UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGACIÓN OPERATIVA

The Gaussian mixture dynamic conditional correlation model: Bayesian estimation, value at risk calculation and portfolio selection

P. Galeano, M. C. Ausín

Report 07-06

Reports in Statistics and Operations Research

The Gaussian Mixture Dynamic Conditional Correlation Model: Bayesian Estimation, Value at Risk calculation and portfolio selection

Pedro Galeano

Departamento de Estadística e Investigación Operativa,

Universidad de Santiago de Compostela, Spain.

M. Concepción Ausín

Departamento de Matemáticas, Universidad de A Coruña, Spain.

Abstract

A multivariate generalized autoregressive conditional heteroscedasticity (MGARCH) model with dynamic conditional correlations where the vector of innovations is assumed to follow a mixture of two Gaussian distributions is analyzed. The Gaussian mixture distribution postulates that a large number of multivariate innovations are generated from a Gaussian distribution with a small covariance matrix, while a small number of multivariate innovations are generated from a Gaussian distribution with a large covariance matrix. It is shown that this specification jointly with a MGARCH model with time varying correlations, can capture the stylized facts usually found in multivariate returns. Inference on the model parameters and prediction of future volatilities is addressed using a Bayesian approach via a Markov Chain Monte Carlo (MCMC) method. Furthermore, the proposed methodology allows us to obtain point estimates and predictive intervals for the Value at Risk (VaR) of a given portfolio, which is strongly affected by the specification of a convenient innovation distribution. Finally, the proposed approach also provides a method for selecting portfolios with a low out-of-sample conditional variance. The good performance of the proposed methodology is illustrated via Monte Carlo experiments and the analysis of the daily closing prices of the Dow Jones and Nasdaq indices.

Keywords: Bayesian inference; Gaussian mixture models; Markov Chain Monte Carlo methods; Multivariate GARCH models; Portfolio selection; Value at Risk.

1 Introduction

The autoregressive conditional heteroscedastic (ARCH) model introduced by Engle (1982) and its generalization, the GARCH model proposed by Bollerslev (1986), have become very popular in modeling financial time series as they are able to deal with several of the main features exhibited by this kind of series. The extension of univariate models to the multivariate framework is important because the estimation of the correlations between different returns is crucial for many issues of financial management such as portfolio analysis, risk management and asset pricing. Several multivariate models have been proposed in the literature since the seminal multivariate ARCH model of Kraft and Engle (1982). For instance, the VEC model proposed by Bollerslev, Engle and Wooldridge (1988) and the BEKK model proposed by Engle and Kroner (1995) are direct generalizations of the univariate ARCH and GARCH models, respectively. The factor models (FM) proposed by Engle, Ng and Rothschild (1990), Bollerslev and Engle (1993) and Vrontos, Dellaportas and Politis (2003*b*), among others, are multivariate models based on the assumption that the co-movements of stock returns are driven by a small number of common factors. The constant conditional correlation (CCC) model proposed by Bollerslev (1990), and the dynamic conditional correlation (DCC) models proposed by Tse and Tsui (2002) and Engle (2002), are nonlinear combinations of univariate GARCH models. Bauwens, Laurent and Rombouts (2006) give a comprehensive survey of multivariate GARCH-type models and their properties. There exists a large number of applications of MGARCH models. Just to mention two of them, Baillie and Myers (1991) estimate time-varying hedge ratios of commodity futures, while Karolyi (1995) analyzes the transmissions between volatilities and co-volatilities of international markets along time through MGARCH models.

Most of the proposed multivariate models have been derived to describe time varying volatilities and correlations of several return series, but other features exhibited by multivariate returns such as long tailed distributions, high multivariate kurtosis and the presence of extremes have received considerably less attention. It is usual to assume that the returns have a conditional multivariate Gaussian or a Student-t distribution. However, it is well known that both distributions are not consistent with the above mentioned features. This problem also appears in the univariate framework in which several authors including Bai, Rusell and Tiao (2003) and Ausín and Galeano (2007), have proposed modeling the standardized innovation distribution with a mixture of two zero mean Gaussian distributions with different variances. Therefore, most of the innovations are generated according to a Gaussian distribution whose variance is slightly less than one, while a few innovations are generated according to a Gaussian distribution whose variance may be much larger than one. These authors have shown that this mixture innovation specification combined with GARCH models, successfully captures volatility clustering, long tails, high kurtosis and the presence

of extreme events. The first contribution of this paper is to analyze a multivariate dynamic conditional correlation model with a multivariate Gaussian mixture innovation distribution, which extends the univariate Gaussian mixture specification to the multivariate framework, showing that it successfully captures the main features of multivariate financial time series.

Inference on multivariate GARCH-type models is usually carried out by maximum likelihood. However, less attention has been made in the analysis of these models from a Bayesian point of view. For instance, Vrontos, Dellaportas and Politis (2003a, 2003b), performed Bayesian inference on some multivariate GARCH models. Bayesian inference is specially well suited for MGARCH models as it offers a natural way to introduce parameter uncertainty in the estimation of volatilities and correlations. Also, predictive distributions of future volatilities and correlations can be obtained which are more informative than simple point forecasts. The second contribution of this paper is to show how to perform Bayesian inference for the proposed time-varying MGARCH model with Gaussian mixture innovations. This is carried out by using Markov Chain Monte Carlo (MCMC) methods which are able to address the complexity of these models. We show how a special reparametrization of the model parameters combined with data augmentation techniques and a block-sampling scheme allow for a straightforward MCMC implementation associated with good mixing performances.

Value at Risk (VaR) is a widely used measure of market risk which has become one of the most important issues in Risk Management, see e.g. Jorion (2006). Value at Risk may be defined as the worst scenario that is expected to occur with a large probability for a portfolio given by a linear combination of the returns of the multivariate series. Clearly, the VaR of a portfolio strongly depends on the assumption made for the innovation distribution and the Gaussian distribution is unable to mimic the heavy-tailedness of financial time series observed in markets, see e.g. Jaschke and Jiang (2002). The third contribution of this paper is to propose a Bayesian procedure for the derivation of predictive distributions for the portfolio VaR based on the individual returns. Under this framework, besides of giving point estimates of VaR, the Bayesian approach provides a measure of precision for VaR estimates via predictive intervals.

The portfolio selection problem may be defined as the determination of the optimal weights which are assigned to each return. There are different answers to the question of what an optimal portfolio means. The classical approach is the mean-variance method proposed by Markowitz (1952) which is based on minimizing the variance for a given expected return using efficient frontiers. However, this methodology has been shown to be extremely unstable due to several causes including the difficulties associated with the large estimation errors of the mean asset returns, see Jagannathan and Ma (2003). These authors, among others have shown that minimum-variance portfolios usually performs better than mean-variance portfolios. Then, we consider here that a portfolio is optimal if it minimizes the risk, measured in terms of the variance. The

fourth contribution of this paper is to propose a Bayesian method to provide point estimates and predictive intervals for the optimal weights of minimum-variance portfolios.

The rest of this paper is organized as follows. Section 2 presents the Gaussian mixture dynamic conditional correlation (GMDCC) model and shows its flexibility in capturing the special features of multivariate financial time series. Section 3 describes how to perform Bayesian inference for the GMDCC model. Non informative priors are chosen for a suitable reparameterization of the model parameters and samples of the posterior distributions are obtained with a Random Walk Metropolis Hastings algorithm, see e.g. Robert and Casella (2004), using a block-sampling scheme. Also, it is shown how to estimate in-sample volatilities and correlations and how to predict future volatilities and correlations. Section 4 deals with the problems of calculation of VaR and determination of optimal portfolios. Section 5 presents a brief Monte Carlo experiment which shows the accuracy in the estimation of the parameters, prediction of volatilities and correlations, calculation of VaR and determination of optimal portfolios. Section 6 illustrates the proposed methodology with the Dow Jones Industrial Average and the Nasdaq composite indices. Finally, Section 7 concludes.

2 The Gaussian mixture dynamic conditional correlation (GMDCC) model

In the following, we assume that the vector of return time series of dimension $K \times 1$ given by $y_t = (y_{1t}, \ldots, y_{Kt})'$ follows a multivariate generalized autoregressive conditionally heteroscedastic (MGARCH) model given by,

$$y_t = \mu + H_t^{\frac{1}{2}} \epsilon_t,$$

where $\mu = (\mu_1, \ldots, \mu_K)'$ is the unconditional mean vector of the process, which is assumed to be constant, H_t is the $K \times K$ positive definite conditional covariance matrix of y_t given the past information $I_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$ and ϵ_t are iid random vectors of dimension $K \times 1$ such that $E[\epsilon_t] = 0$ and $Cov[\epsilon_t] = I_K$, the K-dimensional identity matrix. The (i,j)-th element of the covariance matrix, H_t, is denoted by H_{ijt}, where $i, j = 1, \ldots, K$. Thus, H_{iit}, for $i = 1, \ldots, K$, denotes the individual conditional variances of the components of the return vector y_t, and H_{ijt}, for $i \neq j$ and $i, j = 1, \ldots, K$, denotes the conditional covariances between the i-th and j-th components of the return vector y_t. The conditional covariance matrix can be decomposed as follows,

$$H_t = D_t R_t D_t, \tag{1}$$

where D_t is the $K \times K$ diagonal matrix with the K conditional standard deviations $H_{iit}^{\frac{1}{2}}$, and R_t is the $K \times K$ matrix of conditional correlations. Note that H_t is positive definite if and only if all the conditional variances H_{iit} are positive and the conditional correlation matrix R_t is positive definite.

The conditional variance of each individual return series can be formulated by entertaining a univariate GARCH model, or any other stochastic volatility model allowing for asymmetries, such as the EGARCH model by Nelson (1991), the GJR model by Glosten, Jagannathan and Runkle (1993) or the threshold ARCH model of Zakoian (1994). For simplicity, and without loss of generality, we exemplify our approach by considering the case in which each single return follows a univariate GARCH(p_i, q_i) model given by the equation,

$$H_{iit} = \omega_i + \sum_{l=1}^{p_i} \alpha_{i,l} \left(y_{i,t-l} - \mu_i\right)^2 + \sum_{l=1}^{q_i} \beta_{i,l} H_{ii,t-l},$$

where the parameters of each individual GARCH model verify $\omega_i > 0$, $\alpha_{i,l} \geq 0$, $l = 1, \ldots, p_i$, $\beta_{i,l} \geq 0$, $l = 1, \ldots, q_i$, which ensure positive variances, and consequently, the matrices D_t are positive definite. Additionally we impose that $\sum_{l=1}^{p_i} \alpha_{i,l} + \sum_{l=1}^{q_i} \beta_{i,l} < 1$, so that the GARCH$(p_i, q_i)$ model is covariance stationary (see, Bollerslev, 1986). Note that the orders (p_i, q_i) of the individual GARCH models do not necessarily coincide.

