf{V}_t \mathbf{H}_t'),
\end{aligned}$$

where $\mathbf{u}_t^*$ and $\mathbf{v}_t^*$ are independent. Bos and Shephard (2006) take a Bayesian approach and conduct the MCMC simulation in two blocks. Let $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\lambda})$ where $\boldsymbol{\psi}$ indexes the unknown parameters in $\mathbf{T}_t, \mathbf{Z}_t, \mathbf{G}_t, \mathbf{H}_t$, and $\boldsymbol{\lambda}$ denotes the parameter of the stochastic volatility process of $\mathbf{u}_t$.

1. Sample $\boldsymbol{\theta}, \{\boldsymbol{\alpha}_t\}_{t=1}^n|\{\mathbf{h}_t\}_{t=1}^n, Y_n$.

(a) Sample $\boldsymbol{\theta}|\{\mathbf{h}_t\}_{t=1}^n, Y_n$ using a M-H algorithm or a step from the adaptive rejection Metropolis sampler by Gilks, Best, and Tan (1995) (see Bos and Shephard (2006)).

(b) Sample $\{\boldsymbol{\alpha}_t\}_{t=1}^n|\boldsymbol{\theta}, \{\mathbf{h}_t\}_{t=1}^n, Y_n$ using a simulation smoother for a linear Gaussian state space model (see e.g.de Jong and Shephard (1995), Durbin and Koopman (2002))). We first sample disturbances of the linear Gaussian state space model and obtain samples of $\boldsymbol{\alpha}_t$ recursively.

2. Sample $\{\mathbf{h}_t\}_{t=1}^n|\boldsymbol{\theta}, \{\boldsymbol{\alpha}_t\}_{t=1}^n, Y_n$. For $t = 1, \ldots, n$, we sample $\mathbf{h}_t$ one at a time by the M-H algorithm with the proposal distribution

$$
\begin{aligned}
\mathbf{h}_t|\mathbf{h}_{t-1}, \mathbf{h}_{t+1}, \boldsymbol{\theta} &\sim \mathcal{N}_q(\boldsymbol{\mu} + \mathbf{Q}\boldsymbol{\Phi}'\boldsymbol{\Sigma}_{\eta\eta}^{-1}\left\{(\mathbf{h}_{t+1} - \boldsymbol{\mu}) + (\mathbf{h}_{t-1} - \boldsymbol{\mu})\right\}, \mathbf{Q}), \quad t = 2, \ldots, n-1, \\
\mathbf{h}_n|\mathbf{h}_{n-1}, \boldsymbol{\theta} &\sim \mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\eta\eta}),
\end{aligned}
$$

where $\mathbf{Q}^{-1} = \boldsymbol{\Sigma}_{\eta\eta}^{-1} + \boldsymbol{\Phi}'^{-1}\boldsymbol{\Phi}$.

Although the sampling scheme which samples $\mathbf{h}_t$ at a time is expected to produce highly autocorrelated MCMC samples, the adaptive rejection Metropolis sampling of $\boldsymbol{\theta}$ seems to overcome some of the inefficiencies.

Yu and Meyer (2006) compared several bivariate basic models (without leverage effects) including those in which $\mathbf{h}_t$ follows a VAR(1) process with $\phi_{12} = 0$ to allow Granger causality from one asset to another. Using the popular software WinBUGS, they estimated the model on foreign exchange rate data.

So and Kwok (2006) considered a multivariate stochastic volatility model

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}}), && (16) \\
\mathbf{V}_t^{1/2} &= \mathrm{diag}(\exp(h_{1t}/2), \ldots, \exp(h_{pt}/2)), && (17)
\end{aligned}
$$

where the volatility vector $\mathbf{h}_t - \boldsymbol{\mu}$ follows a vector autoregressive fractionally integrated moving average process, ARFIMA$(p, \mathbf{d}, q)$, such that

$$
\begin{aligned}
\boldsymbol{\Phi}(B)D(B)(\mathbf{h}_{t+1} - \boldsymbol{\mu}) &= \Theta(B)\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta\eta}}), && (18) \\
D(B) &= \mathrm{diag}((1-B)^{d_1}, \ldots, (1-B)^{d_p}), \quad |d_i| < 1/2, && (19) \\
\boldsymbol{\Phi}(B) &= \mathbf{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p, && (20) \\
\Theta(B) &= \mathbf{I} + \boldsymbol{\Theta}_1 B + \cdots + \boldsymbol{\Theta}_q B^q, && (21)
\end{aligned}
$$

where B is a backward operator such that $B^j\mathbf{h}_t = \mathbf{h}_{t-j}$. The $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are assumed to be independent. So and Kwok (2006) investigated statistical properties of the model and proposed a QML estimation method as in Harvey, Ruiz, and Shephard (1994). They linearized the measurement equation by taking the logarithm of the squared returns and considered the linear state space model

$$\begin{aligned}\mathbf{w}_t &= (-1.27)\mathbf{1} + \mathbf{h}_t + \boldsymbol{\xi}_t, \\ \boldsymbol{\Phi}(B)D(B)(\mathbf{h}_{t+1} - \boldsymbol{\mu}) &= \Theta(B)\boldsymbol{\eta}_t,\end{aligned}$$

where $\mathbf{w}_t = (w_{1t}, \ldots, w_{pt})'$, $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{pt})'$, $w_{it} = \log y_{it}^2$, and $\xi_{it} = \log \varepsilon_{it}^2$ for $i = 1, \ldots, n$. The covariance matrix of $\boldsymbol{\xi}_t$ can be obtained as in Harvey, Ruiz, and Shephard (1994). To conduct the QML estimation, So and Kwok (2006) assumed that $\boldsymbol{\xi}_t$ follows a normal distribution and obtained estimates based on the linear Gaussian state space model. However, since $\mathbf{h}_t - \boldsymbol{\mu}$ follows a vector ARFIMA$(p, \mathbf{d}, q)$ process, the conventional Kalman filter is not applicable as the determinant and inverse of large covariance matrix is required to calculate the quasi-log-likelihood function. To avoid this calculation, So and Kwok (2006) approximated the quasi-log-likelihood function by using a spectral likelihood function based on a Fourier transform.

2.2 Leverage effects

We now discuss models in which the basic MSV model is defined to include leverage effects through correlation between $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ (equivalently, $\boldsymbol{\Sigma}_{\varepsilon\eta} \neq \mathbf{O}$). In the analysis of stock returns using univariate stochastic volatility models, there is strong evidence that the leverage effect is an important feature of the data (e.g. Yu (2005), Omori, Chib, Shephard, and Nakajima (2007)). Daníelsson (1998) first mentioned leverage effects in MSV models but the model is not estimated in his empirical study of foreign exchange rates and stock indices. We now follow Chan, Kohn, and Kirby (2006) who considered the model

$$\begin{aligned}\mathbf{y}_t &= \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \\ \mathbf{h}_{t+1} &= \boldsymbol{\mu} + \mathrm{diag}(\phi_1, \ldots, \phi_p)(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\Psi}^{1/2}\boldsymbol{\eta}_t, \\ & \quad \mathbf{h}_1 \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Psi}^{1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Psi}^{1/2}),\end{aligned}$$

where the (i, j) element of $\boldsymbol{\Sigma}_0$ is the (i, j) element of $\boldsymbol{\Sigma}_{\eta\eta}$ divided by $1 - \phi_i\phi_j$ satisfying a stationarity condition such that

$$\boldsymbol{\Sigma}_0 = \boldsymbol{\Phi}\boldsymbol{\Sigma}_0\boldsymbol{\Phi} + \boldsymbol{\Sigma}_{\eta\eta}$$

and

$$\begin{aligned}
\mathbf{V}_t^{1/2} &= \mathrm{diag}\left(\exp(h_{1t}/2), \ldots, \exp(h_{pt}/2)\right), \\
\mathbf{\Psi}^{1/2} &= \mathrm{diag}\left(\sqrt{\psi_1^2}, \ldots, \sqrt{\psi_p^2}\right), \\
\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} &\sim \mathcal{N}_{2p}(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{\varepsilon\varepsilon} & \mathbf{\Sigma}_{\varepsilon\eta} \\ \mathbf{\Sigma}_{\eta\varepsilon} & \mathbf{\Sigma}_{\eta\eta} \end{pmatrix}.
\end{aligned}$$

Actually, the model considered in Chan, Kohn, and Kirby (2006) had correlation between $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_{t-1}$ which is not correctly a model of leverage. Our discussion therefore modifies their treatment to deal with the model just presented, where $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are correlated. Note that $\mathbf{\Sigma}$ is a $2p \times 2p$ correlation matrix with $\mathbf{\Sigma}_{\varepsilon\eta} \neq \mathbf{O}$. Now, following Wong, Carter, and Kohn (2003) and Pitt, Chan, and Kohn (2006), reparameterize $\mathbf{\Sigma}$ such that

$$\mathbf{\Sigma}^{-1} = \mathbf{TGT}, \quad \mathbf{T} = \mathrm{diag}\left(\sqrt{G^{11}}, \ldots, \sqrt{G^{pp}}\right),$$

where $\mathbf{G}$ is a correlation matrix and G^{ii} denotes the (i,i)-th element of the inverse matrix of $\mathbf{G}$. Under this parameterization, we can find the posterior probability that the strict lower triangle of the transformed correlation matrix $\mathbf{G}$ is equal to zero. Let $J_{ij} = 1$ if $G_{ij} \neq 0$ and $J_{ij} = 0$ if $G_{ij} = 0$ for $i = 1, \ldots, 2p, j < i$ and $S(\mathbf{J})$ denote the number of elements in $\mathbf{J} = \{J_{ij}, i = 1, \ldots, 2p,\ j < i\}$. Further let $\mathbf{G}_{\{J=k\}} = \{G_{ij} : J_{ij} = k \in \mathbf{J}\}$ $(k = 0, 1)$ and $\mathcal{A}$ denote a class of $2p \times 2p$ correlation matrices. Wong, Carter, and Kohn (2003) proposed a hierarchical prior for $\mathbf{G}$

$$\begin{aligned}
\pi(d\mathbf{G}|\mathbf{J}) &= V(\mathbf{J})^{-1} d\mathbf{G}_{\{J=1\}} I(\mathbf{G} \in \mathcal{A}), \quad V(\mathbf{J}) = \int_{\mathbf{G}\in\mathcal{A}} d\mathbf{G}_{\{J=1\}}, \\
\pi(\mathbf{J}|S(\mathbf{J}) = l) &= \frac{V(\mathbf{J})}{\displaystyle\sum_{\mathbf{J}^*: S(\mathbf{J}^*) = l} V(\mathbf{J}^*)}, \\
\pi(S(\mathbf{J}) = l|\varphi) &= \begin{pmatrix} p(2p-1) \\ l \end{pmatrix} \varphi^l (1-\varphi)^{p(2p-1)-l}.
\end{aligned}$$

If we assume $\varphi \sim \mathcal{U}(0,1)$, the marginal prior probability $\pi(S(\mathbf{J}) = l) = 1/(p(2p-1)+1)$ (see Wong, Carter, and Kohn (2003) for the evaluation of $V(\mathbf{J})$). Let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_p)'$ $(\psi_j > 0, j = 1, \ldots, p)$.

1. Sample $\boldsymbol{\phi}|\boldsymbol{\mu}, \{\mathbf{h}_t\}_{t=1}^n, \boldsymbol{\psi}, \mathbf{\Sigma}, Y_n$ where $Y_n = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$. Let $\mathbf{\Sigma}^{ij}$ denote the $2p \times 2p$ (i,j)-th block matrix of the $\mathbf{\Sigma}^{-1}$ and $\mathbf{d}$ be a vector consists of the diagonal elements

$$\sum_{t=1}^{n-1} \mathbf{\Psi}^{-1/2}(\mathbf{h}_t - \boldsymbol{\mu}) \left(\mathbf{y}_t' \mathbf{V}_t^{-1/2} \mathbf{\Sigma}^{12} + (\mathbf{h}_{t+1} - \boldsymbol{\mu})' \mathbf{\Psi}^{-1/2} \mathbf{\Sigma}^{22}\right).$$

Propose a candidate

$$
\begin{aligned}
\boldsymbol{\phi} &\sim \mathcal{TN}_R(\boldsymbol{\mu}_{\boldsymbol{\phi}}, \boldsymbol{\Sigma}_{\boldsymbol{\phi}}), \quad R = \{\boldsymbol{\phi} : \phi_j \in (-1,1), j = 1, \ldots, p\}, \\
\boldsymbol{\Sigma}_{\boldsymbol{\phi}}^{-1} &= \boldsymbol{\Sigma}^{22} \odot \left\{ \sum_{t=1}^{n-1} (\mathbf{h}_t - \boldsymbol{\mu})'^{-1} (\mathbf{h}_t - \boldsymbol{\mu}) \right\}, \\
\boldsymbol{\mu}_{\boldsymbol{\phi}} &= \boldsymbol{\Sigma}_{\boldsymbol{\phi}} \mathbf{d},
\end{aligned}
$$

where $\odot$ is the element-by-element multiplication operator (Hadamard product) and apply the M-H algorithm.

