Bayesian estimation of the Gaussian mixture GARCH model

María Concepción Ausín,*

Department of Mathematics, University of A Coruña,

15071 A Coruña, Spain.

Pedro Galeano,

Department of Statistics and Operations Research,

University of Santiago de Compostela,

15782 Santiago de Compostela, A Coruña, Spain.

Abstract

Bayesian inference and prediction for a GARCH model where the innovations are assumed to follow a mixture of two Gaussian distributions is performed. The mixture GARCH model can capture the patterns usually exhibited by many financial time series such as volatility clustering, large kurtosis and extreme observations. A Griddy-Gibbs sampler implementation is proposed for parameter estimation and volatility prediction. Bayesian prediction of the Value at Risk is also addressed providing point estimates and predictive intervals. The method is illustrated using the Swiss Market Index.

Keywords: Bayesian Inference; GARCH models; Griddy-Gibbs sampler; Mixtures models; Value at Risk.

[1]Corresponding author. María Concepción Ausín, Departamento of Matemáticas, Facultad de Informática, Campus de Elviña, Universidad de A Coruña, 15071 A Coruña, Spain. Tel.: +34 981 167 000 (ext. 1301); fax: + 34 981 167 160. E-mail address: mausin@udc.es

1 Introduction

A large amount of theoretical and empirical research has been carried out on analysis of financial time series in the last two decades. The main features exhibited by many of these series are time-varying volatility, heavy-tailed distributions, large kurtosis and extreme events.

Many models have been proposed for modeling the time-varying volatility in financial time series, including the autoregressive conditional heteroskedastic (ARCH) model of Engle (1982), its generalization, the GARCH model of Bollerslev (1986), and the stochastic volatility model of Taylor (1986), see, for example, Shephard (1996) for a review. These models assume that the conditional variance of the series is a function of the current information and have been successful in modeling large periods of tranquility followed by small periods of high volatility.

However, less attention has been paid to explaining heavy-tailed distributions, large kurtosis and extreme events. The usual assumption in fitting models to financial data is that the returns, i.e., the first difference of the logarithm of the series, are conditionally normally distributed. However, the normal GARCH model is known to be inconsistent with high kurtosis, heavy tails and extreme events. The normal stochastic volatility model can capture some leptokurtosis but not large enough to explain the high sample kurtosis found in real data series. To explain these facts, Bollerslev (1987) proposed modeling the innovations of the GARCH model with a t-distribution, Nelson (1991) considered the use of the generalized error distribution and Engle and González-Rivera (1991) applied a non-parametric approach. Alternatively, other authors, including Bauwens, Bos and van Dijk (1999) and Bai, Russell and Tiao (2001, 2003), proposed modeling the innovations distribution with a mixture of two zero mean normal distributions with different variances. This is a model distribution which postulates that a large number of innovations are generated by a normal density with a small variance, while a small number of innovations are generated by a normal density with a large variance. This specification can capture volatility clustering, high kurtosis, heavy tails and the presence of extreme events. Other papers dealing with alternative mixture GARCH models are Wong and Li (2001), Haas, Mittnik and Paolella (2004) and Alexander and Lazar (2004).

Inference on ARCH and GARCH models has been traditionally carried out using maximum likelihood, quasi-maximum likelihood or the generalized method of moments, see e.g. Bollerslev,

Chou and Kroner (1992). There has been much less progress in the analysis of these models from the Bayesian perspective. The Bayesian methodology offers a natural way to introduce parameter uncertainty in the estimation of volatilities. Also, predictive distributions of the future volatilities can be obtained which are more informative than simple point forecasts. Furthermore, predictive distributions of Value at Risk (VaR) can also be obtained. This is one of the most important issues in Risk Management, see Jorion (2000), and the Bayesian approach provides a measure of precision for VaR estimates via predictive intervals. Finally, the recent development of modern Bayesian computational methods based on Markov Chain Monte Carlo (MCMC) can be utilized to address the complexity of these models, see Geweke (1994), Bauwens and Lubrano (1998), Müller and Pole (1998), Bauwens et al. (1999), Nakatsuma (2000) and Vrontos, Dellaportas and Politis (2000).

In this paper, we present a procedure for Bayesian inference and prediction of the GARCH model with Gaussian mixture innovations based on MCMC methods. Although mixture models are intrinsically difficult to analyze, the Bayesian approach allows for data augmentation techniques where indicator variables can be introduced to simplify the likelihood and the derivation of the posterior distributions. Bayesian estimation of mixture models has been broadly studied in non-dynamic settings, see e.g. Diebolt and Robert (1994). Our Bayesian approach combines these ideas with the Griddy-Gibbs sampling algorithm proposed by Ritter and Tanner (1992). This method is a modified Gibbs sampler where the conditional posterior distributions are approximated by numerical integration methods. The Griddy-Gibbs sampler has also been used by Bauwens and Lubrano (1998) for a GARCH model with t-distributed errors. This approach is easier to implement than Metropolis-Hastings algorithms or importance sampling and, although it is more costly in computer time, the problems in finding an appropriate proposal distribution or importance function in these other methods are avoided.

The rest of this paper is organized as follows. Section 2 presents the Gaussian mixture GARCH model and illustrates its flexibility in capturing the patterns exhibited by financial time series. Section 3 describes a Bayesian analysis of this model given an uninformative joint prior distribution for the model parameters and a Griddy-Gibbs algorithm for sampling the posterior distribution. The problem of volatilities and VaR prediction is also addressed. Section 4 presents some Monte Carlo simulations which illustrates the accuracy in the estimation of the parameters and prediction of volatilities and VaR. Various comparisons with other Bayesian and non-Bayesian approaches is

3

also included in this section. Section 5 illustrates our procedure for the log return series of the SMI (Swiss Market Index) which is a clear example of a series with large kurtosis and extreme returns.

