The Gaussian Mixture Dynamic Conditional Correlation Model: Parameter Estimation, Value at Risk Calculation, and Portfolio Selection

Pedro GALEANO and M. Concepción AUSÍN
Departamento de Estadística, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain
(*pedro.galeano@uc3m.es*; *concepcion.ausin@uc3m.es*)

A multivariate generalized autoregressive conditional heteroscedasticity model with dynamic conditional correlations is proposed, in which the individual conditional volatilities follow exponential generalized autoregressive conditional heteroscedasticity models and the standardized innovations follow a mixture of Gaussian distributions. Inference on the model parameters and prediction of future volatilities are addressed by both maximum likelihood and Bayesian estimation methods. Estimation of the Value at Risk of a given portfolio and selection of optimal portfolios under the proposed specification are addressed. 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 indexes.

KEY WORDS: Bayesian inference; Gaussian mixture model; Maximum likelihood estimation; Multivariate generalized autoregressive conditional heteroscedasticity model; Portfolio selection; Value at Risk.

1. INTRODUCTION

The autoregressive conditional heteroscedastic (ARCH) model introduced by Engle (1982) and its generalization, the generalized ARCH (GARCH) model proposed by Bollerslev (1986), have become very popular in modeling financial time series, because 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, including portfolio analysis, risk management, and asset pricing. Several multivariate models have been proposed to handle multiple returns, including the VEC model of Bollerslev, Engle, and Wooldridge (1988); the BEKK model of Engle and Kroner (1995); the factor models proposed by Engle, Ng, and Rothschild (1990), Bollerslev and Engle (1993), and Vrontos, Dellaportas, and Politis (2003a); the CCC model proposed by Bollerslev (1990); and the DCC models proposed by Tse and Tsui (2002) and Engle (2002). Bauwens, Laurent, and Rombouts (2006) have provided a comprehensive survey of multivariate GARCH-type models, including their properties and applications.

Most of the proposed multivariate models have been derived to describe time-varying volatilities and correlations of several return series, but other features have received considerably less attention. For instance, it is well known that negative shocks have a greater impact on volatility than positive shocks, an empirical fact usually known as the leverage effect. In the univariate framework, several models to account for this effect have been proposed, including the exponential GARCH (EGARCH) model of Nelson (1991). But in the multivariate framework, the leverage effect has been taken into account only in some modifications of the VEC and BEKK models (see, e.g., Kroner and Ng 1998). On the other hand, 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 presence of large returns that produce long tails and high excess kurtosis. 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 to model the standardized innovations with a mixture of two zero mean Gaussian distributions. This mixture innovation specification combined with GARCH models successfully captures volatility clustering, long tails, and high excess kurtosis. Other authors who have demonstrated the usefulness of Gaussian mixture models as applied to stock returns include Perez-Quiros and Timmermann (2000, 2001), Wong and Li (2001), Haas, Mittnik, and Paolella (2004), and Bauwens, Hafner, and Rombouts (2007). The first contribution of this article is to propose the Gaussian mixture dynamic conditional correlation (GMDCC) model with conditional volatilities following EGARCH models and with a multivariate Gaussian mixture standardized innovation distribution with unknown number of components, which successfully captures the main features of multivariate financial returns.

Inference on multivariate GARCH-type models is usually carried out by maximum likelihood. Less attention has been paid to the Bayesian point of view (see Vrontos, Dellaportas, and Politis 2003a, 2003b). Bayesian inference is specially well suited for multivariate GARCH-type models, because it provides a natural way to introduce parameter uncertainty in the estimation of volatilities and correlations. In addition, predictive distributions of future volatilities and correlations can be

Journal of Business & Economic Statistics
October 2010, Vol. 28, No. 4
DOI: 10.1198/jbes.2009.07238

obtained, which are more informative than simple point forecasts derived from the maximum likelihood estimates.

The article's second contribution is to demonstrate how to perform both maximum likelihood and Bayesian inference for the proposed GMDCC model. In particular, Bayesian inference is carried out using Markov chain Monte Carlo (MCMC) methods. We show how data augmentation techniques combined with a block-sampling scheme allow for a straightforward MCMC implementation associated with good mixing performance.

Two of the most important issues in risk management are the calculation of the Value at Risk (VaR) of a given portfolio and the portfolio selection problem. The VaR is a widely used measure of the risk of a portfolio, defined as the maximum potential loss over a given time period at a certain confidence level. Statistically speaking, the VaR is the negative value of a quantile of the conditional distribution of the portfolio return; thus its calculation depends strongly on the assumption made for the standardized innovation distribution. The portfolio selection problem is defined as the determination of the optimal weights assigned to each return in the portfolio. The classical approach is the mean-variance method proposed by Markowitz (1952), which assigns the weights that minimize the variance subject to achieving different levels of expected returns. But in nonnormal settings, the variance is a very restrictive measure of risk, and minimizing the VaR seems to be a more natural approach.

The article's third contribution is to show how to handle these problems with our proposed GMDCC model. We provide point estimations of both the VaR of a given portfolio at a given significance level and the weights of the optimal portfolio under the selected choice based on both maximum likelihood and Bayesian approaches. It is important to note that besides giving point estimates, the Bayesian methodology also provides a measure of precision for both VaR and optimal weight estimates via predictive intervals.

The article is organized as follows. Section 2 presents our GMDCC model and shows its flexibility in capturing the special features of multivariate financial time series. Section 3 describes how to perform maximum likelihood and Bayesian inference for the GMDCC model, how to estimate in-sample volatilities and correlations, and how to predict future volatilities and correlations. Section 4 deals with the problems of calculating VaR and determining optimal portfolios. Section 5 presents a brief Monte Carlo experiment that demonstrates our model's accuracy in estimating parameters, predicting volatilities and correlations, calculating VaR, and determining optimal portfolios. Section 6 illustrates our proposed methodology with the Dow Jones Industrial Average and the NASDAQ composite indexes. Section 7 concludes.

2. THE GAUSSIAN MIXTURE DYNAMIC CONDITIONAL CORRELATION MODEL

Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{Kt})'$ be a K-dimensional vector of returns given by $\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t$, where $\boldsymbol{\mu}$ is the unconditionals mean of the process, $\mathbf{H}_t$ is the $K \times K$ positive definite conditional covariance matrix of $\mathbf{y}_t$ given the past information $F_{t-1} = \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots\}$, and the $\boldsymbol{\epsilon}_t$'s are iid random vector of dimension $K \times 1$ such that $E[\boldsymbol{\epsilon}_t] = 0$ and $\operatorname{cov}[\boldsymbol{\epsilon}_t] = \mathbf{I}_K$, the K-dimensional identity matrix. We assume that $\mathbf{y}_t$ is a constant mean process. If the conditional mean of the returns is nonconstant, then we assume that $\mathbf{y}_t$ is a filtered process in which effects due to past returns or other explanatory variables have been consistently eliminated in a first stage by, for example, a VARMA model. This is a usual choice in multivariate GARCH-type modeling (see, e.g., Bauwens, Laurent, and Rombouts 2006), which simplifies the exposition of our ongoing analysis. The conditional covariance matrix can be decomposed as $\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$, where $\mathbf{D}_t$ is the $K \times K$ diagonal matrix with the K conditional standard deviations, denoted by $H_{iit}^{1/2}$, and $\mathbf{R}_t$ is the $K \times K$ matrix of conditional correlations. Note that $\mathbf{H}_t$ is positive definite if and only if all of the conditional volatilities, H_{iit}, are positive and the conditional correlation matrix, $\mathbf{R}_t$, is positive definite.

To take leverage effects into account, the conditional volatility of the ith individual return series is formulated by entertaining a univariate EGARCH(p_i, q_i) model following

$$\ln(H_{iit}) = \omega_i + \sum_{l=1}^{p_i} \alpha_{i,l} \ln(H_{ii,t-l}) + \sum_{l=1}^{q_i} \beta_{i,l}\big(\delta_i \epsilon_{i,t-l} + \gamma_i\big(|\epsilon_{i,t-l}| - E[|\epsilon_{i,t-l}|]\big)\big), \quad (1)$$

where $\beta_{i,1} = 1$, for $i = 1, \ldots, K$, and the term $E[|\epsilon_{i,t-l}|]$ depends on the specification of the standardized innovation distribution (given below). The EGARCH model ensures that the volatilities H_{iit} are all positive, and thus the matrixes $\mathbf{D}_t$ are positive definite. The EGARCH(p_i, q_i) model has strict stationarity and is ergodic (see Nelson 1991) if the roots of $1 - \sum_{l=1}^{p_i} \alpha_{i,l}L^l$ lie outside the unit circle. Note that the orders (p_i, q_i) of the individual EGARCH models do not necessarily coincide, while for symmetric returns, univariate GARCH models can be considered instead of EGARCH models.

