

SMU ECONOMICS & STATISTICS WORKING PAPER SERIES



Multivariate Stochastic Volatility Models: Bayesian Estimation and Model Comparison

Jun Yu, Renate Meyer

November 2004

Paper No. 23-2004

Multivariate Stochastic Volatility Models: Bayesian Estimation and Model Comparison*

Jun Yu[†] and Renate Meyer[‡]

November 22, 2004

Abstract

In this paper we show that fully likelihood-based estimation and comparison of multivariate stochastic volatility (SV) models can be easily performed via a freely available Bayesian software called WinBUGS. Moreover, we introduce to the literature several new specifications which are natural extensions to certain existing models, one of which allows for time varying correlation coefficients. Ideas are illustrated by fitting, to a bivariate time series data of weekly exchange rates, nine multivariate SV models, including the specifications with Granger causality in volatility, time varying correlations, heavy-tailed error distributions, additive factor structure, and multiplicative factor structure. Empirical results suggest that the most adequate specifications are those that allow for time varying correlation coefficients.

JEL classification: C11, C15, C30, G12

Keywords: Multivariate stochastic volatility; Granger causality in volatility; Heavy-tailed distributions; Time varying correlations; Factors; MCMC; DIC.

1 Introduction

Univariate stochastic volatility (SV) models offer powerful alternatives to ARCH-type models in accounting for both the conditional and unconditional properties of volatility. Superior

*Jun Yu gratefully acknowledges financial support from the Wharton-SMU Research Center and computing support from the Center for Academic Computing, both at Singapore Management University. The research of the second author was supported by the Royal Society of New Zealand Marsden Fund. We also wish to thank Manabu Asai, Ching-Fan Chung, Mike McAleer, Yiu Kuen Tse, seminar participants at the Workshop on Econometric Theory and Applications in Taiwan, for helpful discussion. Author for correspondence is Jun Yu. Both WinBUGS code and data used in this paper can be downloaded from Jun Yu's web site <http://www.mysmu.edu/faculty/yujun/default.htm>.

[†]School of Economics and Social Science, Singapore Management University, 469 Bukit Timah Road, Singapore 259756; email: yujun@smu.edu.sg.

[‡]Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand; email: meyer@stat.auckland.ac.nz

performance of univariate SV models over ARCH-type models are documented in Danielsson (1994) and Kim, Shephard and Chib (1998) in terms of in-sample fitting, and in Yu (2002) in terms of out-of-sample forecasting. As a result, the univariate SV model has been the subject of considerable attention in the literature; see, for example, Shephard (2004) for a collection of relevant studies on this topic.

There are both theoretical and empirical reasons why there is a great need to study multivariate volatility models. On the one hand, much of financial decision making, such as portfolio optimization, asset allocation, risk management, and asset pricing, clearly needs to take correlations into account. On the other hand, it is well known that financial market volatilities move together over time across assets. As a result, the multivariate ARCH models (MARARCH) have attracted a lot of attention in modern finance theory and enjoyed voluminous empirical applications; see Bauwen, Laurent and Rombouts (2004) for a survey. Important contributions are Bollerslev, Engle and Woodridge (1988), Diebold and Nerlove (1989), Bollerslev (1990), Engle, Ng and Rothschild (1990), Engle and Kroner (1995), Braun, Nelson and Sunier (1995), Engle (2002), Tse and Tsui (2002), among many others.

Compared to the MARARCH literature, the literature on multivariate SV is much limited (see Asai, McAleer and Yu (2004) for a survey), reflected by much fewer published papers on the topic to date (Harvey, Ruiz and Shephard, 1994; Danielsson, 1998; Pitt and Shephard, 1999; Aguilar and West, 2000; Liesenfeld and Richard, 2003). Yet, the multivariate SV models have certain statistical attractions relative to the MARARCH models (Harvey et al 1994). We believe there are several reasons why the multivariate SV models have had fewer empirical applications. Firstly, the multivariate SV models are more difficult to estimate. Although estimation is already an issue for the MARARCH models, it is believed that estimation is more of an issue for the multivariate SV models. This is because, apart from the inherent problems of multivariate models such as high dimensionality of the parameter space and required positive semi-definiteness of covariance matrices, the likelihood function has no closed form for the multivariate SV model. Secondly, as a result of difficulties with parameter estimation, the computation of model comparison criteria becomes extensive and demanding. Thirdly, compared to abundant alternative specifications in MARARCH, only a handful of multivariate SV model specifications have appeared in the literature. As a result, the existing multivariate SV models may not be able to describe some important stylized features of the data.

A variety of estimation methods have been proposed to estimate the SV models. Less efficient methods include GMM (Melino and Turnbull, 1990 and Andersen and Sorensen, 1996), the quasi maximum likelihood method (Harvey, Ruiz and Shephard, 1994), the method via the empirical characteristic function (Knight, Satchell and Yu, 2002). Fully likelihood-based methods include the simulated maximum likelihood method (SML) (Danielsson, 1994, Richard and Zhang, 2004, Durham, 2004), the numerical maximum likelihood method (Fridman and Harris, 1998), the maximum likelihood Monte Carlo method (Sandmann and Koopman, 1998), and Bayesian Markov Chain Monte Carlo (MCMC) methods (Jacquier, Polson and Rossi, 1994 and Kim et al., 1998). Andersen, Chung and Sorensen (1999) documented a finite sample comparison of various methods in Monte Carlo studies and found that MCMC is one of the most efficient estimation tools. Not surprisingly MCMC is generally regarded in the literature as benchmark for efficiency. Furthermore, as a byproduct of parameter estimation, MCMC methods provide smoothed estimates of latent variables (Jacquier, Poslon and Rossi, 1994). This is because MCMC augments the parameter space by including latent variables. Moreover, unlike most frequentist methods reviewed above whose inference is based on asymptotic arguments, MCMC inference is based on the exact posterior distribution of parameters and latent variables. Another advantage of MCMC is that numerical optimization is not needed in general. This advantage is of practical importance, especially when a model has many estimated parameters. As a result, MCMC has been extensively used to estimate univariate SV models in the literature.

Meyer and Yu (2000) illustrated the ease of implementing Bayesian estimation of univariate SV models based on purpose-built MCMC software called BUGS (**B**ayesian analysis **u**sing **G**ibbs sampler) developed by Spiegelhalter et al. (1996).¹ Since then BUGS has been employed to estimate univariate SV models in a number of studies (for example, Meyer, Fournier and Berg, 2003, Berg, Meyer and Yu, 2004, Lancaster, 2004, Selçuk, 2004, and Yu, 2004). Furthermore, Berg et al. (2004) showed that model selection of alternative univariate SV models is easily performed using the deviance information criterion (DIC) which is computed by BUGS. Arguably, univariate SV models can now be handled routinely in a straightforward fashion.

¹Note that BUGS is available free of charge from

<http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>

for a variety of operating systems such as UNIX, LINUX, and WINDOWS.

Unlike univariate SV models, however, “multivariate stochastic volatility models still pose significant computational challenges to applied researchers” (Chan, Kohn and Kirby, 2003).

One of the main purposes of the present paper is to show that fully likelihood-based estimation and comparison of multivariate SV models can be easily performed via the WINDOWS version of BUGS (WinBUGS) (Spiegelhalter et al 2003). The contribution of our paper is two-fold. First, we extend the literature by offering several interesting extensions to the existing specifications. In particular, we specify a model which allows for Granger causality in volatility and a model with time varying correlations. Second, we extend Meyer and Yu (2000) and Berg et al (2004) to the multivariate setting and show that both estimation and model comparison for multivariate SV models can also be handled in the same way as for the univariate case. We then illustrate the implementation by estimating and comparing nine alternative multivariate SV models in an empirical study. To the best of our knowledge, a comparison of such a rich class of multivariate SV models has not been done before. The comparison results in several interesting empirical findings.

The remainder of the paper is organized as follows. In Section 2, we illustrate the differences among the existing multivariate SV models in a bivariate setting, and propose several new multivariate SV specifications. Section 3 reviews a Bayesian approach for parameter estimation using WinBUGS. Section 4 describes a Bayesian approach for model comparison via DIC. In Section 5, we illustrate the estimation and model comparison using an example of Australian/US dollar and New Zealand/US dollar exchange rates. Section 6 concludes.