2   The GARCH Model with Gaussian mixture errors

The GARCH(p,q) model for a series y_t is given by,

$$y_t = \mu + \sqrt{h_t}\epsilon_t, \tag{1}$$

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \left(y_{t-i} - \mu\right)^2 + \sum_{j=1}^{q} \beta_j h_{t-j},$$

where h_t is the conditional variance of y_t given the previous information $I_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$, and ϵ_t are iid with zero mean and unit variance. We assume that the initial variance h_0 is a known constant and the parameters ω, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)$ follow the restrictions, $\omega > 0, \alpha_1 \geq 0, \ldots, \alpha_p \geq 0, \beta_1 \geq 0, \ldots, \beta_q \geq 0$, to ensure positivity of h_t, for all t, and $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$, to ensure covariance stationarity.

Neither the Gaussian GARCH model, i.e., assuming that ϵ_t is Gaussian distributed, nor the GARCH-t model, i.e., assuming that ϵ_t is t-distributed, are able to match volatility dynamics and large kurtosis, as will be shown below, see e.g. Bai et al. (2003). These authors, among others, suggested the use of a mixture of two Gaussian distributions, that is, $\epsilon_t \sim$ mixture Gaussian (λ, ρ), i.e.,

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

where $0 < \lambda < 1$ and,

$$\sigma^2 = \frac{1}{\rho + \frac{1-\rho}{\lambda}}, \tag{3}$$

so that $var\left(\epsilon_t\right) = 1$. Thus, the innovations ϵ_t are generated from a Gaussian distribution with variance σ^2 with probability ρ, or from a Gaussian distribution with variance σ^2/λ with probability $1 - \rho$. Note that the variance of the first component is always less than one because of (3) and the variance of the second component increases as λ goes to zero. Additionally, we also impose the condition that the probability ρ is restricted to the interval (0.5,1) to ensure that the component

with largest number of elements is the one with smallest variance.

Some of the reasons of using the mixture distribution to model the innovations are as follows. First, this is the distribution used in the variance inflation model of Box and Tiao (1968), which has been shown to be successful in modelling outliers and extreme events in linear models, see e.g. Peña and Guttman (1993). Therefore, it is expected that the extreme returns, which can cause the high sample kurtosis found in practice, are generated by the component with larger variance. Second, this distribution is able to generate high kurtosis. The excess kurtosis, K_y, of a series y_t is defined as the difference between the kurtosis of the series, if it exists, and the kurtosis of the normal distribution which is equal to 3,

$$K_y = \frac{E\left[(y_t - \mu)^4\right]}{E\left[(y_t - \mu)^2\right]^2} - 3, \tag{4}$$

and, if positive, measures how large is the kurtosis compared with the one of the normal distribution. In the case of Gaussian innovations, (4) reduces to,

$$K_g = \frac{6\gamma}{1 - 2\gamma}, \tag{5}$$

where $\gamma = \sum_{i=1}^{\infty} \psi_i$ and ψ_i are the coefficients of the polynomial equation given by $\psi(B)\phi(B) = \beta(B)$, with $\psi(B) = 1 + \sum_{i=1}^{\infty} \psi_i B^i$, $\phi(B) = 1 - \sum_{i=1}^{r} (\alpha_i + \beta_i) B^i$, $r = max(p,q)$ and $\beta(B) = 1 - \sum_{i=1}^{q} \beta_i B^i$, which depends only on the parameters of the volatility equation, see Bai et al. (2003). For non-normal innovations, Bai et al. (2003) showed that,

$$K_y = \frac{K_\epsilon + K_g + \frac{5}{6} K_\epsilon K_g}{1 - \frac{1}{6} K_\epsilon K_g}, \tag{6}$$

where K_ϵ is the excess kurtosis of ϵ_t and K_g is given in (5), provided that K_ϵ and K_g exist. Thus, the overall excess kurtosis of y_t depends symmetrically on the excess kurtosis induced by non-normal innovations, K_ϵ, and the one induced by volatility clustering, K_g.

If ϵ_t follows a Student's t distribution with ν degrees of freedom, then $K_\epsilon = 6/(\nu - 4)$. Note that the second and fourth moments of y_t only exist if $\nu > 4$, implying that the excess kurtosis K_ϵ

should be positive. In this case,

$$K_y = \frac{6 + (\nu + 1)K_g}{\nu - 4 - K_g},$$

which is positive, i.e., y_t is leptokurtic, only if $\nu > 4 + K_g$. In practice, the degrees of freedom parameter, ν, is either fixed to be larger than or equal to 5, in which case the implied kurtosis of the estimated model does not match the observed kurtosis, or it is estimated, in which case its estimate is usually smaller than 5, and the estimated excess kurtosis does not exist, see Bai et al. (2003).

However, if ϵ_t follows a mixture distribution (2), the excess kurtosis of the innovations is given by,

$$K_\epsilon = \frac{3\rho (1 - \rho) \left(\frac{1}{\lambda} - 1\right)^2}{\left(\rho + \frac{1}{\lambda} (1 - \rho)\right)^2},$$

which exists for every value in the domain of ρ and λ and can take any positive value. To show this, note that K_ϵ tends to zero when λ tends to one, and tends to infinity when ρ and λ tend to one and zero, respectively. Figure 1 shows some values of the kurtosis coefficient for values of ρ and λ in the interval $(0.5, 0.99)$ and $(0.01, 0.99)$, respectively. Observe that the value of K_ϵ is larger when ρ and λ are close to one and zero simultaneously. Finally, K_y in (6) is positive if $1 - K_\epsilon K_g/6 > 0$, and, for any possible value of K_g, this condition is verified for certain values of ρ and λ as shown before. This illustrates that model (1) with innovations (2) can capture the large kurtosis typically observed in financial time series. More reasons of using this mixture distribution can be found in Bai et al. (2001, 2003).

3 Bayesian Inference for the Gaussian mixture GARCH model

In this section, we describe how to carry out Bayesian inference for the model (1) using a Gibbs sampling method. Following Bauwens and Lubrano (1998), we make use of the Griddy-Gibbs sampling approach which is based on a combination of a Gibbs sampler with a numerical integration procedure. Several reasons for using this approach are given in Sections 3 and 4, where it is compared with other Bayesian methods, such as Metropolis-Hastings, and classical methods, such as Maximum Likelihood Estimation.