On the other hand, we adopt the specification of the conditional correlation matrix $\mathbf{R}_t$ of Tse and Tsui (2002). For that, let $\mathbf{e}_t = \mathbf{D}_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu})$ be the vector of standardized returns, let $\mathbf{E}_{t-1}$ be the $K \times K$ matrix given by $\mathbf{E}_{t-1} = [\mathbf{e}_{t-1}, \ldots, \mathbf{e}_{t-K}]$, and let $\mathbf{B}_{t-1}$ be the K-dimensional diagonal matrix where the ith element is given by $(\sum_{H=1}^{K} e_{i,t-H}^2)^{1/2}$, for $i = 1, \ldots, K$. Then $\mathbf{R}_t$ is generated from the recursion

$$\mathbf{R}_t = (1 - \theta_1 - \theta_2)\mathbf{R} + \theta_1\mathbf{R}_{t-1} + \theta_2\boldsymbol{\Psi}_{t-1},$$

where θ_1 and θ_2 are nonnegative parameters satisfying $\theta_1 + \theta_2 < 1$, $\mathbf{R}$ is a K-dimensional positive definite matrix with unit diagonal elements and off-diagonal elements denoted by R_{ij} for $i, j = 1, \ldots, K$ with $i \neq j$, and $\boldsymbol{\Psi}_{t-1}$ is a $K \times K$ matrix given by $\boldsymbol{\Psi}_{t-1} = \mathbf{B}_{t-1}^{-1}\mathbf{E}_{t-1}\mathbf{E}_{t-1}'\mathbf{B}_{t-1}^{-1}$. With this specification, the conditional correlation matrixes, $\mathbf{R}_t$, are positive definite, and thus $\mathbf{H}_t$ also is positive definite.

The common distribution for the standardized innovations, $\boldsymbol{\epsilon}_t$, is the multivariate standard Gaussian distribution, because under certain conditions, the quasi-maximum likelihood esti-

mator (QMLE) of the vector of parameters is strongly consistent (see Jeantheau 1998). However, the Gaussianity assumption is rejected for most multivariate residuals, because the number of extremes in the standardized residuals is much larger than the number of extremes that can be generated by the Gaussian distribution. In fact, the multivariate excess kurtosis of the standardized innovations assuming Gaussianity, as defined by Mardia (1970), is $EK_G[\boldsymbol{\epsilon}_t] = 0$. The usual alternative is the standardized Student distribution with ν degrees of freedom, but this also may be problematic, because the excess kurtosis of $\boldsymbol{\epsilon}_t$ exists only if $\nu > 4$, in which case it is given by

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

Note that the fourth moment of $\boldsymbol{\epsilon}_t$ exists only if $\nu > 4$, whereas the second moment of $\boldsymbol{\epsilon}_t$ exists only if $\nu > 2$. In practice, either ν is constrained 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 not constrained, in which case its estimate may be smaller than 4, which implies that the estimated excess kurtosis does not exist. This may be a serious drawback, because Hall and Yao (2003) have shown that in the univariate case, the usual asymptotic consistency and normality properties do not necessarily hold if the standardized innovations do not have moments at least smaller than or equal to four. In the univariate case, a solution to this problem was suggested by Bai, Rusell, and Tiao (2003), who proposed modeling the innovations with a mixture of two zero mean Gaussian distributions. These authors showed that (i) the mixture specification combined with GARCH models can capture the usual patterns exhibited by financial time series, such as volatility clustering and high excess kurtosis; (ii) the Gaussian mixture gives better fits than the standardized Student-t distribution; and (iii) the excess kurtosis implied by the Gaussian mixture is closer to the sample excess kurtosis than that implied by the standardized Student-t distribution. Finally, note that all of the moments of the Gaussian mixture exist, because they are combinations of the moments of Gaussian distributions.

Accordingly, our aim here is to extend this specification to the multivariate framework. For this, we assume that the innovation process, $\boldsymbol{\epsilon}_t$, follows a mixture of S zero mean Gaussian distributions, where S is unknown. Therefore, in the usual notation,

$$\boldsymbol{\epsilon}_t \sim \rho_1 \mathrm{N}(0, \sigma_1^2 \mathbf{I}_K) + \cdots + \rho_S \mathrm{N}(0, \sigma_S^2 \mathbf{I}_K), \tag{2}$$

where $\rho_1, \ldots, \rho_S$ are the weights that sum to 1 and $\rho_1\sigma_1^2 + \cdots + \rho_S\sigma_S^2 = 1$, so that $\mathrm{cov}[\boldsymbol{\epsilon}_t] = \mathbf{I}_K$. Consequently, the standardized innovation vector, $\boldsymbol{\epsilon}_t$, is generated from a Gaussian distribution with covariance $\sigma_s^2 \mathbf{I}_K$ with probability ρ_s for $s = 1, \ldots, S$. It is not difficult to show that (i) the excess kurtosis of the mixture in eq. (2) is given by

$$EK_{GM}[\boldsymbol{\epsilon}_t] = K(K+2)\left(\sum_{s=1}^{S} \rho_s \sigma_s^4 - 1\right), \tag{3}$$

which exists for every value in the domain of $\rho_1, \ldots, \rho_S$ and $\sigma_1^2, \ldots, \sigma_S^2$ and can take any possible positive value, and (ii) the term $E[|\epsilon_{i,t-l}|]$ in eq. (1), is given by

$$E[|\epsilon_{i,t-l}|] = \sqrt{\frac{2}{\pi}} \sum_{s=1}^{S} \rho_s \sigma_s.$$

It is well known that in the mixtures framework, it is convenient to reparameterize both the weights $\rho_1, \ldots, \rho_S$ and the variances $\sigma_1^2, \ldots, \sigma_S^2$ to get a more parsimonious representation of the mixture in eq. (2) (see Frühwirth-Schnatter 2006). To do this, we write

$$\rho_s = (1 - \eta_1) \cdots (1 - \eta_{s-1})\eta_s, \tag{4}$$

where $\eta_s \in (0, 1)$ for $s = 1, \ldots, S-1$ and $\eta_S = 1$, while the variances of the $(s-1)$th and sth components for $s = 2, \ldots, S$, are related by the equation $\sigma_s^2 = \sigma_{s-1}^2/\lambda_{s-1}$, where $\lambda_{s-1} \in (0, 1)$. Note that here we also can identify the components of the mixture, because these are ordered following the ascendant order of their variances. As $\rho_1\sigma_1^2 + \cdots + \rho_S\sigma_S^2 = 1$, it is not difficult to show that σ_s^2 can be written in terms of $\eta_1, \ldots, \eta_{S-1}$ and $\lambda_1, \ldots, \lambda_{S-1}$, as follows:

$$\sigma_s^2 = \frac{\prod_{u=1}^{s-1} 1/\lambda_u}{\eta_1 + \sum_{v=2}^{S} (\prod_{u=1}^{v-1} (1 - \eta_u)/\lambda_u)\eta_v} \tag{5}$$

for $s = 2, \ldots, S$ and $\sigma_1^2 = (\eta_1 + \sum_{v=2}^{S} (\prod_{u=1}^{v-1} \frac{1-\eta_u}{\lambda_u})\eta_v)^{-1}$, while the excess kurtosis in eq. (3) can be written as

$$EK_{GM}[\boldsymbol{\epsilon}_t] = K(K+2) \times \left(\frac{\eta_1 + \sum_{v=2}^{S} (\prod_{u=1}^{v-1} (1 - \eta_u)/\lambda_u^2)\eta_v}{(\eta_1 + \sum_{v=2}^{S} (\prod_{u=1}^{v-1} (1 - \eta_u)/\lambda_u)\eta_v)^2} - 1 \right),$$

which exists for every value in the domain of $\eta_1, \ldots, \eta_{S-1}$ and $\lambda_1, \ldots, \lambda_{S-1}$. In particular, Figure 1 shows the multivariate excess kurtosis of the Gaussian mixture distribution for $K = 2$ and $S = 2$, which depends only on the parameters η_1 and λ_1 and may increase as necessary for values of η_1 and λ_1 close to 1 and 0, respectively.

3. INFERENCE FOR THE GAUSSIAN MIXTURE DYNAMIC CONDITIONAL CORRELATION MODEL

In this section we conduct inference for the parameters of the proposed GMDCC model by both maximum likelihood and Bayesian methods. Maximum likelihood (ML) estimation has become the favorite approach for parameter estimation in multivariate GARCH-type models. This is obtained by maximizing the likelihood function of the model with respect to the model parameters. Assume first that the number of mixture components, S, is fixed. In this case the parameters of the GMDCC model can be summarized in the vector $\boldsymbol{\Phi} = (\boldsymbol{\Omega}_1', \ldots, \boldsymbol{\Omega}_K', \boldsymbol{\Theta}', \boldsymbol{\Lambda}')'$, where $\boldsymbol{\Omega}_i = (\mu_i, \omega_i, \alpha_{i,1}, \ldots, \alpha_{i,p_i}, \beta_{i,2}, \ldots, \beta_{i,q_i}, \delta_i, \gamma_i)'$, for $i = 1, \ldots, K$, are the parameters of the conditional volatilities of the single returns, $\boldsymbol{\Theta} = (\theta_1, \theta_2, R_{12}, \ldots, R_{K-1,K})'$, are the parameters of the conditional correlation matrix, and $\boldsymbol{\Lambda} = (\eta_1, \ldots, \eta_{S-1}, \lambda_1, \ldots, \lambda_{S-1})'$ are the parameters of the Gaussian mixture. Assuming the Gaussian

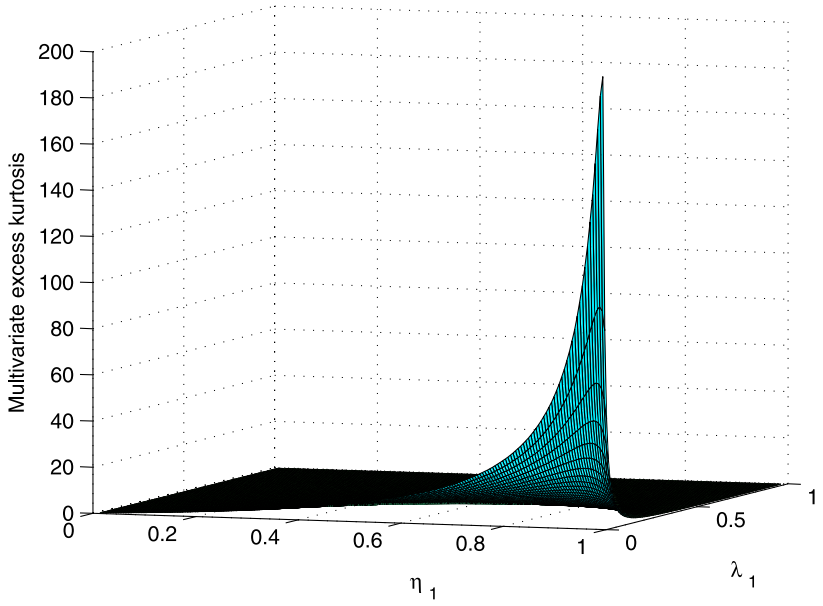

Figure 1. Multivariate excess kurtosis of the Gaussian mixture distributions for $K = 2$ and $S = 2$. The online version of this figure is in color.