2 Multivariate SV Models

2.1 Stylized facts of financial asset returns

Considering that multivariate SV models are most useful for describing the dynamics of financial asset returns, we first summarize some well documented stylized facts of financial asset returns:

1. Asset return distributions are leptokurtic.
2. Asset return volatilities cluster.
3. Returns are cross-dependent.

4. Volatilities are cross-dependent.
5. Sometimes volatility of one asset Granger causes volatility of another asset (that is, volatility spills over from one market to another market).
6. There often exists a lower dimensional factor structure which can explain most of the correlation.
7. Correlations are time varying.

In addition to these seven stylized facts, the issues such as the dimensionality of the parameter space and positive semi-definiteness of the covariance matrix are of practical importance. When we review the existing models and introduce our new models we will comment on their appropriateness for dealing with the stylized facts and the two issues posed above.

2.2 Alternative specifications in a bivariate setting

To illustrate the difference and linkage among alternative multivariate SV models, we focus on the bivariate case in this paper. In particular, we consider nine different bivariate SV models (with acronyms in bold face), two of which are new to the literature.

Let the observed (mean-centered) log-returns at time t be denoted by $\mathbf{y}_t = (y_{1t}, y_{2t})'$ for $t = 1, \dots, T$. Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t})'$, $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t})'$, $\boldsymbol{\mu} = (\mu_1, \mu_2)'$, $\mathbf{h}_t = (h_{1,t}, h_{2,t})'$, $\Omega_t = \text{diag}(\exp(\mathbf{h}_t/2))$, and

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}, \Sigma_\epsilon = \begin{pmatrix} 1 & \rho_\epsilon \\ \rho_\epsilon & 1 \end{pmatrix}, \Sigma_\eta = \begin{pmatrix} \sigma_{\eta_1}^2 & \rho_\eta \sigma_{\eta_1} \sigma_{\eta_2} \\ \rho_\eta \sigma_{\eta_1} \sigma_{\eta_2} & \sigma_{\eta_2}^2 \end{pmatrix}.$$

MODEL 1 (Basic-MSV or **BMSV**):

$$\begin{aligned} \mathbf{y}_t &= \Omega_t \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, I), \\ \mathbf{h}_{t+1} &= \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2)), \end{aligned}$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$. This model is equivalent to stacking two basic univariate SV models together. Clearly, this specification does not allow for correlation across the returns or across the volatilities, nor Granger causality. However, it does allow for leptokurtic return distributions and volatility clustering.

MODEL 2 (Constant Conditional Correlation-MSV or **CCC-MSV**):

$$\begin{aligned}\mathbf{y}_t &= \Omega_t \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_\epsilon), \\ \mathbf{h}_{t+1} &= \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2)),\end{aligned}$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$. In this model, the return shocks are allowed to be correlated and hence the model is similar to the constant conditional correlation (CCC) ARCH model of Bollerslev (1990). As a result, the returns are cross-dependent.

MODEL 3 (MSV with Granger Causality or **GC-MSV**):

$$\begin{aligned}\mathbf{y}_t &= \Omega_t \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_\epsilon), \\ \mathbf{h}_{t+1} &= \boldsymbol{\mu} + \Phi(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2)),\end{aligned}$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$ and $\phi_{12} = 0$. Since ϕ_{21} can be different from zero, the volatility of the second asset is allowed to be Granger caused by the volatility of the first asset. Consequently, both the returns and volatilities are cross-dependent. However, the cross-dependence of volatilities are realized via Granger causality and volatility clustering jointly. Furthermore, when both ϕ_{12} and ϕ_{21} are nonzero, a bilateral Granger causality in volatility between the two assets is allowed. To the best of our knowledge, this specification is new to the SV literature.

MODEL 4 (Generalized CCC-MSV or **GCCC-MSV**):

$$\begin{aligned}\mathbf{y}_t &= \Omega_t \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_\epsilon), \\ \mathbf{h}_{t+1} &= \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_\eta),\end{aligned}$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$. This model was proposed and estimated via QML in Harvey, et al (1994). The same specification was estimated by Danielsson (1998) using SML. In this model, the return shocks are allowed to be correlated, so are the volatility shocks. Consequently, both returns and volatilities are cross-dependent. Obviously, both **GC-MSV** and **GCCC-MSV** can generate cross dependence in volatilities. Which specification is more appropriate is an interesting empirical question.

MODEL 5 (Dynamic Conditional Correlation-MSV or **DCC-MSV**):

$$\begin{aligned} \mathbf{y}_t &= \Omega_t \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t | \Omega_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_{\epsilon,t}), \\ \Sigma_{\epsilon,t} &= \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}, \\ \mathbf{h}_{t+1} &= \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2)), \\ q_{t+1} &= \psi_0 + \psi(q_t - \psi_0) + \sigma_\rho v_t, \quad v_t \stackrel{iid}{\sim} N(0, 1), \quad \rho_t = \frac{\exp(q_t) - 1}{\exp(q_t) + 1}. \end{aligned}$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$, $q_0 = \psi_0$. This model is new to the literature. In this model, not only volatilities but also correlation coefficients are time varying. Of course, ρ_t has to be bounded by -1 and 1 for Σ_ϵ to be a well-defined correlation matrix. This constraint is achieved by using the Fisher transformation, following the suggestion made in Tsay (2002) and Christodoulakis and Satchell (2002) in the MARCh framework. However, it is not easy to generalize the model into higher dimensional situations.

To allow for time-varying correlations in a N -dimensional setting with $N > 2$, one can follow Engle (2002) by constructing a sequence of matrices $\{Q_t\}$ according to

$$Q_{t+1} = S + B \circ (Q_t - S) + A \circ (v_t v_t' - S) = (u' - A - B) \circ S + B \circ Q_t + A \circ v_t v_t', \quad (1)$$

where $v_t \sim N(0, I)$, ι is a vector of ones, and \circ is the Hadamard product.² According to Ding and Engle (2001) and Engle (2002), as long as A , B , and $u' - A - B$ are positive semi-definite, Q_t will be positive semi-definite. As a result, we can obtain Q_t^{-1} and its Choleski decomposition $Q_t^{-1/2}$ (defined by $Q_t^{-1/2}(Q_t^{-1/2})' = Q_t^{-1}$). Finally a sequence of covariance matrices for $\boldsymbol{\epsilon}_t$ is constructed according to

$$\Sigma_{\epsilon,t} = \text{diag}(Q_t^{-1/2}) Q_t \text{diag}(Q_t^{-1/2}). \quad (2)$$

By construction, all the elements in $\Sigma_{\epsilon,t}$ are bounded between -1 and 1 , all the main diagonal elements in $\Sigma_{\epsilon,t}$ are ones, and $\Sigma_{\epsilon,t}$ is positive semi-definite. As a result, Σ_ϵ is a well-defined correlation matrix. This multivariate SV model has not appeared in the literature and is a natural analogue to the DCC-MARCh model of Engle (2002) and the VCC-MARCh model of Tse and Tsui (2002).

²The Hadamard product is defined by a matrix whose elements are obtained by element-by-element multiplication.

MODEL 6 (Heavy-tailed MSV or **t-MSV**):

$$\begin{aligned}\mathbf{y}_t &= \Omega_t \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} t(\mathbf{0}, \Sigma_\epsilon, \nu), \\ \mathbf{h}_{t+1} &= \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2)),\end{aligned}$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$. In this model, a heavy-tailed multivariate Student t distribution for the return shock is used and hence extra excess kurtosis is allowed. The Student t error distribution was first used in Harvey et al (1994) in the multivariate SV context.

To mitigate the computational problem inherent in estimating a large number of parameters in some of the above mentioned multivariate SV models on the one hand and to capture the common feature in asset returns and volatilities on the other, lower-dimensional factor multivariate SV models have been proposed and recently attracted some attention in the literature. Depending on how the factor enters the return equation, factor multivariate SV models can be split into two groups – additive and multiplicative factor multivariate SV models. Let $\mathbf{D} = (1, d)'$, $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{2,t})'$, and $f_t, u_t, h_t, \eta_t, \mu, \phi, \sigma_\eta, h_0, \nu$ be all scalars. The following three specifications belong to the factor multivariate SV family, the first two of which are of additive structure while the last is of multiplicative structure.

