Multivariate Stochastic Volatility

Siddhartha Chib, Yasuhiro Omori and Manabu Asai

Abstract We provide a detailed summary of the large and vibrant emerging literature that deals with the multivariate modeling of conditional volatility of financial time series within the framework of stochastic volatility. The developments and achievements in this area represent one of the great success stories of financial econometrics. Three broad classes of multivariate stochastic volatility models have emerged: one that is a direct extension of the univariate class of stochastic volatility model, another that is related to the factor models of multivariate analysis and a third that is based on the direct modeling of time-varying correlation matrices via matrix exponential transformations, Wishart processes and other means. We discuss each of the various model formulations, provide connections and differences and show how the models are estimated. Given the interest in this area, further significant developments can be expected, perhaps fostered by the overview and details delineated in this paper, especially in the fitting of high-dimensional models.

Siddhartha Chib
Washington University in St. Louis, Campus Box 1133, 1 Brookings Dr., St. Louis, MO 63130, USA, e-mail: `chib@wustl.edu`

Yasuhiro Omori
Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, e-mail: `omori@e.u-tokyo.ac.jp`

Manabu Asai
Faculty of Economics, Soka University, 1-236 Tangi-cho, Hachioji-shi Tokyo, 192-8577, Japan, e-mail: `m-asai@soka.ac.jp`

1 Introduction

A considerable amount of recent literature on financial econometrics has emerged on the modeling of conditional volatility, spurred by the demand for such models in areas such as portfolio and risk management. Much of the early interest centered on multivariate versions of univariate generalized autoregressive conditional heteroscedasticity (GARCH) models. These generalizations have been ably summarized in recent surveys, for example, those of Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2007). More recently, a large and prolific (parallel) body of literature has developed around generalizations of the univariate stochastic volatility (SV) model. A number of multivariate SV (MSV) models are now available along with clearly articulated estimation recipes. Our goal in this paper is to provide a detailed summary of these 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. We note that our treatment does not include any discussion of multivariate modeling of volatility that is relevant for ultra-high-frequency data. Thus, there is no discussion of realized volatility (Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004)).

   To fix notation and set the stage for our developments, the univariate SV model that forms the basis for many MSV models is given by (Ghysels et al. (1996), Broto and Ruiz (2004) and Shephard (2004)):

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

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

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

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} | h_t \sim \mathcal{N}_2(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}, \tag{4}$$

where y_t is a univariate outcome, h_t is a univariate latent variable and $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{N}_m(\mu, \mathbf{\Sigma})$ denote, respectively, a univariate normal distribution with mean μ and variance σ^2, and an m-variate normal distribution with mean vector μ and variance-covariance matrix $\mathbf{\Sigma}$. 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 nonlinear measurement equation (because h_t enters the outcome model nonlinearly). 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 to let the model errors have a non-Gaussian fat-tailed distribution, to permit jumps, and incorporate the leverage effect (through a nonzero 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 been fully resolved by the creative use of Monte Carlo methods, primarily Bayesian Markov chain Monte Carlo (MCMC) methods (Jacquier et al. (1994), Kim et al. (1998), Chib et al. (2002) and Omori et al. (2007)). We refer the readers to Asai et al. (2006) for a discussion of how this problem can be addressed in some special cases by non-Bayesian methods. In this survey, on the other hand, we concentrate on Bayesian methods but mention the full range of methods (Bayesian and non-Bayesian) that have been tried for the various models.

In the multivariate case, 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. A typical starting point is the assumption of series-specific log volatilities h_{tj} $(j \leq p)$ whose joint evolution is governed by a first-order stationary vector autoregressive process:

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

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

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

are both diagonal matrices. We refer to the former specification as the *VAR(1)* model and the latter as the *IAR(1)* (for independent autoregressive) 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

$$\mathbf{y}_t = \mathbf{V}_t^{1/2}\varepsilon_t, \quad \mathbf{V}_t^{1/2} = \text{diag}\left(\exp(h_{1t}/2), \ldots, \exp(h_{pt}/2)\right), \quad t = 1, \ldots, n,$$

with the additional assumptions that

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

and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ is a matrix in correlation (with units on the main diagonal). Thus, conditioned on $\mathbf{h}_t$, $\text{Var}(\mathbf{y}_t) = \mathbf{V}_t^{1/2}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\mathbf{V}_t^{1/2}$ 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

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \mathbf{V}_t^{1/2}\varepsilon_t, \ \mathbf{V}_t^{1/2} = \text{diag}\left(\exp(h_{1t}/2), \ldots, \exp(h_{pt}/2)\right),$$

where $\mathbf{B}$ is a $p \times q$ matrix ($q \leq p$) 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

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

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

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

and as a result the conditional correlation matrix is time-varying.

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

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

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 $\boldsymbol{\Sigma}_{t-1}$. 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 and Section 5 has our conclusions.

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 one approach to modeling the conditional covariance matrix of $\mathbf{y}_t$ is to assume that

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

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

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

where

$$\mathbf{V}_t^{1/2} = \text{diag}\left(\exp(h_{1t}/2), \ldots, \exp(h_{pt}/2)\right),$$

$$\mu = (\mu_1, \ldots, \mu_p)'$$

and

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

Of course, for identification purposes, the diagonal elements of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ must be 1, which means that the matrix $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ is a correlation matrix.

Analyses of this model are given by Harvey et al. (1994), Daníelsson (1998), Smith and Pitts (2006) and Chan et al. (2006). Actually, Harvey et al. (1994) dealt with a special case of this model in which $\mathbf{\Phi} = \text{diag}(\phi_1, \ldots, \phi_p)$. To fit the model, the measurement equation (5) is linearized by letting $w_{it} = \log y_{it}^2$. Because

$$\mathrm{E}(\log \varepsilon_{it}^2) = -1.27, \quad \mathrm{Var}(\log \varepsilon_{it}^2) = \pi^2/2, \tag{8}$$

one now has (a non-Gaussian) linear measurement equation:

$$\mathbf{w}_t = (-1.27)\mathbf{1} + \mathbf{h}_t + \xi_t, \tag{9}$$

where $\mathbf{w}_t = (w_{1t}, \ldots, w_{pt})'$, $\xi_t = (\xi_{1t}, \ldots, \xi_{pt})'$, $\xi_{it} = \log \varepsilon_{it}^2 + 1.27$ and $\mathbf{1} = (1, \ldots, 1)'$. Although the new state error ξ_t does not follow a normal distribution, approximate or quasi maximum likelihood (QML) estimates can be obtained by assuming Gaussianity. Calculation of the (misspecified) Gaussian likelihood also requires the covariance matrix of ξ_t. Harvey et al. (1994) showed that the (i, j)th element of the covariance matrix of $\xi_t = (\xi_{1t}, \ldots, \xi_{pt})'$ is given by $(\pi^2/2)\rho_{ij}^*$, where $\rho_{ii}^* = 1$ and

$$\rho_{ij}^* = \frac{2}{\pi^2} \sum_{n=1}^{\infty} \frac{(n-1)!}{\{\prod_{k=1}^{n}(1/2 + k - 1)\}n} \rho_{ij}^{2n}. \tag{10}$$

The model was applied to four daily foreign exchange rates (pound/dollar, Deutschemark/dollar, yen/dollar and Swiss franc/dollar). As mentioned in Harvey et al. (1994), the preceding fitting method cannot be extended to the leverage model considered below.

