3-2006

Multivariate stochastic volatility: A review

Manabu ASAI
Soka University

Michael McAleer
University of Western Australia

Jun YU
Singapore Management University, yujun@smu.edu.sg

Published in Econometric Reviews, 2006 Mar, Volume 25, Issue 2/3, pp.145-175.
http://doi.org/10.1080/07474930600713564

MULTIVARIATE STOCHASTIC VOLATILITY: A REVIEW

Manabu Asai □ *Faculty of Economics, Soka University, Tokyo, Japan*

Michael McAleer □ *School of Economics and Commerce, University of Western Australia, Perth, Australia*

Jun Yu □ *School of Economics and Social Sciences, Singapore Management University, Singapore*

□ *The literature on multivariate stochastic volatility (MSV) models has developed significantly over the last few years. This paper reviews the substantial literature on specification, estimation, and evaluation of MSV models. A wide range of MSV models is presented according to various categories, namely, (i) asymmetric models, (ii) factor models, (iii) time-varying correlation models, and (iv) alternative MSV specifications, including models based on the matrix exponential transformation, the Cholesky decomposition, and the Wishart autoregressive process. Alternative methods of estimation, including quasi-maximum likelihood, simulated maximum likelihood, and Markov chain Monte Carlo methods, are discussed and compared. Various methods of diagnostic checking and model comparison are also reviewed.*

Keywords Asymmetry; Diagnostic checking; Estimation; Factor models; Leverage; Model comparison; Multivariate stochastic volatility; Thresholds; Time-varying correlations; Transformations.

JEL Classification C11; C15; C32; G12.

1. INTRODUCTION

A wide range of multivariate GARCH and stochastic volatility (SV) models has been developed, analyzed, and applied extensively in recent years to characterize the volatility that is inherent in financial time series data. Bauwens et al. (2006) recently provided a survey of multivariate GARCH models. The GARCH literature has expanded considerably since the univariate ARCH process was developed by Engle (1982). The univariate SV model was proposed by, among others, Taylor (1982, 1986).

Address correspondence to Manabu Asai, Faculty of Economics, Soka University, Hachioji-shi, Tokyo 192-8577, Japan; E-mail: m-asai@soka.ac.jp

The univariate SV literature was surveyed in Ghysels et al. (1996) and Shephard (1996), while estimation methods for univariate SV models were reviewed in Broto and Ruiz (2004). Shephard (2005) collects some of the more important studies on the topic.

Although there have already been many practical and successful applications of multivariate GARCH models, the theoretical literature on multivariate stochastic volatility (MSV) models has developed significantly over the last few years. In GARCH-type models the conditional variance of returns is assumed to be a deterministic function of past returns, whereas in SV models the volatility process is random. The introduction of the additional error term makes the SV model more flexible than the GARCH-type models (see, for example, Kim et al., 1998) and also more directly linked to continuous time models that are often used in asset pricing in finance. Some of the more important existing univariate and multivariate GARCH and discrete time SV models have been analyzed in McAleer (2005). However, a comprehensive review of the important aspects of existing discrete time MSV models in the literature does not yet seem to exist. Owing to the development of a wide variety of discrete time MSV models in recent years, this paper reviews the substantial literature on the specification, estimation, and evaluation of discrete time MSV models.

There are both economic and econometric reasons why multivariate volatility models are important. The knowledge of correlation structures is vital in many financial applications, such as asset pricing, optimal portfolio risk management, and asset allocation, so that multivariate volatility models are useful for making financial decisions. Moreover, as financial volatility moves together across different assets and markets, modeling volatility in a multivariate framework can lead to greater statistical efficiency.

The remainder of the paper is organized as follows. Section 2 presents a range of discrete time MSV models according to various categories, including asymmetric models, factor models, time-varying correlation models, and several alternative specifications, including the matrix exponential transformation, the Cholesky decomposition, and the Wishart autoregressive models. Section 3 compares and discusses alternative methods of estimation, including the quasi-maximum likelihood, simulated maximum likelihood, Monte Carlo likelihood, and Markov chain Monte Carlo techniques. Various methods of diagnostic checking and model comparison are examined in Section 4. Some concluding comments are given in Section 5.

2. MSV MODELS

In what follows, $S = (S_1, \ldots, S_m)'$ denotes a vector of log-prices for m financial assets, and $y = (y_1, \ldots, y_m)'$ denotes a vector of returns

for m financial assets. For expositional purposes, it is assumed that the conditional mean vector of y is zero, although this can easily be relaxed. Moreover, $\exp(\cdot)$ and $\log(\cdot)$ denote the element-by-element exponential and logarithmic operators, respectively, and $\text{diag}\{x\} = \text{diag}\{x_1, \ldots, x_m\}$ denotes the $m \times m$ diagonal matrix with diagonal elements given by $x = (x_1, \ldots, x_m)'$.

In a general level, the continuous time diffusion model for S can be specified by

$$\begin{aligned} dS(t) &= H^{1/2}(t)dW^{(1)}(t) \\ df(vech(H(t))) &= a(vech(H(t)))dt + b(vech(H(t)))dW^{(2)}(t), \end{aligned} \tag{1}$$

where $vech(\cdot)$ is the operator that stacks each column of the lower triangular matrix, $W^{(1)}(t)$ and $W^{(2)}(t)$ are two vectors of Brownian motions whose elements may be cross-correlated, and f, a, b are all known functions. The Euler scheme leads to the discrete time MSV model

$$\begin{aligned} y_t &= H_t^{1/2}\varepsilon_1, \qquad \varepsilon_t \sim N(0, I_m) \\ f(vech(H_t)) &= a(vech(H_{t-1})) + f(vech(H_{t-1})) + b(vech(H_{t-1}))\eta_{t-1}, \\ \eta_{t-1} &\sim N(0, \Sigma_\eta), \end{aligned} \tag{2}$$

where $y_t = S_t - S_{t-1}$.

Without further restrictions, there is no guarantee for the positive definiteness of H_t. As a result, the validity of Equation (2) as a well-defined econometric model is not guaranteed. In order to understand why H_t must be positive definite. Note that, in its mathematical form, the first equation in (2) is identical to the mean equation of the multivariate GARCH model (see, for example, Bauwens et al., 2006, Equation (2)), while H_t in Equation (1) represents the instantaneous covariance matrix. This requirement of positive definiteness is the same as in multivariate GARCH models and becomes nontrivial relative to the univariate case. However, in sharp contrast to multivariate GARCH models, in which H_t is measurable with respect to I_{t-1} as it is specified to be a deterministic function of past returns (Engle and Kroner, 1995), $\eta_{t-1} \notin I_{t-1}$ (where I_t is denoted as the information set observed by the econometrician at t) in MSV models. Consequently, H_t is not measurable with respect to I_{t-1} in MSV models.

The latent feature of H_t makes the positive definiteness a more difficult condition to achieve than in multivariate GARCH models. This is perhaps the reason that, in the discrete time MSV literature, the dynamics of the logarithm transformation of H_t, rather than H_t itself, are almost always modeled. Moreover, without further restrictions, Equation (2) does not warrant identification or parsimony. From the perspective of estimation

and prediction, parsimony is an especially relevant issue. As the dimension of the vector y increases, the number of parameters in Equation (2) increases at a much faster rate, making the model difficult to estimate and not useful for prediction. However, if too many restrictions are imposed, the model may fail to capture some important features in the data. Thus there is a trade-off between the flexibility of specification and the curse of dimensionality.

There are various approaches to simplify Equation (2) in order to address the positive definiteness of H_t and the trade-off between flexibility and the curse of dimensionality. This section reviews these variants of MSV models according to four categories: (i) asymmetric models, (ii) factor models, (iii) time-varying correlation models, and (iv) alternative MSV specifications, including models based on the matrix exponential transformation, the Cholesky decomposition, and the Wishart autoregressive process.

2.1. The Basic Model

The first MSV model proposed in the literature is due to Harvey et al. (1994):

$$\begin{aligned} y_t &= H_t^{1/2}\varepsilon_t, \\ H_t^{1/2} &= \operatorname{diag}\{\exp(h_{1t}/2), \ldots, \exp(h_{mt}/2)\} = \operatorname{diag}\{\exp(h_t/2)\} \end{aligned} \tag{3}$$

$$h_{t+1} = \mu + \phi \circ h_t + \eta_t, \tag{4}$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P_\varepsilon & O \\ O & \Sigma_\eta \end{pmatrix}\right], \tag{5}$$

where $h_t = (h_{1t}, \ldots, h_{mt})'$ is an $m \times 1$ vector of unobserved log-volatility, μ and ϕ are $m \times 1$ parameter vectors, the operator $\circ$ denotes the Hadamard (or element-by-element) product, $\Sigma_\eta = \{\sigma_{\eta,ij}\}$ is a positive-definite covariance matrix, and $P_\varepsilon = \{\rho_{ij}\}$ is the correlation matrix, that is, P_ε is a positive definite matrix with $\rho_{ii} = 1$ and $|\rho_{ij}| < 1$ for any $i \neq j$, $i, j = 1, \ldots, m$. Harvey et al. (1994) also considered the multivariate t distribution for ε_t, as this specification permits greater kurtosis than with the Gaussian assumption. It can be seen that the positive definiteness of H_t and parsimony are achieved by restricting H_t to be a diagonal matrix, with the diagonal element following the exponential function of a Gaussian vector AR(1) process. The number of parameters in the basic model is $2m + m^2$.

