The Structure of Dynamic Correlations in
Multivariate Stochastic Volatility Models

Manabu Asai
Faculty of Economics
Soka University, Tokyo

Michael McAleer
School of Economics and Commerce
University of Western Australia

August 2005

Abstract

This paper proposes two types of stochastic correlation structures for Multivariate Stochastic Volatility (MSV) models, namely the constant correlation (CC) MSV and dynamic correlation (DC) MSV models, from which the stochastic covariance structures could be obtained easily. Both structures can be used for purposes of optimal portfolio and risk management, and for calculating Value-at-Risk (VaR) forecasts and optimal capital charges under the Basel Accord. The choice between the CC MSV and DC MSV models can be made using a deviance information criterion. A technique is developed to estimate the DC MSV model using the Markov Chain Monte Carlo (MCMC) procedure.

Keywords and phrases: Multivariate stochastic volatility, Constant correlation structure, Dynamic correlation structure, Markov chain Monte Carlo.
1. Introduction

Covariance and correlation structures are used routinely for optimal portfolio choice, risk management, obtaining Value-at-Risk (VaR) forecasts, and determining optimal capital charges under the Basel Accord. This issue does not yet seem to have been examined in the Multivariate Stochastic Volatility (MSV) literature.

For multivariate GARCH models, the most general expression is called the ‘vec’ model (see Engle and Kroner (1995)). The vec model parameterizes the vector of conditional covariance matrix of the returns vector, which is determined by its lags and the vector of outer products of the lagged returns vector. A serious issue with the vec model is that it has many parameters to be estimated, and will not guarantee positive definiteness of the conditional covariance matrix without further restrictions. Bollerslev et al. (1988) suggested the diagonal GARCH model, which restricts the off-diagonal elements of the parameters matrices to be zero, and also reduced the number of parameters drastically. Engle and Kroner (1995) proposed the Baba, Engle, Kraft and Kroner (or BEKK) specification that guaranteed the positive definiteness of the conditional covariance matrix. This is essential for obtaining sensible VaR forecasts. Bollerslev (1990) proposed the Constant Conditional Correlation (CCC) model, where the time-varying covariances are proportional to the conditional standard deviation derived from univariate GARCH processes. This specification also guarantees the positive definiteness of the conditional covariance matrix. Recently, Engle (2002) developed the Dynamic Conditional Correlation (DCC) model, which allows the conditional correlation matrix to vary parsimoniously over time. McAleer (2005) provides a comprehensive comparison of a wide range of univariate and multivariate, conditional and stochastic, financial volatility models.

This paper proposes two types of stochastic correlation structures for MSV models, namely the constant correlation (CC) MSV and dynamic correlation (DC) MSV models, from which the stochastic covariance structures could be obtained easily. Both structures can be used for purposes of optimal portfolio and risk management, for calculating VaR forecasts, and for determining optimal Basel Accord capital charges. The choice between the CC MSV and DC MSV models can be made using a deviance information criterion (see Berg et al. (2004)). A technique is developed to estimate the DC MSV model using the Markov Chain Monte Carlo (MCMC) procedure.
2. Constant and Dynamic Correlation Models

This section develops two types of correlation models for MSV models, namely the CC MSV and DC MSV models, from which the stochastic covariance structures may be obtained easily.

2.1 Constant Correlation MSV Models

Consider the constant correlation (CC) MSV model proposed by Harvey, Ruiz and Shephard (1994). Let $y_t$ be an $M \times 1$ vector of stochastic process, as follows:

$$y_t = D_t \varepsilon_t, \quad \varepsilon_t : N(0, \Gamma),$$  

$$D_t = \text{diag} \left\{ \exp \left( h_t / 2 \right) \right\},$$  

$$h_{t+1} = \mu + \phi \sigma (h_t - \mu) + \eta_t, \quad \eta_t : N(0, \Sigma_\eta),$$

where ‘$\circ$’ denotes the Hadamard element-by-element product of two identically-sized matrices or vectors. Here, we use ‘exp’ as the operator for vectors to denote element-by-element exponentiation, and ‘diag’ for vectors to create a diagonal matrix. This model deals with two kinds of correlation matrices, namely the correlations between mean variables, $\Gamma$, and the covariance matrix of volatility, $\Sigma_\eta$. It should be noted that the model is different from the Constant Conditional Correlation (CCC) model of Bollerslev (1990) with respect to two major points, namely (i) the volatilities are stochastic, and (ii) the disturbances of the volatilities are correlated simultaneously.

In order to obtain the VaR forecasts for a given portfolio and to determine the optimal capital charges for a portfolio under the Basel Accord, it would be necessary to calculate the stochastic covariance matrix, $\Sigma_t$, from the correlation matrix, as $\Sigma_t = D_t \Gamma D_t$.

There are two ways in which to deal with the so-called ‘leverage effects’. One approach is to extend the model (1)-(3) in order to assume negative correlations between returns and changes of volatilities as follows:
\[ E(\varepsilon, \eta') = \text{diag}\left\{ \kappa_1 \sigma_{\eta,11}^{1/2}, \ldots, \kappa_m \sigma_{\eta,mn}^{1/2} \right\}, \]

where \( \sigma_{\eta,ii} \) is the \((i, i)\)th element of \( \Sigma_\eta \), and \( \kappa_i \) is restricted to be negative. The model is a multivariate extension of Harvey and Shephard (1996), and has been analysed in Asai and McAleer (2004). The other approach is to incorporate \( y_t \) and \( |y_t| \) in equation (3), as follows:

\[ h_{t+1} = \mu + \phi (h_t - \mu) + \lambda_1 y_t + \lambda_2 |y_t| + \eta_t, \quad \eta_t \sim N\left(0, \Sigma_\eta\right). \tag{4} \]

This model was proposed by Asai and McAleer (2004) to develop the multivariate extension of the SV model suggested by Danielsson (1994). In the Bayesian framework, Yu (2005) and Omori et al. (2004) estimated the univariate model of Harvey and Shephard (1996) using the MCMC technique.