So et al. (1997) provide a similar analysis, but unlike Harvey et al. (1994) the nondiagonal elements of $\mathbf{\Phi}$ are not assumed to equal zero. Estimation of the parameters is again by the QML method which is implemented through a computationally efficient and numerically well-behaved expectation-maximization (EM) algorithm. The asymptotic variance-covariance matrix of the resulting estimates is based on the information matrix. Another related contribution is that of Daníelsson (1998), where the model

$$\mathbf{y}_t = \mathbf{V}_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}_{\varepsilon\varepsilon}),$$
$$\mathbf{h}_{t+1} = \mu + \text{diag}(\phi_1, \ldots, \phi_p)(\mathbf{h}_t - \mu) + \eta_t, \quad \eta_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}_{\eta\eta})$$

is analyzed. The parameters of this model are estimated by the simulated maximum likelihood method. The model and fitting method should be 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). On the basis of 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 et al. (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

$$\mathbf{y}_t = \mathbf{V}_t^{1/2}\varepsilon_t, \quad \mathbf{V}_t^{1/2} = \text{diag}(\exp(h_{1t}/2), \exp(h_{2t}/2)), \quad \varepsilon_t \sim \mathcal{N}_2(\mathbf{0}, \mathbf{\Sigma}_{\varepsilon\varepsilon}),$$
$$\mathbf{h}_{t+1} = \mathbf{Z}_t\alpha + \text{diag}(\phi_1, \phi_2)(\mathbf{h}_t - \mathbf{Z}_{t-1}\alpha) + \eta_t, \quad \eta_t \sim \mathcal{N}_2(\mathbf{0}, \mathbf{\Sigma}_{\eta\eta}),$$
$$\mathbf{h}_1 \sim \mathcal{N}_2(\mathbf{Z}_1\alpha_1, \mathbf{\Sigma}_0),$$

where the (i, j)th element of $\mathbf{\Sigma}_0$ is the (i, j)th element of $\mathbf{\Sigma}_{\eta\eta}$ divided by $1 - \phi_i\phi_j$ to enforce the stationarity of $\mathbf{h}_t - \mathbf{Z}_t\alpha$. 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 MCMC methods (Chib and Greenberg (1996), Chib (2001)). Because the likelihood of the parameters is complex, sampling of the posterior distributions in all applications of MCMC methods in MSV models is indirectly achieved by sampling the posterior distribution of the parameters and each of the latent variables. This tactic circumvents the computation of the likelihood. For this tactic to work it is necessary to efficiently sample the resulting high-dimensional posterior distribution. This is the challenge that has to be surmounted on a model-by-model basis.

To improve the efficiency of the MCMC algorithm, Smith and Pitts (2006) 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 MCMC algorithm of Smith and Pitts (2006) is given by:

1. Sample $\{\mathbf{h}_t\}_{t=1}^n | \rho_{12}, \phi_1, \phi_2, \alpha, \boldsymbol{\Sigma}_{\eta\eta}, Y_n$. Divide $\{\mathbf{h}_t\}_{t=1}^n$ into 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 Metropolis–Hastings (M-H) algorithm using a proposal density of the type introduced by Chib and Greenberg (1994, 1998) and Chib (2001):

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

where

$$l(\mathbf{h}_{a:b}) = \text{const} - \frac{1}{2} \cdot \sum_{t=a}^b \left(\mathbf{1}'\mathbf{h}_t + \mathbf{y}_t' \mathbf{V}_t^{-1/2} \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} \mathbf{V}_t^{-1/2} \mathbf{y}_t \right) - \frac{1}{2} \cdot$$

$$\sum_{t=a}^{b+1} \{\mathbf{h}_t - \mathbf{Z}_t\alpha - \boldsymbol{\Phi}(\mathbf{h}_{t-1} - \mathbf{Z}_{t-1}\alpha)\}' \boldsymbol{\Sigma}_{\eta\eta}^{-1} \{\mathbf{h}_t - \mathbf{Z}_t\alpha - \boldsymbol{\Phi}(\mathbf{h}_{t-1} - \mathbf{Z}_{t-1}\alpha)\}.$$

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.

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

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

$$\boldsymbol{\Sigma}^{-1} = \sum_{t=2}^n (\mathbf{Z}_t - \boldsymbol{\Phi}\mathbf{Z}_{t-1})' \boldsymbol{\Sigma}_{\eta\eta}^{-1}(\mathbf{Z}_t - \boldsymbol{\Phi}\mathbf{Z}_{t-1}) + \mathbf{Z}_1' \boldsymbol{\Sigma}_0^{-1}\mathbf{Z}_1.$$

5. Sample $\boldsymbol{\Sigma}_{\eta\eta} | \{\mathbf{h}_t\}_{t=1}^n, \rho_{12}, \phi_1, \phi_2, \alpha, Y_n$ 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 α_t:

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

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

$$\mathbf{y}_t = \mathbf{Z}_t\alpha_t + \mathbf{u}_t^*, \quad \mathbf{u}_t^* \sim \mathcal{N}_p(\mathbf{0}, \mathbf{G}_t\mathbf{V}_t\mathbf{G}_t'),$$
$$\alpha_{t+1} = \mathbf{T}_t\alpha_t + \mathbf{v}_t^*, \quad \mathbf{v}_t^* \sim \mathcal{N}_r(\mathbf{0}, \mathbf{H}_t\mathbf{V}_t\mathbf{H}_t'),$$

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

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

   – Sample $\theta | \{\mathbf{h}_t\}_{t=1}^n, Y_n$ using a M-H algorithm or a step from the adaptive rejection Metropolis sampler by Gilks et al. (1995); see Bos and Shephard (2006).
   – Sample $\{\alpha_t\}_{t=1}^n | \theta, \{\mathbf{h}_t\}_{t=1}^n, Y_n$ using a simulation smoother for a linear Gaussian state-space model; see de Jong and Shephard (1995) and Durbin and Koopman (2002). We first sample disturbances of the linear Gaussian state-space model and obtain samples of α_t recursively.

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

$$\mathbf{h}_t | \mathbf{h}_{t-1}, \mathbf{h}_{t+1}, \theta \sim \mathcal{N}_q(\mu + \mathbf{Q}\mathbf{\Phi}'\mathbf{\Sigma}_{\eta\eta}^{-1}\{(\mathbf{h}_{t+1} - \mu) + (\mathbf{h}_{t-1} - \mu)\}, \mathbf{Q}),$$
$$t = 2, \ldots, n-1,$$

$$\mathbf{h}_n | \mathbf{h}_{n-1}, \theta \sim \mathcal{N}_q(\mu, \mathbf{\Sigma}_{\eta\eta}),$$

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

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 θ seems to overcome some of the inefficiencies. Yu and Meyer (2006) provide a survey of MSV models that proceed along these lines and illustrate how the Bayesian software program WinBUGS can be used to fit bivariate models.