Let $\boldsymbol{\theta} = (\rho, \lambda, \mu, \omega, \boldsymbol{\alpha}, \boldsymbol{\beta})$, be the parameter vector of model (1). Given a series, y_t, $t = 1, \ldots, T$,
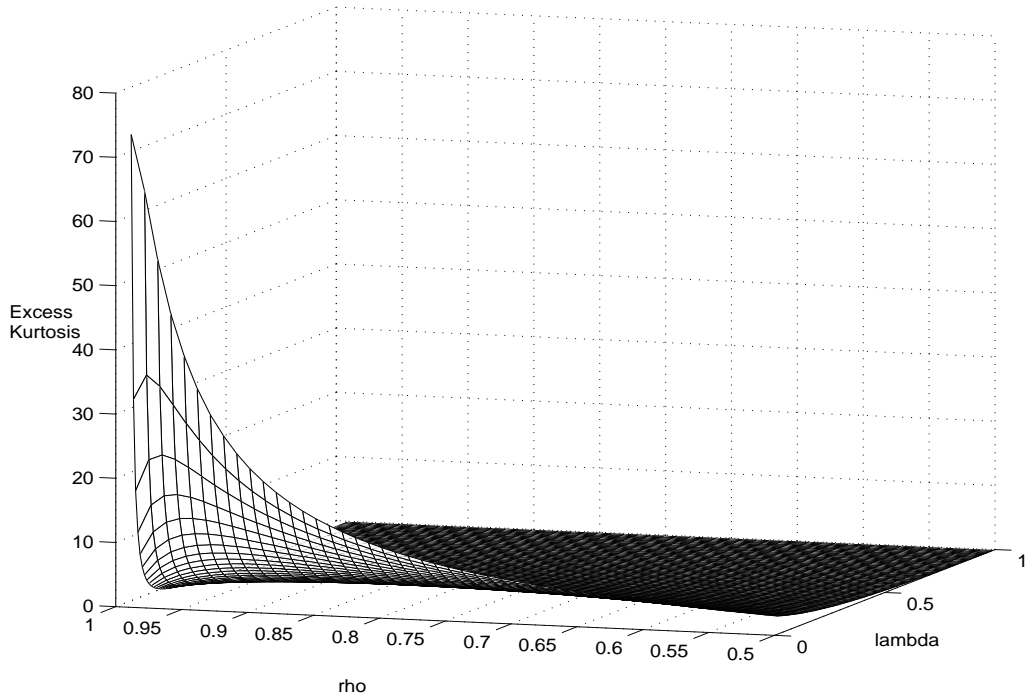
Figure 1: Excess kurtosis of the innovations as a function of the parameters λ and ρ

the likelihood function takes a complicated form but can be simplified by introducing the usual missing data formulation for mixture setups, see e.g. Diebolt and Robert (1994), where a set of latent variables $z_1, \ldots, z_T$ are defined such that,

$$
z_t = \begin{cases} 1, & \text{with probability } \rho, \\ 2, & \text{with probability } 1 - \rho, \end{cases} \tag{7}
$$

for $t = 1, ..., T$. With this approach, the observed series, $\mathbf{y} = (y_1, \ldots, y_T)$, is completed with a missing data set, $\mathbf{z} = (z_1, \ldots, z_T)$, indicating the specific component of the mixture from which every observation is assumed to arise. Then, conditional on these indicators, we have that,

$$
y_t \mid h_t, \ z_t \sim \begin{cases} N\left(\mu, \sigma^2 h_t\right), & \text{if } z_t = 1, \\ N\left(\mu, \frac{\sigma^2}{\lambda} h_t\right), & \text{if } z_t = 2. \end{cases} \tag{8}
$$

Therefore, the likelihood separates into two parts, each one concerning the data assigned to each

of the two mixture components,

$$l\left(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{z}\right) \propto \prod_{t:z_t=1} \left[\rho\left(\sigma^2 h_t\right)^{-1/2} \exp\left\{-\frac{1}{2}\frac{(y_t-\mu)^2}{\sigma^2 h_t}\right\}\right] \times$$

$$\prod_{t:z_t=2} \left[(1-\rho)\left(\frac{1}{\lambda}\sigma^2 h_t\right)^{-1/2} \exp\left\{-\frac{1}{2}\frac{\lambda\left(y_t-\mu\right)^2}{\sigma^2 h_t}\right\}\right].$$

In order to carry out Bayesian inference, we also need to define prior distributions for the model parameters, $\boldsymbol{\theta}$. Let us assume that the prior distributions of the parameters are uniformly distributed over their respective domains, i.e.,

$$f\left(\rho\right) \sim U\left(0.5, 1\right), \qquad f\left(\lambda\right) \sim U\left(0, 1\right), \qquad f\left(\mu\right) \sim U\left(-\infty, \infty\right),$$
$$f\left(\omega\right) \sim U\left(0, \infty\right) \qquad f\left(\boldsymbol{\alpha}\right) \sim U\left[0, 1\right]^p, \qquad f\left(\boldsymbol{\beta}\right) \sim U\left[0, 1\right]^q, \tag{9}$$

and restricted to the stationary region. Note that using model (1), we can consider flat priors for all the parameters. This is not possible for the GARCH model with t-distributed errors where a flat prior for the degrees of freedom parameter leads to an improper posterior distribution as was shown in Bauwens and Lubrano (1998). Moreover, there is high sensitivity to the choice of a proper prior distribution for the degrees of freedom parameter in this model.

Given the data and the priors specified above, it is very complicated to obtain an analytical expression of the posterior distribution of the parameter vector, $p\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right)$. However, Bayesian inference may be performed using the Gibbs sampling procedure, see e.g. Tierney (1994) for an extensive analysis. Under mild conditions, given an initial value $\boldsymbol{\theta}^{(0)}$, the Gibbs sampler can produce a Markov chain $\left\{\boldsymbol{\theta}^{(n)} : n = 0, \ldots, N\right\}$, where $\boldsymbol{\theta}^{(n)} = \left(\rho^{(n)}, \lambda^{(n)}, \mu^{(n)}, \omega^{(n)}, \boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right)$, which has equilibrium distribution $p\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right)$. Gibbs sampling is carried out by cycling repeatedly through draws of each parameter conditional on the remaining parameters.