2. Sample $\boldsymbol{\mu}|\boldsymbol{\phi}, \{\mathbf{h}_t\}_{t=1}^n, \boldsymbol{\psi}, \boldsymbol{\Sigma}, Y_n \sim \mathcal{N}_p(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*)$ where

$$
\begin{aligned}
\boldsymbol{\Sigma}_*^{-1} &= (n-1)(\mathbf{I} - \boldsymbol{\Phi})\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Sigma}^{22}\boldsymbol{\Psi}^{-1/2}(\mathbf{I} - \boldsymbol{\Phi}) + \boldsymbol{\Psi}^{-1/2}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}^{-1/2}, \\
\boldsymbol{\mu}_* &= \boldsymbol{\Sigma}_* \left[(\mathbf{I} - \boldsymbol{\Phi})\boldsymbol{\Psi}^{-1/2} \sum_{t=1}^{n-1} \left\{ \boldsymbol{\Sigma}^{21}\mathbf{V}_t^{-1/2}\mathbf{y}_t + \boldsymbol{\Sigma}^{22}\boldsymbol{\Psi}^{-1/2}(\mathbf{h}_{t+1} - \boldsymbol{\Phi}\mathbf{h}_t) \right\} \right. \\
&\quad \left. + \boldsymbol{\Psi}^{-1/2}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}^{-1/2}\mathbf{h}_1 \right].
\end{aligned}
$$

3. Sample $\boldsymbol{\psi}|\boldsymbol{\phi}, \boldsymbol{\mu}, \{\mathbf{h}_t\}_{t=1}^n, \boldsymbol{\Sigma}, Y_n$. Let $\mathbf{v} = (\psi_1^{-1}, \ldots, \psi_p^{-1})$ and $l(\mathbf{v})$ denote the logarithm of the conditional probability density of $\mathbf{v}$ and $\hat{\mathbf{v}}$ denote the mode of $l(\mathbf{v})$. Then conduct M-H algorithm using a truncated multivariate t-distribution on the region $R = \{\mathbf{v} : v_j > 0, j = 1, \ldots, p\}$ with 6 degrees of freedom, location parameter $\hat{\mathbf{v}}$ and a covariance matrix $-\{\partial^2 l(\mathbf{v})/\partial\mathbf{v}\partial\mathbf{v}'\}_{\mathbf{v}=\hat{\mathbf{v}}}^{-1}$.

4. Sample $\{\mathbf{h}_t\}_{t=1}^n|\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\Sigma}, Y_n$. We divide $\{\mathbf{h}_t\}_{t=1}^n$ in to several blocks, and sample a block at a time given other blocks as in Smith and Pitts (2006). Let $\mathbf{h}_{a:b} = (\mathbf{h}_a', \ldots, \mathbf{h}_b')'$ To sample a block $\mathbf{h}_{a:b}$ given other $\mathbf{h}_j$'s, we conduct a M-H algorithm using a Chib and Greenberg (1994) proposal,

$$
\begin{aligned}
\mathbf{h}_{a:b} &\sim \mathcal{N}_{p(b-a+1)} \left(\hat{\mathbf{h}}_{a:b}, \left[-\frac{\partial l(\mathbf{h}_{a:b})}{\partial \mathbf{h}_{a:b} \partial \mathbf{h}_{a:b}'} \right]_{\mathbf{h}_{a:b} = \hat{\mathbf{h}}_{a:b}}^{-1} \right) \\
l(\mathbf{h}_{a:b}) &= \text{const} - \frac{1}{2} \sum_{t=a}^{b} \mathbf{1}'\mathbf{h}_t - \frac{1}{2} \sum_{t=a}^{b+1} \mathbf{r}_t'^{-1}\mathbf{r}_t \\
\mathbf{r}_t &= \begin{pmatrix} \mathbf{V}_t^{-1/2}\mathbf{y}_t \\ \boldsymbol{\Psi}^{-1/2}\{\mathbf{h}_{t+1} - \boldsymbol{\mu} - \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu})\} \end{pmatrix}
\end{aligned}
$$

a Gaussian approximation of the conditional posterior density based on Taylor expansion of the conditional posterior density around the mode $\hat{\mathbf{h}}_{a:b}$. The mode is found using Newton-Raphson method numerically. The analytical derivatives can be derived similarly as in the Appendix of Chan, Kohn, and Kirby (2006).

5. Sample $\boldsymbol{\Sigma}|\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\psi}, \{\mathbf{h}_t\}_{t=1}^n, Y_n$. Using the parsimonious reparameterization proposed in Wong, Carter, and Kohn (2003), each element G_{ij} is generated one at a time using the M-H algorithm.

Chan, Kohn, and Kirby (2006) applied the proposed estimation method to equities at three levels of aggregation: (i) returns for eight different markets (portfolios of stocks in NYSE, AMEX, NASDAQ and S&P500 index), (ii) returns for eight different industries (portfolios of eight well-known and actively traded stocks in petroleum, food products, pharmaceutical, banks, industrial equipment, aerospace, electric utilities, and department/discount stores) (iii) returns for individual firms within the same industry. They found strong evidence of correlation between $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_{t-1}$ only for the returns of the eight different markets and suggested that this correlation is mainly a feature of market-wide rather than firm-specific returns and volatility.

Asai and McAleer (2006) also analyzed a MSV model with leverage effects letting

$$\begin{aligned}
\boldsymbol{\Phi} &= \mathrm{diag}(\phi_1, \ldots, \phi_p), \\
\boldsymbol{\Sigma}_{\varepsilon\eta} &= \mathrm{diag}\left(\lambda_1\sigma_{1,\eta\eta}, \ldots, \lambda_p\sigma_{p,\eta\eta}\right), \\
\boldsymbol{\Sigma}_{\eta\eta} &= \mathrm{diag}(\sigma_{1,\eta\eta}^2, \ldots, \sigma_{p,\eta\eta}^2),
\end{aligned}$$

The cross asset leverage effects are assumed to be 0 ($\mathrm{Corr}(\varepsilon_{it}, \eta_{jt}) = 0$, for $i \neq j$). As in Harvey and Shephard (1996), they linearized the measurement equations and considered the following state space model conditional on $\mathbf{s}_t = (s_{1t}, \ldots, s_{pt})'$ where $s_{it} = 1$ if y_{it} is positive and $s_{it} = -1$ otherwise:

$$\begin{aligned}
\log y_{it}^2 &= h_{it} + \zeta_{it}, \quad \zeta_{it} = \log \varepsilon_{it}^2, \quad i = 1, \ldots, p, \quad t = 1, \ldots, n, \\
\mathbf{h}_{t+1} &= \tilde{\boldsymbol{\mu}} + \boldsymbol{\mu}_t^* + \mathrm{diag}(\phi_1, \ldots, \phi_p)\mathbf{h}_t + \boldsymbol{\eta}_t^*, \\
\boldsymbol{\mu}_t^* &= \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}_{\varepsilon\eta}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{s}_t, \quad \boldsymbol{\eta}_t^* \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\eta_t^*\eta_t^*}),
\end{aligned}$$

where $\mathrm{E}(\zeta_{it}) = -1.27$, and $\mathrm{Cov}(\zeta_{it}, \zeta_{jt}) = (\pi^2/2)\rho_{ij}^*$ given in (13). The matrix $\boldsymbol{\Sigma}_{\eta_t^*\eta_t^*}$ and $\mathrm{E}(\boldsymbol{\eta}_t^*\boldsymbol{\zeta}_t')$ are given in Asai and McAleer (2006). They also considered an alternative MSV model with leverage effects and size effects given by

$$\begin{aligned}
\mathbf{h}_{t+1} &= \tilde{\boldsymbol{\mu}} + \boldsymbol{\Gamma}_1\mathbf{y}_t + \boldsymbol{\Gamma}_2|\mathbf{y}_t| + \boldsymbol{\Phi}\mathbf{h}_t + \boldsymbol{\eta}_t, \\
\boldsymbol{\Gamma}_1 &= \mathrm{diag}(\gamma_{11}, \ldots, \gamma_{1p}), \quad \boldsymbol{\Gamma}_2 = \mathrm{diag}(\gamma_{21}, \ldots, \gamma_{2p}), \\
|\mathbf{y}_t| &= (|y_{1t}|, \ldots, |y_{pt}|)', \qquad \boldsymbol{\Phi} = \mathrm{diag}(\phi_1, \ldots, \phi_p), \\
\boldsymbol{\Sigma}_{\varepsilon\eta} &= \mathbf{O}, \quad \boldsymbol{\Sigma}_{\eta\eta} = \mathrm{diag}(\sigma_{1,\eta\eta}^2, \ldots, \sigma_{p,\eta\eta}^2)
\end{aligned}$$

This model is a generalization of a univariate model given by Danı́elsson (1994). It incorporates both leverage effects and the magnitude of the previous returns through their absolute values. Asai and McAleer (2006) fit these two models to returns of three stock indices - S&P500 Composite Index, the Nikkei 225 Index, and the Hang Seng Index - by an importance sampling Monte Carlo maximum likelihood estimation method. They find that the MSV model with leverage and size effects is preferred in terms of the AIC and BIC measures.

2.3 Heavy-tailed measurement error models

It has by now quite well established that the tails of the distribution of asset returns are heavier than those of the Gaussian. To deal with this situation it has been popular to employ the Student t distribution as a replacement for the default Gaussian assumption. One reason for the popularity of the Student t distribution is that it has a simple hierarchical form as a scale mixture of normals. Specifically, if T is distributed as standard Student t with ν degrees of freedom then T can be expressed as

$$T = \lambda^{-1/2} Z, \quad Z \sim \mathcal{N}(0,1), \quad \lambda \sim \mathcal{G}(\nu/2, \nu/2).$$

This representation can be exploited in the fitting, especially in the Bayesian context. One early example of the use of the Student t distribution occurs in Harvey, Ruiz, and Shephard (1994) who assumed that in connection with the measurement error ε_{it} that

$$\epsilon_{it} = \lambda_{it}^{-1/2} \varepsilon_{it}, \quad \boldsymbol{\varepsilon}_t \sim \ i.i.d.\ \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon}), \quad \lambda_{it} \sim \ i.i.d.\ \mathcal{G}(\nu_i/2, \nu_i/2),$$

where the mean is $\mathbf{0}$ and the elements of the covariance matrix are given by

$$\begin{aligned} \mathrm{Cov}(\epsilon_{it}, \epsilon_{jt}) \quad &= \quad \begin{cases} \dfrac{\nu_i}{\nu_i - 2}, & i = j, \\ \mathrm{E}(\lambda_{it}^{-1/2})\mathrm{E}(\lambda_{jt}^{-1/2})\rho_{ij}, & i \neq j, \end{cases} \\ \text{and} \quad \mathrm{E}(\lambda_{it}^{-1/2}) &= \frac{(\nu_i/2)^{1/2}\Gamma((\nu_i - 1)/2)}{\Gamma(\nu_i/2)}. \end{aligned}$$

Alternatively, the model can now be expressed as

$$\mathbf{y}_t = \mathbf{V}_t^{1/2} \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\Lambda}_t^{-1/2} = \mathrm{diag}\left(1/\sqrt{\lambda_{1t}}, \ldots, 1/\sqrt{\lambda_{pt}}\right)$$

Taking the logarithm of squared ϵ_{it} one gets

$$\log \epsilon_{it}^2 = \log \varepsilon_{it}^2 - \log \lambda_{it}.$$

They derived the QML estimators using the a mean and covariance matrix of $(\log \varepsilon_{it}^2, \log \varepsilon_{jt}^2)$ using

$$\mathrm{E}(\log \lambda_{it}) = \psi'(\nu/2) - \log(\nu/2), \quad \mathrm{Var}(\log \lambda_{it}) = \psi'(\nu/2),$$

and (10) (13) where ψ and ψ' are the digamma and trigamma functions. On the other hand, Yu and Meyer (2006) considered a multivariate Student t distribution for $\boldsymbol{\varepsilon}_t$ in which case the measurement error has the form

$$\mathbf{T} = \lambda_t^{-1/2}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}), \quad \lambda_t \sim \mathcal{G}(\nu/2, \nu/2).$$

They mentioned that this formulation was empirically better supported than the formulation in Harvey, Ruiz, and Shephard (1994). The model was fit by Bayesian Markov chain Monte Carlo methods.

Another alternative to the Gaussian distribution is the generalized hyperbolic distribution (GH) introduced by Barndorff-Neilsen (1977). This family is also a member of the scale mixture of normals family of distributions. In this case, the mixing distribution is a generalized inverse Gaussian distribution. The generalized hyperbolic distribution is a rich class of distributions that includes the normal, normal inverse Gaussian, reciprocal normal inverse Gaussian, hyperbolic, skewed Student's t, Laplace, normal gamma, and reciprocal normal hyperbolic distributions (e.g. Barndorff-Neilsen and Shephard (2001)). Aas and Haff (2006) have employed the univariate GH distributions (normal inverse Gaussian distributions and univariate GH skew Student's t distributions) and estimated in the analysis of the total index of Norwegian stocks (TOTX), the SSBWG hedged bond index for international bonds, the NOK/EUR exchange rate (NOK is Norwegian kroner), and the EURIBOR 5-year interest rate. They found that the GH skew Student's t distribution is superior to the normal inverse Gaussian distribution for heavy-tailed data, and superior to the skewed t distribution proposed by Azzalini and Capitanio (2003) for very skewed data.