The model of Harvey et al. (1994) can easily be extended to a VARMA structure for h_t:

$$\Phi(L)h_{t+1} = \mu + \Theta(L)\eta_t,$$

where

$$\Phi(L) = I_m - \sum_{i=1}^{p} \Phi_i L^i,$$

$$\Theta(L) = I_m - \sum_{j=1}^{q} \Theta_j L^j,$$

and L is the lag operator.

Assuming that the off-diagonal elements of Σ_η are all equal to zero, the model corresponds to the constant conditional correlation (CCC) model proposed by Bollerslev (1990) in the framework of multivariate GARCH processes. In the CCC model, each conditional variance is specified as a univariate GARCH model, that is, with no spillovers from other assets, while each conditional covariance is given as a constant correlation coefficient times the product of the corresponding conditional standard deviations. If the off-diagonal elements of Σ_η are not equal to zero, then the elements of h_t are not independent.

2.2. Asymmetric Models

It has long been recognized that the volatility of stock returns responds differently to bad news and good news. In particular, while bad news tends to increase the future volatility, good news of the same size will increase the future volatility by a smaller amount, or may even decrease the future volatility. The news impact function (NIF) of Engle and Ng (1993) is a powerful tool for analyzing the volatility asymmetry for GARCH-type models. The idea of the NIF is to examine the relationship between conditional volatility in period $t + 1$ (defined by σ^2_{t+1}) and the standardized shock to returns in period t (defined by ε_t) in isolation.

In both the conditional and the stochastic volatility literature, there has been some confusion regarding the definitions of asymmetry and leverage. The asymmetric effect in volatility is that the effects of positive returns on volatility are different from those of negative returns of a similar magnitude. On the other hand, leverage refers to the negative correlation between the current return and future volatility. Therefore leverage denotes asymmetry, but not all asymmetric effects display leverage. In the class of ARCH specifications that have been developed to capture asymmetric effects, the exponential GARCH (EGARCH) model of Nelson (1991) and the GJR model of Glosten et al. (1993) are widely used. Using the terminology given above, the EGARCH model can describe leverage, whereas the GJR model can capture asymmetric effects but not leverage (for further details, see Asai and McAleer, 2006).

As stated above, a popular explanation for asymmetry is the leverage effect proposed by Black (1976) (see also Christie, 1982). Other forms of asymmetry, such as the asymmetric V-shape, have to be explained by reasons other than the leverage effect. Alternative reasons for the volatility asymmetry that have been suggested in the literature include the volatility feedback effect (Campbell and Hentschel, 1992).

Most asymmetric MSV models are based on the basic SV specifications and hence the positive definiteness of H_t is ensured. Depending on how asymmetry is introduced, we have three types of asymmetric MSV models, as given below.

2.2.1. *Leverage Effect*

A univariate discrete time SV model with the leverage effect was first proposed by Harvey and Shephard (1996), although Wiggins (1987) and Chesney and Scott (1989), among others, considered a continuous time model itself and discretized it for purposes of estimation. The model of Harvey and Shephard (1996) is of the form

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t \\ \ln \sigma_{t+1}^2 &= \alpha + \phi \ln \sigma_t^2 + \sigma_\eta \eta_t \\ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \end{aligned} \tag{6}$$

where all the variables are scalar. Model (6) may be regarded as the Euler approximation to the continuous time SV model that is used widely in the option price literature (see, for example, Hull and White, 1987, who generalized the Black–Scholes option pricing formula to analyze SV and leverage).

In the context of SV, Yu (2005) defined the leverage effect to be a negative relationship between $E(\ln \sigma_{t+1}^2 | y_t)$ and the return in period t (defined by y_t), holding everything else constant. Yu (2005) showed that

$$E(\ln \sigma_{t+1}^2 | y_t) = \alpha + \frac{\alpha\phi}{1-\phi^2} + \rho\sigma_\eta \exp\left(-\frac{\sigma_\eta^4}{4(1-\phi^2)^2} + \frac{\sigma_\eta^2 \alpha}{(1-\phi^2)(1-\phi)} \right) y_t,$$

and hence Model (6) ensures the leverage effect whenever $\rho < 0$. An alternative discrete time SV model with leverage effect was proposed by Jacquier et al. (2004) as

$$\begin{pmatrix} \varepsilon_t \\ \eta_{t-1} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \tag{7}$$

Model (7) differs from (6) in the timing of the correlation. Using the implicit function theorem, one can show that Model (7) implies

$$\frac{\partial \ln \sigma_{t+1}^2}{\partial y_t} = \frac{\rho \sigma_\eta \sigma_{t+1}/\sigma_t}{1 + 0.5\rho\sigma_\eta \varepsilon_{t+1}}. \tag{8}$$

Even when $\rho < 0$, the quantity in (8) can be positive or negative. As a result, the leverage effect is not warranted, and hence it is difficult to interpret the leverage effect (Yu, 2005).

Danielsson (1998) and Chan et al. (2005) considered a multivariate extension of the model of Jacquier et al. (2004). The model is given by Equations (3) and (4), together with

$$\begin{gathered} \begin{pmatrix} \varepsilon_t \\ \eta_{t-1} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P_\epsilon & L \\ L & \Sigma_\eta \end{pmatrix} \right), \\ L = \text{diag}\{\lambda_1 \sigma_{11}^{1/2}, \text{K}, \lambda_m \sigma_{mm}^{1/2}\}, \end{gathered} \tag{9}$$

where the parameter λ_{ij} captures asymmetry. In the empirical analysis, Danielsson (1998) did not estimate the multivariate dynamic leverage (DL) model because the data used in his analysis did not suggest any asymmetry in the estimated univariate models. However, the argument of Yu (2005) regarding the leverage effect also applies to the model in (9). Thus the interpretation of the leverage effect in (9) is unclear and, even if $\lambda_{ij} < 0$, there is no guarantee that there will, in fact, be a leverage effect.

In order to introduce the leverage effect, Asai and McAleer (2005b) considered a multivariate extension of the model of Harvey and Shephard (1996). The model is given by Equations (3) and (4) together with

$$\begin{gathered} \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P_\varepsilon & L \\ L & \Sigma_\eta \end{pmatrix} \right], \\ L = \text{diag}\{\lambda_t \sigma_{\eta,l1}, \ldots, \lambda_m \sigma_{\eta,mm}\}, \end{gathered} \tag{10}$$

where the parameter $\lambda_i, i = 1, \ldots, m$, is expected to be negative. The number of parameters in Model (10) is $3m + m^2$.

2.2.2. *Threshold Effect*

In the GARCH literature, Glosten et al. (1993) proposed modeling the asymmetric responses in conditional volatility using thresholds. In the univariate SV literature, So et al. (2002) proposed a threshold SV model, in which the constant term and the autoregressive parameter in the SV equation changed according to the sign of the previous return.

Although the multivariate threshold SV model has not yet been developed in the literature, it is straightforward to introduce a multivariate

threshold SV model with the mean equation given in (2), together with the volatility equation,

$$h_{t+1} = \mu(s_t) + \phi(s_t) \circ h_t + \eta_t,$$

where

$$\mu(s_t) = (\mu_1(s_{1t}), \ldots, \mu_m(s_{mt}))',$$
$$\phi(s_t) = (\phi_1(s_{1t}), \ldots, \phi_m(s_{mt}))',$$

and s_t is a state vector with elements given by

$$s_{it} = \begin{cases} 0, & \text{if } y_{it} < 0, \\ 1, & \text{otherwise.} \end{cases}$$

2.2.3. *General Asymmetric Effect*

Within the univariate framework, Danielsson (1994) suggested an alternative asymmetric specification to the leverage SV model, which incorporates the absolute value function to capture the sign and magnitude of the previous normalized returns shocks to accommodate asymmetric behavior.

In the model suggested by Danielsson (1994), which was called the SV with leverage and size effect (SV-LSE) model in Asai and McAleer (2005b), two additional terms, $|y_t|$ and y_t, were included in the volatility equation, while the correlation between the two error terms was assumed to be zero. Asai and McAleer (2005a,b) noted that the absolute values of the previous realized returns are included because it was not computationally straightforward to incorporate the previous normalized shocks in the framework of SV models.

In the multivariate context, Asai and McAleer (2005b) suggested an extension of the SV-LSE model of Danielsson (1994) with the mean equation given in (2), together with

$$h_{t+1} = \mu + \gamma_1 \circ y_t + \gamma_2 \circ |y_t| + \phi \circ h_t + \eta_t,$$

where $\eta_t \sim N(0, \Sigma_\eta)$ and γ_1 and γ_2 are $m \times 1$ parameter vectors.