On the other hand, we adopt the specification made by Tse and Tsui (2002) of the conditional correlation matrix R_t. For that, let $e_t = D_t^{-1}\left(y_t - \mu\right)$ be the vector of standardized returns, let E_{t-1} be the $K \times K$ matrix $E_{t-1} = (e_{t-1}, \ldots, e_{t-K})$, and let B_{t-1} be the K-dimensional diagonal matrix where the $i-$th element is given by $\left(\sum_{h=1}^{K} e_{i,t-h}^2\right)^{\frac{1}{2}}$, for $i = 1, \ldots, K$. Then, R_t is generated from the recursion,

$$R_t = \left(1 - \theta_1 - \theta_2\right) R + \theta_1 R_{t-1} + \theta_2 \Psi_{t-1},$$

where θ_1 and θ_2 are non-negative parameters satisfying $\theta_1 + \theta_2 < 1$, R is a K-dimensional positive definite matrix with unit diagonal elements and off-diagonal elements denoted by R_{ij}, for $i \neq j$ and $i, j = 1, \ldots, K$, and Ψ_{t-1} is a $K \times K$ matrix given by $\Psi_{t-1} = B_{t-1}^{-1} E_{t-1} E_{t-1}' B_{t-1}^{-1}$. Therefore, Ψ_{t-1} is a kind of sample correlation matrix of E_{t-1}, as the (i,j)-th element of Ψ_{t-1} is given by,

$$\Psi_{ij,t-1} = \frac{\sum_{h=1}^{K} e_{i,t-h} e_{j,t-h}}{\sqrt{\left(\sum_{h=1}^{K} e_{i,t-h}^2\right)\left(\sum_{h=1}^{K} e_{j,t-h}^2\right)}}, \qquad 1 \leq i < j \leq K.$$

Note also that Ψ_{t-1} can be seen as a kind of multivariate analogous to $\left(y_{i,t-1} - \mu_i\right)^2$ in the univariate GARCH(1,1) model. With this specification, the conditional correlation matrices, R_t, are positive definite, and thus that H_t is also positive definite.

Inference on multivariate GARCH-type models is usually carried out by assuming a distribution for the innovations ϵ_t and maximizing the corresponding log-likelihood function. The common election is the multivariate standard Gaussian distribution because, although the true innovation distribution may be non-Gaussian, the quasi maximum likelihood estimator (QMLE) of the vector of parameters under certain conditions, is strong consistent, as shown by Jeantheau (1998), and asymptotically normal for BEKK models, as shown by Comte and Lieberman (2003). In particular, under Gaussianity, the DCC models proposed by Tse and Tsui (2002) and Engle (2002) can be consistently estimated, although inefficiently, by using the QMLE. These properties justify the use of QMLE if we only want to estimate consistently the model parameters and the first two conditional moments, but the Gaussian assumption has several drawbacks. First, it is well known that the Gaussianity assumption is rejected for most multivariate residuals after fitting a MGARCH model. In fact, after fitting the model via the Gaussian likelihood, the estimated innovation distribution shows long tails and high excess kurtosis, which means that the number of extreme events in the series is much larger than the number of extreme events which may be generated by the Gaussian distribution. Second, if the innovations are not Gaussian, the QMLE is less efficient than the MLE based on the true innovation distribution. Third, many financial applications including VaR calculation deeply rely on the correct assumption on the innovation distribution and has been shown to perform poorly assuming Gaussianity. Finally, in the univariate case, Hall and Yao (2003) have shown that the usual asymptotic consistency and normality properties do not necessarily hold if the innovations do not have at least moments smaller or equal than four. In summary, the Gaussianity assumption is unrealistic in practice.

Mardia (1970) introduced a measure of multivariate kurtosis, which for a K-dimensional multivariate random variable X with mean μ and covariance matrix Σ, is given by,

$$MK\left[X\right] = E\left[\left\{\left(X-\mu\right)'\Sigma^{-1}\left(X-\mu\right)\right\}^{2}\right]. \tag{2}$$

As the multivariate kurtosis (2) of a Gaussian distributed random variable X is $K\left(K+2\right)$, the multivariate excess kurtosis of a given random variable X is defined as follows,

$$EK\left[X\right] = MK\left[X\right] - K\left(K+2\right). \tag{3}$$

Thus, the multivariate excess kurtosis is a generalization of the excess kurtosis in the univariate case, such that, when $EK\left[X\right]$ is positive, the distribution of X has longer tails than the Gaussian distribution. Consequently, if we assume that the innovation process ϵ_t is a sequence of independent and identically distributed Gaussian random variables with zero mean and identity covariance matrix, then, the multivariate excess

kurtosis (3) of ϵ_t is $EK[\epsilon_t] = 0$, as in the univariate case.

The usual alternative distribution for the innovations is the multivariate standardized Student-t distribution with ν degrees of freedom. If we assume that ϵ_t follows this specification, it can be shown that the excess kurtosis of ϵ_t only exists if $\nu > 4$, in which case is given by,

$$EK_T[\epsilon_t] = \frac{2K(K+2)}{\nu - 4}.$$

Note that the fourth moment of ϵ_t only exists if $\nu > 4$, while the second moment of ϵ_t only exists if $\nu > 2$. In practice, ν is fixed to be larger than 4, in which case, the implied multivariate excess kurtosis of the residuals after estimation does not usually match the observed excess kurtosis, or it is estimated, in which case, its estimate may be smaller than 4, which implies that the estimated excess kurtosis does not exist.

In the univariate case, a solution to this problem was suggested by Bai, Rusell and Tiao (2003), who proposed to model the innovations with a mixture of two zero mean Gaussian distributions. This is the distribution used in the variance inflation model of Box and Tiao (1968), which successfully handle extreme events in linear models. Bai, Rusell and Tiao (2003) showed that the Gaussian mixture gives better fits than the Student-t distribution and that the excess kurtosis implied by the Gaussian mixture is closer to the sample excess kurtosis than the implied by the Student-t distribution. Also, these authors and Ausín and Galeano (2007) from a Bayesian point of view, have shown that the mixture specification combined with GARCH models can capture the usual patterns exhibited by financial time series such as volatility clustering, large excess kurtosis and extreme observations. Accordingly with this, we extend this specification to the multivariate framework and assume that the innovation process ϵ_t follows a mixture of two Gaussian distributions as follows,

$$\epsilon_t \sim \begin{cases} N\left(0, \sigma^2 I_K\right), & \text{with probability } \rho, \\ N\left(0, \frac{\sigma^2}{\lambda} I_K\right), & \text{with probability } 1-\rho, \end{cases} \tag{4}$$

where $0 < \lambda < 1$ and $\sigma^2 = \left(\rho + \frac{1-\rho}{\lambda}\right)^{-1}$, so that, $Cov[\epsilon_t] = I_K$. Therefore, the innovation vector, ϵ_t, is generated from a Gaussian distribution with covariance $\sigma^2 I_K$, where σ^2 is always less than 1, with probability ρ, or from a multivariate Gaussian distribution with covariance $\left(\sigma^2/\lambda\right) I_K$, where σ^2/λ is always larger than σ^2 and may increase to infinity if λ goes to 0, with probability $1-\rho$. As our aim is to allow for the presence of several extreme events, we set ρ in the interval $(0.5, 1)$ which ensures that the component with largest number of elements is the one with smallest variance. Note that $1-\rho$ can be seen as the proportion of multivariate extremes in the return series. It is not difficult to show that the excess kurtosis of the mixture

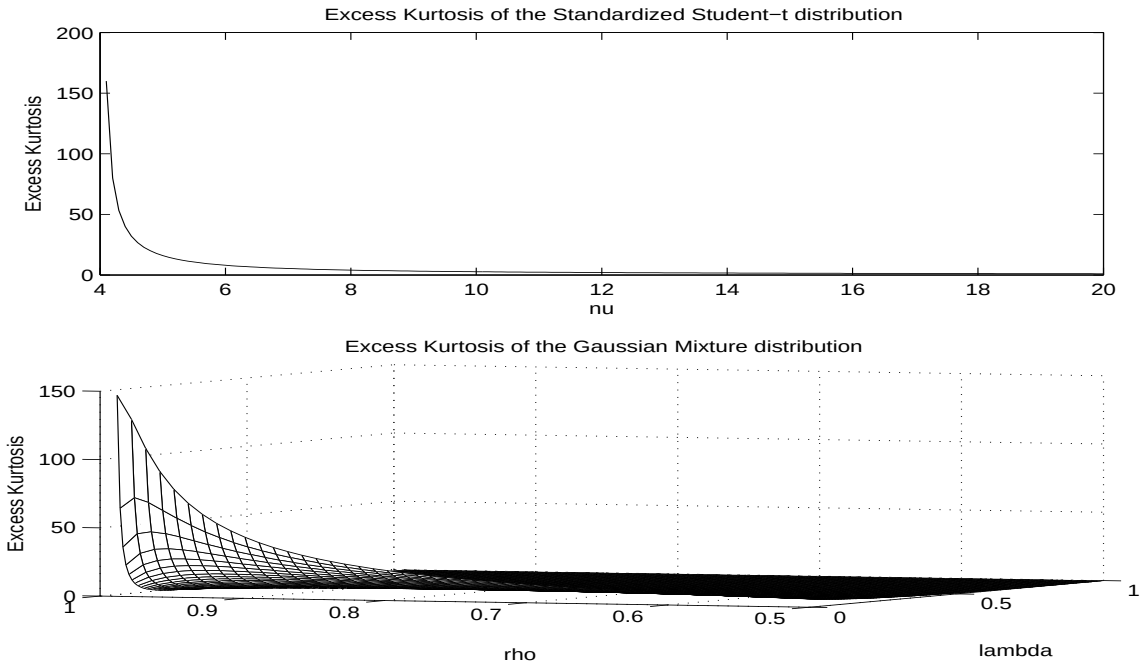

Figure 1: Multivariate excess kurtosis of the Standardized Student-t and the Gaussian Mixture distributions for $K = 2$.

is given by,

$$EK_{GM}\left[\epsilon_t\right] = 3K\frac{\rho\left(1-\rho\right)\left(1-\frac{1}{\lambda}\right)^2}{\left(\rho+\frac{1-\rho}{\lambda}\right)^2},$$

which exists for every value in the domain of ρ and λ and can take any possible positive value. Figure 1 compares the multivariate excess kurtosis of the Student-t and the Gaussian mixture distributions for $K = 2$, respectively. As can be seen, the multivariate excess kurtosis under the Student-t distribution diverges to infinity when ν tends to 4, while the multivariate excess kurtosis under the Gaussian distribution exists for all the values on the range of λ and ρ. Note that when λ tends to one, $EK_{GM}\left[\epsilon_t\right]$ tends to zero, and when λ and ρ tend to one and zero, respectively, $EK_{GM}\left[\epsilon_t\right]$ tends to infinity. In fact, all the moments of the mixture exists as they are a linear combination of the moments of the components of the mixture, which are the moments of Gaussian distributions.

The parameters of the GMDCC model can be summarized in the vector $\Phi = (\Lambda', \Omega_1', \ldots, \Omega_K', \Theta')'$, where $\Lambda = (\rho, \lambda)'$ are the parameters of the Gaussian mixture, $\Omega_i = (\mu_i, \omega_i, \alpha_{i,1}, \ldots, \alpha_{i,p_i}, \beta_{i,1}, \ldots, \beta_{i,q_i})'$, for $i = 1, \ldots, K$, are the parameters of the conditional variances of the single returns and $\Theta = (\theta_1, \theta_2, R_{12}, \ldots, R_{K-1,K})'$ are the parameters of the conditional correlation matrix. In summary, the total number of parameters of the GMDCC model is $K\left(K+3\right)/2 + p_1 + \ldots + p_K + q_1 + \ldots + q_K + 4$. For instance, when $K = 2$, 3 and 4 and entertaining univariate GARCH(1,1) models for the individual returns, the number of parameters of the GMDCC model is 13, 19 and 26, respectively. Note that there are only two and one additional pa-

rameters compared with the DCC model of Tse and Tsui (2002) using Gaussian and Student-t innovations, respectively.