As the purpose of the present paper is to consider the stochastic correlation MSV model, namely DC MSV, we do not consider asymmetric models, although such extensions are straightforward for the CC MSV model.

### 2.2 Dynamic Correlation MSV Models

This subsection develops a dynamic correlation (DC) MSV model, as an extension of the CC MSV model, which is based on the differences between the \( \varepsilon_t \). Our approach is based on the Wishart distribution. It is known that the Wishart distribution of a sample variance covariance matrix computed from i.i.d. multivariate Gaussian observation (see Anderson (1984) and Stuart and Ord (1994)). By considering the serially dependent Wishart Process, we have a process of positive definite matrices, under appropriate conditions. A standardization of the process leads to a serially dependent process of correlation matrices.

Let \( \varepsilon_t \) have a multivariate normal distribution, \( N(0, \Gamma_t) \), conditional on the stochastic correlation matrix, \( \Gamma_t \), where
\[
\Gamma_t = Q_t^{-1}Q_t Q_t^{-1}, \\
Q_{t+1} = \Omega + \psi Q_t + \Xi_t, \\
\Xi_t \sim W_m(\nu, \Lambda), \\
Q_t' = \left(\text{diagonal}\{Q_i\}\right)^{1/2},
\] (5)

and \(W_m(\nu, \Lambda)\) denotes a Wishart distribution, and ‘diagonal’ creates a diagonal matrix by setting the off-diagonal elements of matrices to be zero. The DC MSV model guarantees the positive definiteness of \(\Gamma_t\) under the assumption that \(Q_t\) is positive definite, \(0 < \psi < 1\), and that \(\Omega\) and \(\Lambda\) are symmetric positive definite matrices. The second condition also implies that the time-varying stochastic correlations are mean reverting. Given \(Q_t\) is positive definite, \(Q_t\) is guaranteed to be positive definite (the proof is obtained in a similar manner to that of the DCC model - see below for the case of \(\nu = 1\)). By the positive definiteness of \(Q_t\), \(\Gamma_t\) will also be positive definite. With the above dynamic and stochastic structures, the DC MSV model for \(y_t = D_t \varepsilon_t\) is defined by (2), (3) and (5).

As in the case of the CC MSV model, in order to obtain the VaR forecasts for a given portfolio and to determine the optimal capital charges for a portfolio under the Basel Accord for the DC MSV model, it is necessary to calculate the stochastic covariance matrix, \(\Sigma_t\), from the stochastic correlation matrix, as \(\Sigma_t = D_t \Gamma_t D_t\).

In order to investigate the properties of the DC MSV model, we consider the case of \(\nu = 1\), for simplicity. In this case, \(\Xi_t\) can be expressed as \(\Xi_t = \xi_t \xi_t'\), where \(\xi_t : N(0, \Lambda)\). Then we have

\[
Q_{t+1} = \Omega + \psi Q_t + \xi_t \xi_t',
\] (6)

which is analogous to the DCC model of Engle (2002), namely

\[
Q_{t+1} = \Omega + \alpha Q_t + \beta \varepsilon_t \varepsilon_t'.
\]

Comparing the last terms in the two equations given above, that of the DCC model is
predetermined and can be observed by using estimated conditional volatility, while that of
the DC MSV model is unobservable. Furthermore, letting \( q_t = \text{vec}(Q_t) \), we have a
VAR(1) process for \( q_t \), as follows:

\[
q_{t+1} = \left( \omega + \lambda \right) + \psi q_t + \text{vec}(\xi_t \xi_t' - \Lambda),
\]  

(7)

where \( \omega = \text{vec}(\Omega) \) and \( \lambda = \text{vec}(\Lambda) \). Since the expectation of the last term is zero, it is
possible to obtain \( E(q_t) = (1-\psi)^{-1}(\omega + \lambda) \), or \( E(Q_t) = (1-\psi)^{-1}(\Omega + \Lambda) \). Similarly, it
can be shown that

\[
V(q_t) = (1-\psi)^{-2}\left(I_{m^2} + K_{mm}\right)(\Lambda \otimes \Lambda),
\]

where \( K_{ij} \) is the commutation matrix. Noticing that \( \Xi_t = \xi_t \xi_t' \), the other moments can be
obtained through tedious calculation by using the moment of the multivariate normal
distribution. When \( \nu = 1 \), it may be natural to assume that \( Q_t = (1-\psi)^{-1}(\Omega + \Xi_0) \), since
its mean is equal to the unconditional mean, and as it is a positive definite matrix.

In the following, we consider \( \nu = 1 \) for convenience as (i) it is easy to interpret in this
case, and (ii) an approximated Gaussian process of (7) based on \( \xi_t \) is used to develop the
MCMC technique.