It is worth mentioning that it is possible to relax the assumption that the volatility process is VAR of order 1. In one notable attempt, So and Kwok (2006) consider a MSV model where the volatility vector $\mathbf{h}_t - \mu$ follows a stationary vector autoregressive fractionally integrated moving average process,

ARFIMA$(p, \mathbf{d}, q)$, such that

$$\boldsymbol{\Phi}(B)D(B)(\mathbf{h}_{t+1} - \mu) = \Theta(B)\eta_t, \quad \eta_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\eta}), \tag{11}$$

$$D(B) = \operatorname{diag}((1-B)^{d_1}, \ldots, (1-B)^{d_p}), \quad |d_i| < 1/2, \tag{12}$$

$$\boldsymbol{\Phi}(B) = \mathbf{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p, \tag{13}$$

$$\Theta(B) = \mathbf{I} + \boldsymbol{\Theta}_1 B + \cdots + \boldsymbol{\Theta}_q B^q, \tag{14}$$

where B is a backward operator such that $B^j \mathbf{h}_t = \mathbf{h}_{t-j}$. The ε_t and η_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 et al. (1994). They linearized the measurement equation by taking the logarithm of the squared returns and considered the linear state-space model

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

$$\boldsymbol{\Phi}(B)D(B)(\mathbf{h}_{t+1} - \mu) = \Theta(B)\eta_t,$$

where $\mathbf{w}_t = (w_{1t}, \ldots, w_{pt})'$, $\xi_t = (\xi_{1t}, \ldots, \xi_{pt})'$, $w_{it} = \log y_{it}^2$, and $\xi_{it} = \log \varepsilon_{it}^2$ for $i = 1, \ldots, n$. The covariance matrix of ξ_t can be obtained as in Harvey et al. (1994). To conduct the QML estimation, So and Kwok (2006) assumed that ξ_t follows a normal distribution and obtained estimates based on the linear Gaussian state-space model. However, since $\mathbf{h}_t - \mu$ follows a vector ARFIMA$(p, \mathbf{d}, q)$ process, the conventional Kalman filter is not applicable as the determinant and inverse of a 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

Another extension of the basic MSV model is to allow for correlation between ε_t and η_t by letting $\boldsymbol{\Sigma}_{\varepsilon\eta} \neq \mathbf{O}$. This extension is important because at least for returns on stocks there is considerable evidence that the measurement and volatility innovations are correlated (Yu (2005), Omori et al. (2007)). That this correlation (the leverage effect) should be modeled is mentioned by Daníelsson (1998) but this suggestion is not implemented in his empirical study of foreign exchange rates and stock indices. One compelling work on a type of leverage model is due to Chan et al. (2006), who considered the model

$$\mathbf{y}_t = \mathbf{V}_t^{1/2}\varepsilon_t,$$

$$\mathbf{h}_{t+1} = \mu + \operatorname{diag}(\phi_1, \ldots, \phi_p)(\mathbf{h}_t - \mu) + \boldsymbol{\Psi}^{1/2}\eta_t,$$

$$\mathbf{h}_1 \sim \mathcal{N}_p(\mu, \boldsymbol{\Psi}^{1/2}\boldsymbol{\Sigma}_0 \boldsymbol{\Psi}^{1/2}),$$

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

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

$$\boldsymbol{\Psi}^{1/2} = \mathrm{diag}\left(\sqrt{\psi_1^2}, \ldots, \sqrt{\psi_p^2}\right),$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}_{2p}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & \boldsymbol{\Sigma}_{\varepsilon\eta} \\ \boldsymbol{\Sigma}_{\eta\varepsilon} & \boldsymbol{\Sigma}_{\eta\eta} \end{pmatrix}.$$

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

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

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 that are 1's 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 et al. (2003) proposed a hierarchical prior for $\mathbf{G}$:

$$\pi(d\mathbf{G}|\mathbf{J}) = V(\mathbf{J})^{-1}d\mathbf{G}_{\{J=1\}}I(\mathbf{G} \in \mathcal{A}), \quad V(\mathbf{J}) = \int_{\mathbf{G} \in \mathcal{A}} d\mathbf{G}_{\{J=1\}},$$

$$\pi(\mathbf{J}|S(\mathbf{J}) = l) = \frac{V(\mathbf{J})}{\sum_{\mathbf{J}^*:S(\mathbf{J}^*)=l} V(\mathbf{J}^*)},$$

$$\pi(S(\mathbf{J}) = l|\varphi) = \binom{p(2p-1)}{l} \varphi^l(1-\varphi)^{p(2p-1)-l}.$$

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 et al. (2003) for the evaluation of $V(\mathbf{J})$. Let $\phi = (\phi_1, \ldots, \phi_p)'$ and $\psi = (\psi_1, \ldots, \psi_p)'$ $(\psi_j > 0, j = 1, \ldots, p)$.

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

of the diagonal elements

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

Propose a candidate

$$\phi \sim \mathcal{TN}_R(\mu_\phi, \mathbf{\Sigma}_\phi), \quad R = \{\phi : \phi_j \in (-1, 1), j = 1, \ldots, p\},$$

$$\mathbf{\Sigma}_\phi^{-1} = \mathbf{\Sigma}^{22} \odot \left\{ \sum_{t=1}^{n-1} \mathbf{\Psi}^{-1/2}(\mathbf{h}_t - \mu)(\mathbf{h}_t - \mu)'\mathbf{\Psi}^{-1/2} \right\},$$

$$\mu_\phi = \mathbf{\Sigma}_\phi \mathbf{d},$$

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

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

$$\mathbf{\Sigma}_*^{-1} = (n-1)(\mathbf{I} - \mathbf{\Phi})\mathbf{\Psi}^{-1/2}\mathbf{\Sigma}^{22}\mathbf{\Psi}^{-1/2}(\mathbf{I} - \mathbf{\Phi}) + \mathbf{\Psi}^{-1/2}\mathbf{\Sigma}_0^{-1}\mathbf{\Psi}^{-1/2},$$

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

3. Sample $\psi|\phi, \mu, \{\mathbf{h}_t\}_{t=1}^n, \mathbf{\Sigma}, Y_n$. 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 the M-H algorithm using a truncated multivariate t distribution on the region $R = \{\mathbf{v} : v_j > 0, j = 1, \ldots, p\}$ with six 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 $\{\mathbf{h}_t\}_{t=1}^n|\phi, \mu, \psi, \mathbf{\Sigma}, Y_n$. We divide $\{\mathbf{h}_t\}_{t=1}^n$ into 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,

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

$$l(\mathbf{h}_{a:b}) = \text{const} - \frac{1}{2}\sum_{t=a}^{b} \mathbf{1}'\mathbf{h}_t - \frac{1}{2}\sum_{t=a}^{b+1} \mathbf{r}_t'\mathbf{\Sigma}^{-1}\mathbf{r}_t,$$

$$\mathbf{r}_t = \begin{pmatrix} \mathbf{V}_t^{-1/2}\mathbf{y}_t \\ \mathbf{\Psi}^{-1/2}\{\mathbf{h}_{t+1} - \mu - \mathbf{\Phi}(\mathbf{h}_t - \mu)\} \end{pmatrix},$$

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 the Newton–Raphson method numerically. The analytical derivatives can be derived similarly as in the Appendix of Chan et al. (2006).

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

Chan et al. (2006) applied the proposed estimation method to equities at three levels of aggregation: (1) returns for eight different markets (portfolios of stocks in NYSE, AMEX, NASDAQ and S&P500 indices); (2) 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); (3) returns for individual firms within the same industry. They found strong evidence of correlation between ε_t and η_{t-1} only for the returns of the eight different markets and suggested that this correlation is mainly a feature of marketwide rather than firm-specific returns and volatility.