mixture specification for the standardized innovation distribution, the likelihood function of the model for the observed return series $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$ is given by

$$L(\mathbf{\Phi}|\mathbf{y}) = \prod_{t=1}^{T}\left(\sum_{s=1}^{S}\rho_s p(\mathbf{y}_t|\mathbf{\Phi}_s)\right), \tag{6}$$

where $p(\mathbf{y}_t|\mathbf{\Phi}_s)$ is the probability density function of the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\sigma_s^2\mathbf{H}_t$, and ρ_s and σ_s^2 are given in terms of $\mathbf{\Lambda}$ in eqs. (4) and (5), respectively. In practice, instead of maximizing the likelihood in eq. (6), we obtain the ML estimate of $\mathbf{\Phi}$, denoted by $\widehat{\mathbf{\Phi}}$, by maximizing the logarithm of the likelihood function,

$$l(\mathbf{\Phi}|\mathbf{y}) = -\frac{TK}{2}\log 2\pi + \sum_{t=1}^{T}\log\left(\sum_{s=1}^{S}\rho_s|\sigma_s^2\mathbf{H}_t|^{-1/2} \times \exp\left(-\frac{(\mathbf{y}_t-\boldsymbol{\mu})'\mathbf{H}_t^{-1}(\mathbf{y}_t-\boldsymbol{\mu})}{2\sigma_s^2}\right)\right). \tag{7}$$

Although consistency and asymptotic normality of the QMLE of different multivariate GARCH-type models have been established by several authors, deriving the consistency and the asymptotic distribution of the ML estimate for the GMDCC model is a challenging task that is beyond the scope of this article. To report standard errors of the ML estimate, we conjecture that the usual asymptotic results for Gaussian mixtures hold in this situation and treat $\sqrt{T}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})$ as normally distributed with mean 0 and covariance matrix the inverse of the expected Fisher information matrix. Consequently, the standard deviations of the estimated model parameters should be interpreted carefully, but we present them here for completeness.

Once that the model parameters have been estimated, we may use the MLE to estimate in-sample volatilities and correlations and to predict future volatilities and correlations. First, obtaining estimates of in-sample volatilities and correlations is straightforward simply by replacing $\mathbf{\Phi}$ by its ML estimate in the respective equations for $\mathbf{H}_t$ and $\mathbf{R}_t$, for $t = 1, \ldots, T$. Second, given the ML estimate, predictions of $\mathbf{H}_{T+1}$ and $\mathbf{R}_{T+1}$ can be obtained similar as was done in the case of in-sample estimation. Finally, $\mathbf{y}_{T+1}$ is predicted by using the estimated mean of the process, $\widehat{\boldsymbol{\mu}}$.

Next, we develop a procedure for performing Bayesian inference of the GMDCC model. Bayesian analysis can be greatly simplified by introducing a set of unobserved latent variables, $\mathbf{z} = (z_1, \ldots, z_T)'$, indicating the specific component of the mixture from which each return is assumed to arise. Thus the so-called "complete likelihood function" for the returns and latent variables can be written as

$$L(\mathbf{\Phi}|\mathbf{y}, \mathbf{z}) = \prod_{t=1}^{T}\prod_{s=1}^{S}(\rho_s p(\mathbf{y}_t|\mathbf{\Phi}_s))^{I(z_t=s)}, \tag{8}$$

where $I(z_t = s)$ is an indicator variable set as 1 if $z_t = s$ and 0 otherwise. Under the Bayesian framework, inference on the parameters of the model, $\mathbf{\Phi}$, can be done through the posterior density of $\mathbf{\Phi}$ conditionally on the return series, $\mathbf{y}$, and the latent variables, $\mathbf{z}$, denoted by $p(\mathbf{\Phi}|\mathbf{y}, \mathbf{z})$. Using the Bayes theorem,

$$p(\mathbf{\Phi}|\mathbf{y}, \mathbf{z}) = \frac{L(\mathbf{\Phi}|\mathbf{y}, \mathbf{z})p(\mathbf{\Phi})}{\int L(\mathbf{\Phi}|\mathbf{y}, \mathbf{z})p(\mathbf{\Phi})\, d\mathbf{\Phi}},$$

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

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

almost surely, where b is the number of realizations discarded in a burn-in period. (See, e.g., Robert and Casella 2004 for an overview on MCMC methods from both theoretical and practical standpoints.)

First, we need to specify the expressions of the prior distribution, $p(\boldsymbol{\Phi})$. We use independent uniform prior distributions in the domain of all the bounded parameters of the vector $\boldsymbol{\Phi}$, that is, the parameters $(\alpha_{i,1}, \ldots, \alpha_{i,p_i})'$ of the EGARCH models, the parameters of the correlation matrixes of the vector $\boldsymbol{\Theta}$, and the parameters of the Gaussian mixture of the vector $\boldsymbol{\Lambda}$. On the other hand, we use independent normal distribution with mean 0 and a very large variance (e.g., 100) for all of the unbounded parameters of the vector $\boldsymbol{\Phi}$, that is, the parameters $(\mu_i, \omega_i, \beta_{i,2}, \ldots, \beta_{i,q_i}, \delta_i, \gamma_i)'$ of the EGARCH models. The variance of this Gaussian prior is much larger than the variance of the posterior distribution obtained, so it ensures that we are using noninformative, but proper priors for all of the model parameters.

Now we construct an MCMC algorithm to sample from the posterior distribution $p(\boldsymbol{\Phi}|\mathbf{y}, \mathbf{z})$. As noted by Vrontos, Dellaportas, and Politis (2003a, 2003b), the convergence of this type of algorithm may be accelerated by updating the highly correlated parameters simultaneously using a block-sampling approach. Thus we define the following algorithm scheme whose main steps are elaborated as follows:

1. Set $n = 0$ and initial values $\boldsymbol{\Phi}^{(0)} = (\boldsymbol{\Omega}_1^{(0)\prime}, \ldots, \boldsymbol{\Omega}_K^{(0)\prime}, \boldsymbol{\Theta}^{(0)\prime}, \boldsymbol{\Lambda}^{(0)\prime})'$.
2. Obtain a sample, $\mathbf{z}^{(n+1)}$, of the distribution of $\mathbf{z}|\mathbf{y}, \boldsymbol{\Phi}^{(n)}$.
3. Obtain a sample, $\boldsymbol{\Omega}_i^{(n+1)}$, of the distribution of $\boldsymbol{\Omega}_i|\boldsymbol{\Omega}_1^{(n+1)}, \ldots, \boldsymbol{\Omega}_{i-1}^{(n+1)}, \boldsymbol{\Omega}_{i+1}^{(n)}, \ldots, \boldsymbol{\Omega}_K^{(n)}, \boldsymbol{\Theta}^{(n)}, \boldsymbol{\Lambda}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}$, for $i = 1, \ldots, K$.
4. Obtain a sample, $\boldsymbol{\Theta}^{(n+1)}$, of the distribution of $\boldsymbol{\Theta}|\boldsymbol{\Omega}_1^{(n+1)}, \ldots, \boldsymbol{\Omega}_K^{(n+1)}, \boldsymbol{\Lambda}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}$.
5. Obtain a sample, $\boldsymbol{\Lambda}^{(n+1)}$, of the distribution of $\boldsymbol{\Lambda}|\boldsymbol{\Omega}_1^{(n+1)}, \ldots, \boldsymbol{\Omega}_K^{(n+1)}, \boldsymbol{\Theta}^{(n+1)}, \mathbf{y}, \mathbf{z}^{(n+1)}$.
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 $\mathbf{y}_t$, for $t = 1, \ldots, T$, has been generated from the sth component. These probabilities are given by

$$p(z_t = s \mid \mathbf{y}, \boldsymbol{\Phi}) = \frac{\rho_s/\sigma_s^K \exp(-U_s/(2\sigma_s^2))}{\sum_{v=1}^{S} \rho_v/\sigma_v^K \exp(-U_v/(2\sigma_v^2))}, \qquad s = 1, \ldots, S,$$

where $U_s = \sum_{t:z_t=s} (\mathbf{y}_t - \boldsymbol{\mu})' \mathbf{H}_t^{-1} (\mathbf{y}_t - \boldsymbol{\mu})$, $s = 1, \ldots, S$. In step 3 we sample from the conditional posterior probability of $\boldsymbol{\Omega}_i$ whose kernel is given by

$$\kappa(\boldsymbol{\Omega}_i \mid \boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_{i-1}, \boldsymbol{\Omega}_{i+1}, \ldots, \boldsymbol{\Omega}_K, \boldsymbol{\Theta}, \boldsymbol{\Lambda}, \mathbf{y}, \mathbf{z}) = p(\boldsymbol{\Omega}_i) \times \prod_{t=1}^{T} \left(H_{iit}^{-1/2} |\mathbf{R}_t|^{-1/2}\right) \times \exp\left(-\frac{1}{2} \sum_{s=1}^{S} \frac{U_s}{\sigma_s^2}\right), \quad (9)$$

where $p(\boldsymbol{\Omega}_i)$ is the prior probability of $\boldsymbol{\Omega}_i$. To do that, we make use of the random-walk Metropolis–Hastings method (RWMH) (see, e.g., Robert and Casella 2004) using the following steps:

