A full-factor multivariate GARCH model

I. D. VRONTOS[†], P. DELLAPORTAS[†] AND D. N. POLITIS[‡]

[†]*Department of Statistics, Athens University of Economics and Business, Patission 76, 10434 Athens, Greece*

[‡]*Department of Mathematics, University of California at San Diego, La Jolla, CA 92093-0112, USA*

Summary   A new multivariate time series model with time varying conditional variances and covariances is presented and analysed. A complete analysis of the proposed model is presented consisting of parameter estimation, model selection and volatility prediction. Classical and Bayesian techniques are used for the estimation of the model parameters. It turns out that the construction of our proposed model allows easy maximum likelihood estimation and construction of well-mixing Markov chain Monte Carlo (MCMC) algorithms. Bayesian model selection is addressed using MCMC model composition. The problem of accounting for model uncertainty is considered using Bayesian model averaging. We provide implementation details and illustrations using daily rates of return on eight stocks of the US market.

Keywords: *Autoregressive conditional heteroscedasticity, Bayesian model averaging, Markov chain Monte Carlo model composition, Maximum likelihood estimation.*

1. INTRODUCTION

The volatility of financial time series appears to change over time. One class of models that has been developed to deal with this feature is the autoregressive conditional heteroscedastic (ARCH) model introduced by Engle (1982), and extended by Bollerslev (1986) to the generalized ARCH (GARCH) model. A large number of models has been proposed in the literature based on these models; see, for example, the review article of Bollerslev *et al.* (1992).

The extension from univariate models to a multivariate framework is important because the modelling of variances and covariances is crucial for multivariate financial models such as the capital asset pricing model, portfolio allocation and risk management; see, for example, Bollerslev *et al.* (1988), Ng (1991) and Gourieroux (1997). In the literature, formulations of heteroscedastic multivariate models have been introduced by Kraft and Engle (1982), Engle *et al.* (1984, 1990), Bollerslev *et al.* (1988), Diebold and Nerlove (1989), Bollerslev (1990), King *et al.* (1994), Engle and Kroner (1995), Kroner and Ng (1998), Engle (2000), Klaassen (2000), Alexander (2001) and Engle and Sheppard (2001) among others. Naturally, different multivariate models impose different restrictions on the dynamic behaviour of the variances, covariances and correlations.

Two major problems related to multivariate ARCH and GARCH models are the large number of parameters to be estimated, and the difficulty of estimation due to the positive definiteness restrictions on the covariance matrix. Factor ARCH–GARCH models have been introduced

to solve these problems, providing a parsimonious parametrization and a positive definite covariance matrix. The motivation for the factor models is commonality in the conditional variance movements. Although, in some cases, financial and economic theories suggest such a characteristic, the restrictions imposed by the factor ARCH–GARCH models on the dynamic behaviour of the covariances and the correlations are quite strong, and may depend on the number of factors. For example, in a two factor model, one would expect more dynamics in the covariances and in the correlations than in a one factor model. High-dimensional applications of factor models require Markov chain Monte Carlo (MCMC) methods; see, Chib *et al.* (1998) and Han (2002) where 40 and 36 stocks are modelled respectively.

Recently, Klaassen (2000) and Alexander (2001) have suggested the construction of unconditionally uncorrelated linear combinations of the observed series based on principal component analysis. Engle (2000) and Engle and Sheppard (2001) advocated the direct modelling of conditional correlations. All these approaches are promising in the sense that they can model high-dimensional observed series. For example, Engle and Sheppard (2001) model a 100-dimensional set of stock indices. More details about these models are presented in Section 2.2.

In this study, we propose a new parametrization of a multivariate GARCH model, where the covariance matrix is always positive definite, the number of parameters is relatively small, and the model can be applied very easily to high-dimensional time series data. The estimation of the parameters of the multivariate model is achieved by using classical and Bayesian techniques. Maximum likelihood estimation is implemented using Fisher scoring. Our model construction is such that the expected information matrix is obtained in closed form. According to the Bayesian approach, we construct a Markov chain which has as a stationary distribution the posterior distribution of the model parameters. Simulation of this Markov chain provides, after some burn-in period and adequately many iterations, samples from the posterior distribution of interest; see, for example, Smith and Roberts (1993) or Besag *et al.* (1995). We provide detailed guidelines on how to construct the required Markov chain using a blocking sampling scheme in the Metropolis–Hastings algorithm.

Bayesian model search is addressed using the MCMC model composition (MC^3) method of Madigan and York (1995) together with the delayed rejection algorithm (DRA) of Tierney and Mira (1999). We base our inference on posterior model probabilities and account for model uncertainty using Bayesian model averaging.

The remainder of the paper is organized as follows. The full-factor multivariate GARCH model is introduced in Section 2. Classical and Bayesian implementation is presented in Section 3. Bayesian model comparison and Bayesian model averaging techniques are presented in Section 4. In Section 5 we illustrate the above methods in a dataset of eight stocks from the US market, and we conclude in Section 6 with a brief discussion.

2. A FULL-FACTOR MULTIVARIATE GARCH MODEL

2.1. Description and properties of the model

Throughout the paper we consider having observed data of the form

$$\mathbf{y}_t, \qquad t = 1, \ldots, T,$$

where each $\mathbf{y}_t = (y_{1,t}, \ldots, y_{N,t})'$ is a $N \times 1$ vector. The multivariate full-factor GARCH model is given by the following equations:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{\varepsilon}_t &= W\mathbf{X}_t, \\ \mathbf{X}_t \mid \Phi_{t-1} &\sim N_N(\mathbf{0}, \Sigma_t), \end{aligned} \quad (1)$$

where $\boldsymbol{\mu}$ is a $N \times 1$ vector of constants, $\boldsymbol{\varepsilon}_t$ is a $N \times 1$ innovation vector, W is a $N \times N$ parameter matrix, Φ_{t-1} is the information set up to time $t-1$, $\mathbf{X}_t$ is a $N \times 1$ vector of factors with elements $x_{i,t}$, $i = 1, \ldots, N$ and Σ_t is a $N \times N$ diagonal variance covariance matrix, which is given by $\Sigma_t = \text{diag}(\sigma_{1,t}^2, \ldots, \sigma_{N,t}^2)$ with

$$\sigma_{i,t}^2 = \alpha_i + b_i x_{i,t-1}^2 + g_i \sigma_{i,t-1}^2, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T$$

and $\sigma_{i,t}^2$, $i = 1, \ldots, N$ is the variance of the ith factor at time t, $\alpha_i > 0$, $b_i \geq 0$, $g_i \geq 0$, $i = 1, \ldots, N$. In other words the factors $x_{i,t}$, $i = 1, \ldots, N$ are GARCH(1, 1) processes. According to the above model, the vector $\boldsymbol{\varepsilon}_t$ is a linear combination of the factors $x_{i,t}$, $i = 1, \ldots, N$.

Assuming that the vector $\mathbf{X}_t$ follows a conditional multivariate normal distribution, $\mathbf{X}_t \mid \Phi_{t-1} \sim N(\mathbf{0}, \Sigma_t)$, then the vector $\boldsymbol{\varepsilon}_t \mid \Phi_{t-1} \sim N(\mathbf{0}, H_t)$, where

$$H_t = W\Sigma_t W' = W\Sigma_t^{1/2}\Sigma_t^{1/2}W' = (W\Sigma_t^{1/2})(W\Sigma_t^{1/2})' = LL'. \quad (2)$$

Equation (2) presents the conditional covariance matrix H_t of the vector $\boldsymbol{\varepsilon}_t$. $\Sigma_t^{1/2}$ is the diagonal matrix with elements $\sigma_{1,t}, \sigma_{2,t}, \ldots, \sigma_{N,t}$. It is well known that the decomposition of a positive definite matrix into the product of a triangular matrix and its transpose always exists, and this decomposition is unique if the diagonal elements are restricted to be positive. So we can take W triangular with elements $w_{ij} = 0$ for $j > i$ and $w_{ii} > 0$ for $i = 1, \ldots, N$. In order to reduce the number of parameters in our model, a natural restriction is to assume that $w_{ii} = 1$, for $i = 1, \ldots, N$. Note that similar constraints have been adopted by Geweke and Zhou (1996), Chib *et al.* (1998) and Aguilar and West (2000) among others for the factor model. Under the assumption that the matrix W is triangular with diagonal elements equal to unity, the conditional covariance matrix H_t can be written as

$$H_t = W\Sigma_t W' = \begin{bmatrix} h_{11,t} & h_{12,t} & h_{13,t} & \cdots & h_{1N,t} \\ h_{21,t} & h_{22,t} & h_{23,t} & \cdots & h_{2N,t} \\ h_{31,t} & h_{32,t} & h_{33,t} & \cdots & h_{3N,t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1,t} & h_{N2,t} & h_{N3,t} & \cdots & h_{NN,t} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{1,t}^2 & w_{21}\sigma_{1,t}^2 & w_{31}\sigma_{1,t}^2 & \cdots & w_{N1}\sigma_{1,t}^2 \\ w_{21}\sigma_{1,t}^2 & \sum_{i=1}^{2} w_{2i}^2 \sigma_{i,t}^2 & \sum_{i=1}^{2} w_{2i} w_{3i} \sigma_{i,t}^2 & \cdots & \sum_{i=1}^{2} w_{2i} w_{Ni} \sigma_{i,t}^2 \\ w_{31}\sigma_{1,t}^2 & \sum_{i=1}^{2} w_{3i} w_{2i} \sigma_{i,t}^2 & \sum_{i=1}^{3} w_{3i}^2 \sigma_{i,t}^2 & \cdots & \sum_{i=1}^{3} w_{3i} w_{Ni} \sigma_{i,t}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{N1}\sigma_{1,t}^2 & \sum_{i=1}^{2} w_{Ni} w_{2i} \sigma_{i,t}^2 & \sum_{i=1}^{3} w_{Ni} w_{3i} \sigma_{i,t}^2 & \cdots & \sum_{i=1}^{N} w_{Ni}^2 \sigma_{i,t}^2 \end{bmatrix}. \quad (3)$$

From the construction of the model, the variance covariance matrix H_t is always positive definite if the factor variances $\sigma_{i,t}^2$, $i = 1, \ldots, N$ are well defined. Note also that the factors $x_{i,t}$,

$i = 1, \ldots, N$ have zero idiosyncratic variances so they are not parameters to be estimated but are given by $\mathbf{X}_t = W^{-1}\boldsymbol{\varepsilon}_t$. Therefore, for $N = 1$ the model reduces to the GARCH(1, 1) model.