3 Bayesian Markov Chain Monte Carlo

3.1 Overview

In this section, we develop a technique to estimate the DC MSV model via Markov Chain
Monte Carlo (MCMC) procedure (see, for example, Chib and Greenberg (1996)). For
univariate SV models, there are two kinds of efficient and fast MCMC methods, namely
(i) the integration sampler suggested by Kim et al. (1998) and Chib et al. (2002), and (ii)
the multi-move sampler proposed by Shephard and Pitt (1997, 2004) and Watanabe and Omori (2004). For reasons explained in Subsection 3.2 below, we develop the block samplers which are an extension of Shephard and Pitt (1997, 2004).

Apart from MCMC, competing approaches include the likelihood approach based on importance sampling, such as Sandmann and Koopman (1998) and Liesenfeld and Richard (2003), and the reprojection method proposed by Gallant and Tauchen (1998). Asai and McAleer (2004) and Liesenfeld and Richard (2003) estimated MSV models by using the importance sampling procedures. Although we have to deal with the latent process of \( Q_t \), which has Wishart disturbances, the importance sampling approach is inapplicable when the latent process is non-Gaussian. Furthermore, the reprojection method needs to be extended further for application to MSV models.

Let \( \theta_1 \) denote the vector of parameters to be estimated in the latent process for the dynamic correlations, \( Q_t \), and \( \theta_2 \) the vector of parameters for the volatility processes, \( h_t \). Given the prior density, \( \pi(\theta_1, \theta_2) \), the aim of Bayesian inference is to obtain the parameter vector, \( (\theta_1, \theta_2) \), from the augmented posterior distribution, namely

\[
\pi(\theta_1, \theta_2, Q, h | y) \propto \pi(\theta_1, \theta_2) \prod_{t=1}^{T} f(y_t | Q_t, h_t, \theta_1, \theta_2) f(Q_t | Q_{t-1}, \theta_1) f(h_t | h_{t-1}, \theta_2).
\]

In order to conduct inferences on the parameters, we produce a sample \( (\theta_1^{(s)}, \theta_2^{(s)}, Q^{(s)}, h^{(s)}) \) from this density by the MCMC method. The produced draws \( (\theta_1^{(s)}, \theta_2^{(s)}) \) are taken from the posterior density marginalized over \( (Q, h) \).

The proposed MCMC algorithm is based on the two blocks, \( (\theta_1, Q) \) and \( (\theta_2, h) \):

**Algorithm**

1. Initialize \( Q \) and \( (\theta_1, \theta_2) \).
2. Sample $h$ and $\theta_2$ from $\theta_2, h | y, Q$ through drawing:
   (a) $h$ from $h | y, \theta_2, Q$;
   (b) $\theta_2$ from $\theta_2 | h$.

3. Sample $Q$ and $\theta_1$ from $\theta_1, Q | y, h$ through drawing:
   (a) $Q$ from $Q | y, \theta_1, h$;
   (b) $\theta_1$ from $\theta_1 | Q$.

4. Go to Step 2.

The remainder of this section proposes the sampling method for each step mentioned above. Step 3a is most important. Our method is based on the Metropolis-Hastings (MH) algorithm (see Chib and Greenberg (1995)). Since the process of $q_t$ is linear but non-Gaussian, we consider a linear approximation of the equation, and apply the simulation smoother proposed by de Jong and Shephard (1995) in order to generate candidates for the MH algorithm. Sampling all latent variables as one block will run into large very large rejection frequencies, while generating a single state at a time, which yields a highly autocorrelated sample, and needs a huge number of samples to conduct a statistical inference. We will take an intermediate approach, namely a block sampler such as Shephard and Pitt (1997) and Elerian et al. (2001).

Step 3b is based on the MH algorithm. We will use the Wishart distribution to generate candidates for $\Omega$ and $\Lambda^{-1}$, while we use the gamma distribution for $\psi$. In order to implement Step 2a, we extend the multi-move sampler proposed by Shephard and Pitt (1997, 2004) and Watanabe and Omori (2004). Although their approach was developed specifically for univariate SV models, the extension to MSV models is straightforward. As the model for $h_t$ can be interpreted as a seemingly unrelated regression (SUR) model, in that SUR can be written as a regression model with parameter restrictions, Step 2b follows from the updates of a VAR(1) model with parameter restrictions.

3.2 Simulation of the latent variable $Q_{t+1}$ from $\pi(Q_{t+1} | y, \theta_t, Q_t, h)$

In order to sample $Q$ from $\pi(Q | y, \theta, h)$ by using the MH algorithm, we consider the
process of latent variables, \( q_i = \text{vec}(Q_i) \), as follows;

\[
q_{t+1} = \omega + \psi q_i + \left( \xi_i \otimes \xi_i \right), \quad \xi_i : \mathcal{N}(0, \Lambda),
\]

(8)

for \( t = s, K, s + k - 1 \) where \( q_i = (1 - \psi)^{-1} \left( \omega + \xi_0 \otimes \xi_0 \right) \). In order to construct a proposal density, we first introduce an auxiliary state equation:

\[
p_{t+1} = c_t + \psi p_t + S_t \xi_t, \quad \xi_t : \mathcal{N}(0, \Lambda),
\]

(9)

\[
c_t = \omega - \left( \hat{\xi}_t \otimes \hat{\xi}_t \right), \quad S_t = \left( I_m \otimes \hat{\xi}_t \right) + \left( \hat{\xi}_t \otimes I_m \right),
\]

(10)

for \( t = s, K, s + k - 1 \) with an initial condition \( p_s = \hat{\rho}_s \). Since the state equation is non-Gaussian, we take \( p_s = \hat{\rho}_s = 0 \) for convenience. The derivation of the auxiliary state equation is given in the Appendix.