3.1. Generate a candidate vector $\widetilde{\boldsymbol{\Omega}}_i$ from the multivariate normal distribution $\mathrm{N}(\boldsymbol{\Omega}_i^{(n)}, c\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Omega}_i})$, where c is a constant and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Omega}_i}$ is the covariance matrix of the ML estimate of $\boldsymbol{\Omega}_i$. Let

$$\tau_{\boldsymbol{\Omega}_i}^{(n)} = \min\Big\{1, \big(\kappa\big(\widetilde{\boldsymbol{\Omega}}_i \mid \boldsymbol{\Omega}_1^{(n+1)}, \ldots, \boldsymbol{\Omega}_{i-1}^{(n+1)}, \boldsymbol{\Omega}_{i+1}^{(n)}, \ldots, \boldsymbol{\Omega}_K^{(n)}, \boldsymbol{\Theta}^{(n)}, \boldsymbol{\Lambda}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}\big)\big) / \big(\kappa\big(\boldsymbol{\Omega}_i^{(n)} \mid \boldsymbol{\Omega}_1^{(n+1)}, \ldots, \boldsymbol{\Omega}_{i-1}^{(n+1)}, \boldsymbol{\Omega}_{i+1}^{(n)}, \ldots, \boldsymbol{\Omega}_K^{(n)}, \boldsymbol{\Theta}^{(n)}, \boldsymbol{\Lambda}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}\big)\big)\Big\},$$

where $\kappa(\widetilde{\boldsymbol{\Omega}}_i \mid \boldsymbol{\Omega}_1^{(n+1)}, \ldots, \boldsymbol{\Omega}_{i-1}^{(n+1)}, \boldsymbol{\Omega}_{i+1}^{(n)}, \ldots, \boldsymbol{\Omega}_K^{(n)}, \boldsymbol{\Theta}^{(n)}, \boldsymbol{\Lambda}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)})$ is given in eq. (9).

3.2. Define

$$\boldsymbol{\Omega}_i^{(n+1)} = \begin{cases} \widetilde{\boldsymbol{\Omega}}_i, & \text{with probability } \tau_{\boldsymbol{\Omega}_i}^{(n)} \\ \boldsymbol{\Omega}_i^{(n)}, & \text{with probability } 1 - \tau_{\boldsymbol{\Omega}_i}^{(n)}. \end{cases}$$

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.

Finally, in steps 4 and 5 we sample from the conditional posterior distributions of $\boldsymbol{\Theta}$ and $\boldsymbol{\Lambda}$, respectively, whose kernels are given by

$$\kappa(\boldsymbol{\Theta} \mid \boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_K, \boldsymbol{\Lambda}, \mathbf{y}, \mathbf{z}) = p(\boldsymbol{\Theta}) \times \prod_{t=1}^{T} \left[\left(\prod_{i=1}^{K} H_{iit}^{-1/2}\right) |\mathbf{R}_t|^{-1/2}\right] \times \exp\left(-\frac{1}{2} \sum_{s=1}^{S} \frac{U_s}{\sigma_s^2}\right)$$

and

$$\kappa(\boldsymbol{\Lambda} \mid \boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_K, \boldsymbol{\Theta}, \mathbf{y}, \mathbf{z}) = p(\boldsymbol{\Lambda}) \times \prod_{s=1}^{S} \left(\frac{\rho_s}{\sigma_s^K}\right)^{T_s} \times \prod_{t=1}^{T} \left(\prod_{i=1}^{K} H_{iit}^{-1/2}\right) \times \exp\left(-\frac{1}{2} \sum_{s=1}^{S} \frac{U_s}{\sigma_s^2}\right),$$

respectively, where $T_s = \sum_{t=1}^{T} I(z_t = s)$, with $\sum_{s=1}^{S} T_s = T$, and $p(\boldsymbol{\Theta})$ and $p(\boldsymbol{\Lambda})$ are the prior probabilities of $\boldsymbol{\Theta}$ and $\boldsymbol{\Lambda}$, respectively, which can be performed using a similar RWMH as that described in step 3.

Besides 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, $\boldsymbol{\Phi}^{(n)}$, which is denoted by $H_{iit}^{(n)}$, for $n = b + 1, \ldots, T$. Then the posterior expected value of H_{iit}, $E[H_{iit} \mid y]$, can be approached by the mean of the posterior sample of conditional volatilities, that is,

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

In addition, 95% Bayesian confidence intervals can be obtained by just calculating 0.025 and 0.975 quantiles of each posterior sample, respectively. Similarly, we can estimate in-sample correlations R_{ijt}, using the draws $R_{ijt}^{(n)}$. 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 $\mathbf{y}_{T+1}$ is given by

$$p(\mathbf{y}_{T+1}|\mathbf{y}) = \int_{\boldsymbol{\Phi}} p(\mathbf{y}_{T+1}|\mathbf{y}, \boldsymbol{\Phi}) p(\boldsymbol{\Phi}|\mathbf{y})\, d\boldsymbol{\Phi}, \tag{10}$$

where $p(\mathbf{y}_{T+1}|\mathbf{y}, \boldsymbol{\Phi})$ is a mixture of S multivariate Gaussian distributions with mean $\boldsymbol{\mu}$ and covariances $\sigma_s^2 \mathbf{H}_{T+1}$, for $s = 1, \ldots, S$. Thus the predictive density $p(\mathbf{y}_{T+1}|\mathbf{y})$ in eq. (10) can be estimated by

$$\frac{1}{N-b} \sum_{n=b+1}^{N} p\big(\mathbf{y}_{T+1}|\mathbf{y}, \boldsymbol{\Phi}^{(n)}\big),$$

from which we can obtain point predictions and predictive intervals. The prediction of $H_{ii,T+m}$, $R_{ij,T+m}$, and $\mathbf{y}_{T+m}$ when $m > 1$ is more complicated because the values of $\mathbf{y}_t$ are unknown for $t \geq T+1$. However, we also can generate samples from the predictive distributions of $H_{ii,T+m}$, $R_{ij,T+m}$, and y_{T+m}, when $m > 1$, using a generalization of the sequential procedure proposed by Ausín and Galeano (2007).

Estimating the number of mixture components, S, from the return series is a very difficult task, especially because it involves inference for models that may be overfitted if the true number of components is less than the number of components of the fitted model. From the frequentist standpoint, it has been widely documented that standard test theory breaks down in mixture contexts. For instance, the usual χ^2 asymptotic distribution of the likelihood ratio test statistics does not hold in mixture modeling. On the other hand, several methods are available for selecting the mixture components from a Bayesian standpoint, including the use of marginal likelihoods, Bayes factors, entropy distance, or more complex approaches, such as reversible-jump MCMC algorithms or birth-and-death processes (see Marin, Mengersen, and Robert 2005). However, a simple, plausible method frequently considered from both standpoints is the use of the Bayesian information criterion (BIC). In fact, the BIC has been applied by many authors to finite mixture models, and several simulation studies support its use. Moreover, in the case of iid observations, the BIC has been found to be consistent in selecting the true number of the mixture components (see Keribin 2000). In our problem, the BIC selects the number of mixture components, S, which provides the minimum value of the BIC given by $BIC(s) = -2l(\widehat{\boldsymbol{\Phi}}|\mathbf{y}, s) + \log(T) m_s$, where $l(\widehat{\boldsymbol{\Phi}}|\mathbf{y}, s)$ is the value of the log-likelihood function in eq. (7) evaluated at the ML estimation parameters, $\widehat{\boldsymbol{\Phi}}$, assuming s mixture components, and m_s is the number of parameters of the GMDCC model with s mixture components.

4. VALUE AT RISK CALCULATION AND PORTFOLIO SELECTION

In this section we analyze two of the most important issues in risk management, namely the calculation of the VaR of a given portfolio and the portfolio selection problem, using the information provided by the proposed GMDCC model. In particular, we show how to obtain point estimates of both the portfolio VaR at a given significance level and the weights of the optimal portfolio based on both maximum likelihood and Bayesian approaches. Using the Bayesian methodology, besides of giving point estimates, we provide a measure of precision for both VaR and optimal weight estimates via predictive intervals.

Given a vector return series $\mathbf{y}_t = (\mathbf{y}_{1t}, \ldots, \mathbf{y}_{Kt})'$, a portfolio of the components of $\mathbf{y}_t$, denoted by p_t, is defined as a linear combination of the individual returns (i.e., $p_t = \boldsymbol{\delta}' \mathbf{y}_t$), where the weights $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_K)'$ add to 1. The VaR may be defined as the maximum potential loss expected with probability $1 - \pi$, where π is supposed to be small. Statistically speaking, the one-step-ahead VaR, denoted by VaR_{T+1}, is defined as the negative value of the 100πth quantile of the distribution of the portfolio return, that is, $\Pr(p_{T+1} \leq -\mathrm{VaR}_{T+1}) = \pi$. Assuming that the model parameters are known, the conditional distribution of the one-step-ahead portfolio is

$$p_{T+1}|\boldsymbol{\Phi}, \mathbf{y} \sim \rho_1 \mathrm{N}(\boldsymbol{\delta}'\boldsymbol{\mu}, \sigma_1^2 \boldsymbol{\delta}' \mathbf{H}_{T+1} \boldsymbol{\delta}) + \cdots + \rho_S \mathrm{N}(\boldsymbol{\delta}'\boldsymbol{\mu}, \sigma_S^2 \boldsymbol{\delta}' \mathbf{H}_{T+1} \boldsymbol{\delta}). \tag{11}$$

Thus VaR_{T+1} is the negative value of the 100πth quantile of this univariate mixture, which may be easily obtained using, for instance, the Newton–Raphson method.