The expression of the likelihood function for the observed return series, $y = (y_1, \ldots, y_T)$, assuming the Gaussian mixture specification for the innovation distribution, can be quite simplified using a set of unobserved latent variables given by,

$$z_t = \begin{cases} 1, & \text{with probability } \rho, \\ 2, & \text{with probability } 1-\rho, \end{cases}$$

for $t = 1, \ldots, T$. Thus, the return series is completed with a missing data set, $z = (z_1, \ldots, z_T)$, indicating the specific component of the mixture from which each multivariate innovation is assumed to arise. Then, conditional on these indicators, we have that,

$$y_t \mid H_t,\ z_t \sim \begin{cases} N\left(\mu, \sigma^2 H_t\right) & \text{if } z_t = 1 \\ N\left(\mu, \frac{\sigma^2}{\lambda} H_t\right) & \text{if } z_t = 2, \end{cases}$$

and the likelihood for the completed data can be separated into two blocks, taking into account the observations assigned to each component of the Gaussian mixture,

$$l\left(\Phi|y,z\right) = \prod_{t:z_t=1} \left[\rho\left(2\pi\right)^{-K/2} \left|\sigma^2 H_t\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(y_t-\mu\right)' \left(\sigma^2 H_t\right)^{-1}\left(y_t-\mu\right)\right\}\right] \times$$
$$\prod_{t:z_t=2} \left[\left(1-\rho\right)\left(2\pi\right)^{-K/2} \left|\frac{\sigma^2}{\lambda} H_t\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(y_t-\mu\right)' \left(\frac{\sigma^2}{\lambda} H_t\right)^{-1}\left(y_t-\mu\right)\right\}\right].$$

The likelihood may be further simplified by replacing $H_t = D_t R_t D_t$ and $e_t = D_t^{-1}\left(y_t - \mu\right)$,

$$l\left(\Phi|y,z\right) = \left(2\pi\sigma^2\right)^{-TK/2} \rho^{T_1} \left(1-\rho\right)^{T_2} \lambda^{KT_2/2} \prod_{t=1}^{T} \left[\left(\prod_{i=1}^{K} H_{iit}^{-1/2}\right) \left|R_t\right|^{-1/2}\right] \times \exp\left(-\frac{S_1 + \lambda S_2}{2\sigma^2}\right),$$

where $T_j = \#\left\{z_t = j\right\}$, and $S_j = \sum_{t:z_t=j} e_t' R_t^{-1} e_t$, for $j = 1, 2$ respectively.

3 Bayesian Inference for the GMDCC model

This section describes how to perform Bayesian inference of the GMDCC model introduced in the previous section via Markov Chain Monte Carlo methods. Under the Bayesian framework, inference on the parameters of the model, Φ, is done through the posterior density of Φ conditionally on the return series, y, which is

denoted by $p(\Phi|y)$. Using the Bayes theorem, $p(\Phi|y)$ is given by,

$$p(\Phi|y) = \frac{l(\Phi|y)\, p(\Phi)}{\int l(\Phi|y)\, p(\Phi)\, d\Phi},$$

where $l(\Phi|y)$ is the likelihood function and $p(\Phi)$ is the prior probability of Φ. Although the analytical derivation of $p(\Phi|y)$ for the GMDCC model is extremely difficult, we may rely on Markov Chain Monte Carlo (MCMC) methods to obtain samples of the posterior distribution. The idea is to build an irreducible and aperiodic Markov chain in the parameter space with states $\Phi^{(0)}, \Phi^{(1)}, \ldots, \Phi^{(N)}$, where $\Phi^{(0)}$ is the initial state, such that, under very mild conditions, the chain has equilibrium distribution $p(\Phi|y)$. Therefore, as n goes to infinity, $\Phi^{(n)}$ tends in distribution to a random variable which has density $p(\Phi|y)$. Moreover, if f is a function of the parameters Φ, then, the strong law of large numbers guaranties that,

$$\frac{1}{N-s} \sum_{n=s+1}^{N} f\left(\Phi^{(n)}\right) \rightarrow E\left[f(\Phi)\,|y\right],$$

almost surely, where s is the number of realizations which are discarded in a burn-in period. See, for instance, Robert and Casella (2004) for an overview on MCMC methods from both theoretical and practical points of view.

Firstly, we need to specify the expressions of the prior distributions. We use uniform prior distributions in the domain of all the parameters of the vector Φ, except for the mean parameters, $\mu_1, \ldots, \mu_k$, and the parameters $\omega_1, \ldots, \omega_k$ of the univariate GARCH models. We assume a prior standard normal distribution for each μ_i. On the other hand, we assume a uniform distribution on the interval $\left(0, \widehat{\sigma}_i^2\right)$, where $\widehat{\sigma}_i^2$ is the sample variance of the i-th return, which is a large enough upper bound for the values attained by ω_i. These distributions ensure a non informativeness prior for these parameters, avoiding the use of improper priors.

Next, we reparameterize the model parameters using logit transformations as follows,

$$\begin{array}{ccc}
\rho^* = \ln\left(\frac{\rho - 0.5}{1-\rho}\right), & \omega_i^* = \ln\left(\frac{\omega_i}{\widehat{\sigma}_i^2 - \omega_i}\right), & R_{ij}^* = \ln\left(\frac{1+R_{ij}}{1-R_{ij}}\right), \\
\lambda^* = \ln\left(\frac{\lambda}{1-\lambda}\right), & \alpha_{il}^* = \ln\left(\frac{\alpha_{il}}{1 - \sum_{l=1}^{p_i} \alpha_{i,l} - \sum_{l=1}^{q_i} \beta_{i,l}}\right), & \theta_1^* = \ln\left(\frac{\theta_1}{1-\theta_1-\theta_2}\right), \\
\mu_i^* = \mu_i, & \beta_{il}^* = \ln\left(\frac{\beta_{il}}{1 - \sum_{l=1}^{p_i} \alpha_{i,l} - \sum_{l=1}^{q_i} \beta_{i,l}}\right), & \theta_2^* = \ln\left(\frac{\theta_2}{1-\theta_1-\theta_2}\right).
\end{array}$$

These transformations allow for a better performance of the MCMC algorithm as the parameters taking values in finite ranges are transformed to parameters defined in the interval $(-\infty, \infty)$. This type of reparametrization was also used by Vrontos, Dellaportas and Politis (2003*a*, 2003*b*) in other multivariate GARCH models. But, in addition, the transformations that we propose here automatically impose the

stationarity restrictions on the single GARCH models, $\sum_{l=1}^{p_i} \alpha_{i,l} + \sum_{l=1}^{q_i} \beta_{i,l} < 1$, and on the dynamic correlation model, $\theta_1 + \theta_2 < 1$, while the transformed parameters, α_{il}^*, β_{il}^*, θ_1^* and θ_2^*, remain unconstrained. Note that the means μ_i are not transformed but for notational convenience, we also define $\mu_i^* = \mu_i$. The transformed parameters are summarized into the vector $\Phi^* = (\Lambda^{*\prime}, \Omega_1^{*\prime}, \ldots, \Omega_K^{*\prime}, \Theta^{*\prime})'$, where $\Lambda^*, \Omega_1^*, \ldots, \Omega_K^*$ and Θ^* denote the transformed block vectors. The prior distributions on the transformed parameters are then obtained as the transformation of the prior distributions on the parameter vector Φ.

Now, we construct an MCMC algorithm to sample from the joint posterior distribution. As noted by Vrontos, Dellaportas and Politis (2003*a*, 2003*b*), the convergence of this type of algorithms may be accelerated by updating the highly correlated parameters simultaneously using a blocking sampling approach. Thus, we define the following algorithm scheme whose main steps are elaborated below.

1. Set $n = 0$ and initial values $\Phi^{*(0)} = \left(\Lambda^{*(0)\prime}, \Omega_1^{*(0)\prime}, \ldots, \Omega_K^{*(0)\prime}, \Theta^{*(0)\prime}\right)'$.
2. Update the allocations by sampling from $z^{(n+1)} \sim z|y, \Phi^{*(n)}$.
3. Update the innovation distribution by sampling from $\Lambda^{*(n+1)} \sim \Lambda^*|\Omega_1^{*(n)}, \ldots, \Omega_K^{*(n)}, \Theta^{*(n)}, y, z^{(n+1)}$.
4. For $i = 1, \ldots, K$, update the variance equations by sampling from,

$$\Omega_i^{*(n+1)} \sim \Omega_i^*|\Lambda^{*(n+1)}, \Omega_1^{*(n+1)}, \ldots, \Omega_{i-1}^{*(n+1)}, \Omega_{i+1}^{*(n)}, \ldots, \Omega_K^{*(n)}, \Theta^{*(n)}, y, z^{(n+1)}.$$

5. Update the correlation equations by sampling from $\Theta^{*(n+1)} \sim \Theta^*|\Lambda^{*(n+1)}, \Omega_1^{*(n+1)}, \ldots, \Omega_K^{*(n+1)}, y, z^{(n+1)}$.
6. Define $n = n + 1$ and go to 2, until $n = N$, for a large N.

In step 2, we sample from the conditional posterior probabilities that each multivariate return, y_t, for $t = 1, \ldots, T$, has been generated from the first or the second component of the mixture which are given by,

$$p\left(z_t = 1 \mid y, \Phi^*\right) = \frac{\exp\left(\rho^*\right) \exp\left\{-\frac{e_t' R_t^{-1} e_t}{2\sigma^2}\right\}}{\exp\left(\rho^*\right) \exp\left\{-\frac{e_t' R_t^{-1} e_t}{2\sigma^2}\right\} + \left(\frac{\exp(\lambda^*)}{1+\exp(\lambda^*)}\right)^{K/2} \exp\left\{-\frac{\exp(\lambda^*)}{1+\exp(\lambda^*)} \frac{e_t' R_t^{-1} e_t}{2\sigma^2}\right\}},$$

and $p\left(z_t = 2|y, \Phi^*\right) = 1 - p\left(z_t = 1|y, \Phi^*\right)$, respectively.