Thus, we now obtain the conditional posterior distribution of each parameter. Firstly, from (7) and (8), the conditional posterior probability that the observation y_t has been generated by the first mixture component is,

$$f\left(z_t = 1 \mid y_t, \boldsymbol{\theta}\right) = \frac{\rho \exp\left\{-\frac{1}{2}\frac{(y_t-\mu)^2}{\sigma^2 h_t}\right\}}{\rho \exp\left\{-\frac{1}{2}\frac{(y_t-\mu)^2}{\sigma^2 h_t}\right\} + (1-\rho)\left(\frac{1}{\lambda}\right)^{-1/2} \exp\left\{-\frac{1}{2}\frac{\lambda(y_t-\mu)^2}{\sigma^2 h_t}\right\}}, \tag{10}$$

and clearly, the probability of having been generated by the second component is one minus expression (10).

The conditional posterior density $f\left(\rho \mid \boldsymbol{\theta}_{-\rho}, \mathbf{y}, \mathbf{z}\right)$, where $\boldsymbol{\theta}_{-\rho}$ denotes the remaining parameters except ρ, has the following kernel,

$$\kappa\left(\rho \mid \boldsymbol{\theta}_{-\rho}, \mathbf{y}, \mathbf{z}\right) = \frac{\rho^{T_1}(1-\rho)^{T_2}}{\sigma^T} \exp\left\{-\frac{S_1 + \lambda S_2}{2\sigma^2}\right\}, \tag{11}$$

where $T_i = \#\{z_t = i\}$, the number of observations assigned to the i-th component, and $S_i = \sum_{t=1}^T \left\{(y_t - \mu)^2 / h_t : z_t = i\right\}$, for $i = 1, 2$. Recall that σ is a function of (ρ, λ) as given in (3), and h_t is a function of $(\mu, \omega, \boldsymbol{\alpha}, \boldsymbol{\beta})$ as given in (1).

The conditional posterior density $f\left(\lambda \mid \boldsymbol{\theta}_{-\lambda}, \mathbf{y}, \mathbf{z}\right)$ has a kernel given by,

$$\kappa\left(\lambda \mid \boldsymbol{\theta}_{-\lambda}, \mathbf{y}, \mathbf{z}\right) = \frac{\lambda^{T_2/2}}{\sigma^T} \exp\left\{-\frac{S_1 + \lambda S_2}{2\sigma^2}\right\}, \tag{12}$$

while the kernel of the conditional posterior density $f\left(\mu \mid \boldsymbol{\theta}_{-\mu}, \mathbf{y}, \mathbf{z}\right)$ is given by,

$$\kappa\left(\mu \mid \boldsymbol{\theta}_{-\mu}, \mathbf{y}, \mathbf{z}\right) = \prod_{t:z_t=1}\left[h_t^{-1/2}\exp\left\{-\frac{1}{2}\frac{(y_t-\mu)^2}{\sigma^2 h_t}\right\}\right] \times \prod_{t:z_t=2}\left[h_t^{-1/2}\exp\left\{-\frac{1}{2}\frac{\lambda(y_t-\mu)^2}{\sigma^2 h_t}\right\}\right]. \tag{13}$$

The kernels of the conditional posterior densities $f\left(\omega \mid \boldsymbol{\theta}_{-\omega}, \mathbf{y}, \mathbf{z}\right)$, $f\left(\alpha_i \mid \boldsymbol{\theta}_{-\alpha_i}, \mathbf{y}, \mathbf{z}\right)$, $i = 1, \ldots, p$, and $f\left(\beta_j \mid \boldsymbol{\theta}_{-\beta_j}, \mathbf{y}, \mathbf{z}\right)$, $j = 1, \ldots, q$, have the same expression (13), but restricted to the corresponding stationary region on their domains.

The posterior densities (11), (12) and (13) are not of a simple form and thus, random samples can not be easily generated. The Griddy-Gibbs sampler, introduced by Ritter and Tanner (1992), solves this problem by evaluating each kernel function over a grid of points, approximating the cumulative distribution function using a numerical integration method, and generating a draw from each conditional posterior distribution by inversion of the cumulative distribution function at a random value sampled uniformly in $[0, 1]$. Given these conditional posteriors, we propose the following Griddy-Gibbs sampler:

1. Let $n = 0$. Set initial values $\boldsymbol{\theta}^{(0)}$.

2. Update the indicators, $\mathbf{z}$, by sampling from $\mathbf{z}^{(n+1)} \sim \mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^{(n)}$.

3. Update ρ by sampling from $\rho^{(n+1)} \sim \rho \mid \boldsymbol{\theta}_{-\rho}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}$. For that:

   (a) Approximate the following integrals using a numerical integration method,

   $$\Phi_g \simeq \int_{\rho_1}^{\rho_g} \kappa\left(\rho \mid \theta_{-\rho}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}\right) d\rho, \qquad g = 2, ..., G, \tag{14}$$

   where $\rho_1, \ldots, \rho_G$ is a grid of ordered points on the domain of ρ.

   (b) Generate $u \sim U\left(0, \Phi_G\right)$ and invert $\Phi\left(\rho \mid \boldsymbol{\theta}_{-\rho}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}\right)$ by numerical interpolation to get a draw $\rho^{(n+1)}$.

4. Update λ by sampling from $\lambda^{(n+1)} \sim \lambda | \rho^{(n+1)}, \mu^{(n)}, \omega^{(n)}, \boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}$, as in 3.

5. Update μ by sampling from $\mu^{(n+1)} \sim \lambda | \rho^{(n+1)}, \lambda^{(n+1)}, \omega^{(n)}, \boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}$, as in 3.

6. Update ω by sampling from $\omega^{(n+1)} \sim \omega | \rho^{(n+1)}, \lambda^{(n+1)}, \mu^{(n+1)}, \boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}$, as in 3.

7. Update α_i, for $i = 1, \ldots, p$, by sampling from,

   $$\alpha_i^{(n+1)} \sim \alpha_i | \rho^{(n+1)}, \lambda^{(n+1)}, \mu^{(n+1)}, \omega^{(n+1)}, \alpha_1^{(n+1)}, \ldots, \alpha_{i-1}^{(n+1)}, \alpha_{i+1}^{(n)}, \ldots, \alpha_p^{(n)}, \boldsymbol{\beta}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)},$$

   as in 3.