MODEL 7 (Additive Factor-MSV or **AFactor-MSV**):

$$\begin{aligned}\mathbf{y}_t &= \mathbf{D}f_t + \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\epsilon_1}^2, \sigma_{\epsilon_2}^2)) \\ f_t &= \exp(h_t/2)u_t, u_t \stackrel{iid}{\sim} N(0, 1), \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \eta_t \stackrel{iid}{\sim} N(0, 1),\end{aligned}$$

with $h_0 = 0$. This model was proposed by Jacquier et al (1995, 1999). The first component in the return equation has a smaller number of factors which capture the information relevant to the pricing of all assets while the second one is idiosyncratic noise which captures the asset specific information. Like the univariate SV model, the AFactor-MSV model allows for excess kurtosis and volatility clustering. Clearly, it also allows for cross dependence in both returns and volatilities. Note that in this model and MODEL 8 that will be introduced below, h_t represents the log-volatility of the common factor, f_t . The conditional correlation coefficient between y_{1t} and y_{2t} is given by:

$$\frac{d \exp(h_t)}{\sqrt{(\exp(h_t) + \sigma_{\epsilon_1}^2)(d^2 \exp(h_t) + \sigma_{\epsilon_2}^2)}} = \frac{d}{\sqrt{(1 + \sigma_{\epsilon_1}^2 \exp(-h_t))(d^2 + \sigma_{\epsilon_2}^2 \exp(-h_t))}}.$$

Unless $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2 = 0$, the correlation coefficients are time varying but the dynamics of the correlations depend on the dynamics of h_t . Moreover, correlation is an increasing function of h_t , implying that the higher the volatility of the common factor, the higher the correlation in returns.

MODEL 8 (Heavy-tailed Factor-MSV or **AFactor-t-MSV**):

$$\begin{aligned} \mathbf{y}_t &= \mathbf{D}f_t + \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} t(\mathbf{0}, \mathbf{diag}(\sigma_{\epsilon_1}^2, \sigma_{\epsilon_2}^2), \nu) \\ f_t &= \exp(h_t/2)u_t, u_t \stackrel{iid}{\sim} t(0, 1, \nu), \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \eta_t \stackrel{iid}{\sim} N(0, 1), \end{aligned}$$

with $h_0 = \mu$. In this model, a heavy-tailed Student t distribution for the return shock is used. The correlation structure here is the same as that in MODEL 7 (**AFactor-MSV**). Relative to MODEL 7, extra excess kurtosis is allowed here.

MODEL 9 (Multiplicative Factor-MSV or **MFactor-MSV**):

$$\begin{aligned} \mathbf{y}_t &= \exp(h_t/2)\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_\epsilon), \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \eta_t \stackrel{iid}{\sim} N(0, 1), \\ \Sigma_\epsilon &= \begin{pmatrix} 1 & \rho_\epsilon \sigma_{\epsilon_2} \\ \rho_\epsilon \sigma_{\epsilon_2} & \sigma_{\epsilon_2}^2 \end{pmatrix} \end{aligned}$$

with $h_0 = \mu$. This model, also known as the stochastic discount factor model, was considered in Quintana and West (1987). Compared with MODEL 1 (**BMSV**), this model has even fewer parameters. Obviously it retains all the properties inherent in the univariate SV model such as excess kurtosis and volatility clustering. Cross dependence in returns is induced by the dependence in $\boldsymbol{\epsilon}_t$ but the correlations are time invariant. Moreover, the correlation in log-volatilities is always one but time varying correlation in returns is not allowed.

Most of the models reviewed above are non-nested with each other. For example, MODEL 9 (**MFactor-MSV**) is not nested with MODEL 7 (**MFactor-MSV**) or MODEL 8 (**MFactor-t-MSV**). Neither MODEL 7 nor MODEL 8 are nested or nested within any other models, including MODEL 5 (**DCC-MSV**). However, MODEL 9 (**MFactor-MSV**) can be viewed as a special case of MODEL 2 (**CCC-MSV**), in which $\mu_1 = \mu_2$, $\phi_{11} = \phi_{22}$, $\sigma_{\eta_1} = \sigma_{\eta_2}$, $\eta_{1t} = \eta_{2t}$ and hence $h_{1,t} = h_{2,t}$.

3 Bayesian Estimation Using WinBUGS

The models in Section 2.2 are completed by the specification of a prior distribution for all unknown parameters $\mathbf{a} = (a_1, \dots, a_p)$. For instance, in MODEL 1 (**BMSV**), $p = 6$ and the vector \mathbf{a} of unknown parameters is $\mathbf{a} = (\mu_1, \mu_2, \phi_{11}, \phi_{22}, \sigma_{\eta_1}^2, \sigma_{\eta_2}^2)$. Bayesian inference is based on the joint posterior distribution of all unobserved quantities $\boldsymbol{\theta}$ in the model. The vector $\boldsymbol{\theta}$ comprises the unknown parameters and the vector of latent log-volatilities, i.e. $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{h}_1, \dots, \mathbf{h}_T)$.

In the sequel, let $p(\cdot)$ denote the generic probability density function of a random variable. Using independent priors for the parameters and successive conditioning on the sequence of latent states, the joint prior density of $\boldsymbol{\theta}$ in MODEL 1 is given by

$$p(\mathbf{a})p(\mathbf{h}_0) \prod_{t=1}^T p(\mathbf{h}_t|\mathbf{a}) = p(\mu_1)p(\mu_2)p(\phi_{11})p(\phi_{22})p(\sigma_{\eta_1}^2)p(\sigma_{\eta_2}^2)p(\mathbf{h}_0) \prod_{t=1}^T p(\mathbf{h}_t|\mathbf{a}).$$

After observing the data, this joint prior density is updated to the joint posterior density of all unknown quantities, $p(\boldsymbol{\theta}|\mathbf{y})$ (where $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)$) via Bayes' theorem by multiplying prior $p(\boldsymbol{\theta})$ and likelihood $p(\mathbf{y}|\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}) \propto p(\mathbf{a})p(\mathbf{h}_0) \prod_{t=1}^T p(\mathbf{h}_t|\mathbf{a}) \prod_{t=1}^T p(\mathbf{y}_t|\mathbf{h}_t). \quad (3)$$

To calculate the marginal posterior distribution of the parameters of interest $p(\mathbf{a}|\mathbf{y})$ requires $(p + 2T)$ -dimensional integration to find the normalization constant $p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta}$ followed by $2T$ -dimensional integration over all latent volatilities, as

$$p(\mathbf{a}|\mathbf{y}) = \int_{\mathbf{h}_1} \dots \int_{\mathbf{h}_T} p(\mathbf{a}, \mathbf{h}_1, \dots, \mathbf{h}_T) d\mathbf{h}_T \dots d\mathbf{h}_1. \quad (4)$$

This is neither analytically nor numerically tractable in general. Simulation-based integration techniques have proven to be the most effective methods to deal with this integration problem. Say, a sample of size M , $(\mathbf{a}^{(1)}, \mathbf{h}^{(1)}, \dots, \mathbf{a}^{(M)}, \mathbf{h}^{(M)})$ can be obtained from $p(\boldsymbol{\theta}|\mathbf{y})$. By simply ignoring the sampled latent volatilities, the subvector $(\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots, \mathbf{a}^{(M)})$ constitutes a sample from the marginal posterior distribution (4) of \mathbf{a} and kernel density estimates of each component can be used to estimate the marginal posterior density of each parameter. The usual summary statistics can be calculated to estimate population quantities of interest, eg. the sample mean $\frac{1}{M} \sum_{m=1}^M \mathbf{a}^{(m)}$ is a consistent estimate of the posterior mean $E[\mathbf{a}|\mathbf{y}]$.

Unfortunately, direct independent sampling from a high-dimensional distribution such as in (3) is usually not possible (see Liesenfeld and Richard (2003) and Durham (2004) for counterexamples, however). MCMC techniques overcome this problem by constructing a Markov chain with stationary distribution equal to the target density $p(\boldsymbol{\theta}|\mathbf{y})$ and simulate from this Markov chain. Provided the Markov chain is run long enough to have reached equilibrium, the samples in each iteration can be regarded as (dependent) samples from $p(\boldsymbol{\theta}|\mathbf{y})$. By the ergodic theorem, sample averages are still consistent estimates of the population quantities.

Care needs to be taken in determining the number of iterations to achieve convergence to the stationary distribution. Various convergence diagnostics have been developed and implemented in the CODA package, a collections of SPLUS or R routines. CODA may also be downloaded from the BUGS website.