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

$$\boldsymbol{\Phi} = \text{diag}(\phi_1, \ldots, \phi_p),$$
$$\boldsymbol{\Sigma}_{\varepsilon\eta} = \text{diag}\left(\lambda_1 \sigma_{1,\eta\eta}, \ldots, \lambda_p \sigma_{p,\eta\eta}\right).$$

The cross-asset leverage effects are assumed to be 0 ($\text{Corr}(\varepsilon_{it}, \eta_{jt}) = 0$, 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:

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

where $\text{E}(\zeta_{it}) = -1.27$, and $\text{Cov}(\zeta_{it}, \zeta_{jt}) = (\pi^2/2)\rho_{ij}^*$ given in (10). The matrix $\boldsymbol{\Sigma}_{\eta_t^* \eta_t^*}$ and $\text{E}(\eta_t^* \zeta_t')$ are given in Asai and McAleer (2006). They also considered an alternative MSV model with leverage effects and size effects given by

$$\mathbf{h}_{t+1} = \tilde{\mu} + \boldsymbol{\Gamma}_1 \mathbf{y}_t + \boldsymbol{\Gamma}_2 |\mathbf{y}_t| + \boldsymbol{\Phi}\mathbf{h}_t + \eta_t,$$
$$\boldsymbol{\Gamma}_1 = \text{diag}(\gamma_{11}, \ldots, \gamma_{1p}), \quad \boldsymbol{\Gamma}_2 = \text{diag}(\gamma_{21}, \ldots, \gamma_{2p}),$$
$$|\mathbf{y}_t| = (|y_{1t}|, \ldots, |y_{pt}|)', \quad \boldsymbol{\Phi} = \text{diag}(\phi_1, \ldots, \phi_p),$$
$$\boldsymbol{\Sigma}_{\varepsilon\eta} = \mathbf{O}.$$

This model is a generalization of a univariate model given by Daníelsson (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 found that the MSV model with leverage and size effects is preferred in terms of the Akaike information criterion (AIC) and Bayesian information criterion (BIC) measures.

2.3 Heavy-tailed measurement error models

It has by now been 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 occurred in Harvey et al. (1994), who assumed that in connection with the measurement error ε_{it} that

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

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

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

$$\text{and } \quad \mathrm{E}(\lambda_{it}^{-1/2}) = \frac{(\nu_i/2)^{1/2}\Gamma((\nu_i - 1)/2)}{\Gamma(\nu_i/2)}.$$

Alternatively, the model can now be expressed as

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

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 mean and covariance matrix of $(\log \epsilon_{it}^2, \log \epsilon_{jt}^2)$ using

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

and (8) and (10), 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 ε_t, in which case the measurement error has the form

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

They mentioned that this formulation was empirically better supported than the formulation in Harvey et al. (1994). The model was fit by Bayesian MCMC methods.

Another alternative to the Gaussian distribution is the generalized hyperbolic (GH) distribution introduced by Barndorff-Nielsen (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 GH 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 (Barndorff-Nielsen and Shephard (2001)). Aas and Haff (2006) 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 Norwegian kroner/euro exchange rate and the EURIBOR five-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 $\mathbf{x} \sim \mathcal{GH}(\nu, \alpha, \beta, \mathbf{m}, \delta, \mathbf{S})$ follows a multivariate GH distribution with density

$$f(\mathbf{x}) = \frac{(\gamma/\delta)^\nu K_{\nu - \frac{p}{2}}\left(\alpha\sqrt{\delta^2 + (\mathbf{x} - \mathbf{m})'\mathbf{S}^{-1}(\mathbf{x} - \mathbf{m})}\right) \exp\{\beta'(\mathbf{x} - \mathbf{m})\}}{(2\pi)^{\frac{p}{2}} K_\nu(\delta\gamma) \left\{\alpha^{-1}\sqrt{\delta^2 + (\mathbf{x} - \mathbf{m})'\mathbf{S}^{-1}(\mathbf{x} - \mathbf{m})}\right\}^{\frac{p}{2} - \nu}},$$

$$(15)$$

$$\gamma \equiv \sqrt{\alpha^2 - \beta'\mathbf{S}\beta} \geq 0, \quad \alpha^2 \geq \beta'\mathbf{S}\beta,$$
$$\nu, \alpha \in R, \quad \beta, \mathbf{m} \in R^n, \quad \delta > 0,$$

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 $|\mathbf{S}| = 1$ (Protassov (2004), Schmidt et al. (2006)). It can be shown that $\mathbf{x}$ can be expressed as

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

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

$$\delta > 0, \gamma^2 \geq 0, \text{ if } \nu < 0,$$
$$\delta > 0, \gamma^2 > 0, \text{ if } \nu = 0,$$
$$\delta \geq 0, \gamma^2 > 0, \text{ if } \nu > 0.$$

For the generation of a random sample from $\mathcal{GIG}(\nu, a, b)$, see Dagpunar (1989), Doornik (2002) and Hörmann et al. (2004). The estimation of such a multivariate distribution would be difficult and Protassov (2004) relied on the EM algorithm with ν fixed and fit 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 et al. (2006) proposed an alternative class of distributions, called the multivariate affine GH 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 et al. (2004) and Ferreira and Steel (2004).

3 Factor MSV Model

3.1 Volatility factor model

A weakness of the preceding MSV models is that the implied conditional correlation matrix does not vary with time. One approach for generating time-varying correlations is via factor models in which the factors follow a SV process. One type of factor SV model (that however does not lead to time-varying correlations) was considered by Quintana and West (1987) and by Jungbacker and Koopman (2006), who utilized a single factor to decompose the outcome into two multiplicative components, a scalar common volatility factor and a vector of idiosyncratic noise variables, as

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

where h_t is a scalar. The first element in $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ is assumed to be 1 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 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 US 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 was considered by Harvey et al. (1994), who introduced a common factor in the linearized state-space version of the basic MSV model by letting

$$\mathbf{w}_t = (-1.27)\mathbf{1} + \mathbf{\Theta}\mathbf{h}_t + \overline{\mathbf{h}} + \xi_t, \tag{16}$$

$$\mathbf{h}_{t+1} = \mathbf{h}_t + \eta_t, \quad \eta_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}), \tag{17}$$

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

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

Harvey et al. (1994) estimated the parameters by the QML method. To make the factor loadings interpretable, the common factors are rotated such that $\mathbf{\Theta}^* = \mathbf{\Theta}\mathbf{R}'$ and $\mathbf{h}_t^* = \mathbf{R}\mathbf{h}_t$, where $\mathbf{R}$ is an orthogonal matrix.

Tims and Mahieu (2006) considered a similar but simpler model for the logarithm of the range of the exchange rates in the context of an application involving four currencies. Let w_{ij} denote a logarithm of the range of foreign exchange rate of the currency i relative to the currency j, and $\mathbf{w} = (w_{12}, w_{13}, w_{14}, w_{23}, w_{24}, w_{34})$. Now assume that

$$\mathbf{w}_t = \mathbf{c} + \mathbf{Z}\mathbf{h}_t + \xi_t, \quad \xi_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}_{\xi\xi}),$$

$$\mathbf{h}_{t+1} = \text{diag}(\phi_1, \ldots, \phi_q)\mathbf{h}_t + \eta_t, \quad \eta_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{\Sigma}_{\eta\eta}),$$

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 jth 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 by Kalman filtering methods.