8. Update β_j, for $j = 1, \ldots, q$, by sampling from,

   $$\beta_j^{(n+1)} \sim \beta_j | \rho^{(n+1)}, \lambda^{(n+1)}, \mu^{(n+1)}, \omega^{(n+1)}, \boldsymbol{\alpha}^{(n+1)}, \beta_1^{(n+1)}, \ldots, \beta_{j-1}^{(n+1)}, \beta_{j+1}^{(n)}, \ldots, \beta_q^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)},$$

   as in 3.

9. Let $n = n + 1$ and go to 2 unless $n = N$.

Some comments on the proposed Griddy-Gibbs sampler are in order. Firstly, following Bauwens and Lubrano (1998), we use the trapezoidal rule of integration (see Davis and Rabinowitz, 1975) for approximating the integrals in (14) with a fixed grid of equidistant points. Other alternatives are also possible such as the adaptive Simpson and Lobatto quadratures. Adaptive methods have the advantage of using a variable grid that is modified to have more points where the mass of the

posterior distribution is concentrated. However, we have chosen a fixed grid of points because it is possible to obtain a smooth estimation of the marginal posterior densities of each parameter, as described below. In our examples, we have chosen 40 point grids which when compared with lower and higher numbers of points seem to be accurate enough. Another important issue is the choice of the bounds of integration. These come from the prior restrictions (9) for the parameters ρ, λ, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, but we should also restrict the domain of the parameters ω and μ to some intervals where the value of their posterior densities is big enough to contribute to the integrals. We have taken the sample unconditional variance of the series, $\widehat{\sigma}_y^2$, as the maximum possible value of ω, while we have allowed μ to be in the interval $(\overline{y} - 4\widehat{\sigma}_y/\sqrt{T}, \overline{y} + 4\widehat{\sigma}_y/\sqrt{T})$. We have found that these choices seem to be large enough in practice. Note that the bounds of integration are inappropriate if the tails of a marginal posterior density look truncated. Finally, we use linear interpolation between adjacent points in point 3.b.

The Griddy-Gibbs sampler allows us to obtain a smooth estimation of the marginal posterior density of each parameter. For instance, we estimate the posterior density of ρ for each point, ρ_g, of the grid using Rao-Blackwellization, see Casella and Robert (1996), as follows,

$$f\left(\rho_g \mid \mathbf{y}\right) \simeq \frac{1}{N-s} \sum_{n=s+1}^{N} \frac{\kappa\left(\rho_g \mid \boldsymbol{\theta}_{-\rho}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}\right)}{\Phi_G}, \qquad g = 2, ..., G, \tag{15}$$

where s is the number of burn-in draws required to reach the equilibrium distribution and Φ_G is given in (14). Note that, as commented before, this approximation can be carried out because we have chosen a fixed grid instead of a variable one. In Bauwens and Lubrano (1998), this approach is referred as "conditioning" and it is also used to reduce the variance in the estimation of the moments of the marginal posterior distributions by estimating for example, $E\left[\rho \mid \mathbf{y}\right]$ with $\sum_{n=s+1}^{N} E\left[\rho \mid \theta_{-\rho}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}\right] /(N-s)$ instead of considering the usual sample mean, $\sum_{n=s+1}^{N} \rho^{(n)}/(N-s)$. In practice we have found no significant differences between using the conditioning estimators for the posterior mean and variance or using the usual sample moments of draws.

In GARCH models, the estimation of in-sample volatilities and prediction of future volatilities is essential. Using the MCMC output, we can easily obtain a sample from the posterior distribution of each conditional variance, h_t, for $t = 1, \ldots, T$, by calculating the value of the conditional variance

11

for each draw $\boldsymbol{\theta}^{(n)}$, denoted by $h_t^{(n)}$. With these samples, we can estimate the posterior mean for each conditional variance by,

$$E\left[h_t \mid \mathbf{y}\right] \simeq \frac{1}{N-s} \sum_{n=s+1}^{N} h_t^{(n)}. \tag{16}$$

Also the posterior median and 95% credible intervals can be obtained by just calculating the median and the .025 and .975 quantiles of each posterior sample, respectively.

Of particular interest are the predictive distribution and intervals for the one-step ahead volatility, h_{T+1}. Analogously, we can obtain a sample from the predictive distribution of h_{T+1} and 95% predictive intervals. In a similar way, we can estimate the predictive density of y_{T+1},

$$f\left(y_{T+1}|\mathbf{y}\right) = \int_{\boldsymbol{\theta}} f\left(y_{T+1}|\mathbf{y}, \boldsymbol{\theta}\right) f\left(\boldsymbol{\theta}|\boldsymbol{y}\right) d\boldsymbol{\theta}. \tag{17}$$

The density $f\left(y_{T+1}|\mathbf{y}, \boldsymbol{\theta}\right)$ is a Gaussian mixture with mean μ and variance h_{T+1}. Thus, the predictive density (17) can be estimated as the mean of the density functions obtained for all the draws of the MCMC sample,

$$f\left(y_{T+1}|\boldsymbol{y}\right) \simeq \frac{1}{N-s} \sum_{n=s+1}^{N} f\left(y_{T+1}|\boldsymbol{y}, \boldsymbol{\theta}^{(n)}\right).$$

Suppose now that we are interested in the prediction of $h_{T+2}, \ldots, h_{T+s}$ and $y_{T+2}, \ldots, y_{T+s}$. We cannot obtain samples from their predictive distributions using the same procedure as for h_{T+1} and y_{T+1} because the values of y_t are unknown for $t \geq T+1$. However, given the observed series and the MCMC sample, we can generate values of the predictive distributions of h_{T+i} and y_{T+i}, for $i = 2, \ldots, s$, using the following simulation procedure based on the one proposed in Bauwens and Lubrano (1998, 2002). For each value of the parameters, $\boldsymbol{\theta}^{(n)}$, the value of $h_{T+1}^{(n)}$ is known and $y_{T+1}^{(n)}$ can be generated from its mixture distribution. Sequentially, given $y_{T+i-1}^{(n)}$, the value of $h_{T+i}^{(n)}$ is known and $y_{T+i}^{(n)}$ can be generated, for $i = 2, \ldots, s$. Using this procedure, we obtain samples from the predictive distributions, $f(h_{t+i}|\boldsymbol{y})$ and $f(y_{t+i}|\boldsymbol{y})$, for $i = 2, \ldots, s$, which allow us to estimate their quantiles. For example, the probability that a future volatility is larger than a given threshold can be estimated with the proportion of observations in the sample which are larger than this threshold. Moreover, by replicating M times the proposed simulation procedure, we obtain samples $y_{T+1}^{(n,m)}, \ldots, y_{T+s}^{(n,m)}$ and $h_{T+1}^{(n,m)}, \ldots, h_{T+s}^{(n,m)}$, for $m = 1, \ldots, M$, which allow to construct predictive intervals for this probability. Furthermore, these samples will provide VaR estimates and

predictive intervals, as shown below.