Here, we advocate the software package WinBUGS for posterior computation in multivariate SV models. WinBUGS provides an easy to use and efficient implementation of the Gibbs sampler, a specific MCMC technique that constructs a Markov chain by sampling from all univariate full conditional distributions in a cyclic way. WinBUGS has been successfully applied for a variety of statistical models such as random effects, generalized linear, proportional hazards, latent variable, and frailty models. In particular, state-space models (Harvey, 1990), either linear or nonlinear, either Gaussian or non-Gaussian, either observed state or latent state, either univariate or multivariate, are amenable to a Bayesian analysis via WinBUGS.

Meyer and Yu (2000) described the use of BUGS for Bayesian posterior computation in univariate SV models and emphasized the ease with which BUGS can be used for the exploratory phase of model building as any modifications of a model including changes of priors and sampling error distributions are readily realized with only minor changes of the code. BUGS automates the calculation of the full conditional posterior distributions that are needed for Gibbs sampling using a model representation by directed acyclic graphs. It contains an expert system for choosing an effective sampling method for each full conditional. The reader is referred to Meyer and YU (2000) for a comprehensive introduction on using BUGS for fitting SV models. WinBUGS is a new interactive version of the BUGS program that allows models to be described using a slightly amended version of the BUGS language. The BUGS website contains a short Flash illustration on the basic steps of running WinBUGS. WinBUGS also allows models to be fitted using Doodles (graphical representations of models

by directed acyclic graphs) which can, if desired, be automatically translated into a text-based description. In Meyer and Yu (2000), the Doodle corresponding to a certain BUGS implementation of a univariate SV model is explained in detail.

BUGS can be slow due to the single move Gibbs sampler. However, the new interactive WinBUGS version contains much improved algorithms to sample from the full conditional posterior distributions. WinBUGS contains a small expert system for choosing the best sampling method. For discrete full conditional distributions, WinBUGS uses the inversion method to simulate values. For continuous distributions, it tests first for conjugacy. If it detects conjugacy, then it will use optimized standard simulation algorithms. For logconcave full conditionals, it uses the derivative-free adaptive rejection technique of Gilks (1992). For non-logconcave full conditionals with

restricted range, WinBUGS uses the slice sampling technique of Neal (1997) with an adaptive phase of 500 iterations and a current point Metropolis algorithm for unrestricted non-logconcave full conditionals. The current point Metropolis algorithm is based on a symmetric normal proposal distribution whose standard deviation is tuned over the first 4000 iterations in order to get an acceptance rate between 20% and 40%. Furthermore, it contains the option of using ordered overrelaxation (Neal, 1998) which generates multiple samples at each iteration and then selects one which is negatively correlated with the current value. The time per iteration will be increased but the within-chain correlations should be reduced and hence fewer iterations may be necessary. It also contains a blocking option for multivariate updating, but only for generalized linear model components at this stage. The use of these improved sampling techniques coupled with an increase in computational speed due to advances in computer hardware has made it possible to fit multivariate SV models in WinBUGS.

4 DIC

The Akaike information criterion (AIC; Akaike, 1973) is a popular method for comparing alternative and possibly non-nested models. It trades off a measure of model adequacy, measured by the log-likelihood, against a measure of complexity, measured by the number of free parameters. Obviously the calculation of AIC requires the specification of the number of free parameters. For a non-hierarchical Bayesian model with parameter θ , obtaining the number of free parameters is straightforward. However, for a complex hierarchical model the

specification of the dimensionality of the parameter space is rather arbitrary. This is typically the case for SV models. The reason is that when MCMC is used to estimate SV models, as mentioned above, the parameter space is augmented. For example, in the basic MSV model, we include the $2T$ latent volatilities into the parameter space with T being the sample size. As these volatilities are dependent, they cannot be counted as $2T$ additional free parameters. Consequently, AIC is not applicable for comparing SV models (Berg et al. 2004).

The deviance information criterion (DIC) of Spiegelhalter, Best, Carlin and van der Linde (2002) is intended as a generalization of AIC to complex hierarchical models. Like AIC, DIC consists of two components, a term that measures goodness-of-fit and a penalty term for increasing model complexity:

$$\text{DIC} = \bar{D} + p_D. \quad (5)$$

The first term, \bar{D} , is defined as the posterior expectation of the deviance:

$$\bar{D} = E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] = E_{\boldsymbol{\theta}|\mathbf{y}}[-2 \ln f(\mathbf{y}|\boldsymbol{\theta})]. \quad (6)$$

The ‘better’ the model fits the data, the smaller is the value of \bar{D} .

The second component, p_D , measures the complexity of the model by the *effective number of parameters* and is defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean $\bar{\boldsymbol{\theta}}$ of the parameters:

$$p_D = \bar{D} - D(\bar{\boldsymbol{\theta}}) = E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] - D(E_{\boldsymbol{\theta}|\mathbf{y}}[\boldsymbol{\theta}]) = E_{\boldsymbol{\theta}|\mathbf{y}}[-2 \ln f(\mathbf{y} | \boldsymbol{\theta})] + 2 \ln f(\mathbf{y}|\bar{\boldsymbol{\theta}}). \quad (7)$$

Equation (7) shows that p_D can be regarded as the expected excess of the true over the estimated residual information in data \mathbf{y} conditional on $\boldsymbol{\theta}$. Hence, we can interpret p_D as the expected reduction in uncertainty due to estimation.

Rearranging Equation (7) gives $\bar{D} = D(\bar{\boldsymbol{\theta}}) + p_D$. As a result, DIC can be re-represented as

$$\text{DIC} = D(\bar{\boldsymbol{\theta}}) + 2p_D \quad (8)$$

which can be interpreted as a classical ‘plug-in’ measure of fit plus a measure of complexity.

Spiegelhalter et al. (2002) give an asymptotic justification of DIC in the case where the number of observations T grows with respect to the number of parameters p and where the prior is non-hierarchical and completely specified (i.e. without hyperparameters). In this situation, $\text{AIC} = D(\hat{\boldsymbol{\theta}}) + 2p$, where $\hat{\boldsymbol{\theta}}$ denotes the maximum likelihood (ML) estimate. This

is the same formula as (8) but with the posterior mean $\bar{\theta}$ substituted by $\hat{\theta}$. Similar to AIC, therefore, the model with the smallest DIC is estimated to be the one that would best predict a replicate dataset of the same structure as that observed. This focus of DIC, however, is different from the posterior-odd-based approaches, where how well the prior has predicted the observed data is addressed. Berg et al (2004) examined the performance of DIC relative to two posterior odd approaches – one is based on the harmonic mean estimate of marginal likelihood (Newton and Raftery, 1994) and the other is Chib’s estimate of marginal likelihood (Chib, 1995) – in the context of univariate SV models. They found reasonably consistent performance of these three model comparison methods.

From the definition of DIC it can be seen that DIC is almost trivial to compute and particularly suited to compare Bayesian models when posterior distributions have been obtained using MCMC simulation. Indeed, DIC is automatically computed by WinBUGS1.4. This is in contrast to Chib’s marginal likelihood method where computational cost is more demanding as the likelihood needs to be evaluated using other independent procedures such as the particle filter (Kim et al, 1998). Although Chib, Nardari and Shephard (2002) successfully used Chib’s method to compare several specifications in a family of factor multivariate SV with the additive structure, we believe the computational tractability of DIC would make it feasible to compare a much larger class of specifications.

It should be pointed out that because WinBUGS calculates DIC at the posterior mean, it requires the posterior mean to be a good estimate of the stochastic parameters. Therefore, it is important to check skewness and modality of the posterior distribution when using DIC.

5 Empirical Illustration

5.1 Data

In this section we fit the models introduced in Section 2.3 to actual financial time series data. The data used are 519 weekly mean corrected log-returns of Australian Dollar and New Zealand Dollar, both against the US dollar, from January 1994 to December 2003. Both series are plotted in Fig. 1 where cross-dependence both in returns and volatilities seems strong.³

³The data were obtained from the Sauder School of Business at the University of British Columbia via the URL <http://fx.sauder.ubc.ca/data.html>.