In step 3, we sample from the conditional posterior probability of Λ^* whose kernel is given by,

$$\kappa\left(\Lambda^* \mid \Omega_1^*, \ldots, \Omega_K^*, \Theta, y, z\right) = p\left(\Lambda^*\right) \times \frac{\left(\sigma^2\right)^{-TK/2} \exp\left(\rho^*\right)^{T_1}}{\left(1 + \exp\left(\rho^*\right)\right)^T} \left(\frac{\exp\left(\lambda^*\right)}{1 + \exp\left(\lambda^*\right)}\right)^{KT_2/2} \times \exp\left(-\frac{S_1 + \frac{\exp(\lambda^*)}{1+\exp(\lambda^*)} S_2}{2\sigma^2}\right). \tag{5}$$

where $p(\Lambda^*)$ is the prior probability of Λ^*, i.e., the product of the individual priors of ρ^* and λ^*. In order to do that, we make use of the Random Walk Metropolis Hastings (RWMH) method, see e.g. Robert and Casella (2004), using the following steps,

3.1. Generate a candidate vector $\widetilde{\Lambda}^*$ from a multivariate normal distribution $N\left(\Lambda^{*(n)}, c\widehat{\Sigma}_{\Lambda^*}\right)$ where c is a constant and $\widehat{\Sigma}_{\Lambda^*}$ is the covariance matrix of the MLE of Λ^*. Let,

$$\tau_{\Lambda^*}^{(n)} = \min\left\{1, \frac{\kappa\left(\widetilde{\Lambda}^* \mid \Omega_1^{*(n)}, \ldots, \Omega_K^{*(n)}, \Theta^{*(n)}, y, z^{(n+1)}\right)}{\kappa\left(\Lambda^{*(n)} \mid \Omega_1^{*(n)}, \ldots, \Omega_K^{*(n)}, \Theta^{*(n)}, y, z^{(n+1)}\right)}\right\},$$

where $\kappa\left(\widetilde{\Lambda}^* \mid \Omega_1^{*(n)}, \ldots, \Omega_K^{*(n)}, \Theta^{*(n)}, y, z^{(n+1)}\right)$ is given in (5).

3.2. Define,

$$\Lambda^{*(n+1)} = \begin{cases} \widetilde{\Lambda}^*, & \text{with probability } \tau_{\mu^*}^{(n)}, \\ \Lambda^{*(n)}, & \text{with probability } 1 - \tau_{\mu^*}^{(n)}. \end{cases}$$

The constant c is taken by tuning the acceptance rate to achieve fast convergence. Usually, an acceptance rate lying between 0.2 to 0.5 is plausible and practical for good convergence. We have found that $c = 0.9$ works well in practice.

In step 4, we sample from the conditional posterior distribution of Ω_i^* whose kernel is given by,

$$\kappa\left(\Omega_i^* \mid \Lambda^*, \Omega_1^*, \ldots, \Omega_{i-1}^*, \Omega_{i+1}^*, \ldots, \Omega_K^*, \Theta, y, z\right) = p\left(\Omega_i^*\right) \times \prod_{t=1}^{T}\left[H_{iit}^{-1/2}\left|R_t\right|^{-1/2}\right] \times \exp\left(-\frac{S_1 + \frac{\exp(\lambda^*)}{1+\exp(\lambda^*)}S_2}{2\sigma^2}\right),$$

where $p\left(\Omega_i^*\right)$ is the prior probability of Ω_i^*, which can be performed using a similar RWMH as the described in step 3.

Finally, in step 5, we sample from the conditional posterior distribution of Θ^* whose kernel is given by,

$$\kappa\left(\Theta^* \mid \Lambda^*, \Omega_1^*, \ldots, \Omega_K^*, y, z\right) = p\left(\Theta^*\right) \times \prod_{t=1}^{T}\left[\left(\prod_{i=1}^{K} H_{iit}^{-1/2}\right)\left|R_t\right|^{-1/2}\right] \times \exp\left(-\frac{S_1 + \frac{\exp(\lambda^*)}{1+\exp(\lambda^*)}S_2}{2\sigma^2}\right),$$

where $p\left(\Theta^*\right)$ is the prior probability of Θ^*, using also an analogous RWMH algorithm to the described in step 3.

Besides of making inference on the parameters of the GMDCC model, we may use the Markov chain to estimate in-sample volatilities and correlations and to predict future volatilities and correlations. First, a sample from the posterior distribution of each conditional variance, H_{iit}, for $i = 1, \ldots, K$ and $t = 1, \ldots, T$, can be obtained by calculating the value of each conditional variance for each draw, $\Phi^{(n)}$, which is denoted

by $H_{iit}^{(n)}$, for $n = s+1, \ldots, T$. Then, the posterior expected value of H_{iit}, $E\left[H_{iit} \mid y\right]$, can be approached by the mean of the posterior sample of conditional variances, i.e.:

$$\frac{1}{N-s} \sum_{n=s+1}^{N} H_{iit}^{(n)}.$$

Note that we can also approximate the posterior median of H_{iit} using the median of the posterior sample of conditional variances. Finally, 95% credible intervals can be obtained by just calculating .025 and .975 quantiles of each posterior sample, respectively. Similarly, we can estimate in-sample correlations R_{ijt}, using the draws $R_{ijt}^{(n)}$.

Next, prediction of future volatilities, correlations and returns, are of particular interest. Assuming that the parameters are known, then $H_{ii,T+1}$ and $R_{ij,T+1}$ are also known. Thus, given the MCMC outputs, a sample from the predictive distribution of $H_{ii,T+1}$ and $R_{ij,T+1}$ and 95% predictive intervals can be obtained similarly to the case of in-sample estimation. On the other hand, the predictive density of y_{T+1} is given by,

$$p\left(y_{T+1}|y\right) = \int_{\Phi} p\left(y_{T+1}|y, \Phi\right) p\left(\Phi|y\right) d\Phi, \tag{6}$$

where the distribution of $y_{T+1}|y, \Phi$ is a mixture of two multivariate Gaussian distributions with common mean μ and covariances $\sigma^2 H_{T+1}$ y $\left(\sigma^2/\lambda\right) H_{T+1}$, respectively. Thus, the predictive density $p\left(y_{T+1}|y\right)$ in (6) can be estimated as the mean of the density functions obtained for the draws of the MCMC samples:

$$\frac{1}{N-s} \sum_{n=s+1}^{N} p\left(y_{T+1}|y, \Phi^{(n)}\right),$$

from which we can obtain point predictions and predictive intervals.

The prediction of $H_{ii,T+m}$, $R_{ij,T+m}$ and y_{T+m} when $m > 1$ is more complicated because the values of y_t are unknown for $t \geq T+1$. However, we can also generate samples from the predictive distributions of $H_{ii,T+m}$, $R_{ij,T+m}$ and y_{T+m}, when $m > 1$, using the following sequential procedure, which is an extension of the proposed in Ausín and Galeano (2007) for single returns. For each MCMC output, $\Phi^{(n)}$, $H_{ii,T+1}^{(n)}$ and $R_{ij,T+1}^{(n)}$ are known, and thus, samples $y_{T+1}^{(n)}$ can be generated from the mixture distribution of $y_{T+1}|y, \Phi^{(n)}$. These samples gives two set of samples from the predictive distribution of $H_{ii,T+2}$ and $R_{ij,T+2}$, denoted by $H_{ii,T+2}^{(n)}$ and $R_{ij,T+2}^{(n)}$. Now, the predictive distribution of y_{T+2} is unknown but we can generate samples from the Gaussian mixture distribution with parameters $\rho^{(n)}$ and $\lambda^{(n)}$, $\epsilon_{T+2}^{(n)}$, and define,

$$y_{T+2}^{(n)} = \mu^{(n)} + H_{T+2}^{(n)\frac{1}{2}} \epsilon_{T+2}^{(n)}.$$

By replicating this scheme, given $y_{T+m-1}^{(n)}$, $H_{ii,T+m}^{(n)}$ and $R_{ij,T+m}^{(n)}$, the samples $y_{T+m}^{(n)}$ can be generated. Then, samples from the predictive distributions, $p(H_{ii,T+m}|y)$, $p(R_{ij,T+m}|y)$ and $p(y_{T+m}|y)$, are obtained, which allow us to obtain point estimates and predictive intervals.

4 Value at Risk (VaR) calculation and portfolio selection

In this section, we take advantage of the MCMC outputs in order to estimate several quantities of interest for portfolio management. In particular, we deal with VaR calculation and optimal portfolio selection.

Given a vector return series $y_t = (y_{1t}, \ldots, y_{Kt})'$, a portfolio of the components of y_t is defined as a linear combination of the individual returns, i.e., a portfolio p_t is given by $p_t = \delta' y_t$, where the weights $\delta = (\delta_1, \ldots, \delta_K)'$ add to 1. The VaR is the maximum potential loss expected with probability $1 - \pi$, where π is supposed to be small, for instance, 0.01 or 0.05. As the losses may exceed VaR with small probability, it can be thought as the worst case outcome of the portfolio performance. Thus, the one step ahead VaR, denoted by VaR_{T+1}, is defined as the 100π-th quantile of the distribution of the portfolio return, i.e.,

$$\Pr\left(p_{T+1} \leq \text{VaR}_{T+1}\right) = \pi.$$

Assuming that the model parameters are known, the conditional distribution of the portfolio is,

$$p_{T+1} \sim \begin{cases} N\left(\delta'\mu, \sigma^2 \delta' H_{T+1} \delta\right), & \text{with probability } \rho, \\ N\left(\delta'\mu, \frac{\sigma^2}{\lambda} \delta' H_{T+1} \delta\right), & \text{with probability } 1-\rho. \end{cases} \tag{7}$$

Thus, if the model parameters, Φ, are known, VaR_{T+1} is the $100\pi-$th quantile of this univariate mixture, which may be easily obtained, for instance, using the Newton-Raphson method. Consequently, using the posterior sample, $\Phi^{(n)}$, a consistent estimator of the posterior mean of the one step ahead VaR_{T+1}, $E\left[\text{VaR}_{T+1} \mid y\right]$, is given by the sample mean of the posterior sample of VaR_{T+1}:

$$\frac{1}{N-s} \sum_{n=s+1}^{N} \text{VaR}_{T+1}^{(n)}, \tag{8}$$

where $\text{VaR}_{T+1}^{(n)}$ is the one-step-ahead VaR obtained for each value $\Phi^{(n)}$, $n = s+1, \ldots, N$, of the MCMC output. We can also obtain predictive intervals for VaR_{T+1} using the quantiles of the posterior sample, $\text{VaR}_{T+1}^{(n)}$, for $n = s+1, \ldots, N$.

The m step ahead VaR, denoted by VaR_{T+m}, is defined as the $100\pi-$th quantile of the conditional

distribution of the sum of the portfolio returns,

$$\Pr\left(p_{T+1} + \cdots + p_{T+m} \leq \text{VaR}_{T+m}\right) = \pi.$$

The calculation of VaR_{T+m} when $m > 1$ is more complicated because the values of y_t are unknown for $t > T$. However, we can also generate values of the predictive distributions of VaR_{T+m}, when $m > 1$, using the samples $y_{T+i}^{(n)}$, $i = 1, \ldots, m$, generated to predict y_{T+i}. With these samples, we obtain samples of the predictive distribution of $p_{T+1} + \cdots + p_{T+m}$, from which we can also obtain point predictions and predictive intervals for VaR_{T+m}.