Under the assumption of a multivariate normal distribution for the vector $\mathbf{X}_t$, the likelihood for model (1), for a sample of T observations $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$, can be written as

$$l(\mathbf{y} \mid \boldsymbol{\theta}) = (2\pi)^{-\frac{TN}{2}} \prod_{t=1}^{T} |H_t|^{-1/2} \exp\left[-\tfrac{1}{2} \sum_{t=1}^{T} (\mathbf{y}_t - \boldsymbol{\mu})' H_t^{-1} (\mathbf{y}_t - \boldsymbol{\mu}) \right].$$

In the following, we will assume for convenience that $x_{i,0}$ and $\sigma_{i,0}^2$ are known, $b_i = b$ and $g_i = g$ for $i = 1, \ldots, N$. These assumptions can be relaxed without adding much difficulty in the estimation process. For example, Vrontos *et al.* (2000) discuss Bayesian estimation of GARCH and EGARCH models when the variance at time zero is unknown. Under these assumptions, the number of parameters to be estimated for a N-dimensional problem is $2N + 2 + \frac{N(N-1)}{2}$ and the parameter vector is $\boldsymbol{\theta} = (\mu_1, \mu_2, \ldots, \mu_N, \alpha_1, \alpha_2, \ldots, \alpha_N, b, g, w_{21}, w_{31}, w_{32}, \ldots, w_{N1}, \ldots, w_{N,N-1})'$.

Bollerslev (1986) presents the necessary and sufficient conditions for the existence of the $2m$th moment of the GARCH(1, 1) model and the conditions for the wide-sense stationarity of the GARCH model. We can take advantage of these results for the full-factor multivariate GARCH model we propose. First note that $H_t = W \Sigma_t W'$ and $\text{vec}(H_t) = \text{vec}(W \Sigma_t W') = (W \otimes W)\text{vec}(\Sigma_t)$, where vec(.) denotes the vector operator that stacks the columns of the matrix. That is, the variances and the covariances in our model are linear combinations of the variances $\sigma_{i,t}^2$, $i = 1, \ldots, N$ of the factors. But according to our model $x_{i,t}$, $i = 1, \ldots, N$ are independent GARCH(1, 1) processes. Therefore, the unconditional variances and covariances of the full-factor multivariate GARCH model are linear combinations of the unconditional variances of the GARCH(1, 1) processes $x_{i,t}$, $i = 1, \ldots, N$. Define $\mathbf{w}_t = \text{vech}(H_t)$, where vech(.) denotes the column stacking operator of the lower portion of a symmetric matrix. That is, $\mathbf{w}_t = \text{vech}(H_t) = (h_{11,t}, h_{21,t}, h_{22,t}, h_{31,t}, h_{32,t}, h_{33,t}, \ldots, h_{N1,t}, h_{N2,t}, h_{N3,t}, \ldots, h_{NN,t})'$ is a $N(N + 1)/2 \times 1$ vector. After straightforward calculations and using the results of Bollerslev (1986), the unconditional variances and covariances are given by $E(\mathbf{w}_t) = \frac{1}{1-b-g}\big(\alpha_1, w_{21}\alpha_1, \sum_{i=1}^{2} w_{2i}^2 \alpha_i, w_{31}\alpha_1,$ $\sum_{i=1}^{2} w_{3i} w_{2i} \alpha_i, \quad \sum_{i=1}^{3} w_{3i}^2 \alpha_i, \ldots, w_{N1}\alpha_1, \quad \sum_{i=1}^{2} w_{Ni} w_{2i} \alpha_i, \quad \sum_{i=1}^{3} w_{Ni} w_{3i} \alpha_i, \ldots, \sum_{i=1}^{N} w_{Ni}^2 \alpha_i\big)'.$

2.2. Some comments on the model

The proposed model can be considered as a full-factor model with zero idiosyncratic variances. Different factor models have been proposed in the literature, and have been analysed by many researchers using either an ARCH–GARCH framework (see, for example, Diebold and Nerlove (1989), Engle *et al.* (1990), King *et al.* (1994)) or a stochastic volatility framework (see, for example, Chib *et al.* (1998), Aguilar and West (2000)). In particular, Harvey *et al.* (1994) proposed a factor representation that can be extended to a full-factor structure similar to ours, but they proposed modelling the factors as stochastic volatility rather than GARCH(1, 1) processes. We have not empirically investigated in detail the advantages and disadvantages of the full-factor representation with zero idiosyncratic variances for the factor models above.

The structure of our model resembles that of Alexander (2001) and of Klaassen (2000). Alexander's intent is the classical principal component target of reducing the data from a 'portfolio' of k series to an 'underlying' driving force of 2–3 series of principal components. To achieve

this, Alexander proposes a linear transformation of the data by using an orthogonal matrix of eigenvectors derived from the sample correlation matrix. Thus, the principal components are only unconditionally uncorrelated. In contrast, our intent is to model a GARCH effect on the (conditional) correlations in a parsimonious manner. The linear transformation (1) uses a triangular matrix W (which needs to be estimated jointly with the GARCH processes) which produces conditionally uncorrelated factors $x_{i,t}$, $i = 1, \ldots, N$.

In our model, the conditional correlations are given by

$$\rho_{jk,t} = \frac{\sum_{i=1}^{\min\{j,k\}} w_{ji} w_{ki} \sigma_{i,t}^2}{\left(\sum_{i=1}^{j} w_{ji}^2 \sigma_{i,t}^2\right)^{1/2} \left(\sum_{i=1}^{k} w_{ki}^2 \sigma_{i,t}^2\right)^{1/2}},$$

a non-linear function of all $\sigma_{i,t}^2$, $i = 1, \ldots, \max\{j, k\}$, which preserves the positive-definiteness of H_t. Thus, the model can be viewed as a generalization of the constant correlation coefficient model of Bollerslev (1990).

Denoting by R_t the correlation matrix with elements $\rho_{jk,t}$, (3) can be written in the form advocated by Engle (2000)

$$H_t = D_t R_t D_t,$$

where $D_t = \text{diag}(h_{11,t}^{1/2}, h_{22,t}^{1/2}, \ldots, h_{nn,t}^{1/2})$. Engle suggests modelling D_t and R_t separately imposing GARCH effects on $h_{ii,t}$ and a parsimonious formulation for R_t. Whereas our conditional correlation (3) depends on W and past values of $\mathbf{X}_t$ and σ_t^2 of other returns (depending on the order imposed), Engle's conditional correlation depends on $\rho_{ij,t-1}$, on unconditional correlations and on squares of past standardized residuals.

An empirical comparative study that investigates the extent to which these models capture empirical conditional correlation structures is the subject of current research and will be reported elsewhere.

The structure of the conditional covariance matrix H_t in equation (3) implies that the order of the components of the time series in the $\mathbf{y}_t$ vector affects the conditional variances and covariances. The ordering has an impact in model fitting and assessment; see, for example, Aguilar and West (2000). We solve this important practical issue using the MCMC model composition (MC^3) method that generates a process that moves through model space and the Delayed Rejection Algorithm (DRA). Moreover, these methods provide an idealized way to extract posterior model probabilities and to construct predictive densities that take into account model uncertainty.

3. INFERENCE FOR A GIVEN MODEL

In this section we consider classical and Bayesian techniques for the estimation of the parameters of the full-factor multivariate GARCH model.

3.1. Classical approach

Maximum likelihood estimates for heteroscedastic models are usually obtained by using numerical optimization algorithms such as scoring algorithm, the method proposed by Mak (1993) and developed further by Mak *et al.* (1997) and by Berndt *et al.* (1974). Here, we compute

the maximum likelihood estimates using the Fisher scoring algorithm. The kth iteration of the algorithm takes the form

$$\hat{\boldsymbol{\theta}}^k = \hat{\boldsymbol{\theta}}^{k-1} + \left\{ -E\left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]\right\}^{-1} \frac{\partial L_T}{\partial \boldsymbol{\theta}}, \tag{4}$$

where $\hat{\boldsymbol{\theta}}^{k-1}$ is the estimate of the parameter vector obtained after $k-1$ iterations, L_T is the log-likelihood function, $-E\left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]$ is the expected information matrix $\hat{I}$ computed at $\hat{\boldsymbol{\theta}}^{k-1}$, and $\frac{\partial L_T}{\partial \boldsymbol{\theta}}$ is the gradient computed at $\hat{\boldsymbol{\theta}}^{k-1}$. One of the great advantages of our model is, we believe, the fact that the gradient and the expected information matrix in (4) are available in closed forms. This is not surprising since, by construction, our model consists of a linear combination of univariate GARCH models in which such a property exists. The motivation for using the Fisher scoring algorithm in our full-factor multivariate GARCH model comes from the results of other research into GARCH models. For example, Fiorentini *et al.* (1996) computed the analytic conditional expected information matrix of the parameters for the GARCH(p, q) model and constructed a mixed-gradient algorithm in order to accelerate the convergence. According to their results, the superiority of gradient algorithms, which use the estimated information matrix, is clear in early iterations. Watson and Engle (1983) used the method of scoring and the EM algorithm for the estimation of dynamic factor, mimic and varying coefficient regression models. They suggest, for practical methods, a mixed EM and scoring algorithm, and the use of the scoring algorithm for inference. Similar are the results of the experiments of Demos and Sentana (1998), who present an EM algorithm for conditionally heteroscedastic factor models and propose a quasi-Newton algorithm for later iterations.