Consider the conditional posterior distribution of \( \vec{\xi} = (\xi_{t_s}^T, \xi_{s+k-1}^T)^T \), given \( q_s, q_{s+k}, Q_s, y_s, Q_{s+k}, h_s, K, h_{s+k} \), where \( k \geq 2 \). For the block sampler used here, \( \vec{\xi} \) is sampled from its full conditional distribution, where \( k \) changes stochastically as described below in this subsection. By using the sample of \( \vec{\xi} \), we obtain the sample of

\[
q = (q_{s+k}^T, K, q_{s+k}^T)^T.
\]

Let \( l_s \) denote the logarithm of the conditional likelihood of \( y_s \), given \( Q_s \) and \( h_s \), and let \( L = \sum_{s=1}^{s+k} l_s \). Given \( q_s, q_{s+k+1} \), we expand the logarithm of the conditional posterior density of \( \vec{\xi} \) around the mode \( \hat{\vec{\xi}} \), and we obtain a proposal density of \( \vec{\xi} \) as follows:
\[
\log f(\xi, K, \xi_{s+k-1} | q_s, q_{s+k+1}, \theta_t, y_s, K, y_{s+k-1}, h_s, K, h_{s+k-1})
\]
\[
= \text{const} - \frac{1}{2} \sum_{i=s}^{s+k-1} \xi_i^\prime \xi_i + L + \log p(q_{s+k+1} | q_{s+k})
\]
\[
\approx \text{const} - \frac{1}{2} \sum_{i=s}^{s+k-1} \xi_i^\prime \xi_i + \hat{L} + b'(p - \hat{p}) - \frac{1}{2}(p - \hat{p})^\prime B(p - \hat{p}) + \log p(q_{s+k+1} | q_{s+k})
\]
\[
= \text{const} + \log f^\ast(\xi, K, \xi_{s+k-1} | q_s, q_{s+k+1}, \theta_t, y_s, K, y_{s+k-1}, h_s, K, h_{s+k-1}),
\]

where \( \hat{L} = L \bigg|_{\xi = \xi} \), \( b = \left(b_{s+1}^\prime, K b_{s+k}^\prime \right)^\prime \), \( b_t = \partial l_t / \partial q_t \bigg|_{\xi = \xi} \),

\[
B = \begin{pmatrix}
B_{s+1} & O \\
B_{s+2} & O \\
O & B_{s+k}
\end{pmatrix},
\]

and \( B_t = -E \left[ \partial^2 l_t / \partial q_t \partial q_t^\prime \right]_{\xi = \xi} \).

First, we describe how to find a mode of the conditional posterior density. Consider the approximate linear Gaussian state-space model with the auxiliary state equation:

\[
p_{s+1} = \psi p_s + S_t \xi_t, \quad \xi_t : N(0, \Lambda), \quad t = s, s+1, K, s+k, \quad (11)
\]
\[
\hat{y}_t = p_t + \left( B_{t}^{1/2} \right)^{-1} u_t, \quad u_t : N(0, I_m), \quad t = s+1, K, s+k, \quad (12)
\]

where \( \hat{y}_t = \hat{p}_t + B_t^{-1} b_t \). By repeatedly applying the Kalman filter and a disturbance smoother to the linear Gaussian system (11) and (12), we first obtain a smoothed estimate of \( \xi_t \), and then substitute it recursively into the auxiliary state equation to obtain a smoothed estimate of \( p_t \). Then we replace \( \hat{\xi}_t, \hat{q}_t \) by the obtained smoothed estimates.

By repeating the procedure until the smoothed estimators converge, we obtain the posterior mode of \( \xi_t, p_t \).
In order to sample $\xi$ from the conditional posterior density, we conduct the following MH algorithm. Given the current value of $\xi$, say $\hat{\xi}$, we accept $\xi$ with probability

$$\min \left\{ 1, \frac{f(\xi) f^*(\hat{\xi})}{f(\hat{\xi}) f^*(\xi)} \right\}.$$ 

If it is rejected, accept $\hat{\xi}$ as a sample.

For the block sampler, we divide $h_i, K, h_{i+}$ into $K+1$ blocks, $\left(h_{i(i-j)}, K, h_{j(i)}\right)^T$, with $k_0 = 0, k_{K+1} = T, k(i) - k(i-1)$ $\geq 2$ for $i = 1, K, K+1$. The $K$ knots, $\left(k_1, K, k_K\right)$, are selected randomly. For example, Shephard and Pitt (1997) use a uniform distribution, while Elerian et al. (2001) use a Poisson distribution.

Although the integration sampler suggested by Kim et al. (1998) and Chib et al. (2002) is an efficient method for sampling volatility, it is not applicable to correlation processes or to $Q_0$ in this context. For general nonlinear and non-Gaussian state space models, some authors, including Kitagawa (1996), Hürzeler and Kunsh (1998) and Tanizaki and Mariano (1998), have developed a Monte Carlo filter and smoother in order to approximate the posterior distributions of the states, given the system parameters using many discrete points or particles. Kitagawa (1998) extended this to sample the states from the posterior distribution. However, when the Monte Carlo smoother is used, it is necessary to take the approximate sampling density carefully to obtain an accurate approximation by particles, as an outlier or structural change can lead to a poor approximation.

Block samplers can cope with such problems. With regard to block samplers, there are various block samplers, including Durbin and Koopman (2000), Garmerman (1998) and Shephard and Pitt (1997). The latter two methods, however, are not applicable to non-Gaussian state equations. Although Durbin and Koopman (2000) consider the
non-Gaussian state equation, they assume that each element of the disturbance is independent, and that the density of the each element is a function of its square. These assumptions are too restrictive to apply to our DC MSV model. For these reasons, we develop the block sampler as an extension of Shephard and Pitt (1997, 2004) and Watanabe and Omori (2004).