On the other hand, another important issue of financial management is the optimal portfolio selection problem. As mentioned in the introduction, among all the possibilities of defining what an optimal portfolio is, we assume that a portfolio is optimal if it has minimum variance, i.e., minimum risk. The problem of determining the one step ahead optimal portfolio can be formulated as the solution of the following problem,

$$\delta_{opt} = \arg\min_{\delta} \left\{\delta' H_{T+1} \delta : \delta' 1_K = 1\right\},$$

where $1_K = (1, \ldots, 1)'$. It is well known that the solution of this problem is given by,

$$\delta_{opt} = \frac{1}{1_K' H_{T+1}^{-1} 1_K} H_{T+1}^{-1} 1_K. \tag{9}$$

Note that δ_{opt} depends on the inverse of H_{T+1}, but this matrix is positive definite by construction. Then, the conditional standard deviation of the optimal portfolio at time $T+1$, $p_{opt} = \delta_{opt}' y_{T+1}$, is given by $h_{opt}^{1/2} = \left(\delta_{opt}' H_{T+1} \delta_{opt}\right)^{1/2}$, while the expected gain for the optimal portfolio is given by $g_{opt} = \delta_{opt}' \mu$. From the classical point of view, μ and H_{T+1} are replaced in the expression for δ_{opt}, $h_{opt}^{1/2}$ and g_{opt} by their maximum likelihood estimates and the uncertainty due to the parameter estimates is ignored. The Bayesian approach allows for inclusion of parameter uncertainty through the predictive distributions of μ and H_{T+1} as follows. Given the samples of the posterior distribution of μ and the samples of the predictive distribution of H_{T+1}, $\mu^{(n)}$ and $H_{T+1}^{(n)}$, for $n = s+1, \ldots, N$, respectively, we can obtain: (i) a sample of the posterior distribution of the optimal weights, denoted by $\delta_{opt}^{(n)}$, replacing H_{T+1} by $H_{T+1}^{(n)}$ in the equation (9); (ii) a sample of the posterior distribution of the optimal standard deviation, denoted by $h_{opt}^{(n)1/2}$, replacing H_{T+1} by $H_{T+1}^{(n)}$ in the expression for $h_{opt}^{1/2}$; and (iii) a sample of the posterior distribution of the expected gain, denoted by $g_{opt}^{(n)}$, replacing μ by $\mu^{(n)}$ in the expression for g_{opt}. Finally, using these samples, we can obtain consistent estimators of the posterior mean for the weights, standard deviation and gain of the optimal portfolio using

the following means of the posterior samples:

$$\frac{1}{N-s}\sum_{n=s+1}^{N}\delta_{opt}^{(n)}, \qquad \frac{1}{N-s}\sum_{n=s+1}^{N}h_{opt}^{(n)1/2}, \qquad \text{and} \qquad \frac{1}{N-s}\sum_{n=s+1}^{N}g_{opt}^{(n)}.$$

Also, 95% credible intervals can be obtained by just calculating .025 and .975 quantiles of these posterior samples.

5 Computational issues

In this section, we illustrate some of the examples that we have performed to examine our proposed procedure. We consider three bivariate series, with sample sizes $T = 1000$, 2000 and 3000, simulated from model (2) with: (i) individual conditional variances,

$$\begin{aligned} H_{11t} &= 8\times10^{-7} + 0.15\left(y_{1,t-1} - 9\times10^{-5}\right)^2 + 0.8H_{11,t-1}, \\ H_{22t} &= 8\times10^{-7} + 0.1\left(y_{2,t-1} - 1\times10^{-3}\right)^2 + 0.85H_{22,t-1}, \end{aligned}$$

respectively, so that the univariate GARCH models have parameters $\Omega_1 = (9\times10^{-5}, 8\times10^{-7}, 0.15, 0.8)$ and $\Omega_2 = (10^{-3}, 8\times10^{-7}, 0.1, 0.85)$; (ii) dynamic conditional correlation,

$$R_t = (1 - 0.6 - 0.2)\times 0.5 + 0.6R_{t-1} + 0.2\Psi_{t-1},$$

so that $\Theta = (0.6, 0.2, 0.5)$; and, (iii) a Gaussian mixture distribution (4) with parameters $\Lambda = (0.90, 0.15)$, for the innovation process.

The proposed MCMC algorithm is run for each simulated series using 20000 iterations and the initial 10000 ones are discarded for inference as burn-in iterations. We consider the classical MLE of the parameters as the initial values and use the block-sampling approach, described in Section 2, based on multivariate normal proposal distributions whose covariance matrices are the ones of the MLE. The algorithm is programmed in MATLAB (The MathWorks, Inc.) using the internal Gaussian and uniform random number generators. The MCMC chains present a good mixing performance and fast convergence as illustrated in Figure 2, where the traces and histograms of the posterior samples for each model parameter of the first simulated series with sample size $T = 1000$ are shown. Similar trace plots are obtained for the other simulated series which are not shown here to save space. Observe that the algorithm captures the asymmetry of the posterior distributions of the parameters ω_i, α_i and β_i, for $i = 1, 2$. We have also observed that, as expected, these

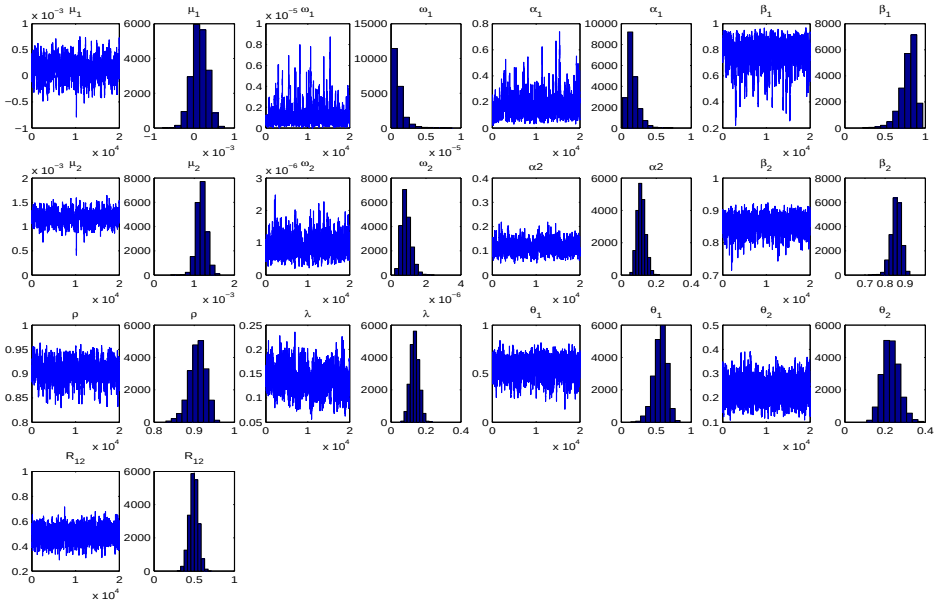

Figure 2: Trace plots and histograms of the posterior MCMC samples for each model parameter corresponding to the first simulated series with $T = 1000$.

posterior distributions become more symmetric for $T = 2000$ and $T = 3000$ (not reported).

Table 1 shows the posterior means and standard deviations of the model parameters for the three experimental setups. Observe the accuracy of the estimations and that, as expected, the posterior standard deviations become smaller as the sample size, T, increases. These Bayesian estimates are compared with the classical MLE leading to very similar results. Observe that, besides of providing point estimates and standard errors, the Bayesian estimation produces posterior densities, as the given in Figure 2, which describe all their uncertainty associated with the model parameters. Moreover, this uncertainty may be introduced in the estimation of volatilities, correlations, VaR, portfolio selection, etc., as it is shown below.

Using the MCMC output, we can estimate in-sample volatilities and correlations as described in Section 3. This is illustrated in Figure 3, where the estimated in-sample volatilities, H_{iit}, for $i = 1, 2$, and the estimated in-sample correlations, R_{12t}, for $t = 1900, \ldots, 2000$, for the second simulated series with sample size $T = 2000$ are presented. Also shown are 95% credible intervals and true values. Observe the accuracy of the estimations and that the Bayesian credible intervals always include the true values of H_{iit} and R_{12t} for all time periods. The same pattern is observed for the estimation of in-sample volatilities and correlations for the remaining simulated series (not reported). Note that, using the MCMC output, we can also make Bayesian predictions for future volatilities and correlations as described in Section 3. In particular, Figure 3 also shows the predictive means and intervals for the one-step ahead volatilities, $H_{ii,T+1}$, for $i = 1, 2$, and

Table 1: Posterior means and standard deviations of the model parameters compared with the maximum likelihood estimations and standard errors for the three simulated series.

	$T = 1000$		$T = 2000$		$T = 3000$	
Parameter	Post. Mean std	MLE std	Post. Mean std	MLE std	Post. Mean std	MLE std
$\rho = 0.9$	0.9079 0.0210	0.9044 0.0212	0.9155 0.0138	0.9106 0.0149	0.9187 0.0124	0.9156 0.0116
$\lambda = 0.15$	0.1305 0.0247	0.1556 0.0244	0.1382 0.0211	0.1591 0.0185	0.1416 0.0187	0.1546 0.0146
$\mu_1 = 9 \times 10^{-5}$	1.27×10^{-4} 1.95×10^{-4}	1.27×10^{-4} 7.35×10^{-5}	9.91×10^{-5} 1.33×10^{-4}	1.00×10^{-4} 4.95×10^{-5}	8.01×10^{-5} 1.12×10^{-4}	8.78×10^{-5} 3.90×10^{-5}
$\omega_1 = 8 \times 10^{-7}$	1.01×10^{-6} 9.16×10^{-7}	6.96×10^{-7} 2.10×10^{-7}	9.67×10^{-7} 6.09×10^{-7}	8.28×10^{-7} 1.60×10^{-7}	8.42×10^{-7} 4.29×10^{-7}	7.44×10^{-7} 1.24×10^{-7}
$\alpha_1 = 0.15$	0.1675 0.0862	0.1153 0.0239	0.1776 0.0599	0.1469 0.0202	0.1557 0.0469	0.1348 0.0162
$\beta_1 = 0.80$	0.7892 0.0977	0.8302 0.0337	0.7742 0.0751	0.7924 0.0260	0.7847 0.0622	0.8010 0.0222
$\mu_2 = 1 \times 10^{-3}$	1.18×10^{-3} 1.30×10^{-4}	1.17×10^{-3} 9.53×10^{-5}	1.02×10^{-3} 8.38×10^{-5}	1.03×10^{-3} 5.67×10^{-5}	1.02×10^{-3} 6.89×10^{-5}	1.02×10^{-3} 4.67×10^{-5}
$\omega_2 = 8 \times 10^{-7}$	8.82×10^{-7} 2.86×10^{-7}	7.25×10^{-7} 1.95×10^{-7}	7.06×10^{-7} 1.75×10^{-7}	6.14×10^{-7} 1.25×10^{-7}	7.13×10^{-7} 1.42×10^{-7}	6.60×10^{-7} 1.10×10^{-7}
$\alpha_2 = 0.10$	0.1160 0.0239	0.0973 0.0181	0.1039 0.0196	0.0921 0.0130	0.0979 0.0154	0.0907 0.0109
$\beta_2 = 0.85$	0.8559 0.0244	0.8678 0.0200	0.8586 0.0205	0.8675 0.0156	0.8566 0.0181	0.8623 0.0143
$\theta_1 = 0.60$	0.5565 0.0980	0.5791 0.0778	0.6199 0.0528	0.6305 0.0441	0.6281 0.0396	0.6322 0.0388
$\theta_2 = 0.20$	0.2279 0.0417	0.2225 0.0377	0.2155 0.0264	0.2124 0.0237	0.1984 0.0192	0.1982 0.0191
$R_{12} = 0.50$	0.4988 0.0544	0.4990 0.0517	0.5373 0.0382	0.5375 0.0379	0.5213 0.0302	0.5236 0.0291