2.3. Factor Models

Although the models reviewed in Section 2.2 can mitigate the problem of the curse of dimensionality, the number of parameters still increases at an increasing speed in m. To afford further parsimonious parameterization, a class of factor MSV models has been proposed in the literature. Unlike the

models reviewed in Section 2.2, the boundary between factor MSV and factor GARCH is blurred. Some factor GARCH models, including Engle et al. (1990), assume that factors are observable, so that the likelihood is readily available. Some other factor models, such as Diebold and Nerlove (1989), assume the factors follow latent ARCH processes. As a result, the latent factor ARCH models can also be classified as factor MSV models.

Although the model of Diebold and Nerlove (1989) can be regarded as a prototype of factor MSV models, the MSV models with the latent SV factor were originally proposed by Harvey et al. (1994), and extended by Shephard (1996), Pitt and Shephard (1999b), Jacquier et al. (1999), and Doz and Renault (2006). Apart from the ability to ensure parsimony, factor MSV models can capture the common features in asset returns and volatilities, and hence have a natural link with the arbitrage pricing theory (APT) of Ross (1976). Owing to these two features, this class of MSV models has received a great deal of attention in the literature. Depending on how factors enter the mean equation, we will classify this class of models into two subclasses.

2.3.1. *Additive Factor Models*

The additive factor MSV model was first introduced by Harvey et al. (1994) and subsequently extended in Jacquier et al. (1995, 1999), Shephard (1996), Pitt and Shephard (1999a), and Aguilar and West (2000). The basic idea is borrowed from the factor multivariate ARCH models and more generally from factor decompositions of covariance structures in multivariate analysis, where the returns are decomposed into two additive components. The first component has a smaller number of factors, which captures the information relevant to the pricing of all assets, while the other component is idiosyncratic noise, which captures the asset-specific information (for further details, see Diebold and Nerlove, 1989).

Denote the $K \times 1$ vector of factors as f_t $(K < m)$ and D is an $m \times K$ dimensional matrix of factor loadings. The additive K factor MSV model presented by Jacquier et al. (1995) can be written as

$$\begin{aligned} y_t &= Df_t + e_t, \\ f_{it} &= \exp(h_{it}/2)\varepsilon_{it}, \\ h_{i,t+1} &= \mu_i + \phi_t h_{it} + \eta_{it}, \qquad i = 1, \ldots, K, \end{aligned} \tag{11}$$

where $e_t \sim N(0, \text{diag}\{\sigma_1^2, \ldots, \sigma_m^2\})$, $\varepsilon_{it} \sim N(0,1)$, $\eta_{it} \sim N(0, \sigma_\eta^2)$ and they are mutually independent. In order to guarantee the identification of D and f_t uniquely, the restrictions $D_{ij} = 0$ and $D_{ii} = 1$ for $i = 1, \ldots, m$ and $j < i$ are usually adopted (see Aguilar and West, 2000). The variance of y_t is $D\Sigma_f D' + \text{diag}\{\sigma_1^2, \ldots, \sigma_m^2\}$, which is always positive definite. In Diebold and

Nerlove (1989), f_t is assumed to follow a latent ARCH process instead of a latent AR(1) process. It is easy to see that the positive definiteness of the conditional variance of y_t is ensured by construction. When $K = 1$, the total number of parameters is $5m - 1$ and hence will only grow linearly with m.

While Model (11) is highly parsimonious and can capture some of the important features in the data, it has several drawbacks. First, since the error e_t is homoskedastic, when the number of assets in the portfolios is greater than the number of factors, it can be shown that there must exist portfolios whose prices involve no heteroskedasticity. This feature is clearly inconsistent with the stylized fact. Second, the diagonality in the covariance of the error e_t is a too strong assumption and is not preserved by portfolio formation, as shown by Doz and Renault (2006).

Model (11) was extended in Pitt and Shephard (1999a) by allowing each element in e_t to evolve according to a univariate SV model, which increases the number of parameters to $8m - 1$. Chib et al. (2005) further extended the model by allowing for jumps and for idiosyncratic errors that follow the Student t SV process. Clearly, these extensions overcome the first drawback while maintaining parsimonious specifications. To deal with the second drawback, Model (11) was extended in Doz and Renault (2006) by allowing $Var(e_t)$ to be a possibly nondiagonal constant definite matrix.

Yu and Meyer (2006) showed that additive factor models accommodate both time-varying volatility and time-varying correlations. In the context of the bivariate one-factor SV model given by

$$\begin{aligned} y_t &= Df_t + e_t, \qquad e_t \sim N\big(0, \operatorname{diag}\{\sigma_{e1}^2, \sigma_{e2}^2\}\big) \\ f_t &= \exp(h_t/2)\varepsilon_t, \qquad \varepsilon_t \sim N(0,1), \\ h_{t+1} &= \mu + \phi h_t + \eta_t, \qquad \eta_t \sim N(0,1), \end{aligned}$$

where h_t is a scalar. Yu and Meyer (2006) derived the conditional correlation coefficient between y_{1t} and y_{2t} as

$$\frac{d}{\sqrt{(1 + \sigma_{e1}^2 \exp(-h_t))(d^2 + \sigma_{e2}^2 \exp(-h_t))}},$$

where $(1,\ d)' = D$. It is clear from the above expression that the correlation depends on the volatility of the factor.

Philipov and Glickman (2006) proposed a high-dimensional additive factor MSV model in which the factor covariance matrices are driven by Wishart random processes, as

$$\begin{aligned} y_t &= Df_t + e_t, \quad e_t \sim N(0, \Omega), \\ f_t \mid V_t &\sim N(0, V_t), \end{aligned}$$

$$V_t^{-1} \mid V_{t-1}^{-1}, A, v, \delta \sim Wish(v, S_{t-1}),$$
$$S_{t-1} = \frac{1}{v} A^{1/2} (V_{t-1}^{-1})^{\delta} (A^{1/2})',$$

where V_t^{-l} is a matrix of factor volatility, A is a symmetric positive definite matrix, δ is a scalar persistence parameter, *Wish* is the Wishart distribution, and v is the degrees of freedom parameter of the Wishart distribution.

2.3.2. *Multiplicative Factor Models*

The multiplicative factor MSV model, also known as the stochastic discount factor model, was considered in Quintana and West (1987). The one-factor model from this class decomposes the returns into two multiplicative components, a scalar common factor and a vector of idiosyncratic noise, as

$$y_t = \exp\left(\frac{h_t}{2}\right)\varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon),$$
$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2),$$

where h_t is a scalar. The first element in Σ_ε is assumed to be one for purposes of identification. By construction, the positivity for the variance of y_t is ensured. Compared with the basic MSV model, this model has a smaller number of parameters, which makes it more convenient computationally. Unlike the additive factor MSV model, however, the correlations are now invariant with respect to time. Moreover, the correlation in log-volatilities is always equal to 1, which is clearly too strong an assumption.

Ray and Tsay (2000) extended the one-factor model to a k-factor model, in which long range dependence is accommodated in the factor volatility:

$$y_t = \exp\left(\frac{h_t' v_t}{2}\right)\varepsilon_t, \quad \varepsilon_t \sim N(0, P_\varepsilon),$$
$$(1 - L)^d h_t = \mu + \eta_t,$$

where v_t is an $(m \times k)$ matrix of rank k, with $k < m$.

2.4. Time-Varying Correlation Models

A consequence of MSV factor models is that conditional correlations are time varying. However, the same set of parameters also determines the time-varying variance. To break the intrinsic tension between the two roles

served by the same set of parameters, one could model the dynamics in volatility and the dynamics in correlation separately, while maintaining the positivity of the covariance matrix. This is the idea behind the time-varying correlation models. To begin, let us relax the assumption of constant correlations in the correlation matrix P_ε in Equation (5) by considering the time-varying correlation matrix, $P_{\varepsilon t} = \{\rho_{ij,t}\}$, where $\rho_{ii,t} = 1$ and $\rho_{ij,t} = \rho_{ji,t}$.

Following the suggestion made by Tsay (2002) and Christodoulakis and Satchell (2002) in the bivariate GARCH framework, Yu and Meyer (2006) proposed that the Fisher transformation of $\rho_{12,t}$ could be modeled in a bivariate SV framework, as

$$\begin{aligned} \rho_{12,t} &= (\exp(v_t) - 1)/(\exp(v_t) + 1), \\ v_{t+1} &= \mu_v + \varphi(v_t - \mu_v) + u_t, \quad u_t \sim N(0, \sigma_u^2). \end{aligned} \tag{12}$$

The first equality in (12) guarantees that $|\rho_{12,t}| < 1$. The obvious drawback with this specification is the difficulty in generalizing it to a higher dimension.

In order to develop an MSV model that accommodates time-varying correlation, Asai and McAleer (2004) proposed a new MSV model based on the Wishart distribution. Suppose that y_t, conditional on Σ_t, has a multivariate normal distribution, $N(0, \Sigma_t)$, where the covariance matrix is given by

$$\Sigma_t = D_t \Gamma_t D_t. \tag{13}$$

In Equation (13), the time-varying correlation matrix is given by Γ_t, while the diagonal matrix D_t is defined by Equations (3) and (4). For the DCC model, Γ_t is specified as

$$\Gamma_t = Q_t^{*-1} Q_t Q_t^{*-1}$$

where $Q_t^* = (\text{diag}\{vecd(Q_t)\})^{1/2}$, where *vecd* creates a vector from the diagonal elements of a matrix, by using some positive definite matrix Q_t.