Consequently, with the ML estimate of the parameters, $\widehat{\boldsymbol{\Phi}}$, one may obtain an estimate of VaR_{T+1} by just simply replacing $\boldsymbol{\mu}$, σ_s^2, and $\mathbf{H}_{T+1}$ by their respective estimates. On the other hand, using the posterior sample, $\boldsymbol{\Phi}^{(n)}$, a consistent estimator of $E[\mathrm{VaR}_{T+1}|\mathbf{y}]$, is given by

$$\frac{1}{N-b} \sum_{n=b+1}^{N} \mathrm{VaR}_{T+1}^{(n)},$$

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

Although other alternatives are plausible, to exemplify the performance of the proposed model for portfolio selection, we assume that the one-step-ahead portfolio is optimal if it has minimum VaR_{T+1}, that is, minimum risk. Therefore, the one-step-ahead optimal portfolio is the solution of

$$\boldsymbol{\delta}_{opt} = \arg\min_{\boldsymbol{\delta}} \{\mathrm{VaR}_{T+1} : \boldsymbol{\delta}' \mathbf{1}_K = 1\}, \tag{12}$$

where $\mathbf{1}_K = (1, \ldots, 1)'$. Given $\boldsymbol{\delta}_{opt}$, the expected gain for the optimal portfolio is $g_{opt} = \boldsymbol{\delta}_{opt}' \boldsymbol{\mu}$. It is well known that the solution of the problem in eq. (12) is not analytically tractable even under the Gaussianity assumption. However, we can make use of numerical optimization procedures to solve it. Given the distribution in eq. (11), the VaR_{T+1} depends on $\boldsymbol{\delta}$ and $\boldsymbol{\Phi}$. From a frequentist standpoint, $\boldsymbol{\Phi}$ is replaced by its ML estimate and the uncertainty due to parameter estimation is ignored. The

Bayesian approach allows for the inclusion of parameter uncertainty through the posterior distribution of $\boldsymbol{\Phi}$. Given the samples from this posterior distribution, we can obtain a sample of the posterior distribution of the optimal weights, denoted by $\boldsymbol{\delta}_{opt}^{(n)}$, and a sample from the posterior distribution of the optimal expected gain, denoted by $g_{opt}^{(n)} = \boldsymbol{\delta}_{opt}^{(n)\prime}\boldsymbol{\mu}^{(n)}$. Finally, we can obtain consistent estimators of the posterior mean of the optimal weights and gain, respectively, using

$$\frac{1}{N-b}\sum_{n=b+1}^{N}\boldsymbol{\delta}_{opt}^{(n)} \qquad \text{and} \qquad \frac{1}{N-b}\sum_{n=b+1}^{N} g_{opt}^{(n)}.$$

Also, Bayesian confidence intervals can be obtained using the quantiles of the 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 $T = 1{,}000$, 2,000, and 3,000, simulated from the GMDCC model with: (i) individual EGARCH(1, 1) models with parameter vectors $\boldsymbol{\Omega}_1 = (0.05, -0.1, 0.93, -0.1, 0.1)'$ and $\boldsymbol{\Omega}_2 = (0.03, -0.11, 0.91, -0.12, 0.11)'$, (ii) dynamic conditional correlation with parameter vector $\boldsymbol{\Theta} = (0.8, 0.1, 0.7)'$, and (iii) a Gaussian mixture distribution with parameters $\boldsymbol{\Lambda} = (0.9, 0.15)'$. We assume that 90% of the innovations are generated from a Gaussian distribution with variance $\sigma_1^2 \approx 0.64$ and that 10% of the innovations are generated from a Gaussian mixture with variance $\sigma_2^2 \approx 4.25$. This second mixture component is designed to contain the extremes.

First, for each of the simulated series, we obtain the ML estimate of the model parameters for $S = 1, \ldots, 5$ components of the mixture as described in Section 3. We then compute the values of the BIC for each simulated series, as shown in Table 1. In every case, the BIC selects the true value $S = 2$ mixture components. Then, assuming that $S = 2$, we run the proposed MCMC algorithm for each simulated series using 20,000 iterations and discard the initial 10,000 iterations for inference as burn-in iterations. We consider the ML estimate of the parameters as the initial values and use the block-sampling approach described in Section 3. The MCMC chains provide good mixing performance and fast convergence. Figure 2 shows the histograms of the posterior samples of each model parameter for the first simulated series with sample size $T = 1{,}000$. We see that the algorithm captures the asymmetry of the posterior distributions of the parameters ω_1, $\alpha_{1,1}$, ω_2, $\alpha_{2,1}$, θ_1, and θ_2. We also observed that, as expected, these posterior distributions became more symmetric for $T = 2{,}000$ and $T = 3{,}000$ (data not reported).

Table 1. Computed values of the BIC for each simulated series for $S = 1, \ldots, 5$ mixture components

BIC	$T = 1{,}000$	$T = 2{,}000$	$T = 3{,}000$
$S = 1$	2,566.0	5,327.4	7,726.8
$S = 2$	2,399.0	4,819.6	6,833.4
$S = 3$	2,412.8	4,834.8	6,848.4
$S = 4$	2,426.6	4,850.0	6,864.4
$S = 5$	2,440.4	4,865.2	6,880.4

Table 2 compares the ML estimates and standard errors with the Bayesian posterior means and standard deviations of the model parameters for the three simulated series. Note that both approaches lead to similar point estimates and that in general, the posterior standard deviations are larger than the standard errors when the posterior distributions are asymmetric. This is

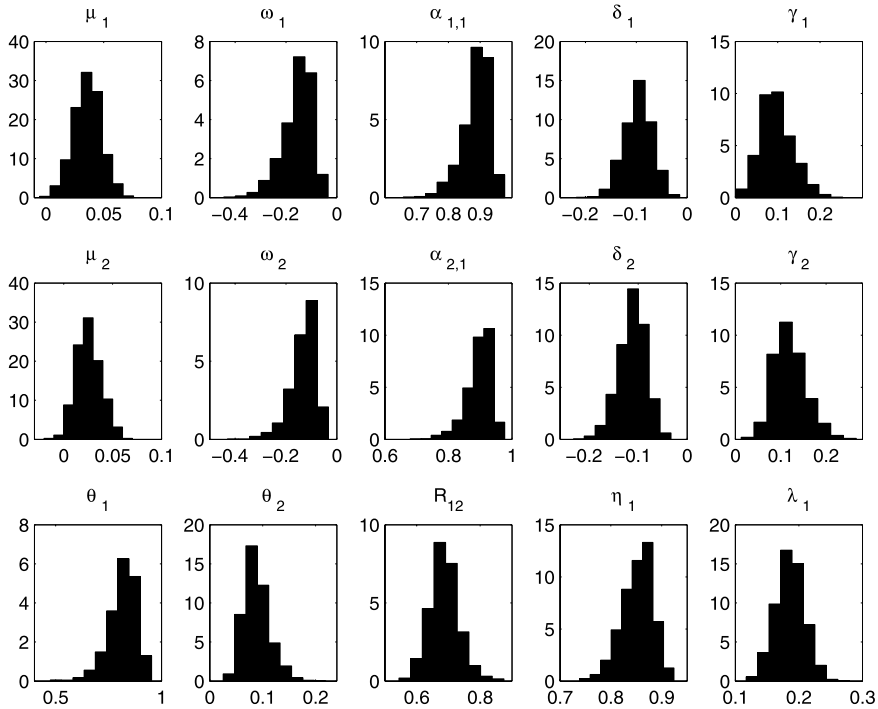

Figure 2. Histograms of the posterior MCMC samples of each model parameter corresponding to the first simulated series with $T = 1{,}000$. The online version of this figure is in color.

Table 2. Maximum likelihood and Bayesian estimates of the model parameters for the three simulated series