3.3 Sampling $\theta_i$ from $\pi(\theta_i | Q)$

Samples of the parameters for volatility, $\Lambda$, $\Omega$ and $\psi$, are generated from their full conditional distribution, as described below. Since $\Lambda$ and $\Omega$ are positive definite symmetric matrices, it is natural to consider the Wishart distribution or the inverse Wishart distribution as priors.

A Wishart distribution is assumed for the prior of $\Lambda^{-1}$, $\left(\Lambda^{-1} : W\left(\nu, S_{\nu}\right)\right)$, so that the log conditional posterior for $\Lambda^{-1}$ is given by

$$\text{const} + \frac{1}{2} \left(\psi^\nu - m - 1\right) \log |\Lambda^{-1}| - \frac{1}{2} \text{tr}\left(\Lambda^{-1} S_{\nu}^\psi\right)$$

where $\psi^\nu = \nu + T$, and $S_{\nu}^\psi = S_{\nu}^{-1} + \sum_{t=1}^{T-1} \left[(Q_{i,t} - \Omega - \psi Q) + (1 - \psi) Q_i - \Omega\right]$. Thus the conditional posterior of $\Lambda^{-1}$ is the Wishart distribution, $W\left(\psi^\nu, S_{\nu}^\psi\right)$. Then we perform the Metropolis-Hastings algorithm using a proposal given by $\Sigma^{-1}_{\eta} \sim W\left(k_1, K_1\right)$. Given the current value $\Sigma^{-1}_{\eta}$, generate $\Sigma^{-1}_{\eta} \sim W\left(k_1, K_1\right)$ and accept it with probability:

$$\min\left\{ \frac{1}{2} \left(\psi^\eta - m - 1\right) \log |\Lambda^{-1}| - \frac{1}{2} \text{tr}\left(\Lambda^{-1} S_{\eta}^\psi\right), 1 \right\}$$
where \( \Sigma_h^{(y)} \) and \( \Sigma_h^{(x)} \) are \( \Sigma_h \) evaluated at \( \Sigma^{-1}_h = \Sigma^{-1}_g^{(y)} \) and \( \Sigma^{-1}_h = \Sigma^{-1}_g^{(x)} \), respectively.

Given \( \Lambda \), a Wishart distribution is assumed for the prior of \( \Omega \) as \( \Omega | \Lambda : W(1, s \Lambda) \), \( s > T^{-1} \), so that the log conditional posterior for \( \Omega \) is given by:

\[
\log f (\Omega | \cdot ) = \text{const}^+ \frac{1}{2} g (\Omega) - \frac{1}{2} \text{tr} \left[ \Lambda^{-1} \left( Q^* - \left\{ T - \frac{1}{s} \right\}_\Omega \right) \right]
\]

\[
\approx \text{const}^+ \frac{1}{2} g^* (\Omega) - \frac{1}{2} \text{tr} \left[ \Lambda^{-1} \left( Q^* - \left\{ T - \frac{1}{s} \right\}_\Omega \right) \right]
\]

\[
= \log f^* (\Omega | \cdot )
\]

where

\[
g (\Omega) = -\frac{m}{2} \left[ \log |\Omega| + \sum_{t=1}^{T-1} \log |\Omega_t + \Omega - \psi \Omega_t| + \log \left( 1 - \psi \right) \Omega_t - \Omega \right],
\]

\[
g^* (\Omega) = \frac{T - m}{2} \log \left| Q^* - \left\{ T - \frac{1}{s} \right\}_\Omega \right|
\]

\[
Q^* = Q_T + (1 - \psi) \sum_{t=1}^{T-1} Q_t - \psi Q_t.
\]

Then we perform the Metropolis-Hastings algorithm using a proposal given by \( \Omega \sim s(sT - 1)^{-1} \left\{ Q^* - W \left( T + 1, \Lambda \right) \right\} \). Given the current value \( \Omega^{(x)} \), generate \( \Omega^{(x)} \sim s(sT - 1)^{-1} \left\{ Q^* - W \left( T + 1, \Lambda \right) \right\} \) and accept it with probability given by:

\[
\min \left\{ 1, \exp \left( g (\Omega^{(x)}) - g (\Omega^{(x)}) + g^* (\Omega^{(x)}) - g^* (\Omega^{(x)}) \right) \right\}.
\]

If it is rejected, accept \( \Omega^{(x)} \) as a sample.

Let \( \pi (\psi) \) denote a prior density for \( \psi \). Then the log conditional posterior density for
\( \psi \) is given by:

\[
\log f (\psi \mid \cdot) = \text{const} \frac{1}{2} g (\psi) - \frac{1}{2} (q^+_a - \psi q^+_b)
\]

\[
\approx \text{const} \frac{1}{2} g^* (\psi) - \frac{1}{2} (q^+_a - \psi q^+_b)
\]

\[
= \log f^* (\psi \mid \cdot),
\]

where

\[
g (\psi) = \log \pi (\psi) - \frac{m}{2} \left\{ \sum_{t=1}^{T-1} \log \left| \Omega_{t-1} - \Omega - \psi \Omega \right| + \log \left| (1 - \psi) \Omega - \Omega \right| \right\},
\]

\[
g^* (\Omega) = \left( \frac{T}{2} - 1 \right) \log \left| \sum_{t=1}^{T} \left( \Omega_t - \Omega \right) - \psi \left( \sum_{t=1}^{T-1} \Omega_t + \Omega \right) \right|,
\]

\[
q^+_a = \text{tr} \left[ \Lambda^{-1} \left\{ \sum_{t=1}^{T} \left( \Omega_t - \Omega \right) \right\} \right],
\]

\[
q^+_b = \text{tr} \left[ \Lambda^{-1} \left\{ \sum_{t=1}^{T-1} \left( \Omega_t + \Omega \right) \right\} \right],
\]