For the full-factor multivariate GARCH model (1) the log likelihood function is given by

$$\begin{aligned}
L_T(\mathbf{y} \mid \boldsymbol{\theta}) &= -\frac{TN}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln|H_t| - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{y}_t - \boldsymbol{\mu})'H_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu}) \\
&= -\frac{TN}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln|W\Sigma_t W'| - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{y}_t - \boldsymbol{\mu})'(W\Sigma_t W')^{-1}(\mathbf{y}_t - \boldsymbol{\mu}) \\
&= -\frac{TN}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln|\Sigma_t| - \frac{1}{2}\sum_{t=1}^{T}\mathbf{X}_t'\Sigma_t^{-1}\mathbf{X}_t, \qquad \text{where } \mathbf{X}_t = W^{-1}(\mathbf{y}_t - \boldsymbol{\mu}), \\
&= -\frac{TN}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left[\sum_{i=1}^{N}[\ln(\sigma_{i,t}^2)]\right] - \frac{1}{2}\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\left[\frac{x_{i,t}^2}{\sigma_{i,t}^2}\right]\right],
\end{aligned}$$

where $\boldsymbol{\theta} = (\mu_1, \mu_2, \ldots, \mu_N, \alpha_1, \alpha_2, \ldots, \alpha_N, b, g, w_{21}, w_{31}, w_{32}, \ldots, w_{N1}, \ldots, w_{N,N-1})'$, and $\alpha_i > 0, i = 1, \ldots, N, b \geq 0, g \geq 0$. In order to avoid these positivity restrictions, we transform the positive parameters using the logarithmic transformation, that is, $\alpha_i^* = \ln(\alpha_i)$, $b^* = \ln(b)$ and $g^* = \ln(g)$. We also divide the parameter vector into three blocks. The first block contains the parameters of the mean equation, that is $\boldsymbol{\theta}_1 = (\mu_1, \mu_2, \ldots, \mu_N)'$, the second block contains the transformed parameters of the variance equation, that is, $\boldsymbol{\theta}_2 = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_N^*, b^*, g^*)'$, and the third block contains the parameters in matrix W, that is, $\boldsymbol{\theta}_3 = (w_{21}, w_{31}, w_{32}, \ldots, w_{N1}, \ldots, w_{N,N-1})'$. The expected information matrix is block diagonal and the three diagonal blocks are estimated by $-E\left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1'}\right]$, $-E\left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2'}\right]$ and $-E\left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta}_3 \partial \boldsymbol{\theta}_3'}\right]$. After the

transformation of the positive parameters, the variances $\sigma_{i,t}^2$ of the factors $x_{i,t}$, $i = 1, \ldots, N$, are given by

$$\sigma_{i,t}^2 = e^{\alpha_i^*} + e^{b^*} x_{i,t-1}^2 + e^{g^*} \sigma_{i,t-1}^2, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T.$$

Some assumptions are also required for the initial values of the variances $\sigma_{i,t}^2$ and the squared factors $x_{i,t}^2$, as the variance equation of GARCH(1, 1) model is dynamic. For $t = 0$, $\sigma_{i,0}^2 = 0$, while the $x_{i,0}^2$ are calculated by using a sufficient number of observations from the sample.

Differentiating with respect to the mean parameters $\boldsymbol{\theta}_1 = (\mu_1, \mu_2, \ldots, \mu_N)'$ yields

$$\frac{\partial L_T}{\partial \boldsymbol{\theta}_1} = \sum_{t=1}^{T} \left\{ \sum_{i=1}^{N} \left[\frac{1}{2\sigma_{i,t}^2} \frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_1} \left(\frac{x_{i,t}^2}{\sigma_{i,t}^2} - 1 \right) - \left(\frac{x_{i,t}}{\sigma_{i,t}^2} \frac{\partial x_{i,t}}{\partial \boldsymbol{\theta}_1} \right) \right] \right\},$$

and the expected Information matrix for the first block is given by

$$\hat{I}_1 = -E \left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1'} \right] = \sum_{t=1}^{T} \left\{ \sum_{i=1}^{N} \left[\frac{1}{2(\sigma_{i,t}^2)^2} \frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_1} \frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_1'} + \frac{1}{\sigma_{i,t}^2} \frac{\partial x_{i,t}}{\partial \boldsymbol{\theta}_1} \frac{\partial x_{i,t}}{\partial \boldsymbol{\theta}_1'} \right] \right\},$$

where

$$\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_1} = 2e^{b^*} x_{i,t-1} \frac{\partial x_{i,t-1}}{\partial \boldsymbol{\theta}_1} + e^{g^*} \frac{\partial \sigma_{i,t-1}^2}{\partial \boldsymbol{\theta}_1}, \qquad i = 1, \ldots, N$$

and the derivatives of $x_{i,t}$, $i = 1, \ldots, N$, with respect to the mean parameters $\boldsymbol{\theta}_1$ are given by the rows of the $-W^{-1}$ matrix. That is, $\frac{\partial x_{1,t}}{\partial \boldsymbol{\theta}_1}$ is given by the first row of the $-W^{-1}$ matrix, $\frac{\partial x_{2,t}}{\partial \boldsymbol{\theta}_1}$ is given by the second row of the $-W^{-1}$ matrix, and so on.

Differentiating with respect to the variance parameters $\boldsymbol{\theta}_2 = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_N^*, b^*, g^*)'$ yields

$$\frac{\partial L_T}{\partial \boldsymbol{\theta}_2} = \sum_{t=1}^{T} \left\{ \sum_{i=1}^{N} \left[\frac{1}{2\sigma_{i,t}^2} \left(\frac{x_{i,t}^2}{\sigma_{i,t}^2} - 1 \right) \frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_2} \right] \right\},$$

while the expected information matrix for the second block is given by

$$\hat{I}_2 = -E \left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2'} \right] = \sum_{t=1}^{T} \left\{ \sum_{i=1}^{N} \left[\frac{1}{2(\sigma_{i,t}^2)^2} \frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_2} \frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_2'} \right] \right\},$$

where

$$\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_2} = \mathbf{c}_{i,t} + e^{g^*} \frac{\partial \sigma_{i,t-1}^2}{\partial \boldsymbol{\theta}_2}, \qquad i = 1, \ldots, N$$

and the vectors $\mathbf{c}_{i,t}$, $i = 1, \ldots, N$, can be calculated very easily. For example, $\mathbf{c}_{1,t} = (e^{\alpha_1^*}, 0, \ldots, 0, e^{b^*} x_{1,t-1}^2, e^{g^*} \sigma_{1,t-1}^2)'$, $\mathbf{c}_{2,t} = (0, e^{\alpha_2^*}, 0, \ldots, 0, e^{b^*} x_{2,t-1}^2, e^{g^*} \sigma_{2,t-1}^2)', \ldots, \mathbf{c}_{N,t} = (0, \ldots, 0, e^{\alpha_N^*}, e^{b^*} x_{N,t-1}^2, e^{g^*} \sigma_{N,t-1}^2)'$.

Differentiating with respect to the parameters in matrix W, that is, with respect to $\boldsymbol{\theta}_3 = (w_{21}, w_{31}, w_{32}, \ldots, w_{N1}, \ldots, w_{NN-1})'$, yields

$$\frac{\partial L_T}{\partial \boldsymbol{\theta}_3} = \sum_{t=1}^{T} \left\{ \sum_{i=1}^{N} \left[\frac{1}{2\sigma_{i,t}^2} \frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_3} \left(\frac{x_{i,t}^2}{\sigma_{i,t}^2} - 1 \right) - \left(\frac{x_{i,t}}{\sigma_{i,t}^2} \frac{\partial x_{i,t}}{\partial \boldsymbol{\theta}_3} \right) \right] \right\},$$

and the expected information matrix for the third block is given by

$$\hat{I}_3 = -E\left[\frac{\partial^2 L_T}{\partial\boldsymbol{\theta}_3\partial\boldsymbol{\theta}'_3}\right] = \sum_{t=1}^{T}\left\{\sum_{i=1}^{N}\left[\frac{1}{2(\sigma_{i,t}^2)^2}\frac{\partial\sigma_{i,t}^2}{\partial\boldsymbol{\theta}_3}\frac{\partial\sigma_{i,t}^2}{\partial\boldsymbol{\theta}'_3} + \frac{1}{\sigma_{i,t}^2}\frac{\partial x_{i,t}}{\partial\boldsymbol{\theta}_3}\frac{\partial x_{i,t}}{\partial\boldsymbol{\theta}'_3}\right]\right\},$$

where

$$\frac{\partial\sigma_{i,t}^2}{\partial\boldsymbol{\theta}_3} = 2e^{b^*}x_{i,t-1}\frac{\partial x_{i,t-1}}{\partial\boldsymbol{\theta}_3} + e^{g^*}\frac{\partial\sigma_{i,t-1}^2}{\partial\boldsymbol{\theta}_3}, \qquad i = 1,\ldots,N.$$

The factors $x_{i,t}$, $i = 1,\ldots,N$, are given by $\mathbf{X}_t = W^{-1}\boldsymbol{\varepsilon}_t$. Thus,

$$\frac{\partial\mathbf{X}_t}{\partial w_{ij}} = \left[-W^{-1}\frac{\partial W}{\partial w_{ij}}W^{-1}\right]\boldsymbol{\varepsilon}_t,$$

and therefore the derivatives of $x_{i,t}$, $i = 1,\ldots,N$, with respect to w_{ij} are given by the ith element of the vector $\left[-W^{-1}\frac{\partial W}{\partial w_{ij}}W^{-1}\right]\boldsymbol{\varepsilon}_t$. For example, the derivative of $x_{1,t}$ with respect to the parameter w_{N1} is given by the first element of the vector $\left[-W^{-1}\frac{\partial W}{\partial w_{N1}}W^{-1}\right]\boldsymbol{\varepsilon}_t$, and so on.