The random variable $\mathbf{x} \sim \mathcal{GH}(\nu, \alpha, \boldsymbol{\beta}, \mathbf{m}, \delta, \mathbf{S})$ follows a multivariate generalized hyperbolic distribution with density

$$\begin{aligned} f(\mathbf{x}) &= \frac{(\gamma/\delta)^{\nu} K_{\nu-\frac{p}{2}}\left(\alpha\sqrt{\delta^2 + (\mathbf{x}-\mathbf{m})'\mathbf{S}^{-1}(\mathbf{x}-\mathbf{m})}\right)\exp\{\boldsymbol{\beta}'(\mathbf{x}-\mathbf{m})\}}{(2\pi)^{\frac{p}{2}} K_{\nu}(\delta\gamma)\left\{\alpha^{-1}\sqrt{\delta^2 + (\mathbf{x}-\mathbf{m})'\mathbf{S}^{-1}(\mathbf{x}-\mathbf{m})}\right\}^{\frac{p}{2}-\nu}}, \qquad (22) \\ & \gamma \equiv \sqrt{\alpha^2 - \boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta}} \geq 0, \quad \alpha^2 \geq \boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta}, \\ & \nu, \alpha \in R, \quad \boldsymbol{\beta}, \mathbf{m} \in R^n, \quad \delta > 0, \end{aligned}$$

where K_ν is a modified Bessel function of the third kind, and $\mathbf{S}$ is a $p \times p$ positive-definite matrix with determinant $|\mathbf{S}| = 1$ (see e.g. Blæsild (1981), Protassov (2004), Schmidt, Hrycej, and Stützle (2006)). It can be shown that $\mathbf{x}$ can be expressed as

$$\mathbf{x} = \mathbf{m} + z_t\mathbf{S}\boldsymbol{\beta} + \sqrt{z_t}\mathbf{S}^{1/2}\boldsymbol{\varepsilon}_t,$$

where $\mathbf{S}^{1/2}$ is a $p \times p$ matrix such that $\mathbf{S} = \mathbf{S}^{1/2}\mathbf{S}^{1/2'}$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$ and $z_t \sim \mathcal{GIG}(\nu, \delta, \gamma)$ follows a generalized inverse Gaussian distribution which we denote $z \sim \mathcal{GIG}(\nu, \delta, \gamma)$ whose density is given by

$$f(z) = \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)} z^{\nu-1} \exp\left\{-\frac{1}{2}\left(\delta^2 z^{-1} + \gamma^2 z\right)\right\}, \quad \gamma, \delta \geq 0, \quad \nu \in R, \quad z > 0,$$

where the range of the parameters given by

$$\begin{array}{lll} \delta > 0, & \gamma^2 \geq 0, & \text{if } \nu < 0, \\ \delta > 0, & \gamma^2 > 0, & \text{if } \nu = 0, \\ \delta \geq 0, & \gamma^2 > 0, & \text{if } \nu > 0, \end{array}$$

(for a generation of a random sample from $\mathcal{GIG}(\nu, a, b)$, see e.g. Dagpunar (1989), Doornik (2002) and Hörmann, Leydold, and Derflinger (2004)). The estimation of such a multivariate distribution would be difficult and Protassov (2004) relied on the EM algorithm with λ fixed and fitted the five dimensional normal inverse Gaussian distribution to a series of returns on foreign exchange rates (Swiss franc, Deutschemark, British pound, Canadian dollar, and Japanese yen). Schmidt, Hrycej, and Stützle (2006) proposed an alternative class of distributions, called the multivariate affine generalized hyperbolic class, and applied it to bivariate models for various asset returns data (Dax, Cac, Nikkei and Dow returns). Other multivariate skew densities have also been proposed for example in Arellano-Valle and Azzalini (2006), Bauwens and Laurent (2005), Dey and Liu (2005) Azzalini (2005), Gupta, González-Farías, and Domínguez-Molina (2004), and Ferreira and Steel (2004).

3 Factor MSV model

3.1 Volatility factor model

A simple factor SV model is considered by Quintana and West (1987), and Jungbacker and Koopman (2006) who utilize a single factor to decompose the outcome into two multiplicative components, a scalar common volatility factor and a vector of idiosyncratic noise variables, as

$$\begin{aligned} \mathbf{y}_t &= \exp\left(\frac{h_t}{2}\right)\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon}), \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2), \end{aligned}$$

where h_t is a scalar. The first element in $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ is assumed to be one for identification reasons. By construction, the positivity of the variance of $\mathbf{y}_t$ is ensured. In comparison with the basic MSV model, this model has fewer parameters, which makes it more convenient to fit. The downside of the model, however, is that unlike the mean factor MSV model which we discuss below, the conditional correlations in this model are time-invariant. Moreover, the correlation between in log-volatilities is 1, which is clearly limiting.

In order to estimate the model, Jungbacker and Koopman (2006) applied a Monte Carlo likelihood method to fit data on exchange rate returns of the British pound, the Deutschemark, and the Japanese yen against the U.S. dollar. They found that the estimate of ϕ is atypically low, indicating that the model is inappropriate for explaining the movements of multivariate volatility.

A more general version of this type is considered by Harvey, Ruiz, and Shephard (1994) who introduced a common factor in the linearized state space version of the basic MSV model by letting

$$\begin{aligned} \mathbf{w}_t &= (-1.27)\mathbf{1} + \mathbf{\Theta}\mathbf{h}_t + \overline{\mathbf{h}} + \boldsymbol{\xi}_t, && (23) \\ \mathbf{h}_{t+1} &= \mathbf{h}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}), && (24) \end{aligned}$$

where $\mathbf{w}_t = (w_{1t}, \ldots, w_{pt})'$, $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{pt})'$ and $\mathbf{h}_t = (h_{1t}, \ldots, h_{qt})'$ $(q \leq p)$. Furthermore,

$$\mathbf{\Theta} = \begin{pmatrix} \theta_{11} & 0 & \cdots & 0 \\ \theta_{21} & \theta_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \theta_{q1} & \cdots & \theta_{q,q-1} & \theta_{qq} \\ \vdots & & \vdots & \vdots \\ \theta_{p,1} & \cdots & \theta_{p,q-1} & \theta_{p,q} \end{pmatrix}, \quad \overline{\mathbf{h}} = \begin{pmatrix} \mathbf{0} \\ \overline{h}_{q+1} \\ \vdots \\ \overline{h}_p \end{pmatrix}.$$

The parameters are estimated by the QML method. To interpret factor loadings, they considered a rotation of the common factors such that $\mathbf{\Theta}^* = \mathbf{\Theta}\mathbf{R}'$ and $\mathbf{h}_t^* = \mathbf{R}\mathbf{h}_t$ where $\mathbf{R}$ is an orthogonal matrix. Harvey, Ruiz, and Shephard (1994) applied it to four daily foreign exchange rates in a model with $q = 2$ factors.

Tims and Mahieu (2006) considered a similar but simpler model for the logarithm of the range of the exchange rate. The daily high and low values were computed over a 24-period for all possible six combinations of four currencies (the US dollar, the UK sterling, the Japanese yen, the euro). Let w_{ij} denote a logarithm of the range of foreign exchange rate of the currency i relative to the currency j, and $\mathbf{w} = (w_{12}, w_{13}, w_{14}, w_{23}, w_{24}, w_{34})$. They assumed that

$$\begin{aligned} \mathbf{w}_t &= \mathbf{c} + \mathbf{Z}\mathbf{h}_t + \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}_{\xi\xi}), \\ \mathbf{h}_{t+1} &= \mathrm{diag}(\phi_1, \ldots, \phi_q)\mathbf{h}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{\Sigma}_{\eta\eta}), \end{aligned}$$

where $\mathbf{c}$ is a 6×1 mean vector, $\mathbf{\Sigma}_{\eta\eta}$ is diagonal, $\mathbf{h}_t = (h_{1t}, \ldots, h_{4t})'$ and h_{jt} is a latent factor for the j-th currency at time t and

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Since this is a linear Gaussian state space model, the estimation of the parameters is straightforward using the Kalman filter.

Ray and Tsay (2000) introduced long range dependence into the volatility factor model using a fractionally integrated process $\mathbf{h}_t$ such that

$$\begin{aligned} \mathbf{y}_t &= \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \mathbf{V}_t^{1/2} = \mathrm{diag}(\exp(\mathbf{z}_1'\mathbf{h}_t/2), \ldots, \exp(\mathbf{z}_q'\mathbf{h}_t/2)), \\ (1-L)^d\mathbf{h}_t &= \boldsymbol{\eta}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon}), \quad \boldsymbol{\eta}_t \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\eta}), \end{aligned}$$

where $\mathbf{z}_i$ $(i = 1, \ldots, q)$ are $q \times 1$ vectors with $q < p$. By taking a logarithm of y_{it}^2 as in Harvey, Ruiz, and Shephard (1994), they considered

$$\mathbf{w}_t = (-1.27)\mathbf{1} + \mathbf{Z}\mathbf{h}_t + \boldsymbol{\xi}_t,$$

where $\mathbf{w}_t = (w_{1t}, \ldots, w_{pt})'$ $(w_{it} = \log y_{it}^2)$, $\mathbf{Z}' = (\mathbf{z}_1, \ldots, \mathbf{z}_q)$, $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{qt})'$ $(\xi_{it} = \log \varepsilon_{it}^2 + 1.27)$. They applied the test statistic proposed by Ray and Tsay (1997) to data on stock returns for groups of companies, randomly selected from those in the S&P 500 index, and found strong evidence in support of common persistence in volatility.

Calvet, Fisher, and Thompson (2006) generalize the univariate Markov-switching multifractal (MSM) model proposed by Calvet and Fisher (2001) to the multivariate MSM and factor MSM models. The univariate model is given by

$$y_t = (M_{1,t}M_{2,t}\cdots M_{k,t})^{1/2}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2),$$

where $M_{j,t}$ $(j \leq k)$ are random volatility components, satisfying $\mathrm{E}(M_{j,t}) = 1$. Given $M_t = (M_{1,t}, M_{2,t}, \ldots, M_{k,t})$, the stochastic volatility of return y_t is given by $\sigma^2 M_{1,t}M_{2,t}\cdots M_{k,t}$. Each $M_{j,t}$ follows a hidden Markov chain as follows;

$$\begin{aligned} &M_{j,t} \text{ drawn from distribution } M, && \text{with probability } \gamma_j, \\ &M_{j,t} = M_{j,t-1}, && \text{with probability } 1-\gamma_j, \end{aligned}$$

where $\gamma_j = 1 - (1-\gamma)^{(b^{j-k})}$, $(0 < \gamma < 1, b > 1)$ and the distribution of M is binomial giving values m or $2-m$ $(m \in [1, 2])$ with equal probability. Thus the MSM model is governed by four parameters (m, σ, b, γ), which is estimated by the maximum likelihood method.

For the bivariate MSM model, we consider the vector of random volatility component $\mathbf{M}_{j,t} = (M^1_{j,t}, M^2_{j,t})'$ $(j \leq k)$. Then, the bivariate model is given by

$$\mathbf{y}_t = (\mathbf{M}_{1,t} \odot \mathbf{M}_{2,t} \odot \cdots \odot \mathbf{M}_{k,t})^{1/2} \odot \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_2(\mathbf{0}, V),$$

where $\odot$ denotes the element-by-element product. For each component $\mathbf{M}_{j,t}$ in the bivariate model, Calvet, Fisher, and Thompson (2006) assume that volatility arrivals are correlated but not necessarily simultaneous. For details, let $s^i_{j,t}$ $(i = 1, 2)$ denote the random variable equal to 1 if there is an arrival on $M^i_{j,t}$ with probability γ_j, and equal to 0 otherwise. Thus, each $s^i_{j,t}$ follows the Bernoulli distribution. At this stage, Calvet, Fisher, and Thompson (2006) introduced the correlation coefficient λ, giving the conditional probability $P(s^2_{j,t} = 1 | s^1_{j,t} = 1) = (1 - \lambda)\gamma_j + \lambda$. They showed that arrivals are independent if $\lambda = 0$, and simultaneous if $\lambda = 1$. Given the realization of the arrival vector $s^1_{j,t}$ and $s^2_{j,t}$, the construction of the volatility components $\mathbf{M}_{j,t}$ is based on a bivariate distribution $\mathbf{M} = (M_1, M_2)$. If arrivals hit both series $(s^1_{j,t} = s^2_{j,t} = 1)$, the state vector $\mathbf{M}_{j,t}$ is drawn from $\mathbf{M}$. If only one series i $(i = 1, 2)$ receives an arrival, the new component $M^i_{j,t}$ is sampled from the marginal M^i of the bivariate distribution $\mathbf{M}$. Finally, $\mathbf{M}_{j,t} = \mathbf{M}_{j,t-1}$ if there is no arrival $(s^1_{j,t} = s^2_{j,t} = 0)$. They assume that $\mathbf{M}$ has a bivariate binomial distribution controlled by m^1 and m^2, in parallel fashion to the univariate case. Again, the closed form solution of the likelihood function is available. This approach can be extended to a general multivariate case. As the number of parameter therefore grows at least as fast as a quadratic function of p, Calvet, Fisher, and Thompson (2006) proposed not only the multivariate MSM model but also the factor MSM model.

The factor MSM model based on q volatility factors $\mathbf{f}^l_t = (f^l_{1,t}, \ldots, f^l_{k,t})'$, $(f^l_{j,t} > 0)$ $(l = 1, 2, \ldots, q)$ is given by

$$\begin{aligned}
\mathbf{y}_t &= (\mathbf{M}_{1,t} \odot \mathbf{M}_{2,t} \odot \cdots \odot \mathbf{M}_{k,t})^{1/2} \odot \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_2(\mathbf{0}, V), \\
\mathbf{M}_{j,t} &= (M^1_{j,t}, M^2_{j,t}, \ldots, M^p_{j,t})', \quad (j \leq k), \\
M^i_{j,t} &= C_i \left(f^1_{j,t}\right)^{w^i_1} \left(f^2_{j,t}\right)^{w^i_2} \cdots \left(f^q_{j,t}\right)^{w^i_q} \left(u^i_{j,t}\right)^{w^i_{q+1}},
\end{aligned}$$

where the weights are non-negative and add up to one, and the constant C_i is chosen to guarantee that $\mathrm{E}(M^i_{j,t}) = 1$, and is thus not a free parameter. Calvet, Fisher, and Thompson (2006) specified the model as follows. For each vector $\mathbf{f}^l_t$, $f^l_{j,t}$ follows an univariate MSM process with parameters (b, γ, m^l). The volatility of each asset i is also affected by an idiosyncratic shock $\mathbf{u}^i_t = (u^i_{1,t}, \ldots, u^i_{k,t})'$, which is specified by parameters (b, γ, m^{q+i}). Draws of the factors $f^l_{j,t}$ and idiosyncratic shocks $u^i_{j,t}$ are independent, but timing of arrivals may be correlated. Factors and idiosyncratic components thus follow univariate MSM with identical frequencies.