5.2 Prior distributions

For the first six models, there are three sets of parameters: parameters in the mean equation (ρ_ϵ, ν) , in the variance equation $(\phi_{11}, \phi_{22}, \phi_{21}, \mu_1, \mu_2, \rho_\eta, \sigma_{\eta_1}, \sigma_{\eta_2})$, and in the correlation equation $(\psi_0, \psi_1, \sigma_\rho)$. We assume that the parameters are mutually independent. The prior distributions are specified as follows:

- $\rho_\epsilon \sim U(-1, 1)$;
- $\nu^* \sim \chi_{(4)}^2$, where $\nu^* = \nu/2$;
- $\mu_1 \sim N(0, 25)$;
- $\mu_2 \sim N(0, 25)$;
- $\phi_{11}^* \sim \text{Beta}(20, 1.5)$, where $\phi_{11}^* = (\phi_{11} + 1)/2$;
- $\phi_{22}^* \sim \text{Beta}(20, 1.5)$, where $\phi_{22}^* = (\phi_{22} + 1)/2$;
- $\phi_{21} \sim N(0, 10)$;
- $\rho_\eta \sim U(-1, 1)$;
- $\sigma_{\eta_1}^2 \sim \text{Inverse-Gamma}(2.5, 0.025)$;
- $\sigma_{\eta_2}^2 \sim \text{Inverse-Gamma}(2.5, 0.025)$;
- $\psi_1^* \sim \text{Beta}(20, 1.5)$, where $\psi_1^* = (\psi_1 + 1)/2$;
- $\psi_0 \sim N(0.7, 10)$;
- $\sigma_\rho^2 \sim \text{Inverse-Gamma}(2.5, 0.025)$.

For the last three models, there are two sets of parameters: parameters in the mean equation $(\rho_\epsilon, d, \nu, \sigma_{\epsilon_1}, \sigma_{\epsilon_2})$, and in the factor equation (ϕ, μ, σ_η) . We assume that the parameters are mutually independent. The prior distributions are specified as follows:

- $\rho_\epsilon \sim U(-1, 1)$;
- $\nu^* \sim \chi_{(4)}^2$, where $\nu^* = \nu/2$;

- $d \sim N(1, 9)$;
- $\sigma_{\epsilon_1}^2 \sim \text{Gamma}(0.3, 0.3)$;
- $\sigma_{\epsilon_2}^2 \sim \text{Gamma}(0.3, 0.3)$;
- $\mu \sim N(0, 25)$;
- $\phi^* \sim \text{Beta}(20, 1.5)$, where $\phi^* = (\phi + 1)/2$;
- $\sigma_\eta^2 \sim \text{Inverse-Gamma}(2.5, 0.025)$.

We report means and standard errors of these prior distributions for the first six models in Table 1 and those for the last three models in Table 2.

5.3 Results

We report means, standard errors, 95% credible intervals of the posterior distributions for the first six models in Table 3 and those for the last three models in Table 4, as well as the computing time to generate 100 iterations for each of the nine models. The computing time is the central processing unit (CPU) time on a HP XW6000 workstation running WinBUGS1.4. For all models, after a burn-in period of 10,000 iterations and a follow-up period of 100,000, we stored every 20th iteration.

The first thing which can be seen from Tables 3-4 is that all nine models can be quickly estimated. The CPU time required for 100 iterations ranges from 2.3 seconds to 9.6 seconds. Moreover, estimating different multivariate SV models in WinBUGS require little effort in coding and often only no more than a few lines of code have to be changed. Second, the estimated means and standard deviations for the parameters appear quite reasonable and in accordance with estimates documented in the literature. For instance, in MODEL 1 (**BMSV**), both volatility processes are estimated to be highly persistent. In MODEL 4 (**GCCC-MSV**) posterior means of both correlation (ρ_ϵ and ρ_η) are high, as already observed in Harvey et al (1994). In MODEL 6 (**t-MSV**) the posterior mean of ν is 23.22, suggesting that a heavy-tailed distribution for errors is not needed. In all three factor models, the factor process is estimated to be highly persistent. In MODEL 7 (**AFactor-MSV**), the factor loading is estimated to be 1.233.

Some interesting empirical results can be found from the two new specifications, MODEL 3 (**GC-MSV**) and MODEL 5 (**DCC-MSV**). In MODEL 3, the posterior mean of ϕ_{12} is 0.4865 with the lower limit of the 95% posterior credibility interval being greater than zero. It suggests that the volatility in Australian dollar Granger causes the volatility in New Zealand dollar, consistent with our expectation. As a result of allowing for Granger causality, the posterior mean of the volatility persistence for New Zealand dollar is reduced from 0.99 to 0.7074. In MODEL 5 (**DCC-MSV**), the correlation process is reasonably highly persistent with a posterior mean of ψ being 0.9814. The posterior mean of the long run mean of the time varying correlation is 0.7195, consistent with what is found in MODEL 4 (**GCCC-MSV**). All these posterior quantities point towards the importance of time varying correlation.

In Table 5 we report DIC together with \bar{D} and p_D for each of the nine models as well as their associated rankings. The most adequate model to describe the bivariate data according to DIC is MODEL 7 (**AFactor-MSV**), followed closely by MODEL 5 (**DCC-MSV**). Figures 2 and 3 show the trace plots and density functions of the parameters d , μ , ϕ , σ_η , σ_{ϵ_1} , and σ_{ϵ_2} in MODEL 7. The models which have the lowest posterior means of the deviance are MODEL 5 and MODEL 8 (**AFactor-t-MSV**). The models which have the smallest effective numbers of parameters are MODEL 7 and MODEL 4 (**GCCC-MSV**). Although the posterior mean of the deviance for MODEL 7 is higher than those of MODEL 8 and MODEL 5, the effective number of parameters is much lower. The effective number of parameters is around 20 for MODEL 7 which is less than one half of those for MODEL 8 and MODEL 5. As both MODEL 7 and MODEL 5 allow for time varying correlations, the message taken from this model comparison exercise is that correlations do indeed vary over time.

To understand the implications of the better specifications, we obtain smoothed estimates of volatilities and correlations from MODEL 7 (**AFactor-MSV**) and MODEL 5 (**DCC-MSV**). In WinBUGS, once the latent processes are sampled and stored it is trivial to obtain the smoothed estimates of them. We plot the estimates of the two volatilities and the correlations from

MODEL 5 in Fig. 4 and the volatilities of the factor and the correlations from MODEL 7 in Fig. 5. Fig. 4 reveals that both Australian dollar and New Zealand dollar experienced a rapid volatility increase over the period from 1995 to 1998. The smoothed estimate of correlations shown in Fig. 4 is interesting. The correlation quickly decreases from 0.75 to 0.45 from the

beginning of the sample and reaches the lowest level in 1995. After that, it steadily increases to 0.8 and stays around that level for the rest of sample period. The correlation reaches the peak in 2002, which corresponds to the period of prolonged depreciation of the two currencies against the US dollar. Fig. 5 tells the same story about the volatilities – the volatilities of the common factor have experienced a rapid volatility increase over the period from 1995 to 1998. However, the implication on the correlations is somewhat different. Compared with Fig. 4, the correlation in Fig. 5 shows more dramatic evidence of non-stationarity in correlations. That is, it seems there is a structural change in the correlation process. The breakdown of the correlation appears to take place at the end of 1998.

6 Conclusion

In this paper we propose to estimate and compare multivariate SV models using Bayesian MCMC techniques via WinBUGS. MCMC is a powerful method and has a number of advantages over alternative methods. Unfortunately, writing the first MCMC program for estimating multivariate SV models is not easy and comparing alternative multivariate SV specifications is computationally costly. WinBUGS imposes a short but sharp learning curve. In the bivariate setting we show that its implementation is easy and computationally reasonably fast. However, since WinBUGS offers single move Gibbs sampling algorithm, one would expect that mixing is generally slow.

We illustrated the implementation in WinBUGS by exploring and comparing nine bivariate models, including Granger causality in volatilities, time varying correlations, heavy-tailed error distributions, additive factor structure, and multiplicative factor structure, two of which are new to the SV literature. Our empirical results based on weekly Australian/US dollar and New Zealand/US dollar exchange rates indicate that the models which allow for time varying coefficients generally fit the data better.