As a result of the presence of extreme events in financial series, VaR has become an useful and widely used measure of market risk. Broadly speaking, VaR is the maximum potential loss associated with an unfavorable movement in market prices during a given time period with certain probability. In other words, the VaR of an asset means that the losses over a time period s should exceed VaR only the π percent of the time. Usually, the probability π takes a small value such as 0.01 or 0.05 (see Jorion (2000) for a more detailed treatment). For a log return series, y_t, the $s-$period $\pi\%$ VaR for an amount of investment A is given by,

$$\text{VaR} = A \times p_\pi,$$

where p_π is the $\pi-$quantile of the distribution of $y_T[s] = y_{T+1} + \ldots + y_{T+s}$, i.e. $\pi = Pr(y_T[s] \leq p_\pi)$.

Now, from the simulation procedure described above, we can construct predictive intervals for the $s-$period $\pi\%$ VaR as follows. For each replication $m = 1, \ldots, M$, we have an estimate of the VaR given by,

$$\text{VaR}^{(m)} = A \times p_\pi^{(m)},$$

where $p_\pi^{(m)}$ is the empirical $\pi-$quantile of the sample of sums $y_{T+1}^{(1,m)} + \ldots + y_{T+s}^{(1,m)}, \ldots, y_{T+1}^{(N,m)} + \ldots + y_{T+s}^{(N,m)}$ for each value of m. Thus, we have a sample of size M from the predictive distribution of the VaR, which allows us to estimate the predictive mean, median and confidence intervals.

4   Computational results

The computational results in this section and the analysis of the real data example in the next one have been carried out by means of various routines written by the authors in MATLAB (The MathWorks, Inc.). We use the Gaussian and uniform random number generators implemented in MATLAB.

4.1 Illustration of the Griddy-Gibbs sampler

We illustrate our MCMC procedure with three artificial series simulated from the following GARCH(1,1) model,

$$
\begin{aligned}
y_t &= 0.01 + \sqrt{h_t}\epsilon_t, & (18) \\
h_t &= 0.001 + 0.15\,(y_{t-1} - 0.01)^2 + 0.7 h_{t-1},
\end{aligned}
$$

where ϵ_t follows a Gaussian mixture as in (2) with parameters $\rho = 0.9$ and $\lambda = 0.15$. Thus, the excess kurtosis of ϵ_t and y_t are given by $K_\epsilon = 3.53$ and $K_y = 8.84$, respectively. We generated three series with 506, 1006 and 2006 observations, respectively, from the model (18) and used the first $T = 500$, 1000 and 2000 observations for estimation. The sample mean, variance and kurtosis of each series are 1.1269×10^{-2}, 4.2763×10^{-3} and 5.1443, 9.3085×10^{-3}, 5.2748×10^{-3} and 6.0114, and 9.4826×10^{-3}, 5.5029×10^{-3} and 7.1317, respectively. For each series, we generated $N = 20000$ runs of the Markov chain with initial values $\boldsymbol{\theta}^{(0)} = (0.5, 0.5, 0, 0.5, 0.5, 0.3)$, (although different initial values give similar results) and discarded the initial 10000 runs. To assess the convergence of the Markov chain, we used the convergence diagnostic proposed in Geweke (1992) based on testing for the equality of the means of the first and last part of the chain. For instance, consider the parameter ρ and let $n_1 = 0.1(N - s)$ and $n_2 = 0.5(N - s)$. Let $\overline{\rho}_1$ and $\overline{\rho}_2$ be the sample mean for the first n_1 runs and the last n_2 runs, respectively, for one of the series. The convergence diagnostic is given by,

$$
CD = \frac{(\overline{\rho}_1 - \overline{\rho}_2)}{\left[\dfrac{\widehat{S}_1^\rho(0)}{n_1} + \dfrac{\widehat{S}_2^\rho(0)}{n_2}\right]^{\frac{1}{2}}}, \tag{19}
$$

where $\widehat{S}_i^\rho(0)$ is the spectral density estimate for n_i runs, $i = 1, 2$. If the chain has converged, the statistic (19) has asymptotically the standard Gaussian distribution. The values of the statistic (19) for each parameter in $\boldsymbol{\theta}$ is shown in Table 1 for each of the three simulated series. These indicate that the chains have converged in all cases.

Figure 2 shows the convergence diagrams of the posterior sample for each parameter for the series with sample size $T = 1000$. Similar plots are obtained for the other simulated series which are not shown here to save space. Figure 3 shows the histograms of the MCMC output for the same

14

Table 1: Values of the Geweke's statistic for each parameter.

	$T = 500$	$T = 100$	$T = 2000$
ρ	1.03	-0.90	-1.65
λ	-0.98	1.04	0.89
μ	0.23	1.22	0.97
ω	-0.62	0.07	-1.69
α	-1.73	0.81	-1.85
β	0.77	-0.43	1.79

series. Also shown are the marginal densities for each parameter obtained using (15). Observe that our Bayesian procedure captures the asymmetry of the posterior distributions of the parameters ρ, ω and β. The asymmetry of these marginal posterior densities is larger for the series with $T = 500$ (not reported) and, in general, the variances are also larger, as expected. On the contrary, the densities are more symmetric and the variances are smaller for the series with $T = 2000$ (not reported). In Table 2, it is shown the parameter estimation results for the three series. Columns 2, 4 and 6 in Table 2 show the mean and standard deviation of the posterior distribution of each parameter and series, respectively. Columns 3, 5 and 7 show the posterior median and mean absolute deviation for each series, respectively. Note that these estimates are close to the true parameters in all cases. In particular, for $T = 500$, when the posterior distribution is asymmetric, the posterior median seems to be a better estimate than the mean, as should be expected.