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

$$\mathbf{y}_t = \mathbf{V}_t^{1/2}\varepsilon_t, \quad \mathbf{V}_t^{1/2} = \text{diag}(\exp(\mathbf{z}_1'\mathbf{h}_t/2), \dots, \exp(\mathbf{z}_q'\mathbf{h}_t/2)),$$
$$(1-L)^d\mathbf{h}_t = \eta_t, \quad \varepsilon_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon}), \quad \eta_t \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_{\eta\eta}),$$

where $\mathbf{z}_i$ $(i = 1, \dots, q)$ are $q \times 1$ vectors with $q < p$. In the fitting, the measurement equation is linearized as in Harvey et al. (1994).

Calvet et al. (2006) generalized 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 $\mathbf{M}_t = (M_{1,t}, M_{2,t}, \dots, M_{k,t})$, 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:

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

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 are estimated by the maximum likelihood method.

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

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

where $\odot$ denotes the element-by-element product. For each component $\mathbf{M}_{j,t}$ in the bivariate model, Calvet et al. (2006) assumed that volatility arrivals are correlated but not necessarily simultaneous. For details, let $s_{j,t}^i$ $(i = 1, 2)$ denote the random variable equal to 1 if there is an arrival on $M_{j,t}^i$ with prob-

ability γ_j, and equal to 0 otherwise. Thus, each $s^i_{j,t}$ follows the Bernoulli distribution. At this stage, Calvet et al. (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 vectors $s^1_{j,t}$ 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 assumed 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 parameters therefore grows at least as fast as a quadratic function of p, Calvet et al. (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

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

where the weights are nonnegative and add up to 1, 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 et al. (2006) specified the model as follows. For each vector f^l_t, $f^l_{j,t}$ follows a 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 the univariate MSM with identical frequencies.

3.2 Mean factor model

Another type of factor MSV model is considered in Pitt and Shephard (1999), who, following a model proposed in Kim et al. (1998), worked with the specification

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \mathbf{V}_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}), \tag{18}$$

$$\mathbf{f}_t = \mathbf{D}_t^{1/2}\gamma_t, \quad \gamma_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}), \tag{19}$$

$$\mathbf{h}_{t+1} = \mu + \mathbf{\Phi}(\mathbf{h}_t - \mu) + \eta_t, \quad \eta_t \sim \mathcal{N}_{p+q}(\mathbf{0}, \mathbf{\Sigma}_{\eta\eta}), \tag{20}$$

where

$$\mathbf{V}_t = \mathrm{diag}(\exp(h_{1t}), \ldots, \exp(h_{pt})), \tag{21}$$

$$\mathbf{D}_t = \mathrm{diag}\left(\exp(h_{p+1,t}), \ldots, \exp(h_{p+q,t})\right), \tag{22}$$

$$\mathbf{\Phi} = \mathrm{diag}(\phi_1, \ldots, \phi_{p+q}), \tag{23}$$

$$\mathbf{\Sigma}_{\eta\eta} = \mathrm{diag}(\sigma_{1,\eta\eta}, \ldots, \sigma_{p+q,\eta\eta}) \tag{24}$$

and $\mathbf{h}_t = (h_{1t}, \ldots, h_{pt}, h_{p+1,t}, \ldots, h_{p+q,t})$. For identification purposes, the $p \times q$ loading matrix $\mathbf{B}$ is assumed to be such that $b_{ij} = 0$ for ($i < j$, $i \le q$) and $b_{ii} = 1$ ($i \le q$) with all other elements unrestricted. Thus, in this model, each of the factors and each of the errors evolve according to univariate SV models. A similar model was also considered by Jacquier et al. (1999) and Liesenfeld and Richard (2003) but under the restriction that $\mathbf{V}_t$ is not time-varying. Jacquier et al. (1999) estimated their model by MCMC methods, sampling h_{it} one at a time from its full conditional distribution, whereas Liesenfeld and Richard (2003) showed how the maximum likelihood estimation can be obtained by the efficient importance sampling method. For the more general model described above, Pitt and Shephard (1999) also employed a MCMC-based approach, now sampling $\mathbf{h}_t$ along the lines of Shephard and Pitt (1997). An even further generalization of this factor model was developed by Chib et al. (2006), who allowed for jumps in the observation model and a fat-tailed t distribution for the errors ε_t. The resulting model and its fitting are explained in Section 3.3. Alternative interesting approaches were also proposed by Diebold and Nerlove (1989) and King et al. (1994) in the framework of GARCH models using Kalman filter algorithms, but we omit the details to focus on the MSV models.

Lopes and Carvalho (2007) considered a general model which nests the models of Pitt and Shephard (1999) and Aguilar and West (2000), and extended it in two directions by (1) letting the matrix of factor loadings $\mathbf{B}$ be time-dependent and (2) allowing Markov switching in the common factor volatilities. The general model is given by (19)–(22), with

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

$$\mathbf{h}_{t+1}^f = \mu_{s_t}^f + \mathbf{\Phi}^f\mathbf{h}_t^f + \eta_t^f, \quad \eta_t^f \sim \mathcal{N}_q\left(\mathbf{0}, \mathbf{\Sigma}_{\eta\eta}^f\right),$$

where $\mathbf{h}_t^f = (h_{p+1,t}, \ldots, h_{p+q,t})'$, $\mu^f = (\mu_{p+1}, \ldots, \mu_{p+q})'$, $\mathbf{\Phi}^f = \mathrm{diag}(\phi_{p+1}, \ldots, \phi_{p+q})$ and $\mathbf{\Sigma}_{\eta\eta}^f$ is the nondiagonal 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 et al. (1998), where the fitting was based on the work of Albert and Chib (1993), they assumed μ_{s_t} followed a Markov switching model, where s_t follows a multistate first-order Markovian process. Lopes and Carvalho (2007) applied this model to two datasets: (1) returns on daily closing spot rates for six currencies relative to the US dollar (Deutschemark, British pound, Japanese yen, French franc, Canadian dollar, Spanish peseta), and (2) 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 (1999) and Chib et al. (2006) by allowing the factors to follow an AR(1) process:

$$\mathbf{f}_t = \mathbf{c} + \mathbf{A}\mathbf{f}_{t-1} + \mathbf{D}_t^{1/2}\gamma_t, \quad \gamma_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}). \tag{25}$$

The model was fit by adapting the approach of Chib et al. (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 et al. (2006). The model is given by

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \mathbf{K}_t\mathbf{q}_t + \mathbf{V}_t^{1/2}\mathbf{\Lambda}_t^{-1}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}), \tag{26}$$

where $\mathbf{\Lambda}_t = \text{diag}(\lambda_{1t}, \ldots, \lambda_{pt})$, $\mathbf{q}_t$ is p independent Bernoulli "jump" random variables and $\mathbf{K}_t = \text{diag}(k_{1t}, \ldots, k_{pt})$ are jump sizes. Assume that each element q_{jt} of $\mathbf{q}_t$ takes the value 1 with probability κ_j and the value 0 with probability $1 - \kappa_j$, and that each element u_{jt} of $\mathbf{u}_t = \mathbf{V}_t^{1/2}\mathbf{\Lambda}_t^{-1}\varepsilon_t$ follows an independent Student t distribution with degrees of freedom $\nu_j > 2$, which we express in hierarchical form as

$$u_{jt} = \lambda_{jt}^{-1/2}\exp(h_{jt}/2)\varepsilon_{jt}, \quad \lambda_{jt} \overset{i.i.d.}{\sim} \mathcal{G}\left(\frac{\nu_j}{2}, \frac{\nu_j}{2}\right), \quad t = 1, 2, \ldots, n. \tag{27}$$

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

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

are conditionally independent Gaussian random vectors. The time-varying variance matrices $\mathbf{V}_t$ and $\mathbf{D}_t$ are defined by (20) and (21). Chib et al. (2006) assumed that the variables $\zeta_{jt} = \ln(1 + k_{jt})$, $j \leq p$, are distributed as $\mathcal{N}(-0.5\delta_j^2, \delta_j^2)$, where $\delta = (\delta_1, \ldots, \delta_p)'$ are unknown parameters.