Having calculated the three blocks $-E\left[\frac{\partial^2 L_T}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}'_i}\right]$, $i = 1, 2, 3$ of the block diagonal expected information matrix $\hat{I}$, and the gradients $\frac{\partial L_T}{\partial\boldsymbol{\theta}_i}$, $i = 1, 2, 3$, one can find the maximum likelihood estimates by applying the Fisher scoring algorithm of equation (4). From our experience of different datasets, the estimates given by the Fisher scoring algorithm (4) are robust to different initial values for the model parameters, and the inverse of the expected information matrix provides estimates of the covariance matrix of the parameters. The above strategy enables us to investigate more deeply the behaviour of covariance matrix estimators. For example, we can very easily evaluate the covariance estimate proposed for dynamic and conditional heteroscedastic models by Bollerslev and Wooldridge (1992). This covariance estimate can be written as

$$BW = (\hat{I}^{-1})(OP)(\hat{I}^{-1}) \tag{5}$$

where $\hat{I}^{-1}$ is the inverse of the expected information matrix, and OP is the matrix of outer products of the first derivatives of the log-likelihood, which can be constructed as

$$OP = \sum_{t=1}^{T}\left[\frac{\partial L_t}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial L_t}{\partial\boldsymbol{\theta}'}\right],$$

where $\frac{\partial L_t}{\partial\boldsymbol{\theta}}$ is the derivative of the log-likelihood with respect to the parameter vector $\boldsymbol{\theta}$ at time t. Under misspecification of the conditional density of the process (non-Gausian distribution) the estimate of the variance covariance matrix given by Bollerslev and Wooldridge (1992) is more robust and superior compared to other covariance estimators; see, for example, Fiorentini *et al.* (1996).

3.2. Bayesian approach

This section presents the Bayesian approach for estimating the parameters of the proposed full-factor multivariate GARCH model using MCMC methods. The general formulation is as follows: suppose that, for a given parameter vector $\boldsymbol{\theta} \in \mathbf{R}^n$ and data $\mathbf{y}$, we want to generate a sample from

the posterior distribution $\pi(\boldsymbol{\theta} \mid \mathbf{y})$, which is known up to a constant of proportionality. The idea is based on the construction of an irreducible and aperiodic Markov chain, which is easily simulated, with realizations $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^t, \ldots$ in the parameter space, equilibrium distribution $\pi(\boldsymbol{\theta} \mid \mathbf{y})$ and a transition probability $K(\boldsymbol{\theta}'', \boldsymbol{\theta}') = \pi(\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}'' \mid \boldsymbol{\theta}^t = \boldsymbol{\theta}')$, where $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}''$ are the realized states at time t and $t + 1$ respectively. Under appropriate regularity conditions, asymptotic results guarantee that as $t \to \infty$, $\boldsymbol{\theta}^t$ tends in distribution to a random variable with density $\pi(\boldsymbol{\theta} \mid \mathbf{y})$, and the ergodic average of an integrable function f of $\boldsymbol{\theta}$ is a consistent estimator of the (posterior) mean of the function. That is,

$$\boldsymbol{\theta}^t \xrightarrow{d} \boldsymbol{\theta} \sim \pi(\boldsymbol{\theta} \mid \mathbf{y}) \qquad \text{and} \qquad \frac{1}{t} \sum_{i=1}^{t} f(\boldsymbol{\theta}^i) \xrightarrow{t \to \infty} E_{\boldsymbol{\theta} \mid \mathbf{y}}[f(\boldsymbol{\theta})] \text{ almost surely.}$$

See, for example, Smith and Roberts (1993), Besag *et al.* (1995) and Gilks *et al.* (1996).

We adopt the Metropolis–Hastings algorithm to obtain a sample from the posterior distribution of interest. There are many possible choices of the transition kernel K which lead to different sampling schemes. A simple strategy is to use n independent Metropolis steps (Tierney (1994), Chib and Greenberg (1995)) for all θ_i, $i = 1, 2, \ldots, n$, and a usual approach is to adopt a random walk chain with an increment normal density $N(0, \sigma^2)$ where the variance σ^2 is appropriately chosen so that the convergence of the MCMC sampler is as fast as possible. In this sampling scheme, the parameters are updated one at a time (single component update), making the algorithm time consuming, particularly in high-dimensional datasets.

In our proposed model, the convergence of the MCMC algorithm is accelerated by reparametrizing the positive parameters to 'near normality' and by using a blocking sampling scheme as in Vrontos *et al.* (2000). We use (as in the maximum likelihood approach) three blocks consisting of the mean parameters, the variance parameters and the parameters in matrix W. In this blocking sampling scheme, we use the results from the classical approach. That is, we start from the maximum likelihood estimates and update the parameters in each block from time t to time $t + 1$ by using three multivariate Normal proposal densities $N(\boldsymbol{\theta}_i^t, c\hat{\Sigma}_{\theta_i})$, $i = 1, 2, 3$ with $\boldsymbol{\theta}_i^t$ denoting the vector of parameters in block i with values at time t, c a constant to tune the acceptance rate, and $\hat{\Sigma}_{\theta_i}$ the variance covariance estimate of the parameters in block i taken from the Fisher scoring algorithm or taken from the method proposed for dynamic and conditional heteroscedastic models by Bollerslev and Wooldridge (1992). We found that for the first two blocks the algorithm works very well with either proposal, whereas for the third block the latter proposal density performed better. This occurred because Fisher scoring produced many zero values in the off-diagonal elements of the covariance matrix of $\boldsymbol{\theta}_3$.

Finally, it is worth mentioning that if filtering, rather than smoothing, of the GARCH process is of interest, our MCMC algorithm can be easily adapted to perform such an exercise; see, for example, Pitt and Shephard (1999) and Kim *et al.* (1998).

4. INFERENCE UNDER MODEL UNCERTAINTY

4.1. Bayesian model comparison

As described at the end of Section 2, the order of the univariate time series in the $\mathbf{y}_t$ vector has an impact on model fitting. We consider the problem of finding the 'best' ordering of the individual time series under the proposed model. That is, the ordering becomes a modelling decision to be

made on the basis of model fit. Given the N observed univariate time series in the $\mathbf{y}_t$ vector, the number of all possible models (all possible different orderings) is $N!$. Let $M = \{m_1, \dots, m_K\}$ be the set of all models, so $K = N!$. A typical approach is to carry out a model selection exercise leading to a single 'best' model ('best' ordering) and then make inferences as if the selected model was the true model. However, this ignores the uncertainty involved in model selection. A Bayesian solution to this problem involves the calculation of the posterior probabilities of all the competing models. The posterior probability of model m_k is given by

$$\pi(m_k \mid \mathbf{y}) = \frac{\pi(\mathbf{y} \mid m_k)\pi(m_k)}{\sum_{i=1}^{K} \pi(\mathbf{y} \mid m_i)\pi(m_i)}, \tag{6}$$

where

$$\pi(\mathbf{y} \mid m_k) = \int \pi(\mathbf{y} \mid \boldsymbol{\theta}_k, m_k)\pi(\boldsymbol{\theta}_k \mid m_k)d\boldsymbol{\theta}_k \tag{7}$$

is the marginal likelihood of model m_k, $\boldsymbol{\theta}_k$ is the vector of parameters of model m_k, $\pi(\mathbf{y} \mid \boldsymbol{\theta}_k, m_k)$ is the likelihood given model m_k, $\pi(\boldsymbol{\theta}_k \mid m_k)$ is the prior density of $\boldsymbol{\theta}_k$ under model m_k, and $\pi(m_k)$ is the prior probability for model m_k. Inference about the model selection problem may be performed using the Bayes factor (BF) of model m_i against model m_j, given by

$$BF = \frac{\pi(\mathbf{y} \mid m_i)}{\pi(\mathbf{y} \mid m_j)} = \frac{\pi(m_i \mid \mathbf{y})}{\pi(m_j \mid \mathbf{y})}\frac{\pi(m_j)}{\pi(m_i)}. \tag{8}$$

4.2. Bayesian model averaging

Having been able to calculate the posterior probabilities (6) of each model, it seems natural to account for model uncertainty in our predictive inferences. Rather than choosing a single 'best' model and then making inferences as if the selected model was the true model, we can use the following model averaging approach, which provides composite predictions. Suppose that we are interested in a quantity Δ. For example, in time varying volatility models this quantity may be the variances and the covariances at a future time period. Then, its posterior distribution given data $\mathbf{y}$ is given by

$$\pi(\Delta \mid \mathbf{y}) = \sum_{i=1}^{K} \pi(\Delta \mid m_i, \mathbf{y})\pi(m_i \mid \mathbf{y}), \tag{9}$$

which is an average of the posterior predictive distributions under each model weighted by their posterior model probabilities. The posterior predictive distribution of Δ given a particular model m_i is found by integrating out the model parameters $\boldsymbol{\theta}_i$:

$$\pi(\Delta \mid m_i, \mathbf{y}) = \int \pi(\Delta \mid \boldsymbol{\theta}_i, m_i, \mathbf{y})\pi(\boldsymbol{\theta}_i \mid m_i, \mathbf{y})d\boldsymbol{\theta}_i.$$

We can also use the maximum likelihood approximation

$$\pi(\Delta \mid m_i, \mathbf{y}) \simeq \pi(\Delta \mid m_i, \mathbf{y}, \hat{\boldsymbol{\theta}}_i), \tag{10}$$

where $\hat{\boldsymbol{\theta}}_i$ is the maximum likelihood estimator of the parameter vector $\boldsymbol{\theta}_i$ of model m_i; see, for example, Volinsky *et al.* (1997). For discussion of the above approach as well as evidence that

accounting for model uncertainty improves predictive performance see Kass and Raftery (1995), Draper (1995) and Raftery *et al.* (1997).

Implementation of Bayesian model comparison and Bayesian model averaging is difficult for two reasons. First, the Bayes factor can be hard to compute, and second, the number of competing models in equations (6) and (9) can be enormous.

4.3. Calculation of the Bayes factor

The Bayes factor requires evaluation of the integrals in the numerator and denominator of (8) which are the marginal densities $\pi(\mathbf{y} \mid m_i)$ and $\pi(\mathbf{y} \mid m_j)$. These integrals are in general difficult to calculate; Kass and Raftery (1995) provide an extensive description and comparison of available numerical strategies. Here, we describe Laplace's method and a variant of Laplace's method, which is used in the full-factor multivariate GARCH model.