3.2 Mean factor model

Pitt and Shephard (1999b), following a model proposed in Kim, Shephard, and Chib (1998), analyzed a factor-based MSV model which extends the general MSV model by including a linear combination of the $q \times 1$ factor vector $\mathbf{f}_t$ in the mean function as follows:

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}), \tag{25}$$

$$\mathbf{f}_t = \mathbf{D}_t^{1/2}\boldsymbol{\gamma}_t, \quad \boldsymbol{\gamma}_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}), \tag{26}$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \mathcal{N}_{p+q}(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\eta}) \tag{27}$$

where

$$\mathbf{V}_t = \mathrm{diag}(\exp(h_{1t}), \ldots, \exp(h_{pt})), \tag{28}$$

$$\mathbf{D}_t = \mathrm{diag}\left(\exp(h_{p+1,t}), \ldots, \exp(h_{p+q,t})\right), \tag{29}$$

$$\boldsymbol{\Phi} = \mathrm{diag}(\phi_1, \ldots, \phi_{p+q}) \tag{30}$$

$$\boldsymbol{\Sigma}_{\eta\eta} = \mathrm{diag}(\sigma_{1,\eta\eta}, ..., \sigma_{p+q,\eta\eta}) \tag{31}$$

and $\mathbf{h}_t = (h_{1t}, \ldots, h_{pt}, h_{p+1,t}, ..., h_{p+q,t})$. For identification purpose, the $p \times q$ loading matrix $\mathbf{B}$ is assumed to be such that $b_{ij} = 0$ for $(i < j,\ i \leq q)$ and $b_{ii} = 1$ $(i \leq q)$ with all other elements unrestricted. Thus, in this model, each of the factors and each of the errors evolve according to univariate SV models. The model is a generalization of the ones considered by Jacquier, Polson, and Rossi (1999) and Liesenfeld and Richard (2003) where $\mathbf{V}_t$ was not time-varying and only the factors followed a univariate SV process. Jacquier, Polson, and Rossi (1999) estimated their model by MCMC methods, sampling h_{it} one at a time from its full conditional distribution, a procedure that is known to be inefficient from Kim, Shephard, and Chib (1998), whereas Liesenfeld and Richard (2003) showed how the MLE could be obtained by the Efficient Importance Sampling method. For the more general model above, Pitt and Shephard (1999b) also employed a MCMC based approach, now sampling $\mathbf{h}_t$ along the lines of Shephard and Pitt (1997). As an application, they considered returns on daily closing prices of five foreign exchange rates (Deutschemark, British pound, Japanese yen, Swiss franc, French franc) quoted in US dollars. The model they fit had one factor. An even further generalization of this factor model was developed by Chib, Nardari, and Shephard (2006) who allowed for jumps in the observation model and a fat-tailed t-distribution for the errors $\boldsymbol{\varepsilon}_t$. The resulting model and its fitting is explained later in Section 3.3.

Lopes and Carvalho (2006) have considered a general model which nests the models of Pitt and Shephard (1999b) and Aguilar and West (2000), and extended it in two directions by (i) letting the matrix of factor loadings $\mathbf{B}$ to be time dependent, and (ii) allowing Markov switching in the common factors volatilities. The general model is given by equations (26)–(29) with

$$\mathbf{y}_t = \mathbf{B}_t \mathbf{f}_t + \mathbf{V}_t^{1/2} \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}),$$
$$\mathbf{h}_{t+1}^f = \boldsymbol{\mu}_{s_t}^f + \boldsymbol{\Phi}^f \mathbf{h}_t^f + \boldsymbol{\eta}_t^f, \quad \boldsymbol{\eta}_t^f \sim \mathcal{N}_q\left(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\eta}^f\right),$$

where $\mathbf{h}_t^f = (h_{p+1,t}, \ldots, h_{p+q,t})'$, $\boldsymbol{\mu}^f = (\mu_{p+1}, \ldots, \mu_{p+q})'$, $\boldsymbol{\Phi}^f = \mathrm{diag}(\phi_{p+1}, \ldots, \phi_{p+q})$, and $\boldsymbol{\Sigma}_{\eta\eta}^f$ is the non-diagonal covariance matrix. Letting the $pq - q(q+1)/2$ unconstrained elements of $\mathrm{vec}(\mathbf{B}_t)$ be $\mathbf{b}_t = (b_{21,t}, b_{31,t}, \ldots, b_{pq,t})'$, they assumed that each element of $\mathbf{b}_t$ follows an AR(1) process. Following So, Lam, and Li (1998), where the fitting was based on the work of Albert and Chib (1993), μ_{s_t} was assumed to follow a Markov switching model, where s_t follows a multi-state first order Markovian process. Lopes and Carvalho (2006) applied this model to two datasets: (i) returns on daily closing spot rates for six currencies relative to US dollar (Deutschemark, British pound, Japanese yen, French franc, Canadian dollar, Spanish peseta), and returns on daily closing rates for four Latin American stock markets indices. In the former application, they used $q = 3$ factors and in the latter case $q = 2$ factors.

Han (2006) modified the model of Pitt and Shephard (1999b) and Chib, Nardari, and Shephard (2006) by allowing the factors to follows an AR(1) process

$$\mathbf{f}_t = \mathbf{c} + \mathbf{A}\mathbf{f}_{t-1} + \mathbf{D}_t^{1/2} \boldsymbol{\gamma}_t, \quad \boldsymbol{\gamma}_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}). \tag{32}$$

The model was fit by adapting the approach of Chib, Nardari, and Shephard (2006) and applied to a collection of 36 arbitrarily chosen stocks to examine the performance of various portfolio strategies.

3.3 Bayesian analysis of mean factor MSV model

We describe the fitting of factor models in the context of the general model of Chib, Nardari, and Shephard (2006). The model is given by

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \mathbf{K}_t \mathbf{q}_t + \mathbf{V}_t^{1/2} \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}), \tag{33}$$

where $\boldsymbol{\Lambda}_t = \mathrm{diag}(\lambda_{1t}, \ldots, \lambda_{pt})$, $\mathbf{q}_t$ is p independent Bernoulli "jump" random variables, and $\mathbf{K}_t = \mathrm{diag}(k_{1t}, \ldots, k_{pt})$ are jump sizes. Assume that each element q_{jt} of $\mathbf{q}_t$ takes the value one with probability κ_j and the value zero with probability $1 - \kappa_j$, and that each element u_{jt} of $\mathbf{u}_t = \mathbf{V}_t^{1/2} \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\varepsilon}_t$ follows an independent Student-t distribution with degrees of freedom $\nu_j > 2$, which we express in hierarchical form as

$$u_{jt} = \lambda_{jt}^{-1/2} \exp(h_{jt}/2) \varepsilon_{jt}, \quad \lambda_{jt} \overset{i.i.d.}{\sim} \mathcal{G}\left(\frac{\nu_j}{2}, \frac{\nu_j}{2}\right), \quad t = 1, 2, \ldots, n. \tag{34}$$

The $\boldsymbol{\varepsilon}_t$ and $\mathbf{f}_t$ are assumed to be independent and

$$\begin{pmatrix} \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t \\ \mathbf{f}_t \end{pmatrix} |\mathbf{V}_t, \mathbf{D}_t, \mathbf{K}_t, \mathbf{q}_t \sim \mathcal{N}_{p+q}\left\{\mathbf{0}, \begin{pmatrix} \mathbf{V}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{D}_t \end{pmatrix}\right\}$$

are conditionally independent Gaussian random vectors. The time-varying variance matrices $\mathbf{V}_t$ and $\mathbf{D}_t$ are defined by equations (27)–(28). Chib, Nardari, and Shephard (2006) assumed that the variable $\zeta_{jt} = \ln(1 + k_{jt})$, $j \le p$, are distributed as $\mathcal{N}(-0.5\delta_j^2, \delta_j^2)$, where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)'$ are unknown parameters.

We may calculate the number of parameters and latent variables as follows. Let $\boldsymbol{\beta}$ denote the free elements of $\mathbf{B}$ after imposing the identifying restrictions. Let $\boldsymbol{\Sigma}_{\eta\eta} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$ and $\boldsymbol{\Sigma}_{\eta\eta}^f = \mathrm{diag}(\sigma_{p+1}^2, \ldots, \sigma_{p+q}^2)$. Then there are $pq - (q^2 + q)/2$ elements in $\boldsymbol{\beta}$. The model has $3(p+q)$ parameters $\boldsymbol{\theta}_j = (\phi_j, \mu_j, \sigma_j)$ $(1 \le j \le p+q)$ in the autoregressive processes (27) of $\{h_{jt}\}$. We also have p degrees of freedom $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)$, p jump intensities $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_p)$, and p jump variances $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)$. If we let $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p, \boldsymbol{\nu}, \boldsymbol{\delta}, \boldsymbol{\kappa})$ denote the entire list of parameters, then the dimension of ψ is 688 when $p = 50$ and $q = 8$. Furthermore, the model contains $n(p+q)$ latent volatilities $\{\mathbf{h}_t\}$ that appears non-linearly in the specification of $\mathbf{V}_t$ and $\mathbf{D}_t$, $2np$ latent variables $\{\mathbf{q}_t\}$ and $\{\mathbf{k}_t\}$ associated with the jump component, and np scaling variables $\{\boldsymbol{\lambda}_t\}$.

To conduct the prior-posterior analysis of this model, Chib, Nardari, and Shephard (2006) focus on the posterior distribution of the parameters and the latent variables

$$\pi\left(\boldsymbol{\beta}, \{\mathbf{f}_t\}, \{\boldsymbol{\theta}_j\}, \{\mathbf{h}_{j\cdot}\}, \{\nu_j\}, \{\boldsymbol{\lambda}_{j\cdot}\}, \{\delta_j\}, \{\kappa_j\}, \{\boldsymbol{\zeta}_{j\cdot}\}, \{\mathbf{q}_{j\cdot}\}|Y_n\right), \tag{35}$$

where the notation $\mathbf{z}_{j\cdot}$ is used to denote the collection $(z_{j1}, \ldots, z_{jn})$. They sample this distribution by MCMC methods through the following steps.

1. Sample $\boldsymbol{\beta}$. The full conditional distribution of $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta}|Y_n, \{\mathbf{h}_{j\cdot}\}, \{\boldsymbol{\zeta}_{j\cdot}\}, \{\mathbf{q}_{j\cdot}\}, \{\boldsymbol{\lambda}_{j\cdot}\}) \propto p(\boldsymbol{\beta}) \prod_{t=1}^{n} \mathcal{N}_p(\mathbf{y}_t|\mathbf{K}_t\mathbf{q}_t, \boldsymbol{\Omega}_t),$$

where $p(\boldsymbol{\beta})$ is the normal prior,

$$\boldsymbol{\Omega}_t = \mathbf{V}_t^* + \mathbf{B}\mathbf{D}_t\mathbf{B}' \quad \text{and} \quad \mathbf{V}_t^* = \mathbf{V}_t \odot \mathrm{diag}(\lambda_{1t}^{-1}, \ldots, \lambda_{pt}^{-1}).$$

To sample from this density, Chib, Nardari, and Shephard (2006) employed the Metropolis-Hastings (M-H) algorithm (Chib and Greenberg (1995)), following Chib and Greenberg (1994) and taking the proposal density to be multivariate-t, $T(\boldsymbol{\beta}|\mathbf{m}, \boldsymbol{\Sigma}, v)$, where $\mathbf{m}$ is the approximate mode of $l = \ln\{\prod_{t=1}^{n} \mathcal{N}_p(\mathbf{y}_t|\mathbf{K}_t\mathbf{q}_t, \boldsymbol{\Omega}_t)\}$, and $\boldsymbol{\Sigma}$ is minus the inverse of the second derivative matrix of l; the degrees of freedom v is set arbitrarily at 15. Let us denote the ij-th free element of $\mathbf{B}$ be denoted by b_{ij} and define $\tilde{\mathbf{y}}_t = \mathbf{y}_t - \mathbf{K}_t\mathbf{q}_t$. We have that

$$l = \sum_{t=1}^{n} \ln \mathcal{N}_p(\mathbf{y}_t|\mathbf{K}_t\mathbf{q}_t, \boldsymbol{\Omega}_t) = \text{const} - \frac{1}{2}\sum_{t=1}^{n} \ln|\boldsymbol{\Omega}_t| - \frac{1}{2}\sum_{t=1}^{n}(\mathbf{y}_t - \mathbf{K}_t\mathbf{q}_t)'\boldsymbol{\Omega}_t^{-1}(\mathbf{y}_t - \mathbf{K}_t\mathbf{q}_t)$$

and

$$\begin{aligned}\frac{\partial l}{\partial b_{ij}} &= \frac{1}{2}\sum_{t=1}^{n}\left\{\tilde{\mathbf{y}}_t'\boldsymbol{\Omega}_t^{-1}\frac{\partial \boldsymbol{\Omega}_t}{\partial b_{ij}}\boldsymbol{\Omega}_t^{-1}\tilde{\mathbf{y}}_t - \text{tr}\left(\boldsymbol{\Omega}_t^{-1}\frac{\partial \boldsymbol{\Omega}_t}{\partial b_{ij}}\right)\right\} \\ &= \sum_{t=1}^{n}\left\{\mathbf{s}_t'\frac{\partial \mathbf{B}}{\partial b_{ij}}\mathbf{D}_t\mathbf{B}'\mathbf{s}_t - \text{tr}\left(\mathbf{E}_t\frac{\partial \mathbf{B}'}{\partial b_{ij}}\right)\right\},\end{aligned}$$

where $\mathbf{s}_t = \boldsymbol{\Omega}_t^{-1}\tilde{\mathbf{y}}_t$, $\mathbf{E}_t = \boldsymbol{\Omega}_t^{-1}\mathbf{B}\mathbf{D}_t$, and

$$\boldsymbol{\Omega}_t^{-1} = (\mathbf{V}_t^*)^{-1} - (\mathbf{V}_t^*)^{-1}\mathbf{B}\left\{\mathbf{D}_t^{-1} + \mathbf{B}'(\mathbf{V}_t^*)^{-1}\mathbf{B}\right\}^{-1}\mathbf{B}(\mathbf{V}_t^*)^{-1}.$$