Asai and McAleer (2004) extended the DCC model by specifying Q_t as

$$\begin{aligned} Q_{t+1} &= (1 - \psi)\overline{Q} + \psi Q_1 + \Xi_1, \\ \Xi_1 &\sim W_m(v, \Lambda), \end{aligned}$$

where $W_m(v, \Lambda)$ denotes a Wishart distribution. This dynamic correlation MSV model guarantees the positive definiteness of Γ_t under the assumption that $\overline{Q}$ is positive definite and $|\psi| < 1$. The latter condition also implies that the time-varying correlations are mean reverting. In the special case where $v = 1$, Ξ_t can be expressed as the cross product of a multivariate normal

variate with mean zero and covariance matrix given by Λ. The total number of parameters in this model is $(3m^2 + 7m + 3)/2$.

Following Engle (2002), Yu and Meyer (2006) proposed an alternative MSV extension of DCC by specifying Q_t as

$$\begin{aligned} Q_{t+1} &= S + B \circ (Q_t - S) + A \circ (e_t e_t' - S) \\ &= S \circ (ll' - A - B) + B \circ Q_t + A \circ e_t e_t', \end{aligned}$$

where $e_t \sim N(0, I_m)$ and l is a vector of unit elements. According to Ding and Engle (2001) and Engle (2002), if A, B and $(ll' - A - B)$ are positive semidefinite, then Q_t will also be positive semidefinite. Moreover, if any of the matrices is positive definite, then Q_t will also be positive definite. The total number of parameters in this model is $2m(m + 2)$.

2.5. Alternative Specifications

This subsection introduces four alternative MSV models based on the matrix exponential transformation, the Cholesky decomposition, the Wishart autoregressive process, and the observed range.

2.5.1. *Matrix Exponential Transformation*

Motivated by the requirement for the positivity of the covariance, Chiu et al. (1996) proposed a general framework for the logarithmic covariance matrix based on the matrix exponential transformation, which is well known in the mathematics literature (see, for example, Bellman, 1970). In this subsection, we denote $\text{Exp}(\cdot)$ as the matrix exponential operation to distinguish it from the standard exponential operation. For any $m \times m$ matrix A, the matrix exponential transformation is defined by the power series expansion

$$\text{Exp}(A) = \sum_{s=0}^{\infty} \left(\frac{1}{S!} \right) A^S,$$

where A^0 reduces to the $m \times m$ identity matrix and A^s denotes the standard matrix multiplication of A s times. Thus, in general, the elements of $\text{Exp}(A)$ do not typically exponentiate the elements of A.

The properties of the matrix exponential and matrix logarithm are summarized in Chiu et al. (1996). For any real symmetric matrix A, we note the singular value decomposition $A = TDT'$, where the columns of the $m \times m$ orthonormal matrix T denote the appropriate eigenvectors of A, and D is an $m \times m$ diagonal matrix with elements equal to the eigenvalues of A. Therefore, $\text{Exp}(A) = T\text{Exp}(D)T'$, where $\text{Exp}(D)$ is an $m \times m$ diagonal matrix, with diagonal elements equal to the exponential of

the corresponding eigenvalues of A. If it is assumed that $\Sigma_t = \text{Exp}(A)$ for any symmetric matrix A_t, then Σ_t is positive definite.

Similarly, the matrix logarithmic transformation, $\text{Log}(B)$, for any $m \times m$ positive definite matrix, B, is defined by using the spectral decomposition of B.

Using the matrix exponential operator, we propose the model

$$\begin{aligned} y_t &\sim N(0, \Sigma_t), \\ \Sigma_t &= \text{Exp}(A_t), \end{aligned}$$

where $\alpha_t = vech(A_t)$ is a vector autoregressive process, as

$$\begin{aligned} \alpha_{t+1} &= \mu + \Upsilon x_t + \phi \circ \alpha_t + \eta_t, \\ \eta_t &\sim N(0, \Sigma_\eta), \end{aligned}$$

with $x_t = (y_t', |y_t'|)'$, $n \times 1$ parameter vectors μ and ϕ, where $n = (m+1)/2$, $n \times n$ covariance matrix Σ_η, and an $n \times 2m$ matrix of parameters Υ. The model may be regarded as an SV generalization of the matrix exponential GARCH model of Kawakatsu (2005). A limitation of this specification is that it is not straightforward to interpret the relationship between the elements of Σ_t and A_t.

2.5.2. *Cholesky Decomposition*

Tsay (2002) advocated an alternative approach to ensuring the positive definiteness of the covariance matrix, which uses the Cholesky decomposition. For a symmetric, positive definite matrix Σ_t, the Cholesky decomposition factors the matrix Σ_t uniquely in the form $\Sigma_t = L_t G_t L_t'$, where L_t is a lower triangular matrix with unit diagonal elements, and G_t is a diagonal matrix with positive elements.

The MSV model of Tsay (2002) is given as

$$\begin{aligned} y_t \mid \Sigma_t &\sim N(0, \Sigma_t), \\ \Sigma_t &= L_t G_t L_t', \\ L_t &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ q_{21,t} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ q_{m1,t} & q_{m2,t} & \cdots & 1 \end{bmatrix}, \\ G_t &= \text{diag}\{g_{11,t}, \ldots, g_{mm,t}\} = \text{diag}\{\exp(h_{1t}), \ldots, \exp(h_{mt})\}, \\ h_{it+1} &= \mu_i + \phi h_{it} + \eta_{it} \qquad i = 1, \ldots, m, \\ q_{ij,t+1} &= \alpha_{ij} + \beta_{ij} q_{ij,t} + u_{ij,t}, \qquad i > j. \end{aligned}$$

The elements in G_t are always positive owing to the exponential transformation. Consequently, the Cholesky decomposition guarantees the positive definiteness of Σ_t. It can be seen that the elements in L_t and G_t are assumed to follow an AR(1) process. Moreover, it is straightforward to derive the relationship between the variances and correlations, on the one hand, and the variables in L_t and G_t, on the other, by

$$\sigma_{ii,t} = \sum_{k=1}^{i} q_{ik,t}^2 g_{kk,t}, \quad i = 1, \ldots, m,$$

$$\sigma_{ij,t} = \sum_{k=1}^{j} q_{ik,t} q_{ik,t} g_{kk,t}, \quad i > j, \;\; i = 2, \ldots, m,$$

$$\rho_{ij,t} = \frac{\sigma_{ij,t}}{\sqrt{\sigma_{ii,t}\sigma_{jj,t}}} = \frac{\sum_{k=1}^{j} q_{ik,t} q_{jk,t} g_{kk,t}}{\sqrt{\sum_{k=1}^{j} q_{ik,t}^2 g_{kk,t} \sum_{k=1}^{i} q_{ik,t}^2 g_{kk,t}}}$$

It is clear from these expressions that the dynamics in $g_{ii,t}$ and $q_{ij,t}$ are the driving forces underlying the time-varying volatility and the time-varying correlation.

However, the dynamics underlying volatility are not determined separately from those associated with the correlations, as both are dependent on their corresponding AR(1) processes. This restriction is, at least in spirit, similar to that associated with factor MSV models. The number of parameters in this model is $3m(m+1)2$.

2.5.3. *Wishart Models*

Gourieroux et al. (2004) and Gourieroux (2006) proposed the Wishart autoregressive (WAR) multivariate process of stochastic positive definite matrices to develop an altogether different type of dynamic MSV model. Let Σ_t denote a time-varying covariance matrix of y_t. Gourieroux et al. (2004) defined the WAR(p) process as

$$\Sigma_t = \sum_{k=1}^{K} x_{kt} x_{kt}', \tag{14}$$

where $K > m - 1$ and each x_{kt} follows the VAR(p) model given by

$$x_{kt} = \sum_{l=1}^{p} A_i x_{k,t-i} + \varepsilon_{kt}, \qquad \varepsilon_{kt} \sim N(0, \Sigma).$$

By using the realized value of volatility, Gourieroux et al. (2004) estimated the parameters of the WAR(1) process using a two-step procedure based

on nonlinear least squares. Gourieroux (2006) proposed the continuous-time WAR process, which is not only a generalization of the Model (14) but also the multivariate extension of the CIR model suggested by Cox et al. (1985). As the CIR process has closed-form solutions for a number of financial problems, such as term structure of T-bonds and corporate bonds, and structural model for credit risk, the WAR process is appropriate for modeling multivariate risk in various financial problems and provides closed-form derivative prices.