The Laplace's method of approximation of (7) is given by

$$\pi(\widehat{\mathbf{y} \mid m_k}) = (2\pi)^{d_{\theta_k}/2} |\tilde{\Sigma}|^{1/2} \pi(\mathbf{y} \mid \tilde{\boldsymbol{\theta}}_k, m_k) \pi(\tilde{\boldsymbol{\theta}}_k \mid m_k), \tag{11}$$

where d_{θ_k} is the dimension of the parameter vector $\boldsymbol{\theta}_k$ of model m_k, $\pi(\mathbf{y} \mid \tilde{\boldsymbol{\theta}}_k, m_k)$ and $\pi(\tilde{\boldsymbol{\theta}}_k \mid m_k)$ are the likelihood and the prior, respectively, evaluated at the posterior mode $\tilde{\boldsymbol{\theta}}_k$, $\tilde{\Sigma} = [-\partial^2 \tilde{l}(\boldsymbol{\theta}_k)]^{-1}$, $\partial^2 \tilde{l}(\boldsymbol{\theta}_k)$ is the Hessian matrix of second derivatives of $\tilde{l}(\boldsymbol{\theta}_k)$, where $\tilde{l}(\boldsymbol{\theta}_k) = \log[\pi(\mathbf{y} \mid \boldsymbol{\theta}_k, m_k) \pi(\boldsymbol{\theta}_k \mid m_k)]$. According to Kass and Raftery (1995), when Laplace's method is applied to both the numerator and denominator of (8), the resulting approximation has relative error of order $O(T^{-1})$.

An important variant of (11) is

$$\pi(\widehat{\mathbf{y} \mid m_k})_{\text{mle}} = (2\pi)^{d_{\theta_k}/2} |\hat{\Sigma}|^{1/2} \pi(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_k, m_k) \pi(\hat{\boldsymbol{\theta}}_k \mid m_k), \tag{12}$$

where $\hat{\Sigma}$ is the inverse of the negative Hessian matrix of the log-likelihood evaluated at the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_k$, $\pi(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_k, m_k)$ and $\pi(\hat{\boldsymbol{\theta}}_k \mid m_k)$ are the likelihood and the prior, respectively, evaluated at $\hat{\boldsymbol{\theta}}_k$. The relative error of this approximation is again of order $O(T^{-1})$. Another alternative is to use the inverse of the expected information matrix in place of $\hat{\Sigma}$ in equation (12). The resulting approximation has a larger asymptotic relative error of order $O(T^{-1/2})$, but (see Kass and Raftery, 1995) it remains sufficiently accurate to be of use in many problems.

4.4. MCMC model search methods

In this section we describe some MCMC methods that provide posterior model probabilities and therefore can account for model uncertainty using Bayesian model averaging.

4.4.1. Markov chain Monte Carlo model composition (MC³).   MCMC model composition (MC³) was introduced by Madigan and York (1995). MC³ generates a stochastic process that moves through model space. We can construct a Markov chain $\{m(t), t = 1, 2, \ldots\}$ with state space M and equilibrium distribution given by (6) and (7). If this Markov chain is simulated for

$t = 1, \ldots, P$, then under certain regularity conditions, for any function $f(m_i)$ defined on the model space, the average

$$\hat{F} = \frac{1}{P} \sum_{i=1}^{P} f(m(i))$$

is a consistent estimate of the $E[f(m)]$; see, for example, Smith and Roberts (1993). To compute (9) in this way set $f(m) = \pi(\Delta \mid m, \mathbf{y})$.

To construct the Markov chain we define a neighbourhood nbd(m) for each model m. We also define a transition matrix q by setting $q(m \to m') = 0$ for all $m' \notin$ nbd(m) and $q(m \to m')$ constant for all $m' \in$ nbd(m). If the current state of the chain is model m then m' is drawn from $q(m \to m')$ and accepted with probability

$$\min\left\{1, \frac{|\text{nbd}(m)|\pi(m' \mid \mathbf{y})}{|\text{nbd}(m')|\pi(m \mid \mathbf{y})}\right\}, \tag{13}$$

where $|\text{nbd}(m)|$ is the number of models that belong in the neighbourhood of model m. Otherwise, the chain stays in state m. Note that, if $|\text{nbd}(m)| = |\text{nbd}(m')|$ and all models are equally likely *a priori* then the probability of acceptance is given by

$$\min\left\{1, \frac{\pi(\mathbf{y} \mid m')}{\pi(\mathbf{y} \mid m)}\right\}. \tag{14}$$

In the application of the MC^3 algorithm in Section 5 we take $|\text{nbd}(m)| = |\text{nbd}(m')|$ and assume that all models are equally likely *a priori*, and therefore the probability of acceptance is given by (14).

Different models come from the particular ordering of the univariate time series in the $\mathbf{y}_t$ vector. After extensive searching for proposal densities or, equivalently, neighbour definitions, that provide an MCMC algorithm with good mixing in a series of problems, we found that multimodalities in the model space is a very frequent phenomenon, so we suggest the following neighbourhood definitions. Assume that we are looking for the neighbours of model $m = \{m_1, \ldots, m_i, \ldots, m_j, \ldots, m_k\}$. Let us denote by nbd$_1^3(m)$ all models of the form $\{m_1, \ldots, m_j, \ldots, m_i, \ldots, m_k\}$ where m_i and m_j have exchanged positions and they are at the most three positions apart. To ensure reversibility, in a cyclic fashion we can also swap m_1 with m_k or m_{k-1} or m_{k-2}, and so on, for $i \leq 3$ and $j \geq k-3$. Moreover, define nbd$_2^3(m)$ as all models in which m_i has been moved at most three positions to the left or right:

$$\text{nbd}_2^3(m) = \{m_1, \ldots, m_{i-2}, m_i, m_{i-1}, m_{i+1}, \ldots, m_k\}$$
$$\cup \{m_1, \ldots, m_{i-1}, m_{i+1}, m_i, m_{i+2}, \ldots, m_k\}$$
$$\cup \{m_1, \ldots, m_{i-3}, m_i, m_{i-2}, m_{i-1}, m_{i+1}, \ldots, m_k\}$$
$$\cup \{m_1, \ldots, m_{i-1}, m_{i+1}, m_{i+2}, m_i, m_{i+3}, \ldots, m_k\}$$
$$\cup \{m_1, \ldots, m_{i-4}, m_i, m_{i-3}, m_{i-2}, m_{i-1}, m_{i+1}, \ldots, m_k\}$$
$$\cup \{m_1, \ldots, m_{i-1}, m_{i+1}, m_{i+2}, m_{i+3}, m_i, m_{i+4}, \ldots, m_k\},$$

dealing in an obvious cyclic fashion with the cases $i \leq 3$ and $i \geq k - 3$. The superscripts in nbd$_1^3(m)$ and nbd$_2^3(m)$ denote that the neighbourhoods are based of distances of length 3. In our example with eight time series we chose a neighbourhood as nbd$^4 = $ nbd$_1^4 \cup$ nbd$_2^4$, proposing each model within nbd^4, and taking care to have $|\text{nbd}^4(m)| = |\text{nbd}(m')|$ in (13). As in every

Metropolis–Hastings algorithm, the scaling of the proposals depends on the particular problem. In our suggested algorithm, this scaling is tuned by the choice of superscript l in nbd_1^l and nbd_2^l.

The neighbourhood $\text{nbd}(m)$ of model m consists of the set of models with either a change in the position of two univariate time series or a move of time series to a different position. We change a randomly chosen time series from one up to four positions to the left or to the right, and we move a randomly chosen time series two or three or four positions to the left or to the right. As an example, suppose that there are eight stocks in the $\mathbf{y}_t$ vector, and that a model m is given by the following ordering of the univariate time series: 12345678. That is, the first stock is at position 1, the second stock is at position 2, and so on. A change in the position of two univariate time series could be, for example, 123**7**56**4**8, where we alter the positions of the time series 4 and 7. A move of time series to a different position is, for example, **2341**567, where we move the time series of position 1 three positions to the right and the time series of positions 2, 3 and 4 one position to the left.

4.4.2. Delayed rejection algorithm. The idea of DRA was proposed by Tierney and Mira (1999). This strategy improves the Metropolis–Hastings algorithm in the (Peskun, 1973) sense that the resulting estimates have smaller asymptotic variance on a sweep by sweep basis. In our case, it is also useful as it increases the probability of moving between local modes of the posterior density.

When a Markov chain remains in the same state over successive iterations, the autocorrelation of the realized chain increases together with the variance of the estimates. In a Metropolis–Hastings algorithm this happens when the proposed candidate is rejected too often. Tierney and Mira (1999) improved the Metropolis–Hastings algorithm by reducing the probability of remaining in the current state using the following idea: when a candidate is rejected, instead of staying at the current state, propose a new candidate from a different proposal distribution, and accept or reject that candidate using an adjusted acceptance probability.

In our model selection exercise, we want to construct a Markov chain that moves through model space. Suppose that the current state of the chain is model m. Then, at the first stage, model m' is drawn from $q(m \rightarrow m')$ and accepted with probability

$$\min\left\{1, \frac{\pi(\mathbf{y} \mid m')}{\pi(\mathbf{y} \mid m)}\right\}.$$

If the candidate model m' is rejected, a new candidate model m'' is proposed from $q(m' \rightarrow m'')$ at the second stage. That is, the new candidate model m'' depends only on the last rejected candidate m'. Note that the neighbourhood of models m and m', and the transitions from model m to m', and from m' to m'' are defined as in the previous section. All models are assumed equally likely *a priori*. Tierney and Mira (1999) and Mira (2000) derived the probability of acceptance for this candidate by imposing detailed balance at each stage, given by

$$\min\left\{1, \frac{\max\{0, [\pi(\mathbf{y} \mid m'') - \pi(\mathbf{y} \mid m')]\}}{\pi(\mathbf{y} \mid m) - \pi(\mathbf{y} \mid m')}\right\}.$$

This is a two-stage symmetric DRA.