With these derivatives, $(\mathbf{m}, \boldsymbol{\Sigma})$ can be found by a sequence of Newton-Raphson iterations. Then the M-H step for sampling $\boldsymbol{\beta}$ is implemented by drawing a value $\boldsymbol{\beta}^*$ from the multivariate-t distribution, namely $T(\mathbf{m}, \boldsymbol{\Sigma}, v)$, and accepting the proposal value with probability

$$\begin{aligned}&\alpha(\boldsymbol{\beta}, \boldsymbol{\beta}^*|\tilde{\mathbf{y}}, \{\mathbf{h}_{j\cdot}\}, \{\boldsymbol{\lambda}_{j\cdot}\}) \\ &= \min\left\{1, \frac{p(\boldsymbol{\beta}^*)\prod_{t=1}^{n}\mathcal{N}_p\left(\tilde{\mathbf{y}}_\mathbf{t}|\mathbf{0}, \mathbf{V}_t^* + \mathbf{B}^*\mathbf{D}_t\mathbf{B}^{*\prime}\right)T(\boldsymbol{\beta}|\mathbf{m}, \boldsymbol{\Sigma}, v)}{p(\boldsymbol{\beta})\prod_{t=1}^{n}\mathcal{N}_p\left((\tilde{\mathbf{y}}_t|\mathbf{0}, \mathbf{V}_t^* + \mathbf{B}\mathbf{D}_t\mathbf{B}'\right)T(\boldsymbol{\beta}^*|\mathbf{m}, \boldsymbol{\Sigma}, v)}\right\},\end{aligned}$$

where $\boldsymbol{\beta}$ is the current value. If the proposal value is rejected, the next item of the chain is taken to be the current value $\boldsymbol{\beta}$.

2. Sample $\{\mathbf{f}_t\}$. The distribution $\{\mathbf{f}_t\}|\tilde{Y}_n, \mathbf{B}, \mathbf{h}, \boldsymbol{\lambda}$ can be divided into the product of the distributions $\mathbf{f}_t|\tilde{\mathbf{y}}_t, \mathbf{h}_t, \mathbf{h}_t^f, \boldsymbol{\lambda}_t, \mathbf{B}$, which have Gaussian distribution with mean $\hat{\mathbf{f}}_t = \mathbf{F}_t\mathbf{B}'(\mathbf{V}_t^*)^{-1}\tilde{\mathbf{y}}_t$ and variance $\mathbf{F}_t = \left\{\mathbf{B}'(\mathbf{V}_t^*)^{-1}\mathbf{B} + \mathbf{D}_t^{-1}\right\}^{-1}$.

3. Sample $\{\boldsymbol{\theta}_j\}$ and $\{\mathbf{h}_{j\cdot}\}$. Given $\{\mathbf{f}_t\}$ and the conditional independence of the errors in (27), the model separates into q conditionally Gaussian state space models. Let

$$z_{jt} = \begin{cases} \ln(y_{jt} - \alpha_{jt} - \exp(\zeta_{jt}) - 1)q_{jt} + c)^2 + \ln(\lambda_{jt}), & j \le p, \\ \ln(f_{j-p,t}^2), & j \ge p+1, \end{cases}$$

where c is an "offset" constant that is set to 10^{-6}. Then from Kim, Shephard, and Chib (1998) it follows that the $p+q$ state space models can be subjected to an independent analysis for sampling the $\{\boldsymbol{\theta}_j\}$ and $\{\mathbf{h}_{j\cdot}\}$. In particular, the distribution of z_{jt}, which is h_{jt} plus a log chi-squared random variable with one degree of freedom, may be approximated closely by a seven component mixture of normal distributions, allowing us to express the model as

$$
\begin{aligned}
&z_{jt}|s_{jt}, h_{jt} \sim \mathcal{N}\left(h_{jt} + m_{s_{jt}}, v_{s_{jt}}^2\right), \\
&h_{j,t+1} - \mu_j = \phi_j\left(h_{j,t} - \mu_j\right) + \eta_{jt}, \quad j \leq p+q,
\end{aligned}
$$

where s_{jt} is a discrete component indicator variable with mass function $\Pr(s_{jt} = i) = q_i$, $i \leq 7$, $t \leq n$, and $m_{s_{jt}}$, $v_{s_{jt}}^2$ and q_i are parameters that are reported in Chib, Nardari, and Shephard (2002). Thus, under this representation, conditioned on the transformed observations we have that

$$
p\left(\{\mathbf{s}_{j\cdot}\}, \boldsymbol{\theta}, \{\mathbf{h}_{j\cdot}\}|\mathbf{z}\right) = \prod_{j=1}^{p+q} p\left(\mathbf{s}_{j\cdot}, \boldsymbol{\theta}_j, \mathbf{h}_{j\cdot}|\mathbf{z}_{j\cdot}\right),
$$

which implies that the mixture indicators, log-volatilities and series specific parameters can be sampled series by series. Now, for each j, one can sample $(\mathbf{s}_{j\cdot}, \boldsymbol{\theta}_j, \mathbf{h}_{j\cdot})$ by the univariate SV algorithm given by Chib, Nardari, and Shephard (2002). Briefly, $\mathbf{s}_{j\cdot}$ is sampled straightforwardly from

$$
p\left(\mathbf{s}_{j\cdot}|\mathbf{z}_{j\cdot}, \mathbf{h}_{j\cdot}\right) = \prod_{t=1}^{n} p\left(s_{jt}|z_{jt}, h_{jt}\right),
$$

where $p(s_{jt}|z_{jt}, h_{jt}) \propto p(s_{jt})\mathcal{N}\left(z_{jt}|\, h_{jt} + m_{s_{jt}}, v_{s_{jt}}^2\right)$ is a mass function with seven points of support. Next, $\boldsymbol{\theta}_j$ is sampled by the M-H algorithm from the density $\pi(\boldsymbol{\theta}_j|\mathbf{z}_{j\cdot}, \mathbf{s}_{j\cdot}) \propto p(\boldsymbol{\theta}_j)p(\mathbf{z}_{j\cdot}|\mathbf{s}_{j\cdot}, \boldsymbol{\theta}_j)$ where

$$
p\left(\mathbf{z}_{j\cdot}|\mathbf{s}_{j\cdot}, \boldsymbol{\theta}_j\right) = p\left(\mathbf{z}_{j1}|\mathbf{s}_{j\cdot}, \boldsymbol{\theta}_j\right) \prod_{t=2}^{n} p\left(\mathbf{z}_{jt}|\mathcal{F}_{j,t-1}^*, \mathbf{s}_{j\cdot}, \boldsymbol{\theta}_j\right) \tag{36}
$$

and $p(z_{jt}|\mathcal{F}_{j,t-1}^*, \mathbf{s}_{j\cdot}, \boldsymbol{\theta}_j)$ is a normal density whose parameters are obtained by the Kalman filter recursions, adapted to the differing components, as indicated by the component vector $\mathbf{s}_{j\cdot}$. Finally, $\mathbf{h}_{j\cdot}$ is sampled from $[\mathbf{h}_{j\cdot}|\mathbf{z}_{j\cdot}, \mathbf{s}_{j\cdot}, \boldsymbol{\theta}_j]$ by the simulation smoother algorithm of de Jong and Shephard (1995).

4. Sample $\{\nu_j\}$, $\{\mathbf{q}_{j\cdot}\}$ and $\{\boldsymbol{\lambda}_{j\cdot}\}$. The degrees of freedom parameters, jump parameters and associated latent variables are sampled independently for each time series. The full conditional distribution of ν_j is given by

$$
\Pr(\nu_j|\mathbf{y}_{j\cdot}, \mathbf{h}_j, \mathbf{B}, \mathbf{f}, \mathbf{q}_{j\cdot}, \boldsymbol{\zeta}_{j\cdot}) \propto \Pr(\nu_j) \prod_{t=1}^{n} T(y_{jt}|\alpha_{jt} + \{\exp(\zeta_{jt}) - 1\}q_{jt}, \exp(h_{jt}), \nu_j), \tag{37}
$$

and one can apply the Metropolis-Hastings algorithm in a manner analogous to the case of $\boldsymbol{\beta}$. Next, the jump indicators $\{\mathbf{q}_{j.}\}$ are sampled from the two-point discrete distribution

$$\Pr(q_{jt}=1|\mathbf{y}_{j.},\mathbf{h}_{j.},\mathbf{B},\mathbf{f},\nu_j,\boldsymbol{\zeta}_{j.},\kappa_j) \propto \kappa_j T(y_{jt}|\alpha_{jt}+\{\exp(\zeta_{jt})-1\},\exp(h_{jt}),\nu_j),$$
$$\Pr(q_{jt}=0|\mathbf{y}_{j.},\mathbf{h}_{j.},\mathbf{B},\mathbf{f},\nu_j,\boldsymbol{\zeta}_{j.},\kappa_j) \propto (1-\kappa_j)T(y_{jt}|\alpha_{jt},\exp(h_{jt}),\nu_j),$$

followed by the components of the vector $\{\boldsymbol{\lambda}_{j.}\}$ from the density

$$\lambda_{jt}|y_{jt},h_{jt},\mathbf{B},\mathbf{f},\nu_j,q_{jt},\psi_{jt} \sim \mathcal{G}\left(\frac{\nu_j+1}{2},\frac{\nu_j+(y_{jt}-\alpha_{jt}-(\exp(\zeta_{jt})-1)q_{jt}))^2}{2\exp(h_{jt})}\right).$$

5. Sample $\{\delta_j\}$ and $\{\boldsymbol{\zeta}_{j.}\}$. For simulation efficiency reasons, δ_j and $\boldsymbol{\zeta}_{j.}$ must also be sampled in one block. The full conditional distribution of δ_j is given by

$$\pi(\delta_j)\prod_{t=1}^{n}\mathrm{N}(\alpha_{jt}-0.5\delta_j^2 q_{jt},\delta_j^2 q_{jt}^2+\exp(h_{jt})\lambda_{jt}^{-1}) \tag{38}$$

by the M-H algorithm. Once δ_j is sampled, the vectors $\boldsymbol{\zeta}_{j.}$ are sampled, bearing in mind that their posterior distribution is updated only when q_{jt} is one. Therefore, when q_{jt} is zero, we sample ζ_{jt} from $\mathcal{N}(-0.5\delta_j^2,\delta_j^2)$, otherwise we sample from the distribution $\mathcal{N}(\Psi_{jt}(-0.5+\exp(-h_{jt})\lambda_{jt}y_{jt}),\Psi_{jt})$, where $\Psi_{jt}=(\delta_j^{-2}+\exp(-h_{jt})\lambda_{jt})^{-1}$. The algorithm is completed by sampling the components of the vector $\boldsymbol{\kappa}$ independently from $\kappa_j|q_{j.}\sim$ $beta(u_{0j}+n_{1j},u_{1j}+n_{0j})$, where n_{0j} is the count of $q_{jt}=0$ and $n_{1j}=n-n_{0j}$ is the count of $q_{jt}=1$.

A complete cycle through these various distributions completes one transition of our Markov chain. These steps are then repeated G times, where G is a large number, and the values beyond a suitable burn-in of say a 1000 cycles, are used for the purpose of summarizing the posterior distribution.

4 Dynamic correlation MSV model

A weakness of the standard MSV model is that it has a conditional correlation matrix that is time-invariant. This weakness is overcome in the mean factor model. For example, consider a bivariate model with a common mean factor

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{b}f_t+\mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \mathbf{V}_t=\mathrm{diag}(\exp(h_{1t}),\exp(h_{2t})), \quad \boldsymbol{\varepsilon}_t\sim\mathcal{N}_2(\mathbf{0},\mathbf{I}),\\
f_t &= \exp(h_{3t}/2)\gamma_t, \quad \gamma_t\sim\mathcal{N}(0,1),\\
\mathbf{h}_{t+1} &= \boldsymbol{\mu}+\boldsymbol{\Phi}(\mathbf{h}_t-\boldsymbol{\mu})+\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t\sim\mathcal{N}_2(\mathbf{0},\mathbf{I}),
\end{aligned}$$

where $\mathbf{b} = (1, b_{21})'$. Then the variance covariance matrix of $\mathbf{y}_t$ given b_{21}, $\mathbf{h}_t$ and h_t^f is

$$\mathrm{Var}(\mathbf{y}_t|\mathbf{h}_t) = \begin{pmatrix} \exp(h_{1t}) & 0 \\ 0 & \exp(h_{2t}) \end{pmatrix} + \exp(h_{3t}) \begin{pmatrix} 1 & b_{21} \\ b_{21} & b_{21}^2 \end{pmatrix},$$

and hence the correlation coefficient is given by

$$\mathrm{Corr}(y_{1t}, y_{2t}|\mathbf{h}_t) = \frac{b_{21}}{\sqrt{\{1 + \exp(h_{1t} - h_{3t})\}\left\{b_{21}^2 + \exp(h_{2t} - h_{3t})\right\}}}$$

which is time varying. Another way of achieving the same end is by modeling the correlation matrix directly. For instance, we may model a time-varying covariance matrix $\mathbf{\Sigma}_t$ and obtain a time-varying correlation matrix using some positive definite matrix $\mathbf{Q}_t$ such that

$$\mathbf{\Sigma}_{\varepsilon\varepsilon,t} = \mathbf{Q}_t^{*-1/2} \mathbf{Q}_t \mathbf{Q}_t^{*-1/2}$$

where $\mathbf{Q}_t^*$ is a diagonal matrix whose (i,i)-th element is the same as that of $\mathbf{Q}_t$. (e.g. Asai, McAleer, and Yu (2006)). We describe several such approaches in this section.