REFERENCES

- Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics* 18, 338–357.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings 2nd International Symposium Information Theory*, eds. B.N. Petrov and

- F. Csaki, Budapest: Akademiai Kiado, pp. 267–281.
- Andersen, T., H. Chung, and B. Sorensen (1999). Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Econometrics* 91, 61–87.
- Asai, M., M. McAleer and J. Yu (2004). Multivariate stochastic volatility models: A survey. Work in progress.
- Andersen, T. and B. Sorensen (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics* 14, 329–352.
- Berg, A., R. Meyer, and J. Yu (2004). Deviance information criterion for comparing stochastic volatility models. *Journal of Business and Economic Statistics* 22, 107–120.
- 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.
- Bollerslev, T., R. Engle and J.M. Wooldridge (1998). A capital asset pricing model with time varying covariances. *Journal of Political Economy* 96, 116–131.
- Braun, P., D. Nelson and A. Sunier (1995). Good news, bad news, volatility and betas. *Journal of Finance* 50, 1575–1603.
- Chan, D., R. Kohn and C. Kirby (2003). Multivariate stochastic volatility with leverage. Unpublished manuscript.
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *The Journal of the American Statistical Association* 90, 1313–1321.
- Chib, S., Nardari, F., and Shephard, N. (2002). Analysis of high dimensional multivariate stochastic volatility models. Unpublished manuscript.
- Christodoulakis, G., and 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.
- 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 comparison with VGARCH models. *Journal of Empirical Finance* 5, 155–173.

- Ding, Z. and Engle, R. (2001). Large scale conditional covariance modelling, estimation and testing. *Academia Economic Papers*. 29, 157-184.
- Durham, G.S. (2004). Monte Carlo Methods for Estimating, Smoothing, and Filtering One and Two-Factor Stochastic Volatility Models. *Journal of Econometrics* forthcoming.
- Engle, R. (2002). Dynamic conditional correlation - A simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*. 17, 239-250.
- Engle, R. and K. Kroner (1995). Multivariate simultaneous GARCH. *Econometric Theory* 11, 122-150.
- Engle, R., V. Ng, and M. Rothschild (1990). Asset pricing with a factor ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics* 45, 213-237.
- Fridman, M. and L. Harris (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business and Economic Statistics* 16, 284-291.
- Gilks, W. (1992). Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics 4* (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds.) Oxford University Press, UK, 642-665.
- Harvey, A.C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, New York.
- Harvey, A.C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies* 61, 247-264.
- Jacquier, E., N.G. Polson, and P.E. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12, 371-389.
- Jacquier, E., N.G. Polson, and P.E. Rossi (1995). Priors and models of stochastic volatility models. Unpublished manuscript, University of Chicago.
- Jacquier, E., N.G. Polson, and P.E. Rossi (1999). Stochastic volatility: Univariate and multivariate extensions. Unpublished manuscript, University of Chicago.
- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361-393.
- Knight, J.L., S.S. Satchell and J. Yu (2002). Estimation of the stochastic volatility model by the empirical characteristic function method. *Australian and New Zealand Journal of Statistics* 44, 319-335.

- Lancaster, T. (2004). *An Introduction to Modern Bayesian Econometrics*. Blackwell: Oxford.
- Liesenfeld, R., and Richard, J. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance* 10, 505-531.
- Meyer, R., D.A. Fournier, and A. Berg (2003). Stochastic volatility: Bayesian computation using automatic differentiation and the extended Kalman filter. *Econometrics Journal* 6, 408-420.
- Meyer, R., and J. Yu (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal* 3, 198-215.
- Melino, A. and S.M. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239-265.
- Neal, R. (1997). Markov Chain Monte Carlo methods based on "slicing" the density function. *Technical Report 9722* Department of Statistics, University of Toronto, Canada.
- Neal, R. (1998). Suppressing random walks in Markov Chain Monte Carlo methods using ordered over-relaxation. In *Learning in Graphical Models*, (M.I. Jordan, ed), Kluwer Academic Publishers, Dordrecht, 205-230.
- Newton, M. and A.E. Raftery (1994). Approximate Bayesian inferences by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B*, 56, 3-48 (with discussion).
- Pitt, M. and N. Shephard (1999). Time varying covariances: A factor stochastic volatility approach. In *Bayesian Statistics 6*, eds. J.M. Bernardo, J.O. Berger, A.P. David and A.F.M. Smith. Oxford, UK: Oxford University Press, pp 547-570
- Quintana, J.M. and M. West (1987). An analysis of international exchange rates using multivariate DLMS. *The Statistician* 36, 275-281.
- Richard, J.F. and W. Zhang (2004). Efficient high-dimensional importance sampling. Working Paper, University of Pittsburgh.
- Sandmann, G. and S.J. Koopman (1998). Maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* 63, 289-306.
- Selçuk, F. (2004). Free float and stochastic volatility: The experience of a small open economy. *Physica A* 342, 693-700.

- Shephard, N. (2004). *Stochastic Volatility: Selected Readings*. Oxford University Press: Oxford
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and A. van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 64, 583-639.
- Spiegelhalter, D.J., A. Thomas, N.G. Best, and W.R. Gilks (2003). *WinBUGS User Manual (Version 1.4)*. MRC Biostatistics Unit, Cambridge, UK.
- Spiegelhalter, D.J., A. Thomas, N.G. Best, and W.R. Gilks (1996). *BUGS 0.5, Bayesian inference using Gibbs sampling. Manual (version ii)*. MRC Biostatistics Unit, Cambridge, UK.
- Tsay, R.S. (2002) *Analysis of Financial Time Series*. Wiley: New York.
- Tse, Y. and A. Tusi (2002) A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* 17, 351-362.
- Yu, J. (2002) Forecasting volatility in the New Zealand stock market. *Applied Financial Economics* 12, 193–202.
- Yu, J. (2004) On leverage in a stochastic volatility model. *Journal of Econometrics* forthcoming.

Table 1: Means and Standard Deviations of Prior Distributions for Parameters in the First Six Models

	ρ_ϵ	ν	ϕ_{11}	ϕ_{22}	ϕ_{21}	μ_1	μ_2	ρ_η	σ_{η_1}	σ_{η_2}	ψ_0	ψ	σ_ρ
Prior Mean	0	8	.86	.86	0	0	0	0	.12	.12	.7	.86	.12
Prior SD	.86	4	.11	.11	.33	5	5	.86	.05	.05	3.3	.11	.05

Table 2: Means and Standard Deviations of Prior Distributions for Parameters in the Last Three Models

	ρ_ϵ	d	ν	σ_{ϵ_1}	σ_{ϵ_2}	ϕ	μ	σ_η
Prior Mean	0	1	8	1	1	.86	0	.12
Prior SD	.86	3	4	1.83	1.83	.11	5	.05

Table 3: Posterior Quantities for Parameters in the First Six Models

		BMSV	CCC-MSV	GC-MSV	GCCC-MSV	DCC-MSV	t-MSV
μ_1	Mean	.0050	.0930	-.1629	-1.094	.0123	.0430
	SD	.2294	.1263	.2417	2.262	.2001	.1108
	95% CI	-.50,.436	-.188,.319	-.689,.211	-6.50,2.45	-.419,.395	-0.20,.242
μ_2	Mean	-.710	-.6562	-.4617	-2.454	-.7595	-.5929
	SD	.4064	.3695	.3725	1.275	.3563	.3873
	95% CI	-1.52,.026	-1.36,.044	-1.25,.112	-5.45,-.37	-1.45,-.101	-1.42,.040
ϕ_{11}	Mean	.9770	.9428	.9788	.9764	.9766	.9252
	SD	.0140	.0418	.0167	.0134	.0135	.0591
	95% CI	.942,.996	.838,.989	.936,.998	.945,.996	.944,.996	.78,.986
ϕ_{22}	Mean	.9920	.9905	.7074	.8622	.9934	.9874
	SD	.0061	.0089	.1272	.0937	.0052	.0117
	95% CI	.976,.999	.966,.999	.423,.920	.625,.982	.980,.999	.957,.999
σ_{η_1}	Mean	.1107	.0967	.0774	.0807	.0973	.0965
	SD	.0265	.0220	.0167	.0148	.0198	.0242
	95% CI	.071,.174	.061,.147	.052,.116	.057,.115	.066,.146	.061,.153
σ_{η_2}	Mean	.1044	.0884	.1262	.1431	.0894	.0881
	SD	.0228	.0203	.0531	.0289	.0187	.0205
	95% CI	.069,.157	.059,.135	.062,.269	.098,.209	.061,.134	.058,.138
ρ_ϵ	Mean		.7439	.7398	.7325		.7471
	SD		.0208	.0228	.0202		.0205
	95% CI		.701,.783	.692,.781	.691,.770		.705,.785
ϕ_{12}	Mean			.4865			
	SD			.2296			
	95% CI			.115,1.00			
ρ_η	Mean				.8363		
	SD				.1240		
	95% CI				.540,.988		
σ_ρ	Mean					.1124	
	SD					.031	
	95% CI					.065,.189	
ψ_0	Mean					1.945	
	SD					.2808	
	95% CI					1.387,2.519	
ψ	Mean					.9814	
	SD					.0122	
	95% CI					.950,.997	
ν	Mean						23.22
	SD						7.174
	95% CI						12.59,40.2
Time (seconds)		3.3	2.7	3.0	9.6	3.5	4.5