5. APPLICATION TO EIGHT STOCKS FROM THE US MARKET

We illustrate the proposed full-factor multivariate GARCH model using 2350 daily observations on eight stocks from the US stock market over the 1/1/1990–1/1/1999 period. If S_t is the value
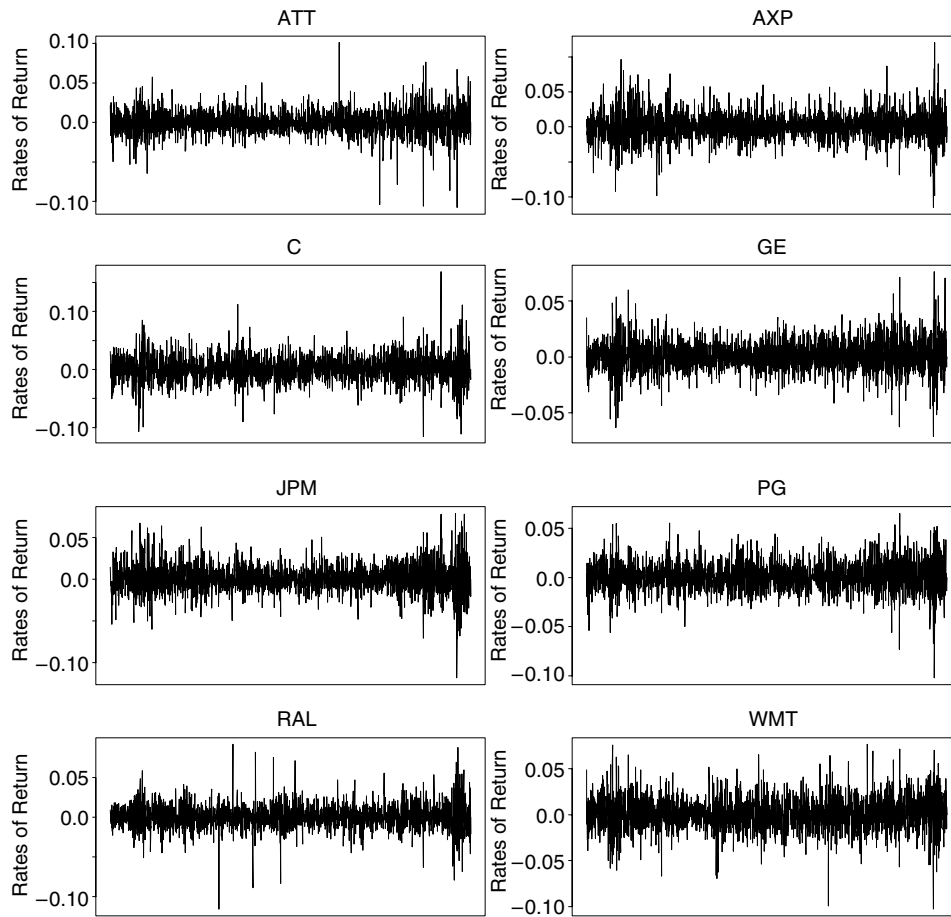
Figure 1. The analysed rates of return for the eight stocks of the US market.

of the stock at time t, then we model the rates of return $y_t = \ln\left(\frac{S_t}{S_{t-1}}\right)$, $t = 1, \ldots, T = 2349$. In Table 1, we present the summary statistics for the rates of return of the analysed stocks, together with the Ljung–Box statistic computed for the rates of return y_t, for the absolute rates and for the squares of the rates. The Ljung–Box statistic is computed using 50 lags and shows high level of autocorrelation in the squares and the absolute values. In Figure 1, we present the analysed rates of return for the eight stocks.

The key steps in our analysis are as follows. First, we ran the MC^3 algorithm without delayed rejection proposals (MC^3 without DRA) and the two stage symmetric DRA (MC^3 with DRA) to find the 'best' ordering of the individual time series and the posterior model probabilities. Second, we estimate the parameters of the 'best' model using classical and Bayesian approach. Finally, we consider the problem of accounting for model uncertainty in the proposed time varying volatility model. We make inferences about quantities of interest such as future variances and covariances by using only the 'best' model, using a set of most probable models and using Bayesian model averaging over all possible models.

**Table 1.** Summary statistics for the rates of return of the analysed stocks.

		Summary statistics					
		Rates of return y_t				$\lvert y_t \rvert$	y_t^2
Order	Stocks	Mean	Stdev	Kurtosis	LB (50)	LB (50)	LB (50)
1	ATT	0.000 374	0.015 23	5.462	60.650	636.265	301.920
2	AXP	0.000 514	0.019 76	3.101	70.564	1107.453	1035.969
3	C	0.001 000	0.021 65	4.118	63.102	599.633	484.489
4	GE	0.000 785	0.013 55	2.536	82.068	928.415	877.401
5	JPM	0.000 371	0.016 50	3.279	80.431	1935.176	1893.695
6	PG	0.000 702	0.014 55	2.390	74.112	731.037	708.877
7	RAL	0.000 443	0.014 77	6.017	99.135	527.327	289.006
8	WMT	0.000 844	0.017 85	2.034	69.490	541.026	498.996

5.1. MCMC model search and convergence assessment

We apply the MC[3] algorithm with and without DRA for 50 000 iterations in order to find the posterior model probabilities. The subsampling methodology proposed by Giakoumatos *et al.* (1999) is used to check the convergence of the MCMC output taken from the above algorithms. The method is based on the use of subsampling for the construction of confidence regions for the mean (in our case) of the unique invariant distribution of the Markov chain. We construct the $(1 - a)100\%$ confidence regions for the mean ($a = 0.05$) based on different (increasing) values $N_j = jN/100$, $j = 1, 2, \dots, 100$, and $N = 50\,000$ iterations. We estimate the 'burn-in' to be N^* if the 'range' of the confidence regions versus $1/\sqrt{N_j}$ is approximately linear for $N > N^*$. Linearity can be checked by using the coefficient of determination of a weighted linear regression between the dependent variable 'range' and $1/\sqrt{N_j}$, $j = 1, 2, \dots, 100$. According to the subsampling convergence diagnostic, we stop the MCMC simulation when the range of this $(1 - a)100\%$ confidence region for the mean is appropriately small, smaller than some prespecified absolute or relative measure of accuracy; see, for details, Giakoumatos *et al.* (1999). We focus our analysis on the 10 most probable models. Using MCMC chains of $N = 50\,000$ iterations and choosing as threshold value $d = 0.999$ for the coefficient of determination in the subsampling convergence diagnostic, we estimate the burn-in period. This turns out to be 17 500 iterations for both the MC[3] with and without DRA (see Figure 2(a)). Note however that the accuracy, that is the range of the 95% confidence region for the mean, is 0.0727 and 0.0640 for MC[3] with and without DRA, respectively. These findings confirm that MC[3] with DRA improves the Metropolis–Hastings algorithm in the sense that the resulting estimates have smaller asymptotic variance (smaller accuracy) on a sweep by sweep basis.

The resulting posterior model probabilities of the MC[3] with and without DRA are presented in Table 2. The posterior model probabilities are calculated using 32 500 iterations; that is, we have discarded the burn-in period of 17 500 iterations. The accuracy is the range of the 95% confidence region of the mean. They are calculated by running the subsampling diagnostic for each model using the MCMC output of 50 000 iterations. The 'best' model (ordering) is 14652873 with posterior probability 0.336 and 0.349 for the MC[3] with and without DRA, respectively. In total, 85 and 73 different models were visited during 50 000 iterations of MC[3]
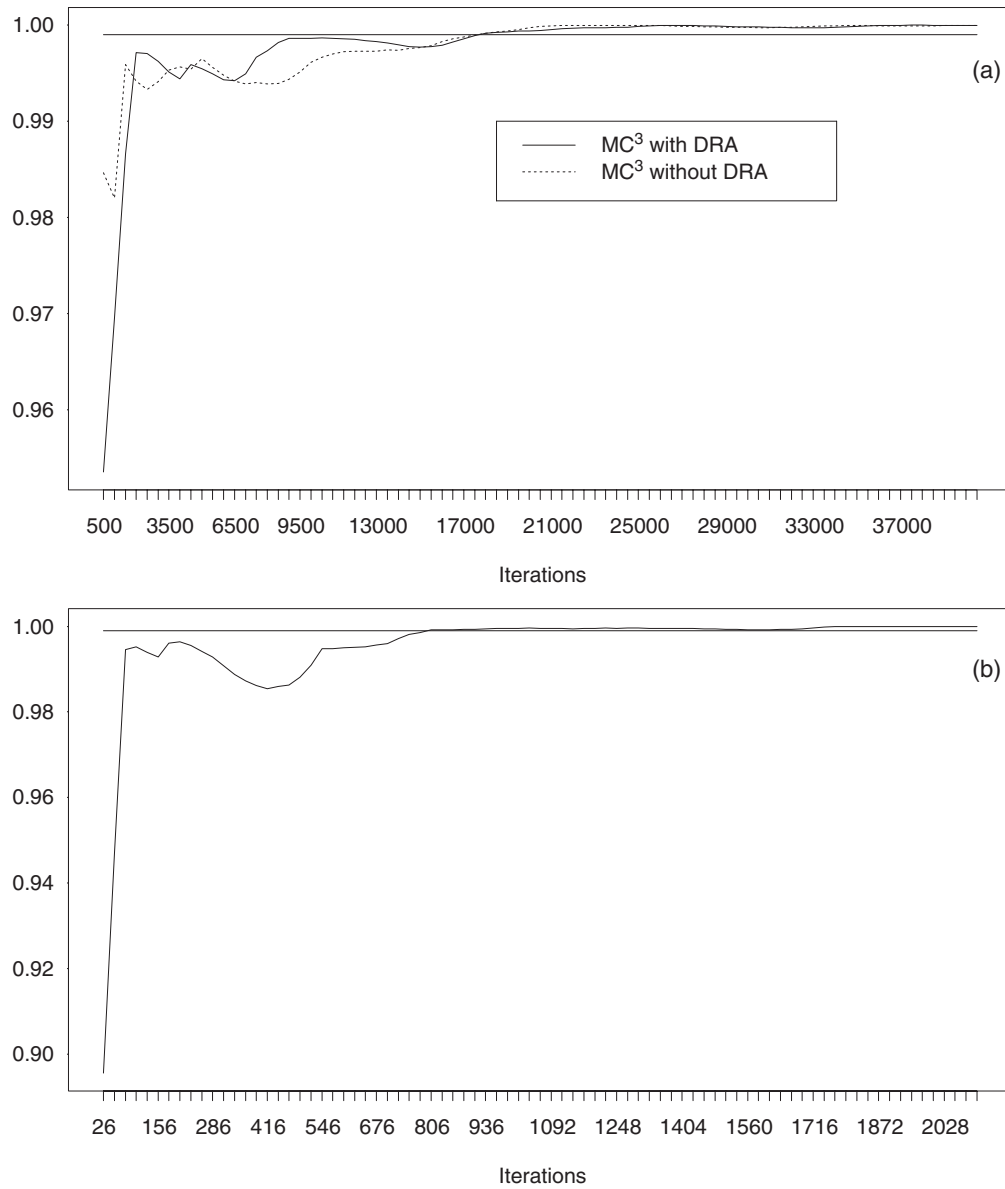
Figure 2. Coefficient of determination in the subsampling convergence diagnostic. The horizontal line represents the threshold value $d = 0.999$. (a) Coefficient of determination for the model search problem; (b) coefficient of determination for the parameters of a given model.