	$T = 1{,}000$		$T = 2{,}000$		$T = 3{,}000$	
Parameter	ML estimate (std)	Posterior mean (std)	ML estimate (std)	Posterior mean (std)	ML estimate (std)	Posterior mean (std)
$\mu_1 = 0.05$	0.0361	0.0359	0.0359	0.0369	0.0463	0.0456
	(0.0123)	(0.0121)	(0.0089)	(0.0087)	(0.0066)	(0.0066)
$\omega_1 = -0.1$	−0.1077	−0.1551	−0.0991	−0.1137	−0.1068	−0.1130
	(0.0424)	(0.0588)	(0.0255)	(0.0273)	(0.0184)	(0.0189)
$\alpha_{1,1} = 0.93$	0.9225	0.8866	0.9249	0.9130	0.9198	0.9143
	(0.0296)	(0.0433)	(0.0190)	(0.0206)	(0.0135)	(0.0143)
$\delta_1 = -0.1$	−0.0884	−0.0967	−0.0960	−0.0969	−0.1109	−0.1138
	(0.0241)	(0.0274)	(0.0174)	(0.0182)	(0.0122)	(0.0133)
$\gamma_1 = 0.1$	0.0708	0.0974	0.1120	0.1243	0.0919	0.0961
	(0.0327)	(0.0384)	(0.0248)	(0.0256)	(0.0159)	(0.0160)
$\mu_2 = 0.03$	0.0247	0.0254	0.0226	0.0237	0.0344	0.0342
	(0.0131)	(0.0126)	(0.0093)	(0.0083)	(0.0072)	(0.0071)
$\omega_2 = -0.11$	−0.0931	−0.1330	−0.1055	−0.1163	−0.1151	−0.1231
	(0.0313)	(0.0517)	(0.0224)	(0.0233)	(0.0215)	(0.0216)
$\alpha_{2,1} = 0.91$	0.9278	0.8952	0.9103	0.9002	0.8970	0.8892
	(0.0241)	(0.0403)	(0.0182)	(0.0187)	(0.0182)	(0.0181)
$\delta_2 = -0.12$	−0.0968	−0.1120	−0.1213	−0.1261	−0.1079	−0.1094
	(0.0244)	(0.0285)	(0.0170)	(0.0175)	(0.0145)	(0.0148)
$\gamma_2 = 0.11$	0.0987	0.1198	0.1033	0.1140	0.1264	0.1337
	(0.0330)	(0.0357)	(0.0247)	(0.0246)	(0.0198)	(0.0209)
$\theta_1 = 0.8$	0.8522	0.8156	0.8307	0.8190	0.7879	0.7799
	(0.0409)	(0.0662)	(0.0345)	(0.0363)	(0.0265)	(0.0265)
$\theta_2 = 0.1$	0.0761	0.0885	0.0743	0.0768	0.1044	0.1070
	(0.0183)	(0.0245)	(0.0136)	(0.0140)	(0.0126)	(0.0126)
$R_{12} = 0.7$	0.7039	0.6874	0.6983	0.6935	0.7382	0.7351
	(0.0407)	(0.0460)	(0.0243)	(0.0241)	(0.0181)	(0.0175)
$\eta_1 = 0.9$	0.8543	0.8522	0.9055	0.9015	0.9076	0.9062
	(0.0338)	(0.0308)	(0.0134)	(0.0130)	(0.0104)	(0.0109)
$\lambda_1 = 0.15$	0.1881	0.1860	0.1379	0.1399	0.1286	0.1289
	(0.0286)	(0.0225)	(0.0155)	(0.0134)	(0.0105)	(0.0102)

caused by the asymptotic Gaussian assumption of the ML estimate. However, as expected, the standard errors and posterior standard deviations become smaller and more similar as the sample size, T, increases. Finally, besides providing point estimates and standard errors, the Bayesian estimation also produces posterior densities that describe all of their uncertainty associated with the model parameters. Moreover, this uncertainty may be introduced in the estimation of volatilities, correlations, VaR, portfolio selection, and so on, as we show later.

Next, we estimate the in-sample volatilities and correlations using both ML estimates and the Bayesian approaches. This is illustrated in Figure 3, which presents the estimated in-sample volatilities, H_{iit}, for $i = 1, 2$, and the estimated in-sample correlations, R_{12t}, for $t = 1{,}900, \ldots, 2{,}000$, for the second simulated series with sample size $T = 2{,}000$ are presented. Also shown are 95% Bayesian confidence intervals and true values. Observe the accuracy of the estimations and note that the Bayesian confidence intervals always include the true values of H_{iit} and R_{12t} for all time periods. We also can make predictions of future volatilities and correlations as shown in Figure 3, which shows the point estimations and Bayesian intervals for the one-step-ahead volatilities, $H_{ii,T+1}$, for $i = 1, 2$, and correlation, $R_{12,T+1}$, where $T + 1 = 2{,}001$, and compares them with their true values.

We next apply the procedures described in Section 4 for VaR calculation and portfolio selection. Figure 4 shows the ML and Bayesian estimations for the one-step-ahead VaR_{T+1} of portfolios of the form $p_{T+1} = \delta \times y_{1,T+1} + (1 - \delta) \times y_{2,T+1}$, as a function of δ, for the three simulated series and for $\pi = 0.05$ and $\pi = 0.01$. Also shown are the Bayesian 95% confidence intervals and true values. Observe the accuracy of the estimations and note that the Bayesian confidence intervals always include the true VaR_{T+1} values. In contrast, Table 3 gives the maximum likelihood and Bayesian estimations of the optimal weight, δ_{opt}, of the portfolio $p_{T+1} = \delta_{opt} \times y_{1,T+1} + (1 - \delta_{opt}) \times y_{2,T+1}$, which minimizes the one-step-ahead VaR_{T+1}, and the corresponding optimal gain, g_{opt}, for the three simulated series. Also shown are the Bayesian 95% confidence intervals and true values. Note that the obtained optimal weights are coherent with the plots shown in Figure 4.

Finally, we also performed simulations with larger-dimensional systems in which $K = 5$ and $K = 10$. In the cases, the only problem found is that the greater computational cost of obtaining both ML and Bayesian estimates compared with that required for smaller systems. Apparently, the accuracy of the estimates is not significantly affected by large dimensions, however.

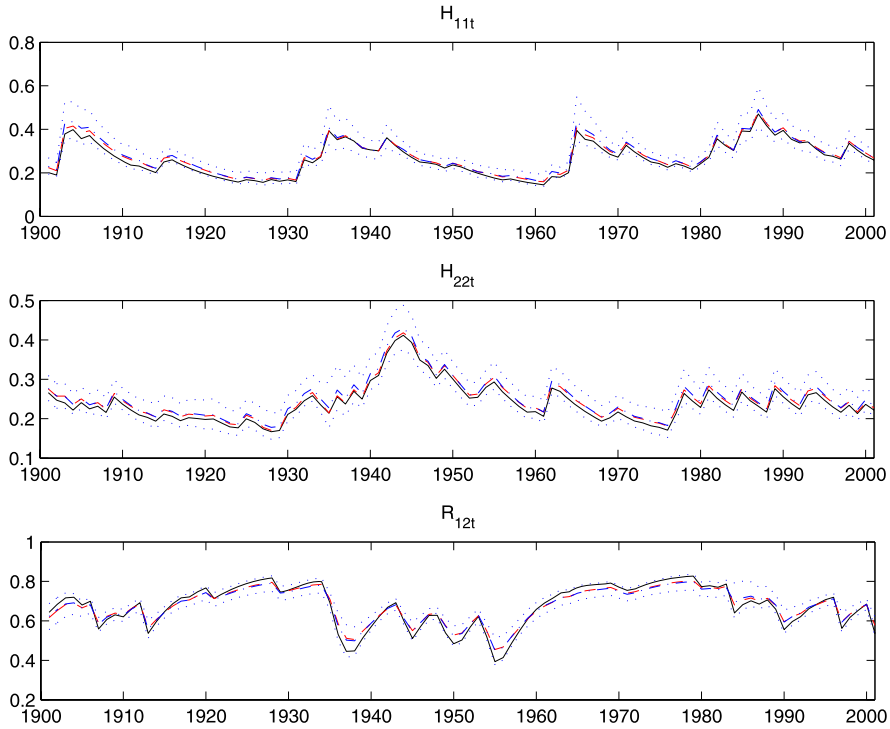

Figure 3. True (solid), ML (dashed), and Bayesian (dashed–dotted) estimates and 95% predictive intervals (dotted) for the volatilities, H_{iit}, for $i = 1, 2$, and correlations, R_{12t}, for $t = 1{,}900, \ldots, 2{,}001$, for the second simulated series with sample size $T = 2{,}000$. The online version of this figure is in color.

6. APPLICATION

As an illustration, we apply the proposed methodology to the daily closing prices of Dow Jones Industrial Average and NASDAQ composite indexes for the period January 2, 1996 to December 29, 2006. The log return series of sample size $T = 2{,}769$ is affected by the presence of several extremes. The sample means, variances, and excess kurtosis of the two log-

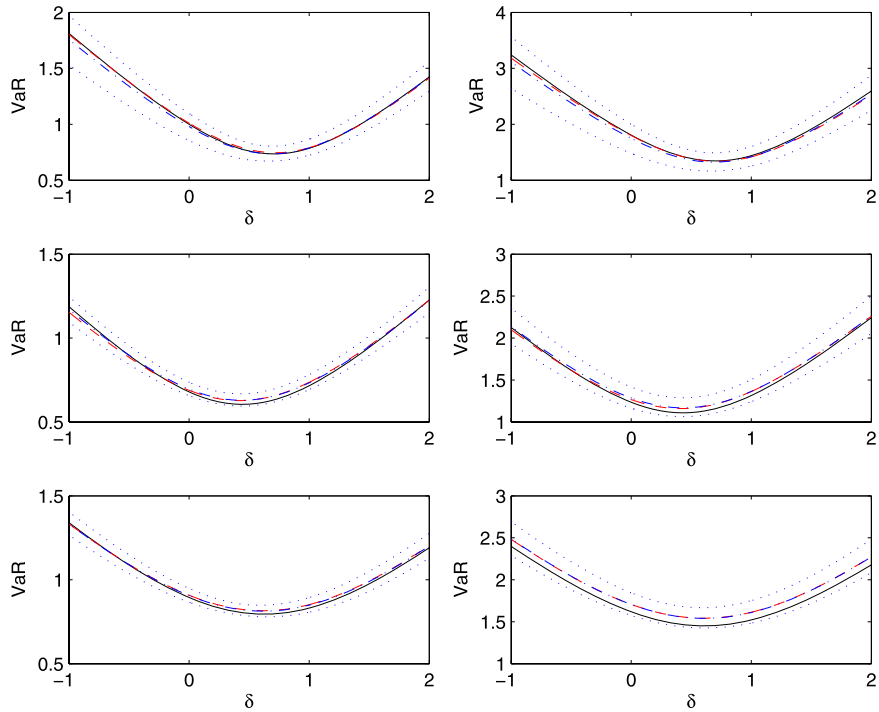

Figure 4. True (solid), ML (dashed), and Bayesian (dashed–dotted) estimates and 95% predictive intervals (dotted) for the VaR_{T+1} of the portfolio $p_{T+1} = \delta \times y_{1,T+1} + (1 - \delta) \times y_{2,T+1}$ as a function of δ, for the series with $T = 1{,}000$ (top), 2,000 (middle), and 3,000 (bottom) and for $\pi = 0.05$ (left) and $\pi = 0.01$ (right). The online version of this figure is in color.