We may calculate the number of parameters and latent variables as follows. Let β 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 β. The model has $3(p + q)$ parameters $\theta_j = (\phi_j, \mu_j, \sigma_j)$ $(1 \le j \le p + q)$ in the autoregressive processes (20) of $\{h_{jt}\}$. We also have p degrees of freedom $\nu = (\nu_1, \ldots, \nu_p)$, p jump intensities $\kappa = (\kappa_1, \ldots, \kappa_p)$ and p jump variances $\delta = (\delta_1, \ldots, \delta_p)$. If we let $\psi = (\beta, \theta_1, \ldots, \theta_p, \nu, \delta, \kappa)$ 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 nonlinearly 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 $\{\lambda_t\}$.

To conduct the prior-posterior analysis of this model, Chib et al. (2006) focused on the posterior distribution of the parameters and the latent variables:

$$\pi\left(\beta, \{\mathbf{f}_t\}, \{\theta_j\}, \{\mathbf{h}_{j\cdot}\}, \{\nu_j\}, \{\lambda_{j\cdot}\}, \{\delta_j\}, \{\kappa_j\}, \{\zeta_{j\cdot}\}, \{\mathbf{q}_{j\cdot}\} | Y_n\right), \qquad (28)$$

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

1. Sample β. The full conditional distribution of β is given by

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

where $p(\beta)$ 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 et al. (2006) employed the M-H algorithm (Chib and Greenberg (1995)), following Chib and Greenberg (1994) and taking the proposal density to be multivariate-t, $T(\beta | \mathbf{m}, \boldsymbol{\Sigma}, v)$, 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. Then, a proposal value β^* is drawn from $T(\mathbf{m}, \boldsymbol{\Sigma}, v)$ and accepted with probability

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

where β is the current value. If the proposal value is rejected, the next item of the chain is taken to be the current value β.

2. Sample $\{\mathbf{f}_t\}$. The distribution $\{\mathbf{f}_t\} | \tilde{Y}_n, \mathbf{B}, \mathbf{h}, \lambda$ can be divided into the product of the distributions $\mathbf{f}_t | \tilde{\mathbf{y}}_t, \mathbf{h}_t, \mathbf{h}_t^f, \lambda_t, \mathbf{B}$, 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 $\{\theta_j\}$ and $\{\mathbf{h}_{j\cdot}\}$. Given $\{\mathbf{f}_t\}$ and the conditional independence of the errors in (20), the model separates into q conditionally Gaussian state-space models. Let

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

where c is an "offset" constant that is set to 10^{-6}. Then from Kim et al. (1998) it follows that the $p+q$ state-space models can be subjected to an independent analysis for sampling the $\{\theta_j\}$ and $\{\mathbf{h}_{j\cdot}\}$. In particular, the distribution of z_{jt}, which is h_{jt} plus a $\log \chi^2$ random variable with one degree of freedom, may be approximated closely by a seven-component mixture of normal distributions:

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

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 et al. (2002). Thus, under this representation, conditioned on the transformed observations

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

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 $(\mathbf{s}_{j\cdot}, \theta_j, \mathbf{h}_{j\cdot})$ by the univariate SV algorithm given by Chib et al. (2002). Briefly, $\mathbf{s}_{j\cdot}$ is sampled 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, θ_j is sampled by the M-H algorithm from the density $\pi(\theta_j|\mathbf{z}_{j\cdot}, \mathbf{s}_{j\cdot}) \propto p(\theta_j)p(\mathbf{z}_{j\cdot}|\mathbf{s}_{j\cdot}, \theta_j)$, where

$$
p\left(\mathbf{z}_{j\cdot}|\mathbf{s}_{j\cdot}, \theta_j\right) = p\left(\mathbf{z}_{j1}|\mathbf{s}_{j\cdot}, \theta_j\right) \prod_{t=2}^{n} p\left(\mathbf{z}_{jt}|\mathcal{F}_{j,t-1}^*, \mathbf{s}_{j\cdot}, \theta_j\right) \tag{29}
$$

and $p(z_{jt}|\mathcal{F}_{j,t-1}^*, \mathbf{s}_{j\cdot}, \theta_j)$ 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 $[\mathbf{h}_{j.}|\mathbf{z}_{j.}, \mathbf{s}_{j.}, \theta_j]$ by the simulation smoother algorithm of de Jong and Shephard (1995).

4. Sample $\{\nu_j\}$, $\{\mathbf{q}_{j.}\}$ and $\{\lambda_{j.}\}$. 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

$$\Pr(\nu_j|\mathbf{y}_{j.}, \mathbf{h}_{j.}, \mathbf{B}, \mathbf{f}, \mathbf{q}_{j.}, \zeta_{j.}) \tag{30}$$

$$\propto \Pr(\nu_j) \prod_{t=1}^{n} T(y_{jt}|\alpha_{jt} + \{\exp(\zeta_{jt}) - 1\}q_{jt}, \exp(h_{jt}), \nu_j),$$

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

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

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

followed by the components of the vector $\{\lambda_{j.}\}$ 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 $\{\zeta_{j.}\}$. For simulation efficiency reasons, δ_j and $\zeta_{j.}$ must also be sampled in one block. The full conditional distribution of δ_j is given by

$$\pi(\delta_j) \prod_{t=1}^{n} \mathrm{N}(\alpha_{jt} - 0.5\delta_j^2 q_{jt}, \delta_j^2 q_{jt}^2 + \exp(h_{jt})\lambda_{jt}^{-1}) \tag{31}$$

by the M-H algorithm. Once δ_j has been sampled, the vectors $\zeta_{j.}$ are sampled, bearing in mind that their posterior distribution is updated only when q_{jt} is 1. Therefore, when q_{jt} is 0, 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 κ 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 1,000 cycles are used for the purpose of summarizing the posterior distribution.

4 Dynamic Correlation MSV Model

Another way to model time-varying correlations is by constructing models that model the correlations (or functions of correlations) directly. We describe several such approaches in this section.