Philipov and Glickman (2004) suggested an alternative model based on Wishart processes

$$y_t \mid \Sigma_t \sim N(0, \Sigma_t),$$
$$\Sigma_t^{-1} \mid v, S_{t-1} \sim W_m(v, S_{t-1}),$$

where v and S_t are the degrees of freedom and the time-dependent scale parameter of the Wishart distribution, respectively. With a time-invariant covariance structure, the above model may be considered as a traditional Normal-Wishart representation of the behavior of multivariate returns. However, Philipov and Glickman (2004) introduced time variation in the scale parameter

$$S_t = \frac{1}{v}(A^{1/2})(\Sigma_t^{-1})^d(A^{1/2}),$$

where A is a positive definite symmetric parameter matrix that is decomposed through a Cholesky decomposition as $A = (A^{1/2})(A^{1/2})'$, and d is a scalar parameter. The quadratic expression ensures that the covariance matrices are symmetric positive definite.

2.5.4. *Range-Based Model*

Tims and Mahieu (2003) proposed a range-based MSV model. As the range can be used as a measure of volatility, which is observed (or realized) when the high and low prices are recorded, Tims and Mahieu (2003) suggested a multivariate model for volatility directly, as

$$\log(range_t) = D' f_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \Sigma),$$
$$f_{t+1} = \Phi f_t + \eta_t, \qquad \eta_t \sim N(0, \Sigma_\eta).$$

As the volatility is not latent in this model, efficient estimation of the parameters is achieved through the use of the Kalman filter. It is not known, however, how to use this model for purposes of asset pricing.

To conclude this section, we summarize some MSV models where the reporting includes the specifications, the number of parameters in each

TABLE 1 Summary of MSV models

Model	Specification	Cited papers	# of parameters	Features
Basic MSV	$y_t = H_t^{1/2}\varepsilon_t, H_t^{1/2} = \mathrm{diag}\{\exp(h_t/2)\}$, $h_{t+1} = \mu + \phi \circ h_t + \eta_t$, $\begin{pmatrix}\varepsilon_t \\ \eta_t\end{pmatrix} \sim N\left[\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}P_\varepsilon & O \\ O & \Sigma_\eta\end{pmatrix}\right]$	Harvey et al. (1994), Jungbacker and Koopman (2006)	$m^2 + 2m$	Volatility clustering, volatility spillover, constant correlations
Leverage effect MSV	$\begin{pmatrix}\varepsilon_t \\ \eta_t\end{pmatrix} \sim N\left[\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}P_\varepsilon & L \\ L & \Sigma_\eta\end{pmatrix}\right]$ $L = \mathrm{diag}\{\lambda_1\sigma_{\eta,11}, \ldots, \lambda_m\sigma_{\eta,mm}\}$	Asai and McAleer (2005b)	$m^2 + 3m$	Volatility clustering, spillover, asymmetric volatility response, constant correlations
Additive factor MSV	$y_t = Df_t + e_t$, $f_t = \exp(h_t/2)\varepsilon_t$, $h_{t+1} = \mu + \phi h_t + \eta_t$	Jacquier et al. (1995)	$5m - 1$	Volatility clustering, dynamic correlations
Generalized additive factor MSV		Pitt and Shephard (1999a)	$8m - 1$	Volatility clustering, dynamic correlations
DCMSV		Asai and McAleer (2004)	$(3m^2 + 7m + 3)/2$	Volatility clustering, dynamic correlations
Cholesky decomposition MSV		Tsay (2002)	$3m(m+1)/2$	Volatility clustering, dynamic correlations

model, which of the cited papers uses such a model, and the main features of each model (see Table 1). For exchange rates, it is said that volatility asymmetry is not an important feature, and hence the asymmetric MSV models, such as the leverage MSV, are not particularly relevant. On the other hand, asymmetric MSV models are useful for stock returns. Some empirical results show that, for exchange rate returns and stock returns, the correlation is time varying. Therefore a DC-MSV model and a factor MSV model for exchange rate returns may be more suitable as far as time-varying correlations are concerned. Using a model selection criterion, the deviance information criterion (DIC), which will be discussed in Section 4, Yu and Meyer (2006) compared the performance of nine alternative MSV models for a bivariate exchange rate series and found that those MSV models that allow for time-varying correlations fit the data better. For stock returns, asymmetric MSV models with time-varying correlations should be developed in future research.

3. ESTIMATION

Unlike the multivariate GARCH model in which the conditional covariance at t is known given the information up to $t-1$, the conditional covariance is latent in the MSV models and has to be integrated out from the joint likelihood function of returns and the conditional covariance. As a result, MSV models do not have a closed-form expression for the likelihood function; hence the exact maximum likelihood is not readily available. Not surprisingly, the estimation methods proposed in the literature are nonstandard.

An important concern for the choice of a particular estimation method lies in its statistical efficiency. In addition to efficiency, other important issues related to estimation include (1) estimation of the latent volatility; (2) determination of the optimal filtering, smoothing, and forecasting methods; (3) computational efficiency; (4) applicability for flexible modeling. Broto and Ruiz (2004) provided a recent survey regarding the numerous estimation techniques for univariate SV models, ranging from inefficient methods, such as GMM, quasi-maximum likelihood, and spectral GMM based on the characteristic function, to more efficient methods, such as EMM, indirect inference, numerical maximum likelihood, simulated maximum likelihood and Markov chain Monte Carlo. While this general set of estimation methods can all be used to estimate certain MSV models, only a small subset of them has been applied in the literature, namely, quasi-maximum likelihood, simulated maximum likelihood, and Markov chain Monte Carlo. Consequently, in this review we will focus on these three estimation methods and refer readers to Broto and Ruiz (2004) for a review of the remaining estimation methods.

3.1. Quasi-Maximum Likelihood

In order to estimate the parameters of the model (3–5), Harvey et al. (1994) proposed a quasi-maximum likelihood (QML) method based on the property that the transformed vector $y_t^* = (\ln y_{1t}^2, \ldots, \ln y_{mt}^2)'$ has a state space form with the measurement equation given by

$$\begin{aligned} y_t^* &= h_t + \xi_t, \\ \xi_t &= \ln \varepsilon_t^2 = (\ln \varepsilon_{1t}^2, \ldots, \ln \varepsilon_{mt}^2)' \end{aligned} \tag{15}$$

and the transition equation (4). The measurement equation errors, ξ_t, are nonnormal, with mean vector $E(\xi_t) = -1.27931$, where 1 is an $m \times 1$ vector of unit elements. Harvey et al. (1994) showed that the covariance matrix of ξ_t, denoted Σ_ξ, is given by,

$$\Sigma_\xi = (\pi^2/2)\{\rho_{ij}^*\}, \quad \text{where } \rho_{ij}^* = 1, \quad \text{and} \quad \rho_{ij}^* = \frac{2}{\pi^2} \sum_{n=1}^{\infty} \frac{(n-1)!}{(1/2)_n n} \rho_{ij}^{2n}, \tag{16}$$

where $(x)_n = x(x+1)\cdots(x+n-1)$ and ρ_{ij} is defined by (5). Treating ξ_t as a Gaussian error term, QML estimates may be obtained by applying the Kalman filter to Equations (4) and (15). Taking account of the nonnormality in ξ_t, the asymptotic standard errors can be obtained by using the results established in Dunsmuir (1979).

If ρ_{ij}^* can be estimated, then it is also possible to estimate the absolute value of ρ_{ij} and the cross correlations between the different values of ε_{it}. Estimation of ρ_{ij} can be obtained by returning to the untransformed observations and noting that the sign of each of the pairs, $\varepsilon_{it}\varepsilon_{jt}$ $(i, j = 1, \ldots, m)$, will be the same as the corresponding pairs of observed values, $y_{it}y_{jt}$. Therefore the sign of ρ_{ij} is estimated as positive if more than one-half of the pairs, $y_{it}y_{jt}$, is positive.

One of the main features of this transformation is that ξ_t and η_t are uncorrelated, even if the original ε_t and η_t are correlated (see Harvey et al., 1994). Since the leverage effects assume a negative correlation between ε_t and η_t, as in Equation (7), the transformation may ignore the information regarding the leverage effects. In the univariate case, Harvey and Shephard (1996) recovered it in the state space form, given the signs of the observed values. As for the multivariate case, Asai and McAleer (2005b) derived the state space form for the leverage effects in the model (3), (4), and (10), based on pairs of the signs of y_{it} and y_{jt}. This representation enables use of the QML method based on the Kalman filter. However, Asai and McAleer (2005b) adopted the Monte Carlo likelihood method (to be discussed below) for purposes of efficient estimation.

The main advantages of the QML method are that it is computationally convenient and also straightforward for purposes of filtering, smoothing,

and forecasting. Unfortunately, the available (though limited) Monte Carlo experiments in the context of the basic univariate SV model suggest that the QML method is generally less efficient than the Bayesian MCMC technique and the likelihood approach based on Monte Carlo simulation (for further details, see Jacquier et al., 1994, and the discussions contained therein). It is natural to believe that inefficiency remains for the QML method relative to the Bayesian MCMC technique and the likelihood approach in the multivariate context, although no Monte Carlo evidence is yet available.