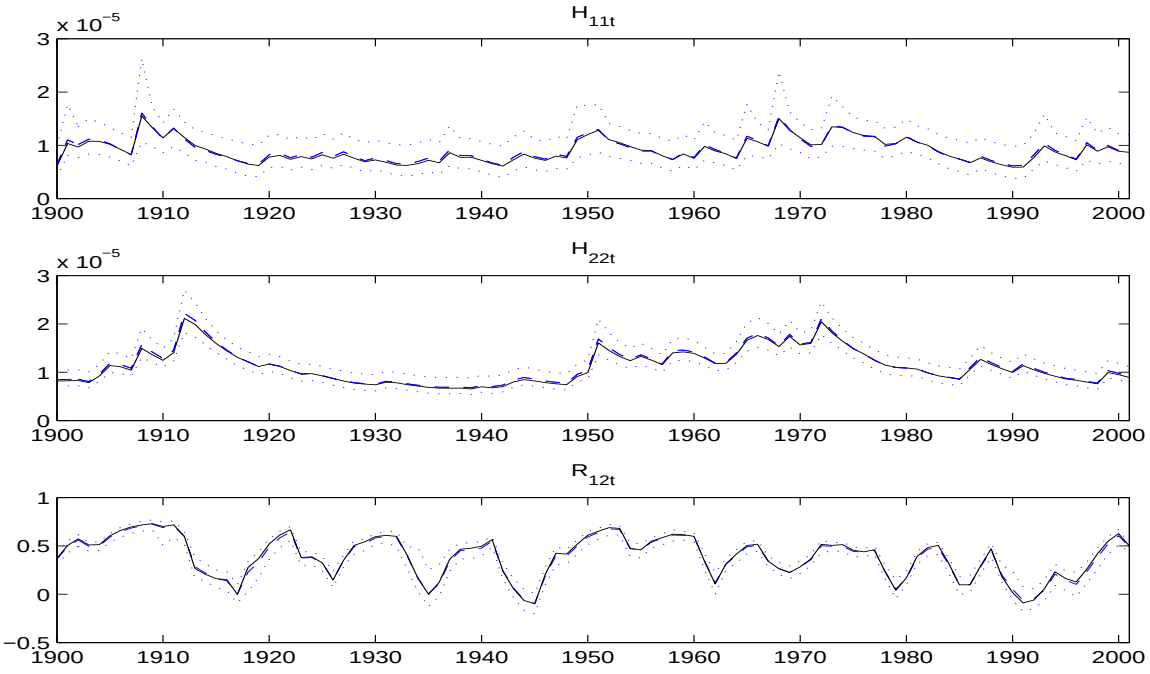

Figure 3: Estimated (dashed) and true (continuous) volatilities, H_{iit}, for $i = 1, 2$, and correlations, R_{12t}, for $t = 1900, \ldots, 2001$, joint with 95% credible intervals (dotted lines) for the second simulated series with sample size $T = 2000$.

Table 2: True values, Bayesian point estimations, MLE, and 95% predictive intervals for the VaR_{T+1} of the portfolio $p_{T+1} = \delta \times y_{1,T+1} + (1-\delta) \times y_{2,T+1}$, for the three simulated series and different values for π and δ.

			$T = 1000$	$T = 2000$	$T = 3000$
$\pi = 0.05$	$\delta = 0.25$	True	-0.00584	-0.00352	-0.00357
		Pred. Mean	-0.00595	-0.00353	-0.00352
		95% interval	$[-0.00661, -0.00527]$	$[-0.00397, -0.00315]$	$[-0.00378, -0.00330]$
		MLE	-0.00583	-0.00350	-0.00350
	$\delta = 0.50$	True	-0.00533	-0.00356	-0.00306
		Pred. Mean	-0.00528	-0.00357	-0.00300
		95% interval	$[-0.00601, -0.00452]$	$[-0.00407, -0.00310]$	$[-0.00332, -0.00275]$
		MLE	-0.00510	-0.00353	-0.00298
	$\delta = 0.75$	True	-0.00612	-0.00393	-0.00297
		Pred. Mean	-0.00601	-0.00392	-0.00290
		95% interval	$[-0.00685, -0.00500]$	$[-0.00445, -0.00339]$	$[-0.00338, -0.00252]$
		MLE	-0.00575	-0.00388	-0.00289
$\pi = 0.01$	$\delta = 0.25$	True	-0.01101	-0.00688	-0.00697
		Pred. Mean	-0.01173	-0.00686	-0.00670
		95% interval	$[-0.01424, -0.00954]$	$[-0.00815, -0.00578]$	$[-0.00773, -0.00596]$
		MLE	-0.01085	-0.00655	-0.00652
	$\delta = 0.50$	True	-0.00994	-0.00678	-0.00588
		Pred. Mean	-0.01027	-0.00675	-0.00563
		95% interval	$[-0.01246, -0.00822]$	$[-0.00809, -0.00561]$	$[-0.00663, -0.00491]$
		MLE	-0.00939	-0.00644	-0.00548
	$\delta = 0.75$	True	-0.01117	-0.00725	-0.00555
		Pred. Mean	-0.01139	-0.00721	-0.00528
		95% interval	$[-0.01376, -0.00892]$	$[-0.00867, -0.00596]$	$[-0.00646, -0.00444]$
		MLE	-0.01032	-0.00688	-0.00514

correlation, $R_{12,T+1}$, where $T + 1 = 2001$, which are compared with their true values.

Next, we apply the procedures described in Section 3 for VaR calculation and portfolio selection. As an illustration, Table 2 shows the Bayesian point estimations and 95% predictive intervals for the one step ahead VaR_{T+1} of portfolios of the form, $p_{T+1} = \delta \times y_{1,T+1} + (1-\delta) \times y_{2,T+1}$, for the three simulated series, two different values for π, $\pi = .05$ and $\pi = .01$, and three different values for δ, $\delta = .25$, $\delta = .5$ and $\delta = .75$. These are compared with the maximum likelihood estimates and true values of VaR_{T+1}, which are obtained using the MLE and true values, respectively, of H_{T+1} and of the model parameters, Φ, and calculating (by Newton-Raphson) the quantile of a two-component normal mixture with mean $\delta'\mu$ and covariances $\sigma^2\delta' H_{T+1}\delta$ and $\left(\sigma^2/\lambda\right)\delta' H_{T+1}\delta$, respectively, as given in (7). Observe the accuracy of the Bayesian estimations and that the Bayesian credible intervals always include the true VaR_{T+1} values in all cases.

Figure 4 shows the histograms of the posterior samples of the first return weight, $\delta_{opt,1}$, standard deviation, $h_{opt}^{1/2}$, and gain, g_{opt}, of the optimal portfolio, $\delta_{opt,1} \times y_{1,T+1} + (1 - \delta_{opt,1})\, y_{2,T+1}$, for the third simulated series with sample size $T = 3000$. The posterior means and confidence intervals obtained from these samples

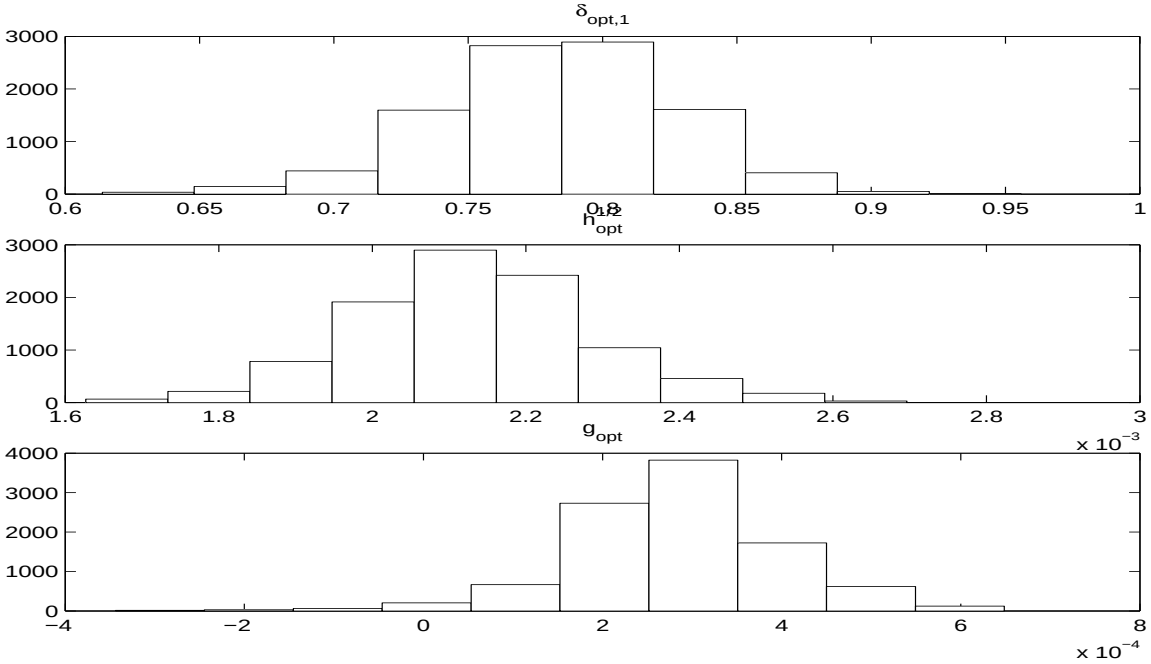

Figure 4: Histograms of the posterior samples of the weight, $\delta_{opt,1}$, standard deviation, $h_{opt}^{1/2}$, and gain, g_{opt}, of the optimal portfolio, $\delta_{opt,1} \times y_{1,T+1} + (1 - \delta_{opt,1})\, y_{2,T+1}$, for the third simulated series with sample size $T = 3000$.

are shown in the third column of Table 3, which also shows these quantities for the remaining simulated series. These are compared with the MLE and true values leading to similar results.

6 Application

As an illustration, in this section we apply the proposed Bayesian procedure to the daily closing prices of Dow Jones Industrial Average and Nasdaq composite indices for the period 2/1/1996-29/12/2006. The log return bivariate series, whose sample size is $T = 2769$, is plotted in Figure 5. Observe that the series is affected by the presence of several extreme values. The sample means, variances and excess kurtosis of both log return series are .0317 and .0297, 1.1833 and 3.1013, and 4.178 and 4.182, respectively. The autocorrelation functions of both returns do not show any significant autocorrelations. In fact, the Ljung-Box statistics for both log returns for lags 5 and 10 are 7.987 and 13.375, and 9.480 and 13.688, respectively, with associated p-values .157 and .203, and .091 and .188, respectively. The sample correlation between both log returns is .705. The final objectives of this application are estimation of the parameters of the model, estimation of in sample volatilities and correlations, calculation of VaR for different return portfolios and selection of the minimum-variance optimal portfolio.

The GMDCC model described in Section 2 is estimated using the Bayesian procedure developed in Section

Table 3: True values, Bayesian point estimations, 95% predictive intervals and MLE for the optimal weight $\delta_{opt,1}$, standard deviation, $h_{opt}^{1/2}$ and gain g_{opt} of the optimal portfolio $\delta_{opt,1} \times y_{1,T+1} + (1 - \delta_{opt,1}) \times y_{2,T+1}$, for the three simulated series.