Then we perform the Metropolis-Hastings algorithm using a proposal given by

\( \psi \sim q^+_a / q^+_b - \text{Gamma} \left( q^+_a / 2, T / 2 \right) \), where \( \text{Gamma} (\cdot, \cdot) \) denotes the gamma distribution. Given the current value \( \psi^{(s)} \), generate

\[
\psi^{(s)} \sim q^+_a / q^+_b - \text{Gamma} \left( q^+_a / 2, T / 2 \right),
\]

and accept it with probability given by:

\[
\min \left\{ 1, \exp \left( g (\psi^{(s)}) - g (\psi^{(s)}) + g^* (\psi^{(s)}) - g^* (\psi^{(s)}) \right) \right\}.
\]

As \( \psi \) is assumed to satisfy \( 0 < \psi < 1 \), it may be useful to consider the beta distribution for the prior of \( \psi \).
3.4 **Simulation of the latent variable** \( h_{s,t} \) **from** \( \pi(h_{s,t} | y, \theta, Q, h_t) \)

In order to generate \( h \) from its full conditional distribution, we use the multi-move sampler proposed by Shephard and Pitt (1997, 2004) and Watanabe and Omori (2004).

Consider the conditional posterior distribution of \( \underline{\eta}=(\eta', K, \eta_{v_k-1}') \), given \( h_s, h_{s+k-1}, \theta, y, K, y_{s+k-1}, Q_s, K, Q_{s+k-1} \), where \( k \geq 2 \). In the multi-move sampler, \( \underline{\eta} \) is sampled from its full conditional distribution, where \( k \) changes stochastically, as described in subsection 3.2. By using the sample of \( \underline{\eta} \), we obtain the sample of \( h=(h_{s+1}', K, h_{s+k}') \).

Let \( l_t \) denote the logarithm of conditional likelihood of \( y_t \) given \( Q_t \) and \( h_t \), and let \( L = \sum_{t=s+1}^{s+k} l_t \), \( \Phi = \text{diag} \{ \phi \} \) and

\[
l_{s+k} = -\frac{1}{2} \left( h_{s+k+1} - \mu - \Phi(h_{s+k} - \mu) \right)' \Sigma^{-1}_{\eta} \left( h_{s+k+1} - \mu - \Phi(h_{s+k} - \mu) \right)
\]

Given \( h_s, h_{s+k-1} \), we expand the logarithm of the conditional posterior density of \( \underline{\eta} \) around the mode \( \hat{\eta} \), and obtain a proposal density of \( \underline{\eta} \), as follows:

\[
\log f(\eta, K, \eta_{s+k-1} | h_s, h_{s+k-1}, y_s, K, y_{s+k-1}) = \text{const} - \frac{1}{2} \sum_{t=s}^{s+k-1} \eta_t' \eta_t + L + l_{s+k} \approx \text{const} - \frac{1}{2} \sum_{t=s}^{s+k-1} \eta_t' \eta_t + \hat{L} + \left. \frac{\partial L}{\partial \eta'} \right|_{\eta=\hat{\eta}} (\eta - \hat{\eta}) + \frac{1}{2} (\eta - \hat{\eta})' \Sigma^{-1}_{\eta} (\eta - \hat{\eta}) + l_{s+k}
\]

\[
= \text{const} - \frac{1}{2} \sum_{t=s}^{s+k-1} \eta_t' \eta_t + \hat{L} + \eta' (\eta - \hat{\eta}) - \frac{1}{2} (h - \hat{h})' A (h - \hat{h})
\]

\[
= \text{const} + \log f^*(\eta, K, \eta_{s+k-1} | h_s, h_{s+k-1}, y_s, K, y_{s+k-1}),
\]
where $\hat{L} = L_{\eta=\hat{\eta}}$, $a = (a'_{s+1}, K' a'_{s+k})'$,

$$A = \begin{pmatrix} A_{s+1} & O \\ A_{s+2} & O \\ O & A_{s+k} \end{pmatrix},$$

$$a_t = \begin{cases} \frac{\partial l}{\partial h_{t|q-\hat{\eta}}} & \text{for } t = s, K, s+k-1 \text{ and } s = T, \\
\frac{\partial^2 l}{\partial h_{t|q-\hat{\eta}}^2} - \frac{\partial^2 l}{\partial h_{t|q-\hat{\eta}} \partial h'_{t|q-\hat{\eta}}} \hat{h}_t + \Phi \Sigma^{-1} \{h_{t+1} - (I_m - \Phi) \mu\}, & \text{for } t = s + k < T, \\
\end{cases}$$

$$A_t = \begin{cases} -\frac{\partial^2 l}{\partial h_{t|q-\hat{\eta}}^2} \hat{h}_t & \text{for } t = s, K, s+k-1 \text{ and } s = T, \\
-\frac{\partial^2 l}{\partial h_{t|q-\hat{\eta}} \partial h'_{t|q-\hat{\eta}}} \hat{h}_t + \Phi \Sigma^{-1} \Phi, & \text{for } t = s + k < T. \\
\end{cases}$$