Finally, the last row of Table 2 shows the computational time required to finish the algorithm for 20000 runs in each case. We note that with a smaller number of iterations, accurate enough estimation results were also obtained. On the other hand, the relationship between the sample size and the computational time is apparently linear. Also, we have observed that the complexity in estimating mixture GARCH models of higher orders is approximately increased by one over six the computational time for a mixture GARCH(1,1), for each new parameter. For example, 16.2 minutes were needed for a mixture GARCH(2,1) simulated series of size $T = 500$ with accurate results.

Table 3 shows the correlations between parameters for the series with length $T = 1000$. Similar correlations were obtained for the other simulated series. Note that although some of the parameters are highly correlated, the benefit of using a grid is that every part of the space is explored, meaning
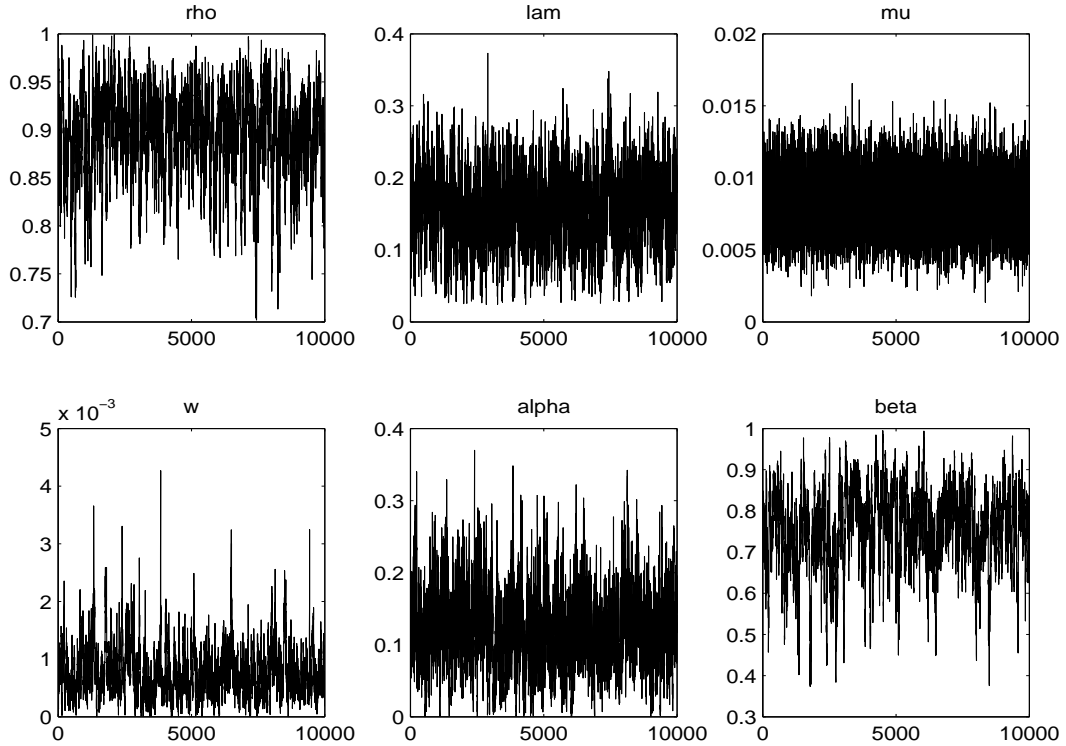
Figure 2: Convergence diagrams of the posterior sample for each parameter for the series of sample size $T = 1000$.

Table 2: Parameter estimation results.

| Parameter	$T = 500$ Mean std	$T = 500$ Median mad	$T = 1000$ Mean std	$T = 1000$ Median mad	$T = 2000$ Mean std	$T = 2000$ Median mad
$\rho = .9$	.8724 .0854	.8932 .0655	.8955 .0446	.9016 .0353	.9163 .0347	.9187 .0275
$\lambda = .15$	.2024 .0796	.1991 .0638	.1526 .0496	.1520 .0398	.1286 .0422	.1293 .0342
$\mu = .01$	.0093 .0026	.0093 .0020	.0083 .0019	.0084 .0015	.0092 .0014	.0092 .0011
$\omega = .001$	.0011 .0006	.0010 .0005	.0007 .0004	.0006 .0003	.0009 .0004	.0009 .0003
$\alpha = .15$	.1582 .0669	.1494 .0520	.1211 .0507	.1178 .0399	.1205 .0472	.1190 .0371
$\beta = .7$	.5879 .1815	.6202 .1459	.7539 .0954	.7647 .0746	.7199 .0987	.7256 .0791
Time	13.99 min.		23.20 min.		41.66 min.	