		$T = 1000$	$T = 2000$	$T = 3000$
$\delta_{opt,1}$	True	0.5170	0.5186	0.7709
	Pred. Mean	0.5450	0.5292	0.7828
	95% interval	$[0.4798, 0.6346]$	$[0.4206, 0.6494]$	$[0.6926, 0.8647]$
	MLE	0.5576	0.5330	0.7795
$h_{opt}^{1/2}$	True	3.90×10^{-3}	2.75×10^{-3}	2.18×10^{-3}
	Pred. Mean	3.98×10^{-3}	2.72×10^{-3}	2.13×10^{-3}
	95% interval	$[3.43 \times 10^{-3}, 4.49 \times 10^{-3}]$	$[2.47 \times 10^{-3}, 3.05 \times 10^{-3}]$	$[1.83 \times 10^{-3}, 2.46 \times 10^{-3}]$
	MLE	3.77×10^{-3}	2.69×10^{-3}	2.10×10^{-3}
g_{opt}	True	5.29×10^{-4}	5.28×10^{-4}	2.98×10^{-4}
	Pred. Mean	6.07×10^{-4}	5.35×10^{-4}	2.83×10^{-4}
	95% interval	$[2.72 \times 10^{-4}, 9.19 \times 10^{-4}]$	$[2.90 \times 10^{-4}, 7.40 \times 10^{-4}]$	$[3.82 \times 10^{-5}, 5.25 \times 10^{-4}]$
	MLE	5.91×10^{-4}	5.34×10^{-4}	2.94×10^{-4}

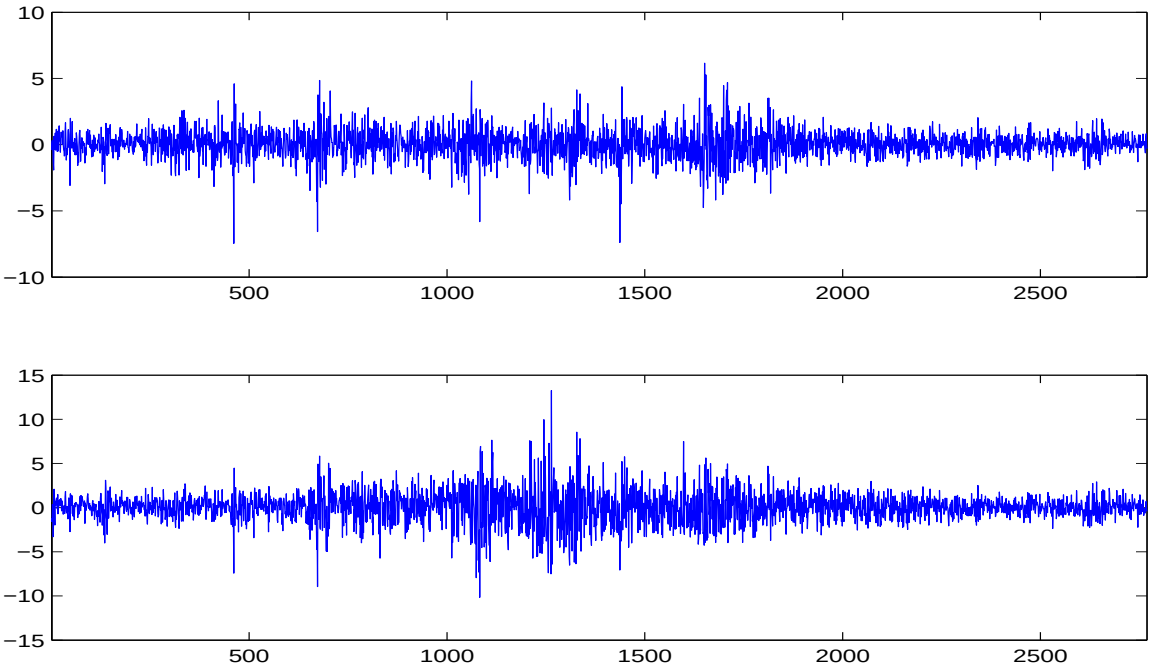

Figure 5: Log return series of the daily closing prices of Dow Jones Industrial Average and Nasdaq composite indices.

Table 4: Posterior means and standard deviations of the model parameters compared with the maximum likelihood estimations and standard errors for the Dow Jones and Nasdaq indices.

Parameter	Post. Mean	Post. std.	MLE	Std. error
ρ	0.9442	0.0265	0.9666	0.0132
λ	0.2695	0.0718	0.2375	0.0406
μ_1	0.0583	0.0499	0.0548	0.0138
ω_1	0.0066	0.0064	0.0070	0.0019
α_1	0.0622	0.0201	0.0558	0.0056
β_1	0.9345	0.0208	0.9377	0.0060
μ_2	0.0763	0.0558	0.0664	0.0196
ω_2	0.0092	0.0054	0.0089	0.0029
α_2	0.0592	0.0111	0.0560	0.0064
β_2	0.9387	0.0112	0.9408	0.0064
θ_1	0.9619	0.0075	0.9559	0.0061
θ_2	0.0244	0.0046	0.0269	0.0037
R_{12}	0.9635	0.0249	0.9286	0.0211

3. The proposed MCMC algorithm is run for this bivariate series using 20000 iterations and the first 10000 are discarded as burn-in iterations. Table 4 shows the posterior means and standard deviations obtained from the MCMC output. Observe that the Bayesian model predicts that about 94.4% of the innovations are generated by the mixture component with smaller covariance matrix, while approximately 5.6% of the innovations are generated by the component with larger covariance matrix. Also, the covariance matrix of this second component, which is designed to include the extreme events, is estimated to be approximately 4.21 times larger than the smaller covariance matrix. Table 4 also compares the Bayesian estimations with the maximum likelihood estimates and standard errors leading to similar results.

Figure 6 illustrates the Bayesian estimations together with 95% credible intervals for the volatilities, H_{iit}, for $i = 1, 2$, and correlations, R_{12t}, of the last 100 observations for the Dow Jones and Nasdaq indices. Note that the Bayesian posterior means are very close to the MLE for the time period. Observe also that the Bayesian credible intervals are not necessarily symmetric as shown for various posterior estimations of the correlations, R_{12t}. Figure 6 also includes the one-step ahead point prediction of $H_{ii,2770}$, for $i = 1, 2$, and the correlation, $R_{12,2770}$, joint with the corresponding predictive intervals. It appears that the Nasdaq index is more volatile than the Dow Jones in this last period of 2006. In fact, this effect is also observed since 1998 (not reported). Also note that the conditional correlations are very high, with values around 0.9 and has a drop in November 2006.

Figure 7 illustrates the Bayesian estimations and 95% credible intervals compared with the MLE for the one-step ahead VaR_{T+1} of the portfolio $p_{T+1} = \delta \times DowJones_{T+1} + (1 - \delta) \times Nasdaq_{T+1}$, as a function of δ, for $\delta \in (-1, 2)$, for $\pi = 0.01$. Observe that the Bayesian credible intervals always include the MLE for each

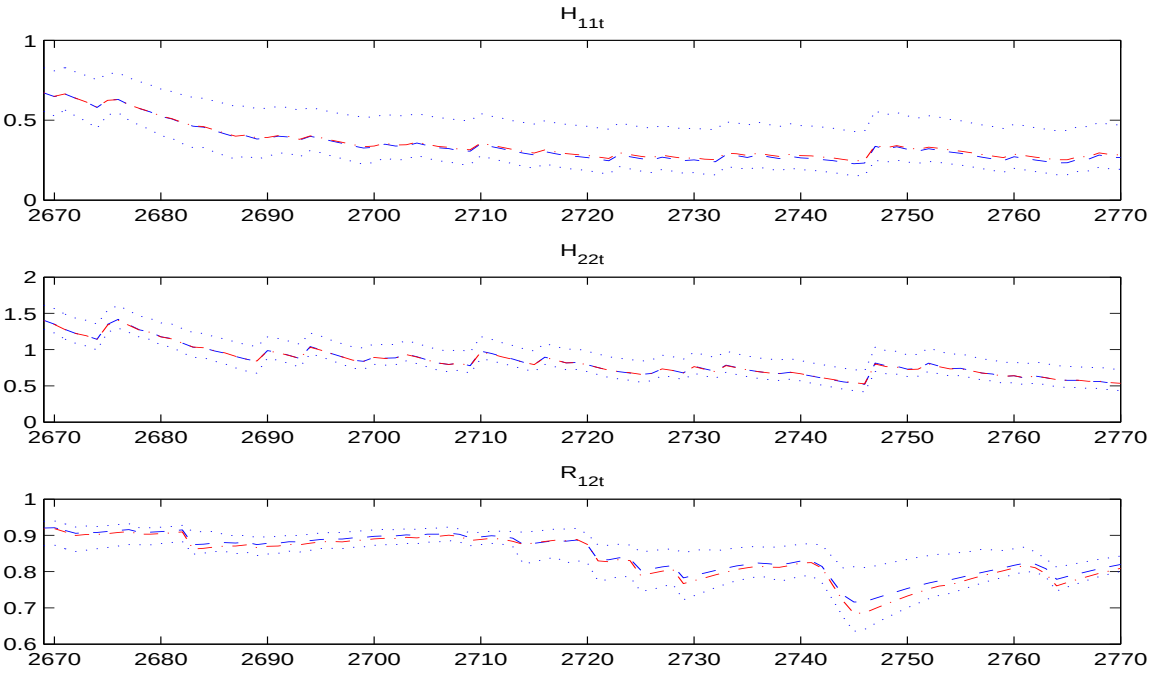

Figure 6: Bayesian estimations (dashed) and 95% intervals (dotted) compared with MLE (dashdot) for the volatilities, H_{iit}, for $i = 1, 2$, and correlations, R_{12t}, of the last 100 observations for the Dow Jones and Nasdaq indices.

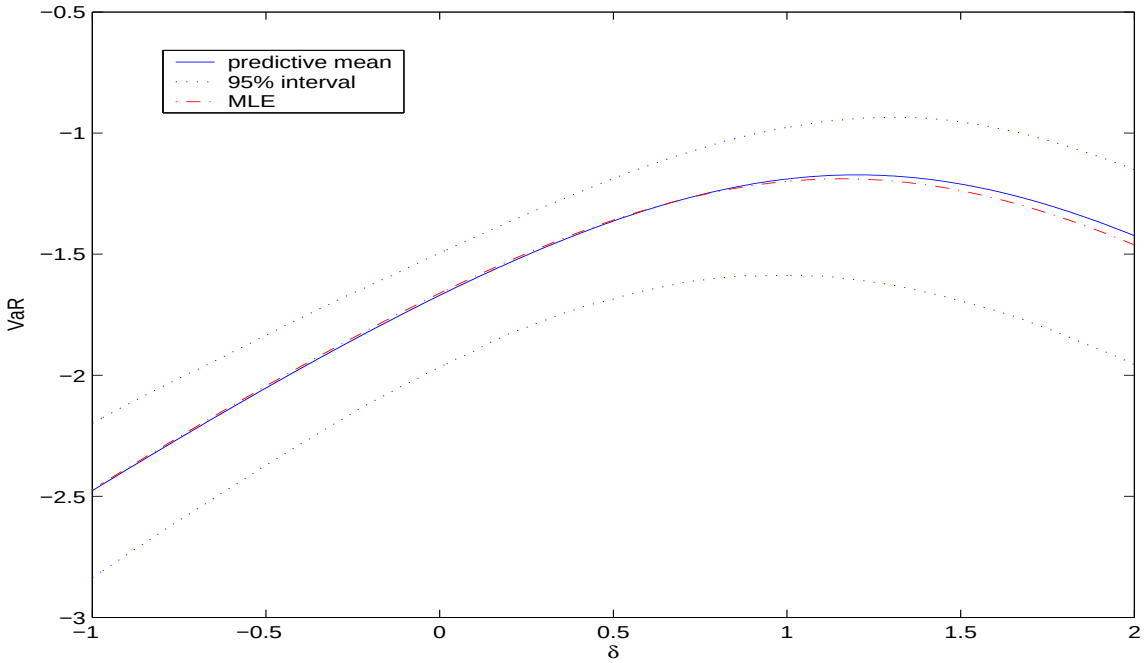

Figure 7: Bayesian estimations (dashed) and 95% intervals (dotted) compared with MLE (dashdot) for the one-step ahead VaR_{T+1} of the portfolio $p_{T+1} = \delta \times DowJones_{T+1} + (1 - \delta) \times Nasdaq_{T+1}$, as a function of δ, using $\pi = 0.01$.