with and without DRA, respectively. For these estimates to be useful, we need to be confident that the reason we are only observing a small fraction of possible models in M, 85 and 73 out of 40 320, is that other models have negligible posterior probability. Thus we ran the MC^3 algorithm starting from 50 different randomly chosen models. We also ran the two-stage symmetric DRA

Table 2. Posterior model probabilities of the 10 most probable models using the MC^3 algorithm with and without DRA. The order of time series is given in Table 1. Figures in brackets are accuracy $\times 100$. Accuracy is the range of the 95% confidence region for the mean.

Model	MC^3 without DRA	MC^3 with DRA
14 652 873	0.34 (7)	0.35 (6)
14 652 738	0.10 (3)	0.10 (3)
14 652 387	0.09 (3)	0.08 (2)
14 653 287	0.06 (2)	0.08 (3)
14 652 837	0.06 (2)	0.06 (2)
14 657 328	0.05 (3)	0.07 (3)
14 652 783	0.05 (2)	0.05 (2)
14 562 873	0.04 (3)	0.03 (2)
14 657 238	0.03 (2)	0.03 (2)
14 657 283	0.02 (1)	0.03 (2)

(MC^3 with DRA) starting from the same 50 models; see, for details, Vrontos (2001). Both methods seem to be very flexible since the algorithms arrive at the 'best' ordering very fast. There was no evidence of the existence of any other regions of model space of high probability. While this is no guarantee that such regions do not exist, it does provide some reassurance. MC^3 with DRA seems to be significantly better than MC^3 without DRA. The mean values of the number of iterations needed for the MC^3 without DRA and for MC^3 with DRA to reach the most probable model 14652873 are approximately 246 and 194 iterations, respectively, and their corresponding standard deviations are approximately 105 and 104. Bearing in mind that multimodalities in model space are a very frequent phenomenon, MC^3 with DRA is useful as it increases the probability of moving between local modes of the posterior density.

5.2. Inference for a given model

Having been able to find the 'best' ordering we present the estimates for the parameters of matrix W and the corresponding standard errors in Table 3. These standard errors (in brackets) are given by the square root of the diagonal elements of the inverse of the expected information matrix. The robust estimates of the covariance matrix (5) are also calculated and the standard errors are presented in Table 3 (in square brackets). We also estimate the parameters of the multivariate model for the 'best' ordering using Bayesian analysis and MCMC methods. We transform the positive parameters to 'near normality' using the logarithmic transformation. These transformations improve the behaviour of our MCMC algorithm. For our illustration we choose non-informative (constant) priors for all the parameters. The blocking sampling scheme was used in order to update the model parameters. We ran the algorithm for 260 000 iterations, and we kept one value every 100 iterations (to save computer space). The resulting samples of 2600 values were checked for convergence using the subsampling diagnostic proposed by Giakoumatos *et al.* (1999). The method is based on the use of subsampling for the construction of confidence regions for the t-quantile ($t = 0.90$) of the unique invariant distribution of the Markov

Table 3. Estimates for the parameters of the full-factor multivariate GARCH model (1). Classical: estimates using Fisher scoring, figures in brackets are standard deviations taken using Fisher scoring, figures in square brackets are standard deviations taken using formula (5); Bayesian: posterior means and posterior standard deviations (in brackets).

	Classical	Bayesian		Classical	Bayesian
w_{21}	0.30 (0.02) [0.02]	0.30 (0.02)	w_{65}	0.10 (0.02) [0.02]	0.10 (0.02)
w_{31}	0.25 (0.02) [0.03]	0.25 (0.02)	w_{71}	0.19 (0.02) [0.02]	0.19 (0.02)
w_{32}	0.42 (0.02) [0.02]	0.42 (0.02)	w_{72}	0.33 (0.02) [0.02]	0.33 (0.02)
w_{41}	0.27 (0.02) [0.03]	0.27 (0.02)	w_{73}	0.22 (0.02) [0.02]	0.22 (0.02)
w_{42}	0.40 (0.02) [0.03]	0.41 (0.02)	w_{74}	0.09 (0.02) [0.02]	0.09 (0.02)
w_{43}	0.14 (0.02) [0.02]	0.14 (0.02)	w_{75}	0.05 (0.02) [0.02]	0.05 (0.02)
w_{51}	0.27 (0.02) [0.03]	0.27 (0.02)	w_{76}	0.08 (0.02) [0.02]	0.08 (0.02)
w_{52}	0.47 (0.03) [0.03]	0.48 (0.03)	w_{81}	0.36 (0.03) [0.04]	0.36 (0.03)
w_{53}	0.20 (0.03) [0.03]	0.21 (0.03)	w_{82}	0.57 (0.03) [0.03]	0.57 (0.03)
w_{54}	0.35 (0.02) [0.02]	0.35 (0.02)	w_{83}	0.30 (0.03) [0.03]	0.30 (0.03)
w_{61}	0.30 (0.02) [0.03]	0.30 (0.02)	w_{84}	0.42 (0.02) [0.04]	0.42 (0.03)
w_{62}	0.50 (0.02) [0.03]	0.50 (0.03)	w_{85}	0.21 (0.02) [0.02]	0.21 (0.02)
w_{63}	0.21 (0.02) [0.03]	0.21 (0.02)	w_{86}	0.07 (0.02) [0.03]	0.07 (0.02)
w_{64}	0.15 (0.02) [0.02]	0.15 (0.02)	w_{87}	0.07 (0.03) [0.03]	0.07 (0.03)

chain. We construct the $(1 - a)100\%$ confidence regions for the 0.90 quantile ($a = 0.05$) based on different (increasing) values $N_j = jN/100$, $j = 1, 2, \ldots, 100$, and $N = 2600$ iterations. As before we estimate the 'burn-in' to be N^* if the 'range' of the confidence regions versus $1/\sqrt{N_j}$ is approximately linear for $N > N^*$, and linearity is checked by using the coefficient of determination of the weighted linear regression between the dependent-variable 'range' and $1/\sqrt{N_j}$, $j = 1, 2, \ldots, 100$. The reason that the t-quantile (with a large t, say $t = 0.90$) is considered, is based on the notion that stabilization of estimates of the invariant distribution of the Markov chain (especially in the tails) is a reliable indicator of the target distribution having been achieved. We stop the MCMC simulation when the range of this $(1 - a)100\%$ confidence region for the mean is appropriately small. Using the MCMC chains of $N = 2600$ iterations and choosing as threshold value $d = 0.999$ for the coefficient of determination in the subsampling convergence diagnostic, we estimate the burn-in period. This turns out to be 780 iterations (see Figure 2(b)). The accuracy, that is the range of the 95% confidence region for the mean, is 0.013. The convergence of the parameters was also checked by using the tests proposed by Heidelberger and Welch (1983) and Raftery and Lewis (1992). The first diagnostic indicates that convergence has been achieved after a burn-in period of 780 iterations, whereas the latter diagnostic gives values for the dependent factor around one. Estimated posterior means and standard deviations for the parameters of matrix W of the latent GARCH model (1) are given in Table 3. Note that, all of the parameters of matrix W are significant.

5.3. Model uncertainty and prediction

In multivariate financial models, prediction of the future covariance matrix is of particular interest. Having been able to calculate the posterior model probabilities, it seems natural to account for model uncertainty in our predictive inferences. Suppose that we are interested in H_{T+1}, the predictive covariance matrix at time $T + 1$. Then, its posterior distribution given data $\mathbf{y}$ is given by