4.1 Modeling by reparameterization

We begin by considering two approaches for modeling time-varying correlations that are based on the dynamic modeling of reparameterized correlations, as in Tsay (2005). The first approach is illustrated by Yu and Meyer (2006) in the context of the bivariate SV model

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_2(\mathbf{0}, \mathbf{\Sigma}_{\varepsilon\varepsilon,t}), \quad \mathbf{\Sigma}_{\varepsilon\varepsilon,t} = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}, \\
\mathbf{h}_{t+1} &= \boldsymbol{\mu} + \mathrm{diag}(\phi_1, \phi_2)(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}_2\left(\mathbf{0}, \mathrm{diag}(\sigma_1^2, \sigma_2^2)\right), \\
q_{t+1} &= \psi_0 + \psi_1(q_t - \psi_0) + \sigma_\rho v_t, \quad v_t \sim \mathcal{N}(0,1), \\
\rho_t &= \frac{\exp(q_t) - 1}{\exp(q_t) + 1},
\end{aligned}$$

where $\mathbf{h}_0 = \boldsymbol{\mu}$ and $q_0 = \psi_0$. The correlation coefficient ρ_t is then obtained from q_t by the Fisher transformation. Yu and Meyer (2006) estimated this model by MCMC methods with the help of WinBUGS program and found that it was superior to other models including the mean factor MSV model. However, the generalization of this bivariate model to the higher dimensions is not easy because it is difficult to ensure the positive definiteness of the correlation matrix $\mathbf{\Sigma}_{\varepsilon\varepsilon,t}$.

The second reparameterization introduced by Tsay (2005) is based on the Choleski decomposition of the time-varying correlation matrix. Specifically, we consider the Choleski decomposition of the correlation matrix $\mathbf{\Sigma}_{\varepsilon\varepsilon,t}$ such that $\mathrm{Cov}(\mathbf{y}_t|\mathbf{h}_t) = \mathbf{L}_t\mathbf{V}_t\mathbf{L}_t'$. The outcome model is then given by

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

As an example, when bivariate outcomes are involved we have

$$\mathbf{L}_t = \begin{pmatrix} 1 & 0 \\ q_t & 1 \end{pmatrix}, \quad \mathbf{V}_t = \mathrm{diag}(\exp(h_{1t}), \exp(h_{2t})),$$

Then,

$$\begin{aligned} y_{1t} &= \varepsilon_{1t} \exp(h_{1t}/2), \\ y_{2t} &= q_t \varepsilon_{1t} \exp(h_{1t}/2) + \varepsilon_{2t} \exp(h_{2t}/2), \end{aligned}$$

which shows that the distribution of $\mathbf{y}_t$ is modeled sequentially. We first let $y_{1t} \sim \mathcal{N}(0, \exp(h_{1t}))$ and then we let $y_{2t}|y_{1t} \sim \mathcal{N}(q_t y_{1t}, \exp(h_{2t}))$. Thus q_t is a slope of conditional mean and the correlation coefficient between y_{1t} and y_{2t} is given by

$$\begin{aligned} \mathrm{Var}(y_{1t}) &= \exp(h_{1t}), \\ \mathrm{Var}(y_{2t}) &= q_t^2 \exp(h_{1t}) + \exp(h_{2t}), \\ \mathrm{Cov}(y_{1t}, y_{2t}) &= q_t \exp(h_{1t}), \\ \mathrm{Corr}(y_{1t}, y_{2t}) &= \frac{q_t}{\sqrt{q_t^2 + \exp(h_{2t} - h_{1t})}} \end{aligned}$$

As suggested in Asai, McAleer, and Yu (2006), we let q_t follow an AR(1) process

$$q_{t+1} = \psi_0 + \psi_1(q_t - \psi_0) + \sigma_\rho v_t, \quad v_t \sim \mathcal{N}(0, 1).$$

The generalization to higher dimensions is straightforward. Let

$$\mathbf{L}_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ q_{21,t} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ q_{p1,t} & \cdots & q_{p,p-1,t} & 1 \end{pmatrix}, \quad \mathbf{V}_t = \mathrm{diag}(\exp(h_{1t}), \ldots, \exp(h_{pt})),$$

and

$$\begin{aligned} y_{1t} &= \varepsilon_{1t} \exp(h_{1t}/2), \\ y_{2t} &= q_{21,t} \varepsilon_{1t} \exp(h_{1t}/2) + \varepsilon_{2t} \exp(h_{2t}/2), \\ &\vdots \\ y_{pt} &= q_{p1,t} \varepsilon_{1t} \exp(h_{1t}/2) + \ldots + q_{p,p-1,t} \varepsilon_{p-1,t} \exp(h_{p-1,t}/2) + \varepsilon_{pt} \exp(h_{pt}/2) \end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(y_{it}) &= \sum_{k=1}^{i} q_{ik,t}^2 \exp(h_{kt}), \quad q_{ii,t} \equiv 1, \quad i = 1, \ldots, p, \\
\mathrm{Cov}(y_{it}, y_{jt}) &= \sum_{k=1}^{i} q_{ik,t} q_{jk,t} \exp(h_{kt}), \quad i < j, \quad i = 1, \ldots, p-1, \\
\mathrm{Corr}(y_{it}, y_{jt}) &= \frac{\displaystyle\sum_{k=1}^{i} q_{ik,t} q_{jk,t} \exp(h_{kt})}{\sqrt{\displaystyle\sum_{k=1}^{i} q_{ik,t}^2 \exp(h_{kt}) \sum_{k=1}^{j} q_{jk,t}^2 \exp(h_{kt})}}, \quad i < j,
\end{aligned}$$

where q_{it} now follows the AR(1) process

$$q_{i,t+1} = \psi_{i,0} + \psi_{i,1}(q_{i,t} - \psi_0) + \sigma_{i,\rho} v_{it}, \quad v_{it} \sim \mathcal{N}(0, 1),$$

Jungbacker and Koopman (2006) considered a similar model with $\mathbf{L}_t = \mathbf{L}$ and estimated the parameters of the model by the Monte Carlo likelihood method. As in the one factor case, they used the data set for the daily exchange rate returns of British pound, the Deutschemark, and the Japanese yen against the U.S. dollar.

4.2 Matrix exponential transformation

For any $p \times p$ matrix $\mathbf{A}$, the matrix exponential transformation is defined by the following power series expansion,

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

where $\mathbf{A}^0$ is equal to a $p \times p$ identity matrix. For any real positive definite matrix $\mathbf{C}$, there exists a real symmetric $p \times p$ matrix $\mathbf{A}$ such that

$$\mathbf{C} = \exp(\mathbf{A}).$$

Conversely, for any real symmetric matrix $\mathbf{A}$, $\mathbf{C} = \exp(\mathbf{A})$ is a positive definite matrix (see e.g. Lemma 1 of Chiu, Leonard, and Tsui (1996), Kawakatsu (2006)). If $\mathbf{A}_t$ is a $p \times p$ real symmetric matrix, there exists a $p \times p$ orthogonal matrix $\mathbf{B}_t$ and a $p \times p$ real diagonal matrix $\mathbf{H}_t$ of eigenvalues of $\mathbf{A}$ such that $\mathbf{A}_t = \mathbf{B}_t \mathbf{H}_t \mathbf{B}_t'$ and

$$\exp(\mathbf{A}_t) = \mathbf{B}_t \left(\sum_{s=0}^{\infty} \frac{1}{s!} \mathbf{H}_t^s \right) \mathbf{B}_t' = \mathbf{B}_t \exp(\mathbf{H}_t) \mathbf{B}_t'$$

Thus we consider the matrix exponential transformation for the covariance matrix $\mathrm{Var}(\mathbf{y}_t) = \mathbf{\Sigma}_t = \exp(\mathbf{A}_t)$ where $\mathbf{A}_t$ is a $p \times p$ real symmetric matrix such that $\mathbf{A}_t = \mathbf{B}_t\mathbf{H}_t\mathbf{B}_t'$ ($\mathbf{H}_t = \mathrm{diag}(h_{1t}, \ldots, h_{pt})$). Note that

$$
\begin{aligned}
\mathbf{\Sigma}_t &= \mathbf{B}_t\mathbf{V}_t\mathbf{B}_t', \quad \mathbf{V}_t = \mathrm{diag}(\exp(h_{1t}), \ldots, \exp(h_{pt})), \\
\mathbf{\Sigma}_t^{-1} &= \mathbf{B}_t'\mathbf{V}_t^{-1}\mathbf{B}_t, \quad |\mathbf{\Sigma}_t| = \exp\left(\sum_{i=1}^{p} h_{it}\right),
\end{aligned}
$$

We model the dynamic structure of covariance matrices through $\boldsymbol{\alpha}_t = \mathrm{vech}(\mathbf{A}_t)$. We may consider a first order autoregressive process for $\boldsymbol{\alpha}_t$

$$
\begin{aligned}
\mathbf{y}_t|\mathbf{A}_t &\sim \mathcal{N}_p(\mathbf{0}, \exp(\mathbf{A}_t)), \\
\boldsymbol{\alpha}_{t+1} &= \boldsymbol{\mu} + \mathbf{\Phi}(\boldsymbol{\alpha}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad (\mathbf{\Phi} : \text{diagonal}), \\
& \boldsymbol{\alpha}_t = \mathrm{vech}(\mathbf{A}_t), \quad \boldsymbol{\eta}_t \sim \mathcal{N}_{p(p+1)/2}(\mathbf{0}, \mathbf{\Sigma}_{\eta\eta}),
\end{aligned}
$$

as suggested in Asai, McAleer, and Yu (2006). The estimation of this model can be done using MCMC or a simulated maximum likelihood estimation, but it is not straightforward to interpret the parameters.

4.3 Wishart Process

4.3.1 Standard model

Philipov and Glickman (2006b) and Philipov and Glickman (2006a) considered a dynamic asset covariance structure and assumed that the conditional covariance matrix $\mathbf{\Sigma}_t$ follows an inverted Wishart distribution whose parameter depends on the past covariance matrix $\mathbf{\Sigma}_{t-1}$. That is

$$
\begin{aligned}
\mathbf{y}_t|\mathbf{\Sigma}_t &\sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}_t), \\
\mathbf{\Sigma}_t|\nu, \mathbf{S}_{t-1} &\sim \mathcal{IW}_p(\nu, \mathbf{S}_{t-1}),
\end{aligned}
$$

where $\mathcal{IW}(\nu_0, \mathbf{Q}_0)$ denotes an inverted Wishart distribution with parameters $(\nu_0, \mathbf{Q}_0)$,

$$
\begin{aligned}
\mathbf{S}_{t-1} &= \frac{1}{\nu}\mathbf{A}^{1/2}\left(\mathbf{\Sigma}_{t-1}^{-1}\right)^{d}\mathbf{A}^{1/2\prime}, \\
\mathbf{A} &= \mathbf{A}^{1/2}\mathbf{A}^{1/2\prime},
\end{aligned}
\tag{39}
$$

and $\mathbf{A}^{1/2}$ is a Choleski decomposition of a positive definite symmetric matrix $\mathbf{A}$ and $-1 < d < 1$. Asai and McAleer (2007) point out that it also possible to parameterize $\mathbf{S}_{t-1}$ as $\nu^{-1}\left(\mathbf{\Sigma}_{t-1}^{-1}\right)^{d/2}\mathbf{A}\left(\mathbf{\Sigma}_{t-1}^{-1}\right)^{d/2\prime}$.