Table 4: Posterior Quantities for Parameters in the Last Three Models

		AFactor-MSV	AFactor-t-MSV	MFactor-MSV
μ	Mean	-1.299	-1.244	1.816
	SD	0.4024	0.4405	0.3041
	95% CI	-2.078,-0.5338	-2.145,-0.4314	1.266,2.469
ϕ	Mean	0.9942	0.9938	0.9804
	SD	0.0051	0.0052	0.014
	95% CI	0.981,0.9998	0.9803,0.9997	0.9463,0.9987
σ_η	Mean	0.1055	0.1075	0.09777
	SD	0.02417	0.02688	0.0225
	95% CI	0.070,0.165	0.0682,0.1676	0.0622,0.1523
d	Mean	1.233	1.201	
	SD	0.0678	0.0693	
	95% CI	1.1,1.355	1.064,1.328	
σ_{ϵ_1}	Mean	0.6799	0.65	
	SD	0.0329	0.0358	
	95% CI	0.611,0.739	0.5754,0.7138	
σ_{ϵ_2}	Mean	0.2087	0.2226	0.9646
	SD	0.1189	0.1206	0.0318
	95% CI	0.0029,0.4048	0.0033,0.4149	0.9035,1.029
ρ_ϵ	Mean			0.7302
	SD			0.0246
	95% CI			0.6801,0.7772
ν	Mean		26.44	
	SD		8.092	
	95% CI		14.40, 46	
Time (seconds)		5.5	6.0	2.3

Table 5: DIC for All Models

Model	DIC		\bar{D}	p_D
	Value	Ranking		
BMSV	2997.270	9	2958.960	38.320
CCC-MSV	2622.090	6	2581.960	40.125
GC-MSV	2616.290	5	2578.890	37.393
GCCC-MSV	2608.060	4	2581.110	26.941
DCC-MSV	2579.970	2	2524.450	55.523
t-MSV	2624.880	7	2546.940	77.938
AFactor-MSV	2577.750	1	2557.270	20.481
AFactor-t-MSV	2583.530	3	2525.900	57.631
MFactor-MSV	2626.660	8	2599.340	27.326

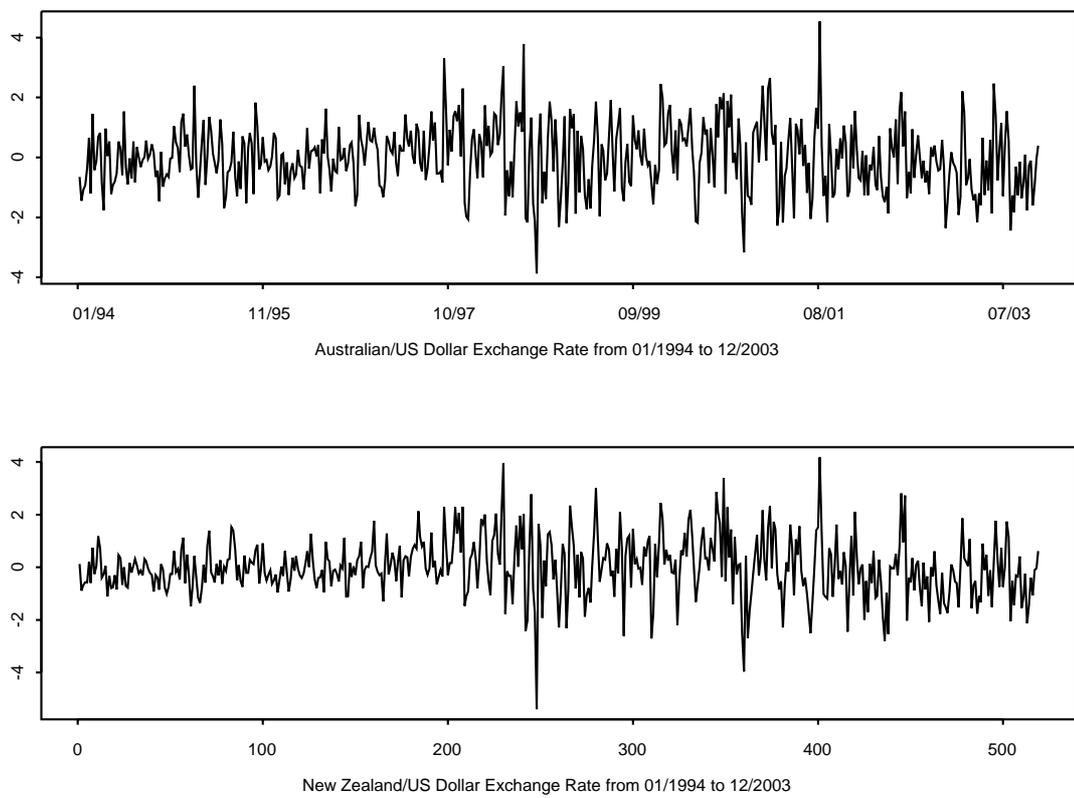


Figure 1: Time series plots for Australian dollar and New Zealand dollar/US dollar exchange rate returns.

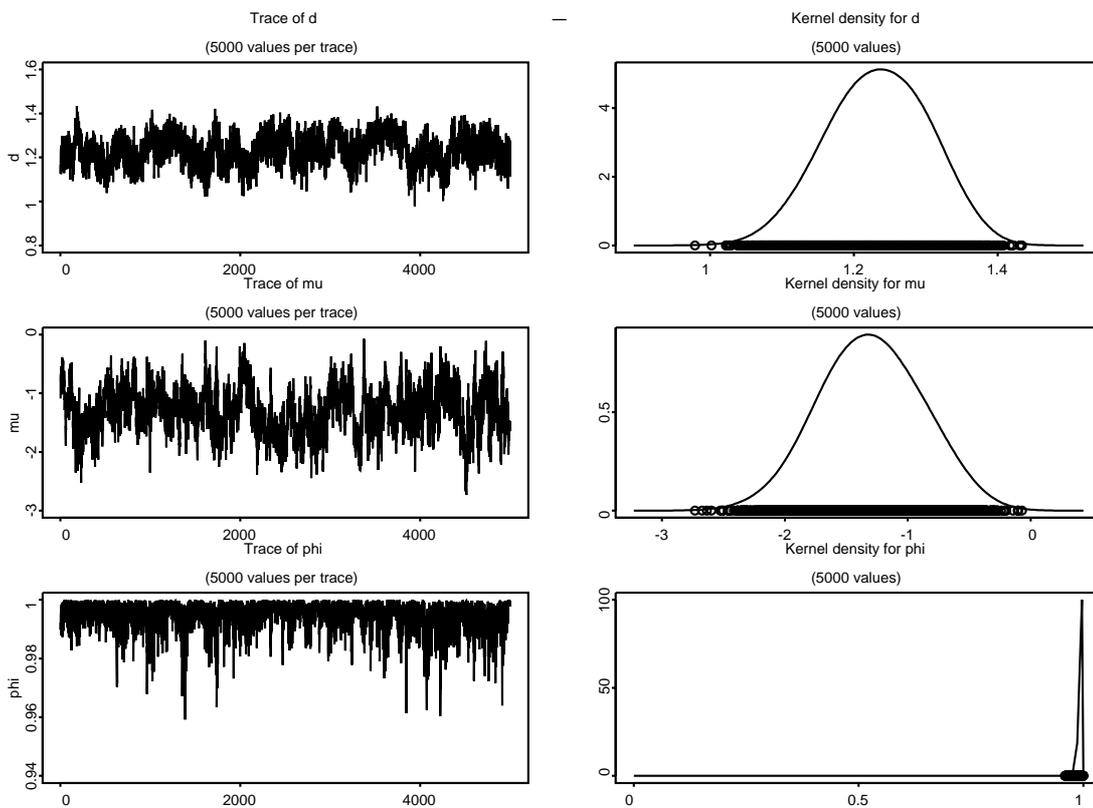


Figure 2: Trace plots and density estimates of the marginal distribution of d , μ , and ϕ in MODEL 7 (AFactor-MSV).