Table 3. True values and ML and Bayesian point estimates together with 95% predictive intervals for the optimal weight, δ_{opt}, of the optimal portfolio $\delta_{opt,1} \times y_{1,T+1} + (1 - \delta_{opt,1}) \times y_{2,T+1}$, and for the corresponding expected gain, for the three simulated series

			$\pi = 0.05$	$\pi = 0.01$
$T = 1{,}000$	δ_{opt}	True	0.7080	0.7008
		MLE	0.7215	0.7173
		Pred. mean	0.6914	0.6876
		95% interval	[0.5636, 0.7916]	[0.5607, 0.7944]
	g_{opt}	True	0.0442	0.0440
		MLE	0.0329	0.0328
		Pred. mean	0.0328	0.0327
		95% interval	[0.0110, 0.0550]	[0.0109, 0.0549]
$T = 2{,}000$	δ_{opt}	True	0.4434	0.4323
		MLE	0.4223	0.4139
		Pred. mean	0.4392	0.4310
		95% interval	[0.3691, 0.5047]	[0.3637, 0.4935]
	g_{opt}	True	0.0389	0.0386
		MLE	0.0282	0.0281
		Pred. mean	0.0296	0.0294
		95% interval	[0.0149, 0.0447]	[0.0148, 0.0445]
$T = 3{,}000$	δ_{opt}	True	0.6250	0.6084
		MLE	0.6173	0.6064
		Pred. mean	0.6193	0.6089
		95% interval	[0.5225, 0.7237]	[0.5140, 0.7123]
	g_{opt}	True	0.0425	0.0422
		MLE	0.0417	0.0416
		Pred. mean	0.0414	0.0412
		95% interval	[0.0299, 0.0542]	[0.0297, 0.0541]

return series are 0.0317 and 0.0297, 1.1833 and 3.1013, and 4.178 and 4.182, respectively. The autocorrelation functions of both returns show no significant autocorrelations. In fact, the Ljung–Box statistics for both log returns for lag 5 are 7.987 and 13.375, and those for lag 10 are 109.480 and 13.688, with associated p-values of 0.157 and 0.203, and 0.091 and 0.188, respectively. The sample correlation between both log returns is 0.705.

First, we estimate the GMDCC model using both ML and the Bayesian procedure. We obtain the ML estimates of the model parameters for $S = 1, \ldots, 4$ and compute the value of the BIC. These are given by 1.5016×10^4, 1.4939×10^4, 1.4943×10^4, and 1.4959×10^4. Therefore, the BIC selects $S = 2$ components in the mixture. Table 4 gives the ML estimates of the model parameters for $S = 2$. Figure 5 compares the empirical cumulative distribution function of the standardized residuals with the cumulative distribution function of the estimated Gaussian mixture, showing a good fit. In fact, the Kolmogorov–Smirnov test does not reject the null hypothesis of equality of distributions for significance levels $\alpha = 0.05$ and $\alpha = 0.01$. Then, we run the proposed MCMC algorithm for the bivariate series with 20,000 iterations, using the first 10,000 as burn-in iterations. Table 4 also gives the posterior means and standard deviations obtained from the MCMC output, leading to similar results as for ML estimation. Both approaches predict that about 97% of the innovations are generated by the mixture component with smaller covariance matrix, whereas approximately the 3% are generated by the component with larger covariance matrix. In addition, the covariance matrix of this second component, which may be viewed as the component with the extremes, is estimated to be approximately 4.5 times larger than the smaller covariance matrix.

Figure 6 illustrates the ML and Bayesian estimates together with 95% Bayesian confidence intervals for the volatilities, H_{iit}, for $i = 1, 2$, and correlations, R_{12t}, of the last 100 returns. The ML estimates are very close to the Bayesian posterior means for

Table 4. ML and Bayesian estimates of the model parameters for the Dow Jones and NASDAQ indexes

	μ_1	ω_1	$\alpha_{1,1}$	δ_1	γ_1
MLE	0.0412	0.0023	0.9908	−0.0322	0.0952
(std)	(0.0139)	(0.0013)	(0.0021)	(0.0069)	(0.0101)
Post. mean	0.0479	0.0024	0.9900	−0.0342	0.1020
(std)	(0.0139)	(0.0016)	(0.0024)	(0.0082)	(0.0119)
	μ_2	ω_2	$\alpha_{2,1}$	δ_2	γ_2
MLE	0.0562	−0.0003	0.9941	−0.0147	0.0929
(std)	(0.0199)	(0.0014)	(0.0016)	(0.0059)	(0.0111)
Post. mean	0.0661	−0.0001	0.9936	−0.0152	0.0999
(std)	(0.0184)	(0.0016)	(0.0017)	(0.0062)	(0.0109)
	θ_1	θ_2	R_{12}	η_1	λ_1
MLE	0.9711	0.0148	0.8704	0.9741	0.2134
(std)	(0.0067)	(0.0032)	(0.0231)	(0.0102)	(0.0327)
Post. mean	0.9667	0.0164	0.8651	0.9607	0.2443
(std)	(0.0087)	(0.0040)	(0.0210)	(0.0228)	(0.0579)

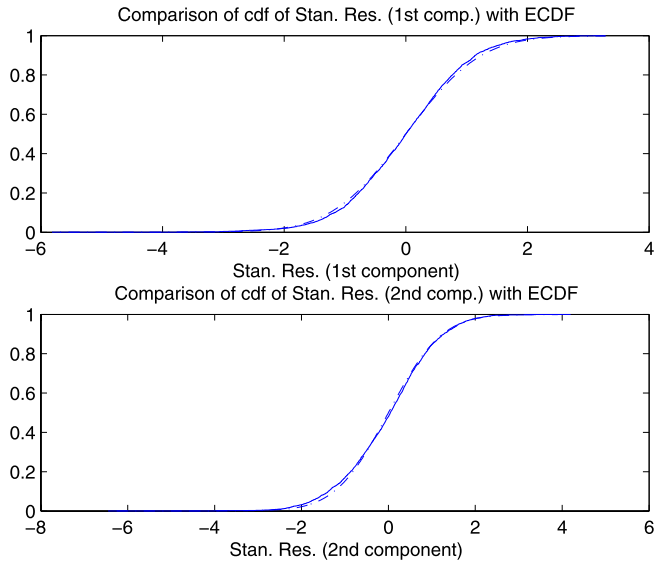

Figure 5. Empirical cumulative distribution function of the standardized residuals compared with the cumulative distribution function of the Gaussian mixture with parameters ρ_1 and λ_1 estimated by ML estimation for the GMDCC model. The online version of this figure is in color.

the all time periods. Note that the Bayesian confidence intervals of H_{11t}, H_{22t} , and R_{12t} are not necessarily symmetric. Figure 6 also includes the one-step-ahead point predictions of $H_{ii,2,770}$, for $i = 1, 2$, and the correlation, $R_{12,2,770}$, along with the corresponding predictive intervals. It appears that the NASDAQ index is more volatile than the Dow Jones index in this last period of 2006. In fact, this effect was seen since 1998 (data not reported). Also note that the conditional correlations are very high, with values around 0.9 and a drop at the end of November 2006.

Figure 7 illustrates both ML and Bayesian estimates along with 95% confidence intervals for the one-step-ahead VaR_{T+1} of the portfolio $p_{T+1} = \delta \times DowJones_{T+1} + (1 - \delta) \times Nasdaq_{T+1}$, for $\delta \in (-1, 2)$, and $\pi = 0.05$ and $\pi = 0.01$. Observe that the Bayesian confidence intervals always include the ML estimate for each weight δ. The one-step-ahead VaR minimizing portfolio weight seems to outweigh the Dow Jones index. This effect can be clearly seen in Table 5, which gives both the ML and Bayesian estimates with a 95% predictive interval for the optimal weight, δ_{opt}, of the optimal portfolio, $\delta_{opt,1} DowJones_{T+1} + (1 - \delta_{opt,1}) Nasdaq_{T+1}$, and the corresponding expected gain, g_{opt}, for $\pi = 0.05$ and $\pi = 0.01$. Both approaches predict an optimal portfolio weight of approximately 1.25 on the Dow Jones index. Finally, note that the gains from using the optimal portfolio weight are statistically significant.