3.2. Simulated Maximum Likelihood

As discussed earlier, the likelihood function of MSV models involves high-dimensional integration, which is difficult to calculate numerically. Nevertheless, estimation of the parameters can be based on evaluating high-dimensional integrals with simulation methods and then maximizing the likelihood function, resulting in the so-called simulated maximum likelihood (SML) estimators. There are several ways to perform SML estimation for MSV models.

Perhaps the most usual approach to SML is the importance sampling method. The basic idea of this method is to approximate first the integrand by a multivariate normal distribution using the so-called Laplace approximation and then to draw samples from this multivariate normal distribution. To fix the idea, let λ_t denote a q-dimensional vector of latent variables, and let $f(Y, \Lambda; \theta)$ be the joint density of $Y = \{y_t\}_{t=1}^T$ and $\Lambda = \{\lambda_t\}_{t=1}^T$. The likelihood function associated with the observable variables, Y, is given by the $(T \times q)$-dimensional integral $L(\theta; Y) = \int f(Y, \Lambda; \theta) d\Lambda$. The Monte Carlo estimator of the likelihood function L based on importance sampling is then given by

$$L(\theta, Y) = \int f(Y, \Lambda; \theta) d\Lambda = \int \frac{f(Y, \Lambda; \theta)}{g(\Lambda \mid Y, \theta)} \, dG(\Lambda \mid Y, \theta),$$

where g is an importance density and G is an importance distribution function. The idea of this SML method is to draw samples $\Lambda^{(1)}, \ldots, \Lambda^{(S)}$ from g so that we can approximate $L(\theta; Y)$ by $\frac{1}{S} \sum \frac{f(Y, \Lambda^{(s)}, \theta)}{g(\Lambda^{(s)} \mid Y, \theta)}$. The key to this SML method is to match $f(Y, \Lambda; \theta)$ and $g(\Lambda \mid Y, \theta)$ as closely as possible, while ensuring that it is easy to simulate from g. To achieve this goal, Shephard and Pitt (1997), Durbin and Koopman (1997), and Durham (2005) proposed basing the importance sampler on the Laplace approximation to $f(Y, \Lambda; \theta)$. In particular, we choose the mean of g to be the mode of $\ln f(Y, \Lambda; \theta)$ with respect to Λ and the variance of g to be the negative of the inverse of the second derivative of $\ln f(Y, \Lambda; \theta)$ with

respect to Λ evaluated at the mode. The convergence of $\frac{1}{S}\sum \frac{f(Y,\Lambda^{(s)},\theta)}{g(\Lambda^{(s)}\mid Y,\theta)}$ to $L(\theta; Y)$ can be justified by the central limit theorem, though it requires the variance of $\frac{f(Y,\Lambda^{(s)},\theta)}{g(\Lambda^{(s)}\mid Y,\theta)}$ to be finite. Koopman and Shephard (2004) proposed a test to check the validity of such an importance sampler.

We will now explain three SML methods used in the MSV literature. Apart from the importance sampling methods, it may be possible to use the numerical integration techniques, such as Kitagawa's (1987) extended Kalman filter.

The first SML method is the accelerated Gaussian importance sampling (AGIS) approach, as developed in Danielsson and Richard (1993). This is designed to estimate dynamic latent variable models, where the latent variable follows a Gaussian process. Danielsson's comments on Jacquier et al. (1994) show that the finite sample property of this SML estimator is close to that of the Bayesian MCMC method in the context of the basic univariate SV model. As for MSV models, Danielsson (1998) applied the AGIS approach to estimate the parameters of the MSV model in Equations (3–5). It seems difficult to extend the AGIS approach to accommodate more flexible SV models, as the method is specifically designed for models with a latent Gaussian process.

While the AGIS technique has limited applicability, a closely related method, namely the efficient importance sampling (EIS) procedure, as proposed by Liesenfeld and Richard (2003, 2006), is applicable to models with more flexible classes of distributions for the latent variables. As in the case of AGIS, EIS is a Monte Carlo technique for the evaluation of high-dimensional integrals. The EIS relies on a sequence of simple low-dimensional least squares regressions to obtain a very accurate global approximation of the integrand. This approximation leads to a Monte Carlo sampler, which produces highly accurate Monte Carlo estimates of the likelihood. Using EIS, Liesenfeld and Richard (2003) successfully estimated the parameters of the additive factor model (11) with one factor.

The last SML method, also known as Monte Carlo likelihood (MCL), was proposed to evaluate the likelihood function of linear non-Gaussian state-space models. The MCL method can approximate the likelihood function to an arbitrary degree of accuracy by decomposing it into a Gaussian part, which is constructed by the Kalman filter, and a remainder function, whose expectation is evaluated through simulation. Asai and McAleer (2005b) and Jungbacker and Koopman (2006) used this importance sampling method to estimate various MSV models. While Jungbacker and Koopman (2006) considered the basic MSV model and the factor MSV models and documented the good performance of the SML method, Asai and McAleer (2005b) dealt with the basic MSV model and asymmetric MSV models.

Durbin and Koopman (1997) demonstrated that the log-likelihood function of state space models with non-Gaussian measurement disturbances could be expressed simply as

$$\ln L(y \mid \theta) = \ln L_G(y \mid \theta) + \ln \left[\frac{p_\xi(\xi \mid \theta)}{p_G(\xi \mid y, \theta)} \right], \tag{17}$$

where $y = (y_1, \ldots, y_T)'$, $y_t = (y_{1t}, \ldots, y_{mt})'$, and $\xi = (\xi_t, \ldots, \xi_T)'$ and $\ln L_G(y \mid \theta)$ are the vectors of measurement disturbances and the log-likelihood function of the approximating Gaussian model, respectively, $p_\xi(\xi \mid \theta)$ is the true density function, $p_G(\xi \mid y, \theta)$ is the Gaussian density of the measurement disturbances of the approximating model, and E_G refers to the expectation with respect to the so-called importance density $p_G(\xi \mid y, \theta)$ associated with the approximating model. Equation (17) shows that the non-Gaussian log-likelihood function can be expressed as the log-likelihood function of the Gaussian approximating model plus a correction for the departures from the Gaussian assumptions relative to the true model.

A key feature of the MCL method is that only the minor part of the likelihood function requires simulations, unlike other SML methods. Therefore, the method is computationally efficient in the sense that it needs only a small number of simulations to achieve the desirable accuracy for empirical analysis. The drawback is that it requires a linear non-Gaussian state-space representation for the model.

The MCL estimates of the parameters, θ, are obtained by numerical optimization of the unbiased estimate of Equation (17). The log-likelihood function of the approximating model, $\ln L_G(y \mid \theta)$, can be used to obtain the starting values. Sandmann and Koopman (1998) are the first to have used this MCL approach in the SV literature. Asai and McAleer (2005b) developed the MCL method for asymmetric MSV models. This MCL method is also able to accommodate the additive factor MSV model.

Being a maximum likelihood method, SML inherits the asymptotic properties of exact maximum likelihood and hence is asymptotically efficient. Therefore relative to QML, SML is statistically more efficient but computationally less efficient. It should be stressed that SML methods require a lower computational burden than the Markov chain Monte Carlo methods explained below.

3.3. Markov Chain Monte Carlo

Much of the focus in the SV literature is on the development and the use of Markov chain Monte Carlo (MCMC) methods. In the context of the basic SV model, Andersen et al. (1999) documented a finite sample comparison of various methods and found that MCMC is the most efficient

estimation tool, while Meyer and Yu (2000) discuss the flexibility of modeling modifications of the basic SV model. Moreover, as a by-product of parameter estimation, MCMC methods can provide estimates of latent volatility and predictive distributions for volatility (see Jacquier et al., 1994).

MCMC was originally developed in the statistical physics literature. The earliest contributions in the context of the basic univariate SV include Jacquier et al. (1994). Useful refinements were made in Shephard and Pitt (1997) and Kim et al. (1998) in the context of univariate SV. As a Bayesian approach, the idea behind MCMC methods is to produce variates from a given multivariate density (the posterior density in Bayesian applications) by repeatedly sampling a Markov chain whose invariant distribution is the target density of interest. One MCMC method proposed in the literature focuses on the density $\pi(\theta, h \mid y)$ instead of the usual posterior density, $\pi(\theta \mid y)$, since the latter requires computation of the likelihood function $f(y \mid \theta) = \int f(y \mid h, \theta) f(h \mid \theta) dh$. As a result, the parameter space is augmented by including all the latent variables. The MCMC procedure only requires alternating back and forth between drawing from $f(h \mid \theta, y)$ and $f(\theta \mid h, y)$. This process of alternating between conditional distributions produces a cyclic chain. See Tierney (1994) for conditions under which a realization of the Markov chain produces the target stationary distribution.

Regarding the property of sample variates from an MCMC algorithm, they are a high-dimensional sample from the target density of interest. These draws can be used as the basis for drawing inferences by appealing to suitable ergodic theorems for Markov chains. For example, posterior moments and marginal densities can be estimated (or simulated consistently) by averaging the relevant function of interest over the sampled variates. The posterior mean of θ is estimated simply as the sample mean of the simulated θ values. These estimates can be made arbitrarily accurate by increasing the simulation sample size.