4.1 Modeling by reparameterization

One approach is illustrated by Yu and Meyer (2006) in the context of the bivariate SV model:

$$\mathbf{y}_t = \mathbf{V}_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}_2(\mathbf{0}, \mathbf{\Sigma}_{\varepsilon\varepsilon,t}), \quad \mathbf{\Sigma}_{\varepsilon\varepsilon,t} = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix},$$

$$\mathbf{h}_{t+1} = \mu + \mathrm{diag}(\phi_1, \phi_2)(\mathbf{h}_t - \mu) + \eta_t, \quad \eta_t \sim \mathcal{N}_2\left(\mathbf{0}, \mathrm{diag}(\sigma_1^2, \sigma_2^2)\right),$$

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

$$\rho_t = \frac{\exp(q_t) - 1}{\exp(q_t) + 1},$$

where $\mathbf{h}_0 = \mu$ 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 the 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 higher dimensions is not easy because it is difficult to ensure the positive-definiteness of the correlation matrix $\mathbf{\Sigma}_{\varepsilon\varepsilon,t}$.

Another approach, introduced by Tsay (2005), is based on the Choleski decomposition of the time-varying correlation matrix. Specifically, one can 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 $\mathbf{y}_t = \mathbf{L}_t\mathbf{V}_t^{1/2}\varepsilon_t$, $\varepsilon_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$. 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,

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

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

$$\text{Var}(y_{1t}) = \exp(h_{1t}),$$
$$\text{Var}(y_{2t}) = q_t^2 \exp(h_{1t}) + \exp(h_{2t}),$$
$$\text{Cov}(y_{1t}, y_{2t}) = q_t \exp(h_{1t}),$$
$$\text{Corr}(y_{1t}, y_{2t}) = \frac{q_t}{\sqrt{q_t^2 + \exp(h_{2t} - h_{1t})}}.$$

As suggested in Asai et al. (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

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

and

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

$$\text{Var}(y_{it}) = \sum_{k=1}^{i} q_{ik,t}^2 \exp(h_{kt}), \quad q_{ii,t} \equiv 1, \quad i = 1, \ldots, p,$$

$$\text{Cov}(y_{it}, y_{jt}) = \sum_{k=1}^{i} q_{ik,t} q_{jk,t} \exp(h_{kt}), \quad i < j, \quad i = 1, \ldots, p-1,$$

$$\text{Corr}(y_{it}, y_{jt}) = \frac{\displaystyle\sum_{k=1}^{i} q_{ik,t} q_{jk,t} \exp(h_{kt})}{\sqrt{\displaystyle\sum_{k=1}^{i} q_{ik,t}^2 \exp(h_{kt}) \sum_{k=1}^{j} q_{jk,t}^2 \exp(h_{kt})}}, \quad i < j,$$

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 the British pound, the Deutschemark and the Japanese yen against the US 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 Lemma 1 of Chiu et al. (1996) and 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 $\text{Var}(\mathbf{y}_t) = \mathbf{\Sigma}_t = \exp(\mathbf{A}_t)$, where $\mathbf{A}_t$ is a $p \times p$ real symmetric matrix such that $\mathbf{A}_t = \mathbf{B}_t \mathbf{H}_t \mathbf{B}_t'$ ($\mathbf{H}_t = \text{diag}(h_{1t}, \ldots, h_{pt})$). Note that

$$\mathbf{\Sigma}_t = \mathbf{B}_t \mathbf{V}_t \mathbf{B}_t', \quad \mathbf{V}_t = \mathrm{diag}(\exp(h_{1t}), \ldots, \exp(h_{pt})),$$

$$\mathbf{\Sigma}_t^{-1} = \mathbf{B}_t' \mathbf{V}_t^{-1} \mathbf{B}_t, \quad |\mathbf{\Sigma}_t| = \exp\left(\sum_{i=1}^{p} h_{it}\right).$$

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

$$\mathbf{y}_t|\mathbf{A}_t \sim \mathcal{N}_p(\mathbf{0}, \exp(\mathbf{A}_t)),$$
$$\alpha_{t+1} = \mu + \mathbf{\Phi}(\alpha_t - \mu) + \eta_t, \quad (\mathbf{\Phi} : \mathrm{diagonal}),$$
$$\alpha_t = \mathrm{vech}(\mathbf{A}_t), \quad \eta_t \sim \mathcal{N}_{p(p+1)/2}(\mathbf{0}, \mathbf{\Sigma}_{\eta\eta}),$$

as suggested in Asai et al. (2006). The estimation of this model can be done using a 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

Another way to obtain a time-varying correlation matrix is by the approach of Philipov and Glickman (2006a, 2006b), who assumed that the conditional covariance matrix $\mathbf{\Sigma}_t$ follows an inverted Wishart distribution with parameters that depend on the past covariance matrix $\mathbf{\Sigma}_{t-1}$. In particular,

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

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

$$\mathbf{S}_{t-1} = \frac{1}{\nu} \mathbf{A}^{1/2} \left(\mathbf{\Sigma}_{t-1}^{-1}\right)^d \mathbf{A}^{1/2\prime}, \qquad (32)$$
$$\mathbf{A} = \mathbf{A}^{1/2} \mathbf{A}^{1/2\prime},$$

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) pointed out that it is 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

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

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

respectively. Thus, the scale parameter d expresses the overall strength of the serial persistence in the covariance matrix over time. On the basis of the process of the logarithm of the determinant and asymptotic behavior of expectation of the determinant, they assumed 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

$$\mathrm{E}\left(\mathbf{\Sigma}_t^{-1}|\nu, \mathbf{S}_{t-1}\right) = \mathbf{A},$$

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

The matrix $\mathbf{A}$ in this model is a measure of the intertemporal 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 a 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 $(\tilde{\nu}, \tilde{\mathbf{S}}_{t-1})$,

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

and accept it with probability

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

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

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

4. Sample d from

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

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 $\pi(d|\{\mathbf{\Sigma}_t\}_{t=1}^{n}, \mathbf{A}, \nu, \mathbf{y})$ and

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

5. Sample ν from

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

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 $\pi(\nu|\{\mathbf{\Sigma}_t\}_{t=1}^{n}, \mathbf{A}, d, \mathbf{y})$ and

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

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 the 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 models is given by

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

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 $\mathbf{Q}_{t+1} = (\mathbf{1}\mathbf{1}' - \boldsymbol{\Psi}) \odot \bar{\mathbf{Q}} + \boldsymbol{\Psi} \odot \mathbf{Q}_t + \boldsymbol{\Xi}_t$, which corresponds to a generalization of the dynamic conditional correlation (DCC) model of Engle (2002).

The second DC MSV model is given by

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

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 (2006a, 2006b), the MCMC fitting procedures are different. Asai and McAleer (2007) estimated these models using returns of the Nikkei 225 Index, the Hang Seng Index and the Straits Times Index.