The conditional expected values of $\mathbf{\Sigma}_t^{-1}$ and $\mathbf{\Sigma}_t$ are

$$\begin{aligned} \mathrm{E}\left(\mathbf{\Sigma}_t^{-1}|\nu,\mathbf{S}_{t-1}\right) &= \nu\mathbf{S}_{t-1} = \mathbf{A}^{1/2}\left(\mathbf{\Sigma}_{t-1}^{-1}\right)^d\mathbf{A}^{1/2\prime}, \\ \mathrm{E}\left(\mathbf{\Sigma}_t|\nu,\mathbf{S}_{t-1}\right) &= \frac{1}{\nu-p-1}\mathbf{S}_{t-1}^{-1} = \frac{\nu}{\nu-p-1}\mathbf{A}^{-1/2}\left(\mathbf{\Sigma}_{t-1}\right)^d\mathbf{A}^{-1/2\prime}, \end{aligned}$$

respectively. Thus the scale parameter d expresses the overall strength of the serial persistence in the covariance matrix over time. Based on the process of the logarithm of the determinant, and asymptotic behavior of expectation of the determinant, they assume that $|d| < 1$ although it is natural to assume that $0 < d < 1$. Notice that when $d = 0$, for example, the serial persistence disappears and we get that

$$\begin{aligned} \mathrm{E}\left(\mathbf{\Sigma}_t^{-1}|\nu,\mathbf{S}_{t-1}\right) &= \mathbf{A}, \\ \mathrm{E}\left(\mathbf{\Sigma}_t|\nu,\mathbf{S}_{t-1}\right) &= \frac{\nu}{\nu-p-1}\mathbf{A}^{-1}. \end{aligned}$$

The matrix $\mathbf{A}$ in this model is a measure of the inter-temporal sensitivity and determines how the elements of the current period covariance matrix $\mathbf{\Sigma}_t$ are related to the elements of the previous period covariance matrix. When $\mathbf{A} = \mathbf{I}$, we note that

$$\mathrm{E}\left(\mathbf{\Sigma}_t^{-1}|\nu,\mathbf{S}_{t-1}\right) = \begin{cases} \mathbf{\Sigma}_{t-1}^{-1}, & d = 1, \\ \mathbf{I}, & d = 0, \\ \mathbf{\Sigma}_{t-1}, & d = -1. \end{cases}$$

Philipov and Glickman (2006b) estimated this model from a Bayesian approach and proposed an MCMC algorithm to estimate their models using monthly return data of five industry portfolios (Manufacturing, Utilities, Retail/Wholesale, Financial and Other) in NYSE, AMEX and NASDAQ stocks. Under the prior

$$\mathbf{A} \sim \mathcal{IW}_p(\nu_0, \mathbf{Q}_0), \quad d \sim \pi(d), \quad \nu - p \sim \mathcal{G}(\alpha, \beta)$$

with $\mathbf{\Sigma}_0$ assumed known, the MCMC algorithm is implemented as follows:

1. Sample $\mathbf{\Sigma}_t|\{\mathbf{\Sigma}_s\}_{s\neq t}, \mathbf{A}, \nu, d, Y_n$ $(t = 1, \ldots, n-1)$ where $Y_n = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$. Given a current sampler $\mathbf{\Sigma}_t$, we generate a candidate $\mathbf{\Sigma}_t^* \sim \mathcal{W}_p(\tilde{\nu}, \tilde{\mathbf{S}}_{t-1})$ where $\mathcal{W}_p(\tilde{\nu}, \tilde{\mathbf{S}}_{t-1})$ denotes a Wishart distribution with parameters $(\tilde{\nu}, \tilde{\mathbf{S}}_{t-1})$,

$$\begin{aligned} \tilde{\nu} &= \nu(1-d)+1, \\ \tilde{\mathbf{S}}_{t-1} &= \mathbf{S}_{t-1}^{-1} + \mathbf{y}_t\mathbf{y}_t', \\ \mathbf{S}_{t-1} &= \frac{1}{\nu}(\mathbf{A}^{1/2})\left(\mathbf{\Sigma}_{t-1}^{-1}\right)^d(\mathbf{A}^{1/2})', \end{aligned}$$

and accept it with probability

$$\frac{|\boldsymbol{\Sigma}_t^*|^{(\nu d-1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left\{\nu\mathbf{A}^{-1}\left(\boldsymbol{\Sigma}_t^*\right)^{-d}\boldsymbol{\Sigma}_{t+1}^{-1}\right\}\right]}{|\boldsymbol{\Sigma}_t|^{(\nu d-1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left\{\nu\mathbf{A}^{-1}\left(\boldsymbol{\Sigma}_t\right)^{-d}\boldsymbol{\Sigma}_{t+1}^{-1}\right\}\right]}$$

2. Sample $\boldsymbol{\Sigma}_n|\{\boldsymbol{\Sigma}_t\}_{t=1}^{n-1},\mathbf{A},\nu,d,Y_n\sim\mathcal{W}_p(\tilde{\nu},\tilde{\mathbf{S}}_{n-1})$.

3. Sample $\mathbf{A}|\{\boldsymbol{\Sigma}_t\}_{t=1}^{n},\nu,d,\mathbf{y}\sim\mathcal{IW}_p(\tilde{\gamma},\tilde{\mathbf{Q}})$, where $\tilde{\gamma}=n\nu+\nu_0$, and

$$\tilde{\mathbf{Q}}^{-1}=\nu\left\{\sum_{t=1}^{n}\left(\boldsymbol{\Sigma}_t^{-1}\right)^{-d/2}\boldsymbol{\Sigma}_t^{-1}\left(\boldsymbol{\Sigma}_{t-1}^{-1}\right)^{-d/2}\right\}+\mathbf{Q}_0^{-1},$$

4. Sample d from

$$\pi(d|\{\boldsymbol{\Sigma}_t\}_{t=1}^{n},\mathbf{A},\nu,\mathbf{y})\propto\pi(d)\exp\left[\frac{\nu d}{2}\sum_{t=1}^{n}\log|\boldsymbol{\Sigma}_t|-\frac{1}{2}\sum_{t=1}^{n}\mathrm{tr}\left\{\mathbf{S}_t^{-1}\left(\boldsymbol{\Sigma}_{t-1}^{-1}\right)^{-d}\right\}\right].$$

To sample d, Philipov and Glickman (2006b) suggested discretizing the conditional distribution (see Appendix A.2 of Philipov and Glickman (2006b)). Alternatively, we may conduct an independent M-H algorithm using a candidate from a truncated normal distribution $\mathcal{TN}_{(0,1)}(\hat{d},\hat{V}_d)$ where $\mathcal{TN}_{(a,b)}(\mu,\sigma^2)$ denote a normal distribution with mean μ and variance σ^2 truncated on the interval (a,b), $\hat{d}$ is a mode of conditional posterior probability density $\pi(d|\{\boldsymbol{\Sigma}_t\}_{t=1}^{n},\mathbf{A},\nu,\mathbf{y})$ and

$$\hat{V}_d=\left\{-\left.\frac{\partial^2\log\pi(d|\{\boldsymbol{\Sigma}_t\}_{t=1}^{n},\mathbf{A},\nu,Y_n)}{\partial d^2}\right|_{d=\hat{d}}\right\}^{-1}.$$

5. Sample ν from

$$\begin{aligned}\pi(\nu|\{\boldsymbol{\Sigma}_t\}_{t=1}^{n},\mathbf{A},d,\mathbf{y})\quad\propto\quad&(\nu-p)^{\alpha-1}\exp\{-\beta(\nu-p)\}\left\{\frac{|\nu\mathbf{A}^{-1}|^{\nu/2}}{2^{\nu p}\prod_{j=1}^{p}\Gamma(\frac{\nu+j-1}{2})}\right\}^{n}\\&\times\exp\left[-\frac{\nu}{2}\sum_{t=1}^{n}\left\{\log|\mathbf{Q}_t|+\mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{Q}_t^{-1}\right)\right\}\right].\end{aligned}$$

As in the previous step, we may discretize the conditional distribution or conduct an independent M-H algorithm using a candidate from a truncated normal distribution $\mathcal{TN}_{(p,\infty)}(\hat{\nu},\hat{V}_\nu)$ where $\hat{\nu}$ is a mode of conditional posterior probability density $\pi(\nu|\{\boldsymbol{\Sigma}_t\}_{t=1}^{n},\mathbf{A},d,\mathbf{y})$ and

$$\hat{V}_\nu=\left\{-\left.\frac{\partial^2\log\pi(\nu|\{\boldsymbol{\Sigma}_t\}_{t=1}^{n},\mathbf{A},d,Y_n)}{\partial\nu^2}\right|_{\nu=\hat{\nu}}\right\}^{-1}.$$

Asai and McAleer (2007) proposed two further models that are especially useful in higher dimensions. Let $\mathbf{Q}_t$ be a sequence of positive definite matrices, which is used to define correlation matrix $\boldsymbol{\Sigma}_{\varepsilon\varepsilon,t} = \mathbf{Q}_t^{*-1/2}\mathbf{Q}_t\mathbf{Q}_t^{*-1/2}$ where $\mathbf{Q}_t^*$ is a diagonal matrix whose (i,i)-th element is the same as that of $\mathbf{Q}_t$. Then the first of their Dynamic Correlation (DC) MSV model is given by:

$$\begin{aligned}
&\mathbf{y}_t = \mathbf{V}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon,t}), \quad \boldsymbol{\Sigma}_{\varepsilon\varepsilon,t} = \mathbf{Q}_t^{*-1/2}\mathbf{Q}_t\mathbf{Q}_t^{*-1/2}, \\
&\mathbf{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}_p\left(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\eta}\right), \quad (\boldsymbol{\Phi} \text{ and } \boldsymbol{\Sigma}_{\eta\eta} : \text{diagonal}) \\
&\mathbf{Q}_{t+1} = (1-\psi)\bar{\mathbf{Q}} + \psi\mathbf{Q}_t + \boldsymbol{\Xi}_t, \quad \boldsymbol{\Xi}_t \sim \mathcal{W}_p(\nu, \boldsymbol{\Lambda})
\end{aligned}$$

Thus, in this model the MSV shocks are assumed to follow a Wishart process, where $\mathcal{W}_p(\nu, \boldsymbol{\Lambda})$ denotes a Wishart distribution with degrees of freedom parameter ν and scale matrix $\boldsymbol{\Lambda}$. The model guarantees that $\mathbf{P}_t$ is symmetric positive definite under the assumption that $\bar{\mathbf{Q}}$ is positive definite and $|\psi| < 1$. It is possible to consider a generalization of the model by letting $\mathbf{Q}_{t+1} = (\mathbf{11}' - \boldsymbol{\Psi}) \odot \bar{\mathbf{Q}} + \boldsymbol{\Psi} \odot \mathbf{Q}_t + \boldsymbol{\Xi}_t$, which corresponds to a generalization of the Dynamic Conditional Correlation (DCC) model of Engle (2002).

The second DC MSV model is given by

$$\boldsymbol{\Sigma}_{t+1}|\nu, \mathbf{S}_t \sim \mathcal{IW}_p(\nu, \mathbf{S}_t), \quad \mathbf{S}_t = \frac{1}{\nu}\boldsymbol{\Sigma}_t^{-d/2}\mathbf{A}\boldsymbol{\Sigma}_t^{-d/2},$$

where ν and $\mathbf{S}_t$ are the degrees of freedom and the time-dependent scale parameter of the Wishart distribution, respectively, $\mathbf{A}$ is a positive definite symmetric parameter matrix, d is a scalar parameter, and $\mathbf{Q}_t^{-d/2}$ is defined by using a singular value decomposition. The quadratic expression, together with $\nu \geq p$, ensures that the covariance matrix is symmetric and positive definite. For convenience, it is assumed that $\mathbf{Q}_0 = \mathbf{I}_p$. Although their model is closely related to the models of Philipov and Glickman (2006b) and Philipov and Glickman (2006a), the MCMC fitting procedures are different. Asai and McAleer (2007) estimated these models using returns of the Nikkei 225 Index, Hang Seng Index and Straits Times Index.

Gourieroux, Jasiak, and Sufana (2004) and Gourieroux (2006) take an alternative approach and derived a Wishart autoregressive process. Let $\mathbf{Y}_t$ and $\boldsymbol{\Gamma}$ denote respectively a stochastic symmetric positive definite matrices of dimension $p \times p$ and a deterministic symmetric matrix of dimension $p \times p$. A Wishart autoregressive process of order 1 is defined to be a matrix process (denoted by $WAR(1)$ process) with conditional Laplace transform:

$$\begin{aligned}
\Psi_t(\boldsymbol{\Gamma}) &= \mathrm{E}_t\left[\exp\{\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{Y}_{t+1})\}\right] \\
&= \frac{\exp\left[\mathrm{tr}\left\{\mathbf{M}'^{-1}\mathbf{M}\mathbf{Y}_t\right\}\right]}{|\mathbf{I} - 2\boldsymbol{\Sigma}\boldsymbol{\Gamma}|^{k/2}} \qquad (40)
\end{aligned}$$

where k is a scalar degree of freedom ($k < p-1$), $\mathbf{M}$ is an $p \times p$ matrix of autoregressive parameters, and $\boldsymbol{\Sigma}$ is a $p \times p$ symmetric and positive definite matrix such that the maximal eigenvalue of $2\mathbf{\Sigma\Gamma}$ is less than 1. Here E_t denotes the expectation conditional on $\{\mathbf{Y}_t, \mathbf{Y}_{t-1}, \ldots, \}$. It can be shown that

$$\mathbf{Y}_{t+1} = \mathbf{M}\mathbf{Y}_t\mathbf{M}' + k\mathbf{\Sigma} + \boldsymbol{\eta}_{t+1},$$

where $\mathrm{E}(\boldsymbol{\eta}_{t+1}) = \mathbf{O}$. The conditional probability density function of $\mathbf{Y}_{t+1}$ is given by

$$\begin{aligned} f(\mathbf{Y}_{t+1}|\mathbf{Y}_t) &= \frac{|\mathbf{Y}_{t+1}|^{(k-p-1)/2}}{2^{kp/2}\Gamma_p(k/2)|\mathbf{\Sigma}|^{k/2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma}^{-1}(\mathbf{Y}_{t+1} + \mathbf{M}\mathbf{Y}_t\mathbf{M}')\right\}\right] \\ &\quad \times {}_0F_1(k/2; (1/4)\mathbf{M}\mathbf{Y}_t\mathbf{M}'\mathbf{Y}_{t+1}) \end{aligned}$$

where Γ_p is the multidimensional gamma function and ${}_0F_1$ is the hypergeometric function of matrix augment (see Gourieroux, Jasiak, and Sufana (2004) for details). When K is an integer and $\mathbf{Y}_t$ is a sum of outer products of k independent vector AR(1) processes such that

$$\begin{aligned} \mathbf{Y}_t &= \sum_{j=1}^{k} \mathbf{x}_{jt}\mathbf{x}_{jt}', \\ \mathbf{x}_{jt} &= \mathbf{M}\mathbf{x}_{j,t-1} + \boldsymbol{\varepsilon}_{jt}, \quad \boldsymbol{\varepsilon}_{jt} \sim N_p(\mathbf{0}, \mathbf{\Sigma}), \end{aligned} \tag{41}$$

we obtain the Laplace transform $\Psi_t(\mathbf{\Gamma})$ is given by (40). Gourieroux, Jasiak, and Sufana (2004) also introduced a Wishart autoregressive process of higher order. They estimate the $WAR(1)$ using a series of intra-day historical volatility-covolatility matrices for three stocks traded on the Toronto Stock Exchange. Finally, Gourieroux (2006) introduced the continuous time Wishart process as the multivariate extension of the Cox-Ingersoll-Ross (CIR) model in Cox, Ingersoll, and Ross (1985).