Table 5: Bayesian estimation and 95% predictive interval for the weight, $\delta_{opt,1}$, standard deviation, $h_{opt}^{1/2}$, and gain, g_{opt}, of the optimal portfolio, $\delta_{opt,1} DowJones_{T+1} + (1 - \delta_{opt,1}) Nasdaq_{T+1}$, compared with the MLE.

	$\delta_{opt,1}$	$h_{opt,1}$	$g_{opt,1}$
Pred. Mean	1.2100	0.5139	0.0544
95% interval	$[0.9011, 1.3563]$	$[0.4216, 0.6836]$	$[-0.0489, 0.1384]$
MLE	1.1711	0.5255	0.0528

weight vector δ. Note also that this kind of plot can be used for portfolio selection in the case of considering that the optimal portfolio is the one which maximizes the VaR as it minimizes the maximum potential loss with probability π. In this case, the optimal portfolio is approximately $p_{T+1} = 1.2 \times DowJones_{T+1} - 0.2 \times Nasdaq_{T+1}$.

Finally, Table 5 shows the Bayesian estimation and 95% predictive interval for the weight, $\delta_{opt,1}$, standard deviation, $h_{opt}^{1/2}$, and gain, g_{opt}, of the optimal portfolio, $\delta_{opt,1} DowJones_{T+1} + (1 - \delta_{opt,1}) Nasdaq_{T+1}$. These are compared with the maximum likelihood estimations which are always included in the predictive intervals. Note that the posterior mean of the optimal weight is 1.21 which is very similar to the optimal weight obtained previously by maximizing the VaR as shown in Figure 7.

7 Conclusions

In this paper, we have proposed a multivariate GARCH model with time-varying correlations in which the innovations are assumed to follow a mixture of two multivariate zero mean Gaussian distributions. This specification extends the Gaussian mixture innovation distribution proposed by Bai, Rusell and Tiao (2003) to the multivariate framework. We have shown how to perform Bayesian inference on this model through MCMC methods, which allows us to estimate and predict conditional variances and correlations. In particular, we have proposed a Random Walk Metropolis Hastings algorithm which is straightforward to implement and has been shown to work well with both simulated and real data examples. Also, we have developed a Bayesian procedure for the derivation of predictive distributions for the portfolio VaR and a method for the determination of optimal portfolios.

Note that more complex approaches for the innovation distribution may be also entertained. For instance, we can use a Gaussian mixture with a larger number of components, but this has the cost of adding two more additional parameters for each new component, which will increase the computational cost of the MCMC algorithm. Also, we can try other distributions with fat tails, such as the multivariate generalized extreme value distribution, but this has the cost of losing the simplicity and the interpretability of the

Gaussian expressions. Finally, an alternative approach may be to allow for different univariate GARCH model parameters depending on which component of the mixture the innovation belongs to, but the same problems as the mentioned above are found. All these alternative approaches are currently under research.

Acknowledgment

We acknowledge financial support by Xunta de Galicia under the Isidro Parga Pondal Program and MEC project SEJ2004-03303. The first author also acknowledges financial support by Xunta de Galicia project PGIDIT06PXIB207009PR.

References

Ausin, M. C. and Galeano, P. (2007) Bayesian estimation of the Gaussian mixture GARCH model. Computational Statistics and Data Analysis, 51, 2636-2652.

Bai, X., Rusell, J. R., and Tiao, G. C. (2003) Kurtosis of GARCH and stochastic volatility models with non-normal innovations. Journal of Econometrics, 114, 349-360.

Baillie, R. T. and Myers, R. J. (1991) Bivariate GARCH estimation of the optimal commodity futures hedge. Journal of Applied Econometrics, 6, 109-124.

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

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307-327.

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

Bollerslev, T. and Engle, R. F. (1993) Common persistence in conditional variances. Econometrica, 61, 167-186.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988) A capital asset pricing model with time-varying covariances. Journal of Political Economy, 96, 116-131.

Box, G. E. P. and Tiao, G. C. (1968) A Bayesian approach to some outlier problems. Biometrika, 55, 119-129.

Comte, F., and Lieberman, O. (2003) Asymptotic theory for multivariate GARCH models. Journal of Multivariate Analysis, 84, 61-84.

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

Engle, R. F. (2002) Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business and Economic Statistics, 20, 339-350.

Engle, R. F. and Kroner, K. (1995) Multivariate simultaneous GARCH. Econometric Theory, 11, 122-150.

Engle, R. F., Ng, V., and Rothschild, M. (1990) Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. Journal of Econometrics, 45, 213-238.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance, 48, 1779-1801.

Hall, P. and Yao, Q. (2003) Inference in ARCH and GARCH models with heavy-tailed errors. Econometrica, 71, 285-317.

Jagganathan, R. T. and Ma, T. (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps. Journal of Finance, 58, 1651-1684.

Jaschke, S. and Jiang, Y. (2002) Approximating value at risk in conditional Gaussian models. In: W. Härdle, T. Kleinow, G. Stahl (Eds.), Applied Quantitative Finance, Springer, Berlin.

Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models. Econometric Theory, 14, 70-86.

Jorion, P. (2006) Value at Risk: The new benchmark for managing financial risk, third ed. McGraw-Hill, NewYork.

Karolyi, G. A. (1995) A multivariate GARCH model of international transmission of stock returns and volatility: the case of the United States and Canada. Journal of Business and Economic Statistics, 13, 11-25.

Kraft, D. F. and Engle, R. F. (1982) Autoregressive conditional heteroscedasticity in multiple time series models. Working paper, University of California at San Diego.

Mardia, K. V. (1970) Measures of multivariate skewness and kurtosis with applications. Biometrika, 57, 519-530.

Markowitz, H. (1952) Portfolio selection. Journal of Finance, 8, 77-91.

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

Robert, C. P. and Casella, G. (2004) Monte Carlo statistical methods, second ed. Springer, New York.

Tse, Y. K. and Tsui, A. K. C. (2002) A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. Journal of Business and Economic Statistics, 20, 351-362.

Vrontos, I. D., Dellaportas, P., and Politis, D. N. (2003*a*) Inference for some multivariate ARCH and GARCH models. Journal of Forecasting, 22, 427-446.

Vrontos, I. D., Dellaportas, P., and Politis, D. N. (2003*b*) A full-factor multivariate GARCH model. Econometrics Journal, 6, 312-334.

Zakoian, J. M. (1994) Threshold heteroskedasticity models. Journal of Economic Dynamics and Control, 15, 931-955.

Reports in Statistics and Operations Research

2004

04-01 Goodness of fit test for linear regression models with missing response data. *González Manteiga, W., Pérez González, A.*
Canadian Journal of Statistics (to appear).

04-02 Boosting for Real and Functional Samples. An Application to an Environmental Problem. *B. M. Fernández de Castro and W. González Manteiga.*

04-03 Nonparametric classification of time series: Application to the bank share prices in Spanish stock market. *Juan M. Vilar, José A. Vilar and Sonia Pértega*.

04-04 Boosting and Neural Networks for Prediction of Heteroskedatic Time Series. *J. M. Matías, M. Febrero, W. González Manteiga and J. C. Reboredo.*

04-05 Partially Linear Regression Models with Farima-Garch Errors. An Application to the Forward Exchange Market. *G. Aneiros Pérez, W. González Manteiga and J. C. Reboredo Nogueira.*

04-06 A Flexible Method to Measure Synchrony in Neuronal Firing. *C. Faes, H. Geys, G. Molenberghs, M. Aerts, C. Cadarso-Suárez, C. Acuña and M. Cano.*

04-07 Testing for factor-by-curve interactions in generalized additive models: an application to neuronal activity in the prefrontal cortex during a discrimination task. *J. Roca-Pardiñas, C. Cadarso-Suárez, V. Nacher and C. Acuña.*

04-08 Bootstrap Estimation of the Mean Squared Error of an EBLUP in Mixed Linear Models for Small Areas. *W. González Manteiga, M. J. Lombardía, I. Molina, D. Morales and L. Santamaría.*

04-09 Set estimation under convexity type assumptions. *A. Rodríguez Casal.*

2005

05-01 SiZer Map for Evaluating a Bootstrap Local Bandwidth Selector in Nonparametric Additive Models. *M. D. Martínez-Miranda, R. Raya-Miranda, W. González-Manteiga and A. González-Carmona.*

05-02 The Role of Commitment in Repeated Games. *I. García Jurado, Julio González Díaz*.

05-03 Project Games. *A. Estévez Fernández, P. Borm, H. Hamers*

05-04 Semiparametric Inference in Generalized Mixed Effects Models. *M. J. Lombardía, S. Sperlich*

2006

06-01 A unifying model for contests: effort-prize games. J. González Díaz

06-02 The Harsanyi paradox and the "right to talk" in bargaining among coalitions. J. J. Vidal Puga

06-03 A functional analysis of NOx levels: location and scale estimation and outlier detection. M. Febrero, P. Galeano, W. González-Manteiga

06-04 Comparing spatial dependence structures. R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga

06-05 On the spectral simulation of spatial dependence structures. R. M. Crujeiras, R. Fernández-Casal

06-06 An L_2-test for comparing spatial spectral densities. R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga.

2007

07-01 Goodness-of-fit tests for the spatial spectral density. R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga.

07-02 Presmothed estimation with left truncated and right censores data. M. A. Jácome, M. C. Iglesias-Pérez

07-03 Robust nonparametric estimation with missing data. G. Boente, W. González-Manteiga, A. Pérez-González

07-04 k-Sample test based on the common area of kernel density estimators, P. Martínez-Camblor, J. de Uña Álvarez, N. Corral-Blanco

07-05 A bootstrap based model checking for selection-biased data, J. L. Ojeda, W. González-Manteiga, J . A. Cristobal

07-06 The Gaussian mixture dynamic conditional correlation model: Bayesian estimation, value at risk calculation and portfolio selection, P. Galeano, M. C. Ausín

Previous issues (2001 – 2003):
http://eio.usc.es/pub/reports.html