As a Bayesian likelihood-based approach, MCMC relies on posterior distributions to make statistical inference. Therefore the full statistical efficiency is automatically achieved. The real concerns include (1) the computational efficiency, that is, how fast the chain can converge; and (2) the flexibility of the MCMC algorithm. Jacquier et al. (1994) proposed to update each element in the vector of latent variables h, one at a time. As the vector h is often highly persistent, this single-move sampler is numerically inefficient and typically requires a large number of draws before the chain converges. In Kim et al. (1998), a multimove sampler, which updates the whole state vector at once, was proposed. It was shown that computational efficiency can be greatly improved relative to the single-move sampler in the context of the basic univariate SV model.

One particularly important technical advantage of the Bayesian MCMC method over classical inferential techniques is that MCMC does not need to use numerical optimization. This advantage becomes especially

important when the number of parameters to be estimated is large, as in the application of MSV models to the analysis of financial data. It should be noted that it is not necessary to apply numerical optimization for the Bayesian MCMC, but that it is possible to include some optimization steps in order to improve computational efficiency (see Chib et al., 2005).

In the context of MSV, Jacquier et al. (1999), Pitt and Shephard (1999a), and Aguilar and West (2000) have applied the MCMC procedure to estimate additive factor MSV models. Chan et al. (2005) estimated an asymmetric MSV model. Philipov and Glickman (2006) estimated a Wishart MSV model. Moreover, Yu and Meyer (2006) estimated and compared nine MSV models, covering most of the model classes reviewed in Section 2. By doing so, they have demonstrated the ease with which different MSV models can be studied routinely with MCMC. In particular, Yu and Meyer (2006) employed the purpose-built Bayesian software package called BUGS (Bayesian analysis using the Gibbs sampler). However, all these MCMC algorithms are based on single-move algorithms, and hence can be numerically inefficient.

In order to improve the simulation efficiency, Chib et al. (2005) developed a multi-move MCMC algorithm that greatly improves the simulation efficiency for a factor MSV model augmented with jumps. Nardari and Scruggs (2003) and Han (2006) applied this estimation methodology to address different empirical issues. Liesenfeld and Richard (2006) proposed an alternative multimove MCMC method to estimate a factor MSV model based on EIS, which can be used to estimate SV models by maximum likelihood as well as simulation smoothing.

Bos and Shephard (2006) modeled the Gaussian errors in the standard Gaussian linear state space model as an SV process and showed that conventional MCMC algorithms for this class of models are ineffective. Rather than sampling the unobserved variance series directly, Bos and Shephard (2006) sampled in the space of the disturbances, which decreased the correlation in the sampler and increased the quality of the Markov chain. Using the reparameterized MCMC sampler, they showed how to estimate an unobserved factor model.

Smith and Pitts (2006) used a bivariate SV model to measure the effects of intervention in stabilization policy. Missing observations were accommodated in the model and a data-based Wishart prior for the precision matrix of the errors in the central bank intervention equation was suggested. A threshold model for the transition equation was estimated by MCMC jointly with the bivariate SV model.

It should be noted that in the models estimated in Chan et al. (2005) and Nardari and Scruggs (2003), and especially in Chib et al. (2005) and Han (2006), there is a large number of parameters, ranging from dozens to hundreds, excluding the latent variables. This level of feasibility, together

with the full statistical efficiency of the resulting estimators, seems difficult to match on the basis of the alternative estimation methods.

4. DIAGNOSTIC CHECKING AND MODEL COMPARISON

Although standard diagnostic check methods for specification are based on the residuals, since MSV models involve the latent variables, it is not obvious how to retrieve the residuals. Following Kim et al. (1998), Pitt and Shephard (1999a) proposed diagnostic checking using a particle filter, a simulation method designed to deal with nonlinear non-Gaussian state-space models. It is known that, for nonlinear non-Gaussian state-space models, one cannot find the closed-form expression for the distribution in either the updating stage or the in-sampling prediction stage. The idea of the particle filter is to draw "particles" from these distributions. The particle filter is one class of simulation filtering methods. Other simulation filtering techniques, such as the EIS filter of Liesenfeld and Richard (2003) and the reprojection technique of Gallant and Tauchen (1998), may also be applicable. By using these filtering methods, we can obtain samples from the prediction density, $f(h_{t+1} \mid Y_t; \theta)$, where $Y_t = (y_1, \ldots, y_t)'$. Pitt and Shephard (1999a) focus on four quantities for assessing overall model fit, outliers and observations which have substantial influences on the fitted model.

The first quantity is the log-likelihood for $t+1$, $l_{t+1} = \log f(y_{t+1} \mid Y_t; \theta)$. As we have

$$f(y_{t+1} \mid Y_t; \theta) = \int f(y_{t+1} \mid h_{t+1}; \theta)\, dF(h_{t+1} \mid Y_t; \theta),$$

Monte Carlo integration may be used as

$$\hat{f}(y_{t+1} \mid Y_t; \theta) = \frac{1}{M} \sum_{i=1}^{M} f\left(y_{t+1} \mid h_{t+1}^i; \theta\right),$$

where $h_{t+1}^i \sim f(h_{t+1} \mid Y_t; \theta)$. It is possible to evaluate the log-likelihood at the ML (SML) estimates or at the posterior means.

The second quantity is the normalized log-likelihood, l_t^n. Pitt and Shephard (1999a) used samples from z^j $(j = 1, \ldots, S)$, where $z^j \sim f(y_{t+1} \mid Y_t; \theta)$, to obtain samples l_{t+1}^i using the above method. Denote the sample mean and standard deviation of the samples of log-likelihood as μ_{t+1}^l and σ_{t+1}^l, respectively. The normalized log-likelihood at $t+1$ may be computed as $l_{t+1}^n = (l_{t+1} - \mu_{t+1}^l)/\sigma_{t+1}^l$. If the model and parameters are correct, then this statistic should have zero mean and unit variance. Large

negative values indicate that an observation is less likely than would be expected from the model.

The third quantity is the uniform residual, $u_{t+1} = F(l_{t+1} \mid Y_t; \theta)$, which may be estimated as

$$\hat{u}_{t+1} = \widehat{F}(l_{t+1}) = (1/S) \sum_{j=1}^{S} I(l_{t+1}^{j} < l_{t+1}),$$

where the l_{t+1}^{j} are constructed as above. Assuming that the parameter vector θ is known, under the null hypothesis that the model is correct, it follows that $\hat{u}_{t+1} \sim UID(0, 1)$.

Finally, the fourth quantity is the distance measure, d_t, which may be computed by

$$\Sigma_{t+1} = V(y_{t+1} \mid Y_t; \theta) \doteq \left(\frac{1}{M}\right) \sum_{i=1}^{M} V(y_{t+1} \mid h_{t+1}^{i}; \theta),$$

where $h_{t+1}^{i} \sim f(h_{t+1} \mid Y_t; \theta)$. If the conditional distribution of y_t is multivariate normal, then the quantity $d_t = y_t' \Sigma_t^{-1} y_t$ is independently distributed as χ_m^2 under the null hypothesis that the parameters and model are correct. Therefore we may use $\sum_{t=1}^{T} d_t \sim \chi_{mT}^2$ as a test statistic.

It should be noted that all the diagnostic checking methods are computationally expensive. This is the case for the basic MSV model, and even more so for more flexible MSV models. Further research is needed to develop methods that are easier to use.

When the MCMC procedure is used, it may require checking the convergence of the Markov chains and prior sensitivities. The former can be assessed by correlograms, and the latter by using alternative priors (for further details, see Chib, 2001; Chib et al., 2005; Kim et al., 1998).

Turning to model selection, we may use the likelihood ratio test for nested models and Akaike information criterion (AIC) or Bayesian information criterion (BIC) for the nonnested models, in the context of the likelihood-based methods, such as SML and MCL. In the Bayesian framework, model comparison can be conducted via the posterior odds ratio or Bayes factor. For both values, the marginal likelihood needs to be calculated, for which estimation is based on the procedure proposed by Chib (1995) and its various extensions.

The AIC is inappropriate for the MCMC method because, when MCMC is used to estimate the SV models, as mentioned above, the parameter space is often augmented. For example, in the basic univariate SV model, we include the T latent volatilities in the parameter space, with T being the sample size. As these volatilities are dependent, they cannot be counted as T additional free parameters. Consequently, AIC is not applicable for

comparing SV models. Recently, Berg et al. (2004) showed that model selection of alternative univariate SV models can be performed easily using the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002), while Yu and Meyer (2006) compared alternative MSV models using DIC.

5. CONCLUDING REMARKS

Relative to the extensive theoretical and empirical multivariate conditional volatility (or GARCH) literature, the MSV literature is still in its infancy. The majority of existing research in the MSV literature deals with specifications and/or estimation techniques, which are often illustrated by fitting a particular set of MSV models to financial returns series. In terms of model specification, we share the view of Bauwens et al. (2006) about multivariate modeling, especially that providing "a realistic and parsimonious specification of the variance matrix ensuring its positivity" is crucial. As a related point, model diagnostic checking methods that are computationally inexpensive are needed so that the limitations of various models can be more easily explored and improved.