First, we describe how to find a mode of the conditional posterior density. Consider the linear Gaussian state-space model given by:

$$h_{t+1} = (I_m - \Phi) \mu + \Phi h_t + \eta_t, \quad \eta_t : N\left(0, \Sigma_{\eta}\right), \quad t = s, s+1, K, s+k,$$

$$\hat{y}_t = \hat{h}_t + (A_t^{1/2})^{-1} u_t, \quad u_t : N\left(0, I_m\right), \quad t = s+1, K, s+k,$$

where $\hat{y}_t = \hat{h}_t + A_t^{-1} a_t$. Thus, applying the Kalman filter and a disturbance smoother to the linear Gaussian state space model, we first obtain a smoothed estimate of $\eta_t$, and then substitute it recursively into the state equation to obtain a smoothed estimate of $h_t$. Then we replace $\hat{\eta}_t, \hat{h}_t$ by the obtained smoothed estimates. By repeating the procedure until the smoothed estimates converge, we obtain the posterior mode of $\eta_t, h_t$. This is equivalent to the method of scoring to maximize the logarithm of the conditional posterior density.
In order to sample $\eta$ from the conditional posterior density, we conduct the following Acceptance-Rejection Metropolis-Hastings (AR-MH) algorithm (see Tierney (1994)). Given the current value of $\eta$, say $\eta_c$:

(i) AR step. Sample $\eta_c : \min\left(f(\eta_c), f^*(\eta_c)\right)$ using the Acceptance-Rejection algorithm.

(a) Generate $\eta_c : f^*$. To generate a candidate $\eta_c$ from $f^*$, we use the simulation smoother proposed by de Jong and Shephard (1995).

(b) Accept $\eta_c$ with probability:

$$\frac{\min\left(f(\eta_c), f^*(\eta_c)\right)}{f^*(\eta_c)}.$$

If it is rejected, go back to (a).

(ii) MH step. Given the current value $\eta_c$, accept $\eta_c$ with probability:

$$\min\left\{1, \frac{f(\eta_c)\min\left(f(\eta_c), f^*(\eta_c)\right)}{f(\eta_c)c\min\left(f(\eta_c), f^*(\eta_c)\right)}\right\},$$

where the proposal density is proportional to $\min\left(f(\eta_c), f^*(\eta_c)\right)$. If it is rejected, accept $\eta_c$ as a sample.

The selection of the knots, $k$, is carried out randomly and independently by using the method described in Subsection 3.2.

3.5 Sampling $\theta_2$ from $\pi(\theta_2 | h)$

Samples of parameters for volatility, $\mu, \phi, \Sigma_\eta$, are generated from their full conditional
distribution, as described below.

We take a multivariate normal distribution for the prior of $\mu$, $(\mu : N(m_0, M_0))$. Then, the conditional posterior for $\mu$ is given by $N(m, M)$, where

$$
\hat{M} = M_0^{-1} + (T-1)(I_m - \Phi)\Sigma^{-1}_\eta (I_m - \Phi) + \Sigma^{-1}_h,
$$

$$
\hat{m} = \hat{M}^{-1}m_0 + (I_m - \Phi)\Sigma^{-1}_\eta \sum_{t=1}^{T-1}(h_{t+1} - \Phi h_t) + \Sigma^{-1}_h h_t,
$$

and $\Phi = \text{diag}\{\phi\}$. It would be convenient to set $m_0 = 0$ and $M_0 = 10I_m$ as a (slightly) informative prior for empirical analysis.

Let $\pi(\phi)$ denote a prior probability density for $\phi$. Then, the log conditional posterior density for $\phi$ is given by:

$$
\text{const} + \log \pi(\phi) - \frac{1}{2} \log |\Sigma_h| - \frac{1}{2} (h_i - \mu)' \Sigma^{-1}_h (h_i - \mu)
$$

$$
- \frac{1}{2} \sum_{t=1}^{T-1} (h_{t+1} - \mu - \Phi (h_t - \mu))' \Sigma^{-1}_\eta (h_{t+1} - \mu - \Phi (h_t - \mu))'.
$$

The MH algorithm can be used with a multivariate truncated normal distribution, $\phi : N(\beta_{\phi}, \Sigma_{\phi}) I_S(\phi)$, where

$$
\Sigma_{\phi}^{-1} = \left[\Sigma^{-1}_\eta \circ X'X\right]^{-1},
$$

$$
\beta_{\phi} = \Sigma_{\phi} \left[\Sigma^{-1}_\eta \circ X'H^+ \right] t_m,
$$

$$
X = [h_1 - \mu, L, h_{T-1} - \mu]',
$$

$$
H^+ = [h_2 - \mu, L, h_T - \mu]',
$$
and \( I_s(\phi) = \prod_{i=1}^{m} I_{[1,1]}(\phi) \). Given the current sample, \( \phi^{(x)} \), generate 

\( \phi^{(y)} : N(\beta \phi, \Sigma^2_y) I_s(\phi) \), and accept it with probability given by:

\[
\min\left(1, \frac{\pi(\phi^{(y)}) |\Sigma_h^{(y)}|^{-1/2} \exp\left(-\frac{1}{2}(h - \mu)' \Sigma_h^{(y)-1}(h - \mu)\right)}{\pi(\phi^{(x)}) |\Sigma_h^{(x)}|^{-1/2} \exp\left(-\frac{1}{2}(h - \mu)' \Sigma_h^{(x)-1}(h - \mu)\right)}\right),
\]

where \( \Sigma_h^{(y)} \) and \( \Sigma_h^{(x)} \) are \( \Sigma_h \) evaluated at \( \phi = \phi^{(y)} \) and \( \phi = \phi^{(x)} \), respectively. As each element of \( \phi \) must satisfy the stationary condition, \( |\phi_i| < 1 \), it is assumed that

\[ \pi(\phi) = \prod_{i=1}^{m} \pi(\phi_i), \]

and that \( (1 + \phi_i)/2 \) follows a beta distribution, which is a natural extension of the prior considered by Kim et al. (1998) for the univariate case.