4.3.2 Factor model

Philipov and Glickman (2006a) propose an alternative factor MSV model that assumes that the factor volatilities follow an unconstrained Wishart random process. Their model has close ties to the model in Philipov and Glickman (2006b), and is given by

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \mathbf{V}^{1/2}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}),$$
$$\mathbf{f}_t|\mathbf{\Sigma}_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{\Sigma}_t), \quad \mathbf{\Sigma}_t|\nu, \mathbf{S}_{t-1} \sim \mathcal{IW}_q(\nu, \mathbf{S}_{t-1}),$$

where $\mathbf{S}_{t-1}$ is defined by (39). In other words, the conditional covariance matrix $\mathbf{\Sigma}_t$ of the factor $\mathbf{f}_t$ follows an inverse Wishart distribution whose parameter depends on the past covariance matrix $\mathbf{\Sigma}_{t-1}$. They implemented the model with $q = 2$ factors on return series data of 88 individual companies from the S&P500.

In another development, Carvalho and West (2006) proposed dynamic matrix-variate graphical models, which are based on dynamic linear models accommodated with the hyper-inverse

Wishart distribution that arises in the study of graphical models (Dawid and Lauritzen (1993) and Carvalho and West (2006)). The starting point is the dynamic linear model

$$\mathbf{y}_t' = \mathbf{X}_t'\boldsymbol{\Theta}_t + \mathbf{u}_t', \quad \mathbf{u}_t \sim \mathcal{N}_p(\mathbf{0}, v_t\boldsymbol{\Sigma}),$$
$$\boldsymbol{\Theta}_t = \mathbf{G}_t\boldsymbol{\Theta}_{t-1} + \boldsymbol{\Omega}_t, \quad \boldsymbol{\Omega}_t \sim \mathcal{N}_{q\times p}(O, \mathbf{W}_t, \boldsymbol{\Sigma}),$$

where $\mathbf{y}_t$ is the $p \times 1$ vector of observations, $\mathbf{X}_t$ is a known $q \times 1$ vector of explanatory variables, $\boldsymbol{\Theta}_t$ is the $q \times p$ matrix of states, $\mathbf{u}_t$ is the $p \times 1$ innovation vector for observation, $\boldsymbol{\Omega}_t$ is the $q \times p$ innovation matrix for states, $\mathbf{G}_t$ is a known $q \times q$ matrix, and $\boldsymbol{\Sigma}$ is the $p \times p$ covariance matrix. $\boldsymbol{\Omega}_t$ follows a matrix-variate normal with mean $\mathbf{O}$ $(q \times p)$, left covariance matrix $\mathbf{W}_t$ and right covariance matrix $\boldsymbol{\Sigma}$; in other words, any row $\boldsymbol{\omega}_{it}$ of $\boldsymbol{\Omega}_t$ has a multivariate normal distribution $\mathcal{N}_q(\mathbf{0}, \sigma_{ii}\mathbf{W}_t)$, while any row $\boldsymbol{\omega}_t^i$ of $\boldsymbol{\Omega}_t$, $\boldsymbol{\omega}_t^{i\prime}$ has a multivariate normal distribution $\mathcal{N}_p(\mathbf{0}, w_{ii,t}\boldsymbol{\Sigma})$. Next, we suppose that $\boldsymbol{\Sigma} \sim \mathcal{HIW}_p(b, \mathbf{D})$, the hyper-inverse Wishart distribution with a degree-of-freedom parameter b and location matrix $\mathbf{D}$. It should be noted that the dynamic linear model with $\boldsymbol{\Sigma} \sim \mathcal{HIW}_p(b, \mathbf{D})$ can be handled from the Bayesian perspective without employing simulation-based techniques. Finally, instead of time-invariant $\boldsymbol{\Sigma}$, Carvalho and West (2006) suggested a time-varying process given by

$$\boldsymbol{\Sigma}_t \sim \mathcal{HIW}_p(b_t, \mathbf{S}_t),$$
$$b_t = \delta b_{t-1} + 1,$$
$$\mathbf{S}_t = \delta\mathbf{S}_{t-1} + \mathbf{v}_t\mathbf{v}_t',$$

where $\mathbf{v}_t$ is defined by Theorem 1 of Carvalho and West (2006). Intuitively, $\mathbf{v}_t$ is the residual from the observation equation. As $\boldsymbol{\Sigma}_t$ appears in both of the observation and state equations, the proposed dynamic matrix-variate graphical model can be considered as a variation of the "Factor MSV model with MSV error." Setting $\delta = 0.97$, Carvalho and West (2006) applied the dynamic matrix-variate graphical models to two datasets; namely (i) 11 international currency exchange rates relative to US dollar, and (ii) 346 securities from the S&P500 stock index.

References

Aas, K. and I. H. Haff (2006). The generalized hyperbolic skew student's t-distribution. *Journal of Financial Econometrics 4*, 275–309.

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics 18*, 338–357.

Albert, J. H. and S. Chib (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics 11*, 1–15.

Anderson, B. D. O. and J. B. Moore (1979). *Optimal filtering*. Englewood Cliffs, New Jersey: Prentice-Hall.

Arellano-Valle, R. B. and A. Azzalini (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics 33*, 561–574.

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

Asai, M. and M. McAleer (2007). The structure of dynamic correlations in multivariate stochastic volatility models. Unpublished paper: Faculty of Economics, Soka University.

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

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics 32*, 159–188.

Azzalini, A. and A. Capitanio (2003). Distributions generated by perterbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society, Series B 65*, 367–389.

Barndorff-Neilsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of the particle size. *Proceedings of the Royal Society. London. Series A. Mathematical and Physical Sciences 353*, 401–419.

Barndorff-Neilsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Ulhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B 63*, 167–241.

Bauwens, L. and S. Laurent (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics 23*, 346–354.

Blæsild, P. (1981). The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data. *Biometrika 68*, 251–263.

Bollerslev, T. (1990). Modelling the coherence in the short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics 72*, 498–505.

Bollerslev, T., R. F. Engle, and J. Woodridge (1988). A capital asset pricing model with time varying covariances. *Journal of Political Economy 96*, 116–131.

Bos, C. S. and N. Shephard (2006). Inference for adaptive time series models: stochastic volatility and conditionally Gaussian state space form. *Econometric Reviews 25*, 219–244.

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

Calvet, L. E. and A. J. Fisher (2001). Forecasting multifractal volatility. *Journal of Econometrics 105*, 27–58.

Calvet, L. E., A. J. Fisher, and S. B. Thompson (2006). Volatility comovement: a multifrequency approach. *Journal of Econometrics 131*, 179–215.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*, 541–553.

Carvalho, C. M. and M. West (2006). Dynamic matrix-variate graphical models. *Bayesian analysis 1*, 1–29.

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 (1994). Bayes inference for regression models with ARMA(p, q) errors. *Journal of Econometrics 64*, 183–206.

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistican 49*, 327–35.

Chib, S. and E. Greenberg (1998). Analysis of multivariate probit models. *Biometrika 85*, 347–61.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for generalized stochastic volatility models. *Journal of Econometrics 108*, 281–316.

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

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

Cox, J., J. Ingersoll, and S. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*, 385–407.

Dagpunar, J. S. (1989). An easily implemented generalized inverse Gaussian generator. *Communications in Statistics-Simulations 18*, 703–710.

Daníelsson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics 64*, 375–400.

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

Daníelsson, J. and J.-F. Richard (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics 8*, 153–173.

Dawid, A. P. and S. L. Lauritzen (1993). Hyper-Markov laws in the statistical analysis. *Annals of Statistics 3*, 1272–1317.

de Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics 19*, 1073–1083.

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

Dey, D. and J. Liu (2005). A new construction for skew multivariate distributions. *Journal of Multivariate Analysis 95*, 323–344.

Diebold, F. X., T. A. Gunther, and T. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*, 863–883.

Doornik, J. A. (2002). *Object-Oriented Matrix Programming Using Ox, 3rd ed.* London: Timberlake Consultants Press and Oxford. www.nuff.ox.ac.uk/Users/Doornik.

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

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika 84*, 669–684.

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.

Ferreira, J. T. A. S. and M. F. J. Steel (2004). Bayesian multivariate regression analysis with a new class of skewed distributions. Statistics Research Report 419, University of Warwick.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*, 183–202.

Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In G. S. M. Rao, C. R. (Ed.), *Statistical Models in Finance (Handbook of Statistics)*, pp. 119–191. Amsterdam: North-Holland.

Gilks, W. R., N. G. Best, and K. K. C. Tan (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics 44*, 455–472.

Godsill, S. J., A. Doucet, and M. West (2004). Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association 99*, 156–168.

Gourieroux, C. (2006). Continuous time Wishart process for stochastic risk. *Econometric Reviews 25*, 177–217.

Gourieroux, C., J. Jasiak, and R. Sufana (2004). The Wishart autoregressive process of multivariate stochastic volatility. Discussion paper: University of Toronto.

Gupta, A. K., G. González-Farías, and J. A. Domínguez-Molina (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis 89*, 181–190.

Han, Y. (2006). The economics value of volatility modelling: Asset allocation with a high dimensional dynamic latent factor multivariate stochastic volatility model. *Review of Financial Studies 19*, 237–271.

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

Harvey, A. C. and N. Shephard (1996). Estimation of asymmetric stochastic volatility model for asset returns. *14*, 429–434.

Hörmann, W., J. Leydold, and G. Derflinger (2004). *Automatic Nonuniform Random Variate Generation*. Berlin: Springer.

Hürzeler, M. and H. R. Künsch (1998). Monte Carlo approximations for general state-space models. *Journal of Computational and Graphical Statistics 7*, 175–193.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics 12*, 371–389.

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

Jungbacker, B. and S. J. Koopman (2005). On importance sampling for state space models. Tinbergen Institute Discussion papers, 05-117/4, Tinbergen Institute.

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.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics 5*, 1–25.

Kitagawa, G. (1998). A self-organizing state-space model. *Journal of the American Statistical Association 93*, 1203–1215.

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

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

Lopes, H. F. and C. M. Carvalho (2006). Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference*. forthcoming.

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

Omori, Y. and T. Watanabe (2003). Block sampler and posterior mode estimation for a nonlinear and non-Gaussian state-space model with correlated errors. Discussion paper series, CIRJE-F-221, Faculty of Economics, Universty of Tokyo.

Philipov, A. and M. E. Glickman (2006a). Factor multivariate stochastic volatility via Wishart processes. *Econometric Reviews 25*, 311–334.

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

Pitt, M. K., D. Chan, and R. Kohn (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika 93*, 537–554.

Pitt, M. K. and N. Shephard (1999a). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association 94*, 590–599.

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

Protassov, R. S. (2004). EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ. *Statistics and Computing 14*, 67–77.

Quintana, J. M. and M. West (1987). An analysis of international exchange rates using multivariate DLMs. *The Statistician 36*, 275–281.

Ray, B. K. and R. S. Tsay (1997). Identifying common long-range dependence in a vector time series. Unpublished paper: University of Chicago.

Ray, B. K. and R. S. Tsay (2000). Long-range dependence in daily stock volatilities. *Journal of Business and Economic Statistics 18*, 254–262.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics 23*, 470–2.

Schmidt, R., T. Hrycej, and E. Stützle (2006). Multivariate distribution models with generalized hyperbolic margins. *Computational Statistics and Data Analysis 50*, 2065–20096.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika 81*, 115–131.

Shephard, N. (2004). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.

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

Smith, J. Q. (1985). Diagnostic checks of non-standard time series models. *Journal of Forecasting 4*, 283–91.

Smith, M. and A. Pitts (2006). Foreign exchange intervention by the Bank of Japan: Bayesian analysis using a bivariate stochastic volatility model. *Econometric Reviews 25*, 425–451.

So, M. K. P. and W. Y. Kwok (2006). A multivariate long memory stochastic volatility model. *Physica A 362*, 450–464.

So, M. K. P., K. Lam, and W. K. Li (1998). A stochastic volatility model with Markov switching. *JBES 16*, 244–253.

So, M. K. P., W. K. Li, and K. Lam (1997). Multivariate modelling of the autoregressive random variance process. *Journal of Time Series Analysis 18*, 429–446.

Tanizaki, H. (2001). Nonlinear and non-Gaussian state space modeling using sampling techniques. *Annals of the Institute of Statistical Mathematics 53*, 63–81.

Tanizaki, H. and R. S. Mariano (1998). Nonlinear and non-Gaussian state space modeling with Monte Carlo simulations. *Journal of Econometrics 83*, 263–290.

Tims, B. and R. Mahieu (2006). A range-based multivariate stochastic volatility model for exchange rates. *Econometric Reviews 25*, 409–424.

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

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

Wong, F., C. Carter, and R. Kohn (2003). Efficient estimation of covariance matrix selection models. *Biometrika 90*, 809–830.

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.