$$\pi(H_{T+1} \mid \mathbf{y}) = \sum_{i=1}^{K} \pi(H_{T+1} \mid m_i, \mathbf{y})\pi(m_i \mid \mathbf{y}), \tag{15}$$
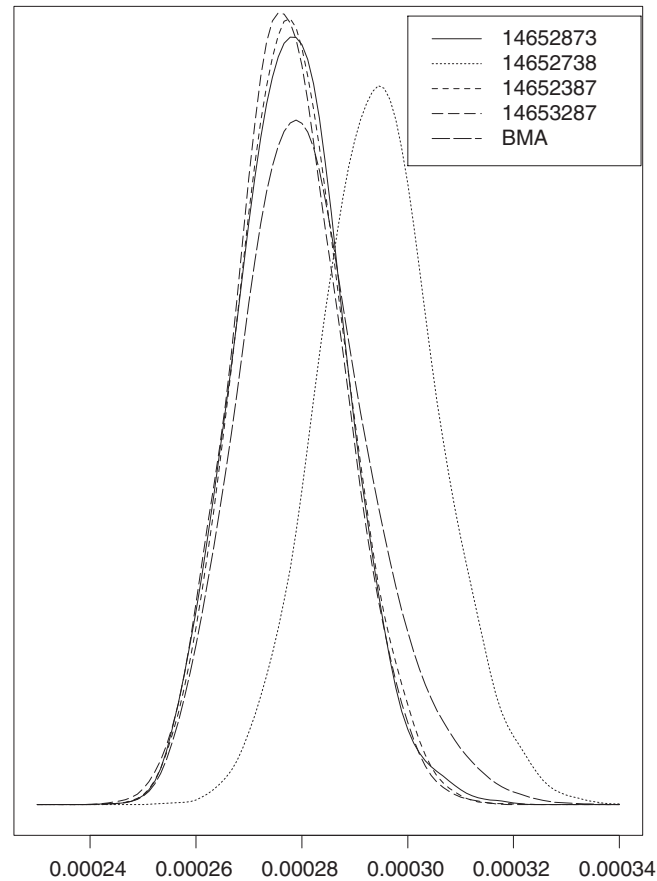
which is an average of the posterior predictive distributions under each model weighted by their posterior model probabilities. Computation of (15) is straightforward after the implementation of MC^3 or of DRA. First, given a model m_i, a posterior sample of $\pi(H_{T+1}|m_i, \mathbf{y})$ is obtained just by calculating, for each sampled point in $\boldsymbol{\theta}$, the covariance matrices $H_1, H_2, \ldots, H_{T+1}$. Then (15) suggests that in order to obtain a sample of $\pi(H_{T+1} \mid \mathbf{y})$, each sampled point under model m_i should be taken with probability $\pi(m_i \mid \mathbf{y})$. Thus, the derived sample of $\pi(H_{T+1} \mid \mathbf{y})$ is obtained by weighting all samples of $\pi(H_{T+1} \mid m_i, \mathbf{y})$ by the corresponding $\pi(m_i \mid \mathbf{y})$.

We made inference about the predictive covariance matrix at time $T + 1$ using the 'best' ordering, Bayesian model averaging based on the four most probable models and Bayesian model average based on all possible models using equations (9) and (10). We ran the algorithms for 50 000 iterations, discarded the first 10 000 iterations as burn-in, and for the remaining 40 000 iterations kept one value in every 10 iterations, obtaining a sample of 4000 iterations. A posterior sample of $\pi(H_{T+1} \mid m, \mathbf{y})$ is obtained using the 'best' model by calculating, for each one of the 4000 sampled points in $\boldsymbol{\theta}$, the covariance matrix H_{T+1}. Then, we calculated the predictive density of H_{T+1} based on the four most probable models. To achieve this, we constructed the predictive densities $\pi(H_{T+1} \mid m_i, \mathbf{y})$, $i = 1, \ldots, 4$, under each model, for each one of the 4000 sampled points in $\boldsymbol{\theta}$, and then we weighted all samples of $\pi(H_{T+1} \mid m_i, \mathbf{y})$ by the corresponding normalized posterior model probabilities. Finally, we calculated the predictive covariance matrix using Bayesian model averaging over all models that were visited during 50 000 iterations of MC^3 algorithm. For each one of these models we estimated their parameters, evaluated the predictive covariance matrix H_{T+1} based on the maximum likelihood estimates, and then weighted these values using the posterior model probabilities $\pi(m_i \mid \mathbf{y})$. We present in Figure 3 the posterior predictive density of the one step-ahead forecast for the volatility of the third stock, based on the four most probable models and on the Bayesian model average of the four most probable models.

6. DISCUSSION

In this study, we propose a new multivariate GARCH model where the covariance matrix is always positive definite and the number of parameters is relatively small with respect to other multivariate models. The model can be thought as a factor model with full-factor representation and zero idiosyncratic variances.

The estimation of the parameters of the multivariate model is done by using both classical and Bayesian techniques. Maximum likelihood estimation is implemented by using the method of Fisher scoring, while the MCMC algorithm is based on a blocking sampling scheme which accelerates convergence of the model parameters. Due to the fact that the covariance matrix is

Figure 3. Posterior predictive density of the one step-ahead forecast for the volatility of the third stock, based on the four most probable models and on the Bayesian model average of the four most probable models.

guaranteed to be positive definite, and that estimation of the parameters is easily implemented, we believe that the model can be applied very easily to high dimensional problems.

We address the problem of model selection among different models (orderings) of the analysed time series. We apply a MCMC model composition (MC^3) method with and without a second delayed rejection stage. This improves the Metropolis–Hastings algorithm in the sense that the resulting estimates have smaller asymptotic variance (smaller accuracy) on a sweep by sweep basis. We believe that this algorithm is very flexible and useful as it increases the probability of moving between local modes of the posterior density.

We also consider the problem of accounting for model uncertainty in the proposed multivariate GARCH model. Conditioning on a single selected model ignores model uncertainty. We make inferences about quantities of interest such as prediction of future variances and covariances using the 'best' model, Bayesian model averaging over a set or over all possible models.

ACKNOWLEDGEMENTS

REFERENCES

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics 18*, 338–57.

Alexander, C. (ed.) (2001). Orthogonal GARCH. *Mastering Risk*, Financial Times vol. 2, pp. 21–38. Prentice-Hall.

Berndt, E. K., B. H. Hall, R. E. Hall and J. A. Hausman (1974). Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement 3/4*, 653–65.

Besag, J., P. J. Green, D. Higdon and K. Mengersen (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science 10*, 3–66.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–27.

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

Bollerslev, T., R. Y. Chou and K. F. Kroner (1992). ARCH modeling in finance—a review of the theory and empirical evidence. *Journal of Econometrics 52*, 5–59.

Bollerslev, T., R. Engle and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy 96*, 116–31.

Bollerslev, T. and J. M. Wooldridge (1992). Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews 11*, 143–72.

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

Chib, S., F. Nardari and N. Shephard (1998). Analysis of high dimensional multivariate stochastic volatility models, unpublished paper, Olin School of Business, Washington University, St. Louis.

Demos, A. and E. Sentana (1998). An EM algorithm for conditionally heteroscedastic factor models. *Journal of Business and Economic Statistics 16*, 357–61.

Diebold, F. X. and M. Nerlove (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics 4*, 1–21.

Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, B 57*, 45–97.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica 50*, 987–1008.

Engle, R. F. (2000). Dynamic conditional correlation—a simple class of multivariate GARCH models. Discussion Paper, University of California, San Diego, CA.

Engle, R. F., C. W. J. Granger and D. F. Kraft (1984). Combining competing forecasts of inflation using a bivariate ARCH model. *Journal of Economic Dynamics and Control 8*, 151–65.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*, 122–50.

Engle, R. F., V. K. Ng and M. Rothschild (1990). Asset pricing with a factor-ARCH covariance structure:empirical estimates for treasury bills. *Journal of Econometrics 45*, 213–37.

Engle, R. F. and K. Sheppard (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Discussion Paper, University of California, San Diego, CA.

Fiorentini, G., G. Calzolari and L. Panattoni (1996). Analytic derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics 11*, 399–417.

Geweke, J. and G. Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies 9*, 557–87.

Giakoumatos, S. G., I. D. Vrontos, P. Dellaportas and D. N. Politis (1999). A Markov chain Monte Carlo convergence diagnostic using subsampling. *Journal of Computational and Graphical Statistics 8*, 431–51.

Gilks, W. R., S. Richardson and D. J. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Gourieroux, C. (1997). *ARCH Models and Financial Applications*. New York: Springer.

Han, Y. (2002). The economic value of volatility modeling: asset allocation with a high dimensional dynamic latent factor multivariate stochastic volatility model, unpublished paper, Olin School of Business, Washington University, St. Louis.

Harvey, A., E. Ruiz and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*, 247–64.

Heidelberger, P. and P. W. Welch (1983). Simulation run length control in the presence of an initial transient. *Operations Research 31*, 1109–44.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*, 773–95.

Kim, S., N. Shephard and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–93.

King, M., E. Sentana and S. Wadhwani (1994). Volatility and links between national stock markets. *Econometrica 62*, 901–33.

Klaassen, F. (2000). Have exchange rates become more closely tied? Evidence from a new multivariate GARCH model. Discussion Paper, University of Amsterdam, Amsterdam, The Netherlands.

Kraft, D. F. and R. F. Engle (1982). Autoregressive conditional heteroscedasticity in multiple time series models. Discussion Paper 82-23, University of California, San Diego, CA.

Kroner, K. F. and L. Ng (1998). Modeling asymmetric comovements of asset returns. *The Review of Financial Studies 11*, 817–44.

Madigan, D. and J. York (1995). Bayesian graphical models for discrete data. *International Statistical Review 63*, 215–32.

Mak, T. K. (1993). Solving non-linear estimation equations. *Journal of Royal Statistical Society B 55*, 945–55.

Mak, T. K., H. Wong and W. K. Li (1997). Estimation of nonlinear time series with conditional heteroscedastic variances by iteratively weighted least squares. *Journal of Computational Statistics and Data Analysis 24*, 169–78.

Mira, A. (2000). On Metropolis–Hastings algorithms with delayedrejection. *Technical Report*, Univerita dell'Insubria,Varese, Italy.

Ng, L. (1991). Tests of the CAPM with time-varying covariances: a multivariate GARCH approach. *Journal of Finance 46*, 1507–21.

Peskun, P. H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika 60*, 607–12.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association 94*, 590–9.

Raftery, A. E. and S. Lewis (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics,* vol. 4. pp. 775–82. Oxford: Oxford University Press.

Raftery, A. E., D. Madigan and J. A. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association 92*, 179–91.

Smith, A. F. M. and G. O. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B 55*, 3–24.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics 22*, 1701–62.

Tierney, L. and A. Mira (1999). Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine 18*, 2507–15.

Volinsky, C. T., D. Madigan, A. E. Raftery and R. A. Kronmal (1997). Bayesian model averaging in proportional hazard models: assessing the risk of a stroke. *Applied Statistics 46*, 433–48.

Vrontos, I. D. (2001). MCMC applications in time-varying volatility models, Ph.D. Thesis, Department of Statistics, Athens University of Economics and Business, Athens, Greece.

Vrontos, I. D., P. Dellaportas and D. N. Politis (2000). Full Bayesian inference for GARCH and EGARCH models. *Journal of Business and Economics Statistics 18*, 187–98.

Watson, W. M. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics 23*, 385–400.