A few papers have directly addressed important economic issues using MSV models, but no paper has yet examined the predictive power of MSV models for volatility. To the best of our knowledge, Nardari and Scruggs (2003), Han (2006), and Smith and Pitts (2006) are the only studies that have examined economic issues in the context of MSV models. Nardari and Scruggs (2003) used MSV models to address the restrictions in the APT theory. Han (2006) examined the economic values of MSV models. Smith and Pitts (2006) analyzed central bank intervention. Clearly, further serious empirical applications of MSV models are warranted. Useful applications of MSV include the study of volatility comovement, volatility causality, volatility contagion, the computation of hedge ratios, and the examination of relative performance for predicting future volatility.

ACKNOWLEDGMENTS

The authors wish to acknowledge the helpful comments of two referees and the insightful discussions with Yoshi Baba, Jiti Gao, Clive Granger, Suhejla Hoti, Essie Maasoumi, Kimio Morimune, Les Oxley, Peter Phillips, Peter Robinson, Neil Shephard, and Ruey Tsay. The first author acknowledges the financial support of the Japan Society for the Promotion of Science, the Australian Academy of Science, and the Grant-in-Aid for the 21st Century COE program "Microstructure and Mechanism Design in Financial Markets" from the Ministry of Education, Culture, Sports, Science and Technology of Japan. The second author is grateful for the financial

support of the Australian Research Council. The third author gratefully acknowledges financial support from the Office of Research at Singapore Management University.

REFERENCES

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

Andersen, T., Chung, H., Sorensen, B. (1999). Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Econometrics* 91:61–87.

Asai, M., McAleer, M. (2004). Dynamic correlations in stochastic volatility models. Unpublished paper, Faculty of Economics, Tokyo Metropolitan University.

Asai, M., McAleer, M. (2005a). Dynamic asymmetric leverage in stochastic volatility models. *Econometric Reviews* 24:317–332.

Asai, M., McAleer, M. (2005b). Asymmetric multivariate stochastic volatility. *Econometric Reviews* 25(2–3):453–473.

Asai, M., McAleer, M. (2006). Asymmetry, leverage and size effects in stochastic volatility models. To appear in *Finance Letters*.

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

Bellman, R. (1970). *Introduction to Matrix Analysis*. New York: McGraw-Hill.

Berg, A., Meyer, R., Yu, J. (2004). Deviance information criterion for comparing stochastic volatility models. *Journal of Business and Economic Statistics* 22:107–120.

Black, F. (1976). Studies of stock market volatility changes. Proceedings of the American Statistical Association, Business and Economic Statistics Section, pp. 177–181.

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

Bos, C. S., Shephard, N. (2006). Inference for adaptive time series models: stochastic volatility and conditionally gaussian state space form. *Econometric Reviews* 25(2–3):219–244.

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

Campbell, J. Y., Hentschel, L. (1992). No news is good news: an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31:281–318.

Chan, D., Kohn, R., Kirby, C. (2005). Multivariate stochastic volatility models with correlated errors. *Econometric Reviews* 25(2–3):245–274.

Chesney, M., Scott, L. O. (1989). Pricing European currency options: a comparison of the modified black-scholes model and a random variance model. *Journal of Financial and Quantitative Analysis* 24:267–284.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 90:1313–1321.

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

Chib, S., Nardari, F., Shephard, N. (2005). Analysis of high dimensional multivariate stochastic volatility models. To appear in *Journal of Econometrics*.

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

Christie, A. A. (1982). The stochastic behavior of common stock variances: value, leverage and interest rate effects. *Journal of Financial Economics* 10:407–432.

Christodoulakis, G., Satchell, S. E. (2002). Correlated ARCH (CorrARCH): modelling the time-varying conditional correlation between financial asset returns. *European Journal of Operational Research* 139:351–370.

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

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

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

Danielsson, J., Richard, J.-F. (1993). Quadratic acceleration for simulated maximum likelihood evaluation. *Journal of Applied Econometrics* 8:153–173.

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

Ding, Z., Engle, R. (2001). Large scale conditional covariance modelling, estimation and testing. *Academia Economic Papers* 29:157–184.

Doz, C., Renault, E. (2006). Factor stochastic volatility in mean models: a GMM approach. *Econometric Reviews* 25(2–3):275–309.

Dunsmuir, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Annals of Statistics* 7:490–506.

Durbin, J., Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika* 84:669–684.

Durham, G. (2005). Monte Carlo methods for estimating, smoothing, filtering one- and two-factor stochastic volatility models. To appear in *Journal of Econometrics*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* 50:987–1007.

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., Ng, V. (1993). Measuring and testing the impact of news in volatility. *Journal of Finance* 43:1749–1778.

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

Engle, R. F., Ng, V., Rothschild, M. (1990). Asset pricing with a factor-ARCH covariance structure. *Journal of Econometrics* 45:213–237.

Gallant, A. R., Tauchen, G. E. (1998). Reprojecting partially observed systems with application to interest rate diffusions. *Journal of the American Statistical Association* 93:10–24.

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

Glosten, L., Jagannathan, R., Runkle, D. (1993). On the relation between the expected value and volatility of nominal excess returns on stocks. *Journal of Finance* 46:1779–1801.

Gourieroux, C. (2006). Continuous time wishart process for stochastic risk. *Econometric Reviews* 25(2–3):177–217.

Gourieroux, C., Jasiak, J., Sufana, R. (2004). The wishart autoregressive process of multivariate stochastic volatility. Unpublished paper, *CREST and CEPREMAP*, France.

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

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

Hull, J., White, A. (1987). The pricing of options on assets with stochastic volatility. *Journal of Finance* 42:281–300.

Jacquier, E., Polson, N. G., 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., Rossi, P. E. (1995). Models and priors for multivariate stochastic volatility. CIRANO Working Paper No. 95s-18, Montreal.

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

Jacquier, E., Polson, N. G., Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122:185–212.

Jungbacker, B., Koopman, S. J. (2006). Monte Carlo likelihood estimation for three multivariate stochastic volatility models. *Econometric Reviews* 25(2–3):385–408.

Kawakatsu, H. (2005). Matrix exponential GARCH. To appear in *Journal of Econometrics*.

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

Kitagawa, G. (1987). Non-Gaussian state-space modelling of non-stationary time series (with discussion). *Journal of American Statistical Association* 82:1032–1063.

Koopman, S. J., Shephard, N. (2004). Estimating the likelihood of the stochastic volatility model: testing the assumption behind importance sampling. Working paper, Nuffield College.

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

Liesenfeld, R., Richard, J.-F. (2006). Classical and bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews* 25(2–3):335–360.

McAleer, M. (2005). Automated inference and learning in modeling financial volatility. *Econometric Theory* 21:232–261.

Meyer, R., Yu, J. (2000). BUGS for a bayesian analysis of stochastic volatility models. *Econometrics Journal* 3:198–215.

Nardari, F., Scruggs, J. (2003). Analysis of linear factor models with multivariate stochastic volatility for stock and bond returns. Unpublished paper, Department of Finance, Arizona State University.

Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* 59:347–370.

Philipov, A., Glickman, M. E. (2004). Multivariate stochastic volatility via Wishart processes. To appear in *Journal of Business and Economic Statistics*.

Philipov, A., Glickman, M. E. (2006). Factor multivariate stochastic volatility via Wishart processes. *Econometric Reviews* 25(2–3):311–334.

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

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

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

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

Ross, S. A. (1976). The arbitrage theory of asset pricing. *Journal of Economic Theory* 13:641–660.

Sandmann, G., Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics* 87:271–301.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In: Cox, D. R., Hinkley, D. V., Barndorff-Nielsen, O. E., eds. *Time Series Models in Econometrics, Finance and Other Fields*. London: Chapman & Hall, pp. 1–67.

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.

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

Smith, M., Pitts, A. (2006). Foreign exchange intervention by the bank of Japan: Bayesian analysis using a bivariate stochastic volatility model. *Econometric Reviews* 25(2–3):425–451.

So, M. K. P., Li, W. K., Lam, K. (2002). A threshold stochastic volatility model. *Journal of Forecasting* 21:473–500.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*. 64:583–639.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes—a study of the daily sugar prices 1961–75. In: Anderson, O. D., ed. *Time Series Analysis: Theory and Practice*. Vol. 1. North-Holland, pp. 203–226.

Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester: John Wiley.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics* 22:1701–1762.

Tims, B., Mahieu, R. (2003). A range-based multivariate stochastic volatility model for exchange rates. *Econometric Reviews* 25(2–3):409–424.

Tsay, R. S. (2002). *Analysis of Financial Time Series: Financial Econometrics*. New York: John Wiley.

Wiggins, J. B. (1987). Option values under stochastic volatility: theory and empirical estimates. *Journal of Financial Economics* 19:351–372.

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics* 127:165–178.

Yu, J., Meyer, R. (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews* 25(2–3):361–384.