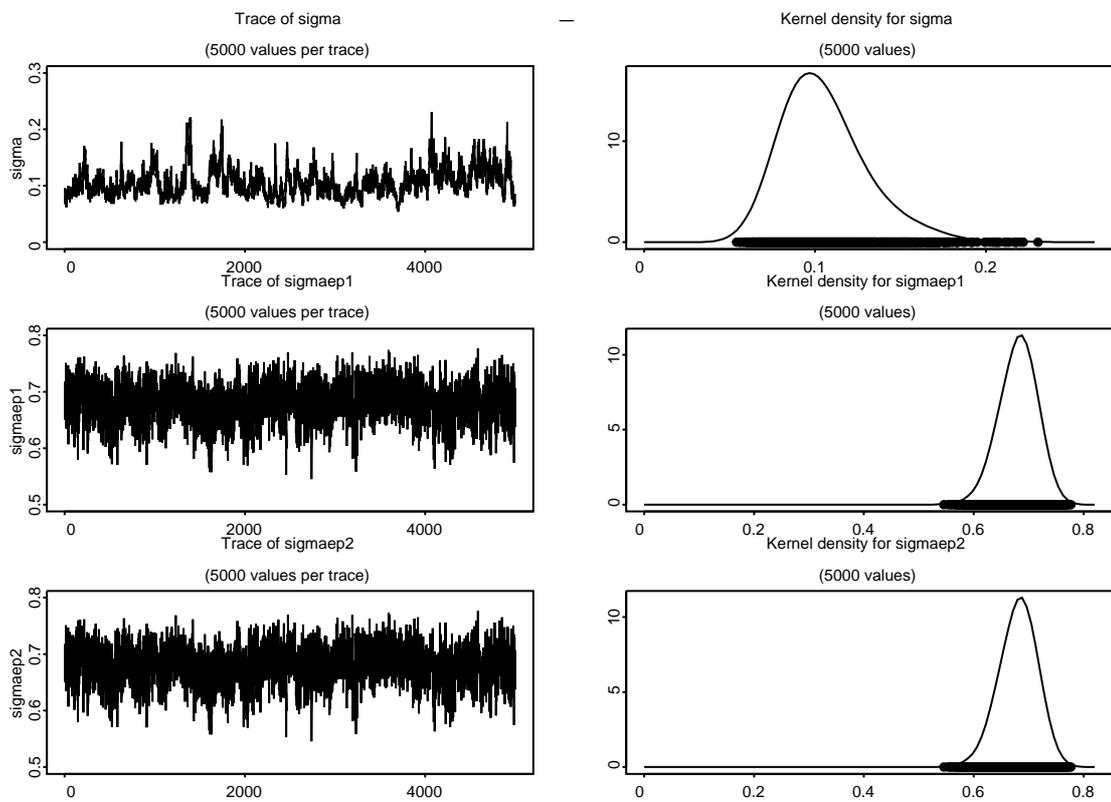


Figure 3: Trace plots and density estimates of the marginal distribution of σ_η , σ_{ϵ_1} , and σ_{ϵ_2} in MODEL 7 (AFactor-MSV).

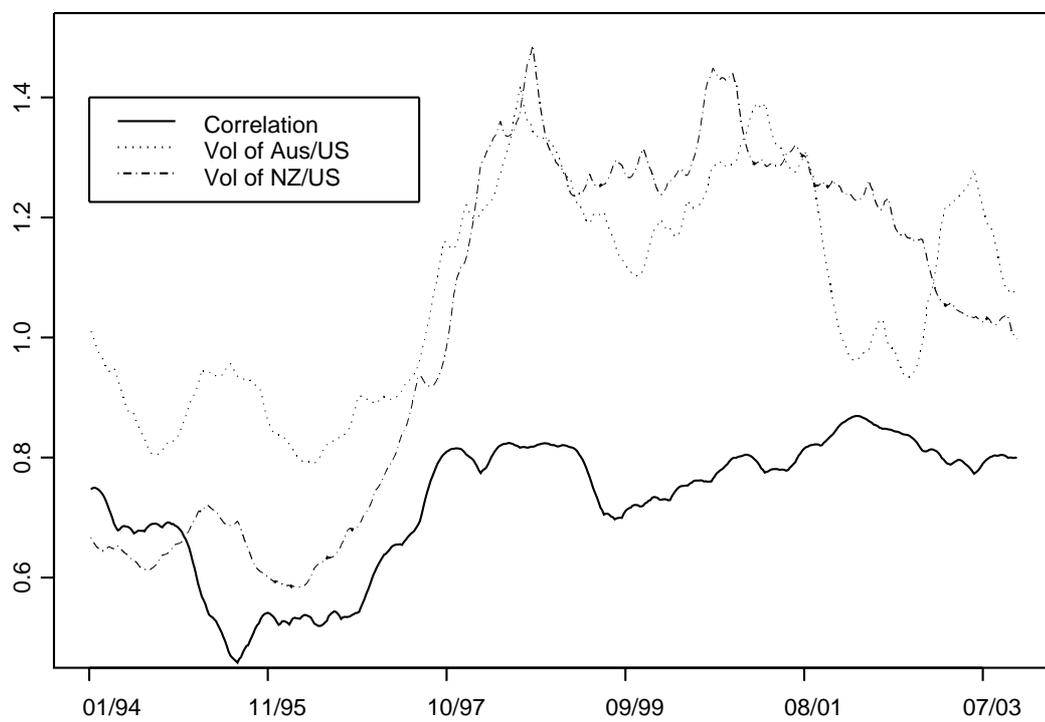


Figure 4: Smoothed estimates of volatilities of exchanges rates and time varying correlations from MODEL 5 (DCC-MSV).

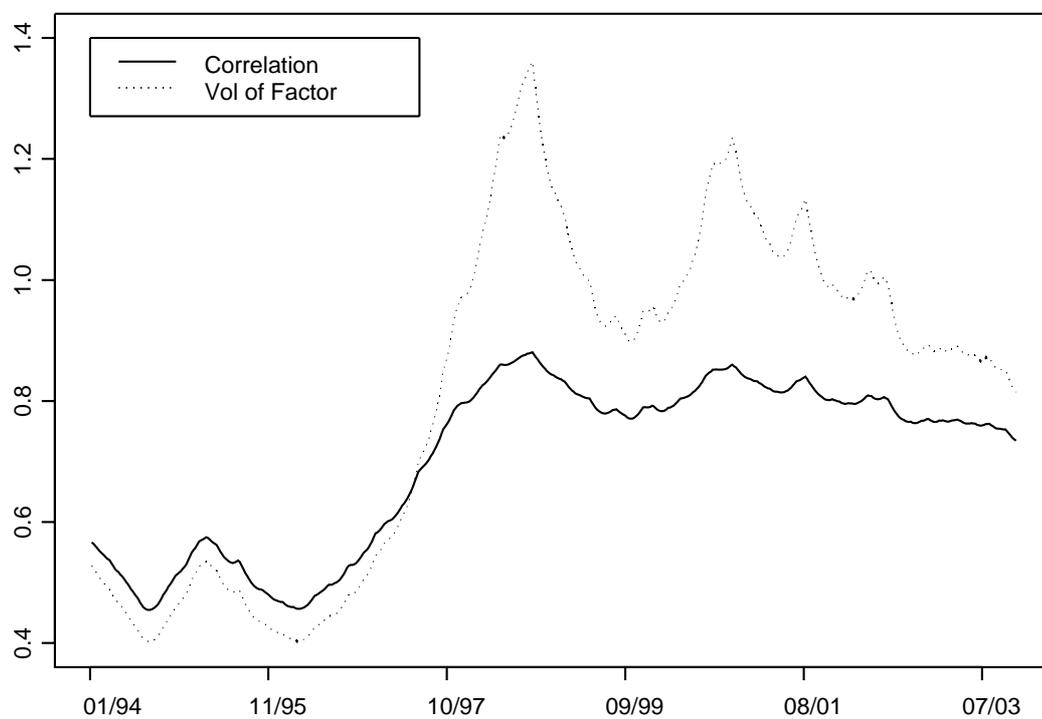


Figure 5: Smoothed estimates of volatilities of the factor and time varying correlations from MODEL 7 (Factor-MSV).