7. CONCLUSIONS

In this article we have proposed a multivariate GARCH model with time-varying correlations in which the individual conditional volatilities follow a univariate EGARCH model and the standardized innovations are assumed to follow a mixture of multivariate zero mean Gaussian distributions. This specification extends the Gaussian mixture innovation distribution proposed by Bai, Rusell, and Tiao (2003) to the multivariate framework with an unknown number of mixture components. We have shown how to perform both ML and Bayesian inference on this model. In particular, Bayesian inference is done using MCMC methods, allowing estimation and prediction of

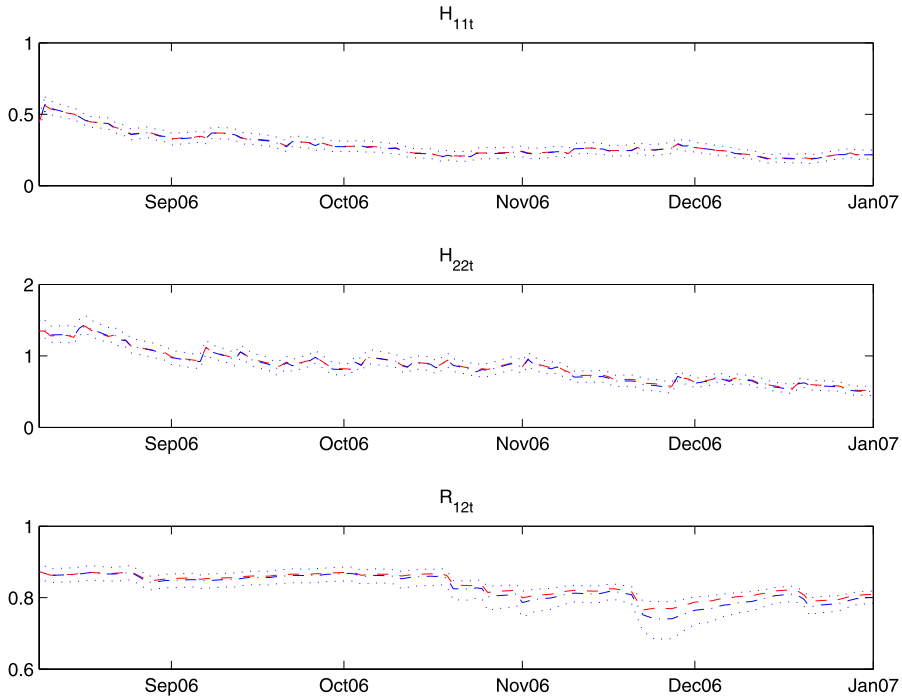

Figure 6. Bayesian estimates (dashed) and 95% intervals (dotted) compared with ML estimates (dashed–dotted) for the volatilities, H_{iit}, for $i = 1, 2$, and correlations, R_{12t}, of the last 100 observations for the Dow Jones and NASDAQ indexes. The online version of this figure is in color.

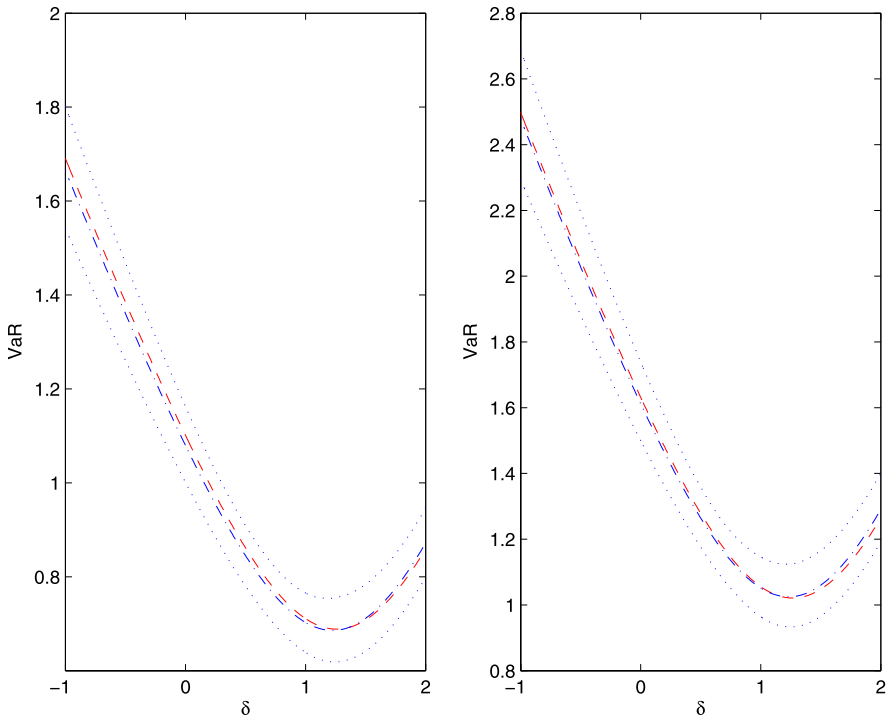

Figure 7. ML (dashed) and Bayesian (dashed–dotted) estimates and 95% predictive intervals (dotted) for the VaR_{T+1} of the portfolio $p_{T+1} = \delta \times DowJones_{T+1} + (1-\delta) \times Nasdaq_{T+1}$, as a function of δ, for $\pi = 0.05$ (left) and $\pi = 0.01$ (right). The online version of this figure is in color.

conditional volatilities and correlations. The two approaches perform similarly, and both have been shown to work well with both simulated and real data examples. We also have developed a procedure for deriving predictive distributions for the portfolio VaR, as well as a method for determining optimal portfolios.

ACKNOWLEDGMENTS

The authors thank the editor, the associate editor, and an anonymous referee for their many helpful suggestions, which led to a much-improved article. The work of Galeano was supported by MEC project MTM2008-03010 and Xunta de Galicia project PGIDIT06PXIB207009PR. The work of Concepción Ausín was supported by MEC project MTM2008-00166.

[Received September 2007. Revised December 2008.]

Table 5. ML and Bayesian estimates for the weight, δ_{opt}, and expected gain, g_{opt}, of the optimal portfolio, $\delta_{opt} DowJones_{T+1} + (1-\delta_{opt}) Nasdaq_{T+1}$, joint with 95% predictive intervals

		$\pi = 0.05$	$\pi = 0.01$
δ_{opt}	MLE	1.2589	1.2659
	Pred. mean	1.2232	1.2316
	95% interval	[1.1430, 1.2929]	[1.1572, 1.2983]
g_{opt}	MLE	0.0373	0.0372
	Pred. mean	0.0440	0.0438
	95% interval	[0.0159, 0.0721]	[0.0156, 0.0720]

REFERENCES

Ausín, M. C., and Galeano, P. (2007), "Bayesian Estimation of the Gaussian Mixture GARCH Model," *Computational Statistics and Data Analysis*, 51, 2636–2652. [559,564]

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. [559,561,569]

Bauwens, L., Hafner, C. M., and Rombouts, J. V. K. (2007), "Multivariate Mixed Normal Conditional Heteroskedaticity," *Computational Statistics and Data Analysis*, 51, 3551–3566. [559]

Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006), "Multivariate GARCH Models: A Survey," *Journal of Applied Econometrics*, 21, 79–109. [559,560]

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327. [559]

——— (1990), "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics*, 72, 498–505. [559]

Bollerslev, T., and Engle, R. F. (1993), "Common Persistence in Conditional Variances," *Econometrica*, 61, 167–186. [559]

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. [559]

Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of the U.K. Inflation," *Econometrica*, 50, 987–1008. [559]

——— (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business & Economic Statistics*, 20, 339–350. [559]

Engle, R. F., and Kroner, K. (1995), "Multivariate Simultaneous GARCH," *Econometric Theory*, 11, 122–150. [559]

Engle, R. F., Ng, V. K., and Rothschild, M. (1990), "Asset Pricing With a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics*, 45, 213–238. [559]

Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models*, New York: Springer. [561]

Haas, M., Mittnik, S., and Paolella, M. S. (2004), "Mixed Normal Conditional Heteroskedasticity," *Journal of Financial Econometrics*, 2, 211–250. [559]

Hall, P., and Yao, Q. (2003), "Inference in ARCH and GARCH Models With Heavy-Tailed Errors," *Econometrica*, 71, 285–317. [561]

Jeantheau, T. (1998), "Strong Consistency of Estimators for Multivariate ARCH Models," *Econometric Theory*, 14, 70–86. [561]

Keribin, C. (2000), "Consistent Estimation of the Order of Mixture Models," *Sankhya, Ser. A*, 62, 49–66. [564]

Kroner, F. K., and Ng, V. K. (1998), "Modelling Asymmetric Comovements of Asset Returns," *The Review of Financial Studies*, 11, 817–844. [559]

Mardia, K. V. (1970), "Measures of Multivariate Skewness and Kurtosis With Applications," *Biometrika*, 57, 519–530. [561]

Marin, J.-M., Mengersen, K. L., and Robert, C. P. (2005), "Bayesian Modelling and Inference on Mixtures of Distributions," in *Bayesian Modelling and Inference on Mixtures of Distributions. Handbook of Statistics*, Vol. 25, eds. D. Dey and C. R. Rao, Amsterdam: North-Holland. [564]

Markowitz, H. (1952), "Portfolio Selection," *Journal of Finance*, 8, 77–91. [560]

Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370. [559,560]

Perez-Quiros, G., and Timmermann, A. (2000), "Firm Size and Cyclical Variations in Stock Returns," *The Journal of Finance*, 55, 1229–1262. [559]

——— (2001), "Business Cycle Asymmetries in Stock Returns: Evidence From Higher Order Moments and Conditional Densities," *Journal of Econometrics*, 103, 259–306. [559]

Robert, C. P., and Casella, G. (2004), *Monte Carlo Statistical Methods* (2nd ed.), New York: Springer. [563]

Tse, Y. K., and Tsui, A. K. C. (2002), "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations," *Journal of Business & Economic Statistics*, 20, 351–362. [559,560]

Vrontos, I. D., Dellaportas, P., and Politis, D. N. (2003a), "A Full-Factor Multivariate GARCH Model," *Econometrics Journal*, 6, 312–334. [559,563]

——— (2003b), "Inference for Some Multivariate ARCH and GARCH Models," *Journal of Forecasting*, 22, 427–446. [559,563]

Wong, C., and Li, W. (2001), "On a Mixture Autoregressive Conditional Heteroscedastic Model," *Journal of the American Statistical Association*, 96, 982–995. [559]