A Wishart distribution is assumed for the prior of \( \Sigma^{-1}_\eta \), \( (\Sigma^{-1}_\eta : W(\nu_\eta, S_\eta)) \), as a standard assumption of multivariate regression models. Then, the log conditional posterior for \( \Sigma^{-1}_\eta \) is given by:

\[
\text{const} - \frac{1}{2} \log |\Sigma_h| - \frac{1}{2}(h - \mu)' \Sigma_h^{-1} (h - \mu) + \frac{1}{2}(\theta_\eta - m - 1) \log |\Sigma_\eta| - \frac{1}{2} \text{tr}\left( \bar{S}_\eta \Sigma^{-1}_\eta \right),
\]

where \( \theta_\eta = \nu + T - 1 \), and \( \bar{S}_\eta = S_\eta^{-1} + \sum_{r=1}^{T-1} \{h_{r+1} - \mu - \Phi(h_r - \mu)\} \{h_{r+1} - \mu - \Phi(h_r - \mu)\}^\prime \).

Then we perform the MH algorithm using a proposal given by \( \Sigma^{-1}_\eta \sim W(\theta_\eta, \bar{S}_\eta) \). With the current value \( \Sigma^{-1}_\eta^{(x)} \), generate \( \Sigma^{-1}_\eta^{(y)} \sim W(\theta_\eta, \bar{S}_\eta) \), and accept it with probability given by:
\[
\min \left\{ \frac{1}{\sqrt{\Sigma_{h}^{(y)}}} \exp \left( -\frac{1}{2} (h_{i} - \mu)^{\prime} \Sigma_{h}^{(y)-1} (h_{i} - \mu) \right) \right\},
\]

where \( \Sigma_{h}^{(y)} \) and \( \Sigma_{h}^{(x)} \) are \( \Sigma_{h} \) evaluated at \( \Sigma_{\eta} = \Sigma_{\eta}^{(y)} \) and \( \Sigma_{\eta} = \Sigma_{\eta}^{(x)} \), respectively.

5 Conclusion

As covariance and correlation structures are used routinely for optimal portfolio choice, risk management, obtaining Value-at-Risk (VaR) forecasts, and determining optimal capital charges under the Basel Accord, it is essential to model the covariances and correlations accurately. This issue does not yet seem to have been examined in the Multivariate Stochastic Volatility (MSV) literature.

This paper proposed two types of stochastic correlation structures for MSV models, namely the constant correlation (CC) MSV and dynamic correlation (DC) MSV models, from which the stochastic covariance structures could be obtained easily. The choice between the CC MSV and DC MSV models was shown to be based on a deviance information criterion. A technique was developed to estimate the DC MSV model using the Markov Chain Monte Carlo (MCMC) procedure.

Acknowledgements

The authors wish to acknowledge very helpful discussions with Christian Gourieroux, Neil Shephard, George Tauchen and Jun Yu. The first author appreciates the financial support of the Japan Society for the Promotion of Science and the Australian Academy of Science. The second author is grateful for the financial support of the Australian Research Council. An earlier version of the paper was presented to the First Symposium on Econometric Theory and Applications (SETA), Institute of Economics, Academia Sinica, Taipei, Taiwan.
Appendix

In this appendix, we derive an approximation of the non-Gaussian state equation:

\[ q_{t+1} = (\omega + \lambda) + \psi q_t + \text{vec}(\Xi_t - \Lambda), \]

where \( \Xi_t \sim W_m(1, \Lambda) \), to a Gaussian process. Fahrmeir and Künstler (1999) suggested a method of approximating non-Gaussian state equations. Their method applies the second-order Taylor expansion to the logarithm of the distribution of the non-Gaussian disturbance to obtain an approximated Gaussian distribution. In this case, it may not be helpful to consider the Gaussian approximation of the distribution of \( \text{vec}(\Xi_t) \) or \( \text{vech}(\Xi_t) \) as the positive semi-definiteness of \( \Xi_t \) may be destroyed.

The approach taken here uses the property of the Wishart distribution, namely \( \Xi_t = \xi_t \xi_t' \), where \( \xi_t : N(0, \Lambda) \). The first-order Taylor series expansion of

\[ \text{vec}(\Xi_t - \Lambda) = \xi_t \otimes \xi_t - \lambda \]

around \( \xi_t \) gives

\[ \xi_t \otimes \xi_t - \lambda \approx (\xi_t', \otimes \xi_t - \lambda) + (I_m \otimes \xi_t' + \xi_t \otimes I_m)(\xi_t - \xi_t'), \]

which leads to the approximated state equation given by:

\[ p_{t+1} = \left[ \omega - \left( \xi_t ' \otimes \xi_t \right) \right] + \psi p_t + \left[ (I_m \otimes \xi_t') + \left( \xi_t \otimes I_m \right) \right] \xi_t', \]

where \( \xi_t : N(0, \Lambda) \).
References