Gourieroux et al. (2004) and Gourieroux (2006) used an alternative approach and derived a Wishart autoregressive process. Let $\mathbf{Y}_t$ and $\boldsymbol{\Gamma}$ denote, respectively, a stochastic symmetric positive-definite matrix 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 a conditional Laplace transform:

$$\Psi_t(\boldsymbol{\Gamma}) = \mathrm{E}_t\left[\exp\{\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{Y}_{t+1})\}\right] = \frac{\exp\left[\mathrm{tr}\left\{\mathbf{M}'^{-1}\mathbf{M}\mathbf{Y}_t\right\}\right]}{|\mathbf{I} - 2\boldsymbol{\Sigma}\boldsymbol{\Gamma}|^{k/2}}, \tag{33}$$

where k is a scalar degree of freedom ($k < p - 1$), $\mathbf{M}$ is a $p \times p$ matrix of autoregressive parameters and $\boldsymbol{\Sigma}$ is a $p \times p$ symmetric and positive-definite

matrix such that the maximal eigenvalue of $2\boldsymbol{\Sigma}\boldsymbol{\Gamma}$ is less than 1. Here E_t denotes the expectation conditional on $\{\mathbf{Y}_t, \mathbf{Y}_{t-1}, \dots, \}$. It can be shown that

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

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

$$f(\mathbf{Y}_{t+1}|\mathbf{Y}_t) = \frac{|\mathbf{Y}_{t+1}|^{(k-p-1)/2}}{2^{kp/2}\Gamma_p(k/2)|\boldsymbol{\Sigma}|^{k/2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_{t+1} + \mathbf{M}\mathbf{Y}_t\mathbf{M}')\right\}\right]$$
$$\times_0 F_1(k/2; (1/4)\mathbf{M}\mathbf{Y}_t\mathbf{M}'\mathbf{Y}_{t+1}),$$

where Γ_p is the multidimensional gamma function and $_0F_1$ is the hypergeometric function of matrix augment; see Gourieroux et al. (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

$$\mathbf{Y}_t = \sum_{j=1}^{k} \mathbf{x}_{jt}\mathbf{x}_{jt}', \tag{34}$$
$$\mathbf{x}_{jt} = \mathbf{M}\mathbf{x}_{j,t-1} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}),$$

we obtain the Laplace transform $\Psi_t(\boldsymbol{\Gamma})$ given by (33). Gourieroux et al. (2004) also introduced a Wishart autoregressive process of higher order. They estimated the WAR(1) using a series of intraday 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 et al. (1985).

4.3.2 Factor model

Philipov and Glickman (2006a) proposed 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

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

where $\mathbf{S}_{t-1}$ is defined by (32). In other words, the conditional covariance matrix $\boldsymbol{\Sigma}_t$ of the factor $\mathbf{f}_t$ follows an inverse Wishart distribution whose parameter depends on the past covariance matrix $\boldsymbol{\Sigma}_{t-1}$. 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 hyperinverse Wishart distribution that arises in the study of graphical models (Dawid and Lauritzen (1993), Carvalho and West (2006)). The starting point is the dynamic linear model

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

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, $\mathbf{\Theta}_t$ is the $q \times p$ matrix of states, $\mathbf{u}_t$ is the $p \times 1$ innovation vector for observation, $\mathbf{\Omega}_t$ is the $q \times p$ innovation matrix for states, $\mathbf{G}_t$ is a known $q \times q$ matrix and $\mathbf{\Sigma}$ is the $p \times p$ covariance matrix. $\mathbf{\Omega}_t$ follows a matrix-variate normal with mean $\mathbf{O}$ $(q \times p)$, left covariance matrix $\mathbf{W}_t$ and right covariance matrix $\mathbf{\Sigma}$; in other words, any column ω_{it} of $\mathbf{\Omega}_t$ has a multivariate normal distribution $\mathcal{N}_q(\mathbf{0}, \sigma_{ii} \mathbf{W}_t)$, while any row ω_t^i of $\mathbf{\Omega}_t$, $\omega_t^{i'}$ has a multivariate normal distribution $\mathcal{N}_p(\mathbf{0}, w_{ii,t} \mathbf{\Sigma})$. Next, we suppose that $\mathbf{\Sigma} \sim \mathcal{HIW}_p(b, \mathbf{D})$, the hyperinverse Wishart distribution with a degree-of-freedom parameter b and location matrix $\mathbf{D}$. It should be noted that the dynamic linear model with $\mathbf{\Sigma} \sim \mathcal{HIW}_p(b, \mathbf{D})$ can be handled from the Bayesian perspective without employing simulation-based techniques. Finally, instead of time-invariant $\mathbf{\Sigma}$, Carvalho and West (2006) suggested a time-varying process given by

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

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 $\mathbf{\Sigma}_t$ 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, (1) 11 international currency exchange rates relative to the US dollar and (2) 346 securities from the S&P500 stock index.

5 Conclusion

We have conducted a comprehensive survey of the major current themes in the formulation of MSV models. In time, further significant developments can be expected, perhaps fostered by the overview and details delineated in this paper, especially in the fitting of high-dimensional models. Open problems

remain, primarily in the modeling of leverage effects, especially in relation to general specifications of cross-leverage effects embedded within multivariate heavy-tailed or skewed error distributions. We also expect that interest in the class of factor-based MSV models and DC models will grow as these approaches have shown promise in the modeling of high-dimensional data.

References

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

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

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

Andersen, T., Bollerslev, T., Diebold, F. X. and Labys, P. (2003): Modeling and forecasting realized volatility. *Econometrica* **71**, 579–625.

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

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

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

Asai, M., McAleer, M. and Yu, J. (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 Capitanio, A. (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-Nielsen, 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-Nielsen, O. E. and Shephard, N. (2001): Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society Series B* **63**, 167–241.

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

Bauwens, L. and Laurent, S. (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.

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

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

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

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

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

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

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

Chan, D., Kohn, R. and Kirby, C. (2006): Multivariate stochastic volatility models with correlated errors. *Econometric Reviews* **25**, 245–274.

Chib, S. (2001): Markov chain Monte Carlo methods: Computation and inference. *In: Heckman, J. J. and Leamer, E. (Eds.): Handbook of Econometrics* **5**, 3569–3649. North-Holland, Amsterdam.

Chib, S. and Greenberg, E. (1994): Bayes inference for regression models with ARMA(p,q) errors. *Journal of Econometrics* **64**, 183–206.

Chib, S. and Greenberg, E. (1995): Understanding the Metropolis-Hastings algorithm. *The American Statistician* **49**, 327–335.

Chib, S. and Greenberg, E. (1996): Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* **12**, 409–431.

Chib, S. and Greenberg, E. (1998): Analysis of multivariate Probit models. *Biometrika* **85**, 347–361.

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

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

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

Cox, J., Ingersoll, J. and Ross, S. (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.

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

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

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

Diebold, F. X. and Nerlove, M. (1989): The dynamics of exchange rate volatility: A multivariate latent-factor ARCH model. *Journal of Applied Econometrics* **4**, 1–22.

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

Durbin, J. and Koopman, S. J. (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 Kroner, K. F. (1995): Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**, 122–150.

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

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

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

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

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

Gupta, A. K., González-Farías, G. and Domínguez-Molina, J. A. (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., Ruiz, E. and Shephard, N. (1994): Multivariate stochastic variance models. *Review of Economic Studies* **61**, 247–264.

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

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

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

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

Jungbacker, B. and Koopman, S. J. (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., Shephard, N. and Chib, S. (1998): Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65**, 361–393.

King, M., Sentana, E. and Wadhwani, S. (1994): Volatility and links between national stock markets. *Econometrica* **62**, 901–933.

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

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

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

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

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

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

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

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

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

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

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

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

Smith, M. and Pitts, A. (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 Kwok, W. Y. (2006): A multivariate long memory stochastic volatility model. *Physica A* **362**, 450–464.

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

So, M. K. P., Lam, K. and Li, W. K. (1998): A stochastic volatility model with Markov switching. *Journal of Business and Economic Statistics* **16**, 244–253.

Tims, B. and Mahieu, R. (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* (2nd ed.). Wiley, New York.

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

Wong, F., Carter, C. and Kohn, R. (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 Meyer, R. (2006): Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews* **25**, 361–384.