16
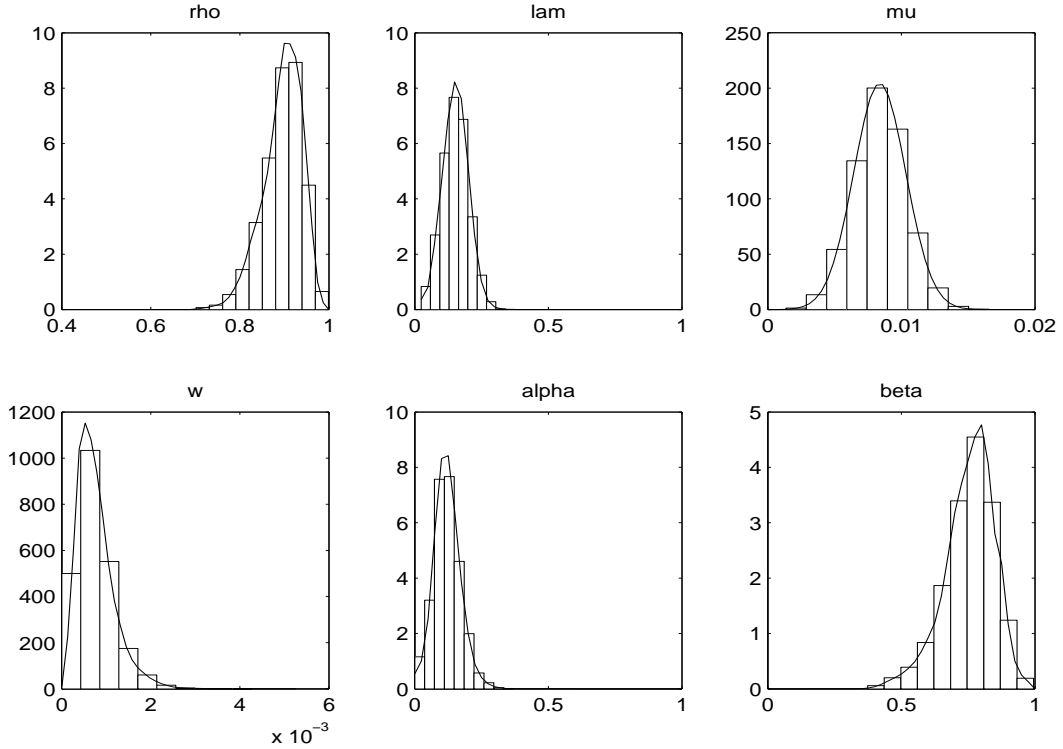
Figure 3: Histograms and marginal posterior densities of the MCMC output for each parameter for the series of sample size $T = 1000$.

that the sampler does not get trapped in any particular region. Scatter plots of the draws after different number of runs (not reported here) suggest that there is no trapping.

Figure 4 shows the true and posterior mean (16) for the last 100 in-sample volatilities of the three series with their 95% credible intervals as described in Section 3. Note the accuracy of the estimation of these unobserved volatilities and that the Bayesian credible intervals includes the true generated volatilities for all time periods. Figure 4 also shows the predicted mean of the one-step

Table 3: Correlations between parameters for the series of size $T = 1000$.

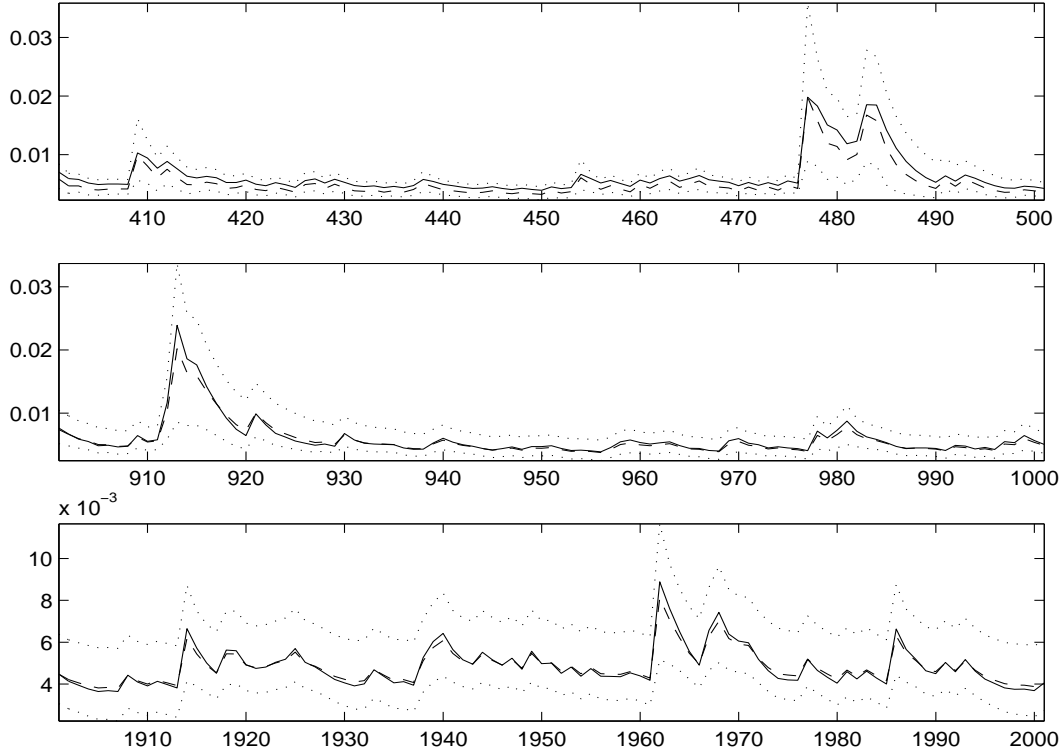|	ρ	λ	μ	ω	α	β
ρ	1	-0.59	-0.01	-0.05	-0.00	0.04
λ		1	-0.00	-0.11	-0.15	-0.05
μ			1	0.01	-0.07	0.02
ω				1	0.37	-0.88
α					1	-0.59
β						1

17

Figure 4: True (solid line) and predictive mean (dashed line) of the last 100 conditional volatilities with 95% credible intervals (dotted lines) for the simulated series with $T = 500$ (top), $T = 1000$ (center) and $T = 2000$ (bottom).

ahead volatility, h_{T+1}, for each series, where $T + 1 = 501$, 1001 and 2001, respectively, and their predictive intervals. Their numeric values can be examined in Table 4, where predictive means and 95% intervals for $h_{T+1}, \ldots, h_{T+6}$, are also shown. These estimations have been computed with the samples obtained using the simulation procedure described in Section 3. Histograms of these samples for the series with $T = 1000$ are plotted in Figure 5. Note that the predictive distribution of h_{T+1} is apparently symmetric, while the distribution of h_{T+i}, for $i = 2, ..., 6$ are highly skewed to the right. As a matter of fact, they are so right-skewed that only the data under the 99% percentile are considered for the histograms to get a better illustration of the shape of these distributions. The same features are observed in the histograms of the future volatilities for the other simulated series.

Finally, Table 5 shows the predictive means and 95% predictive intervals of VaR at future times $T + 1, \ldots, T + 6$ with probability $\pi = 0.01$ and initial investment of $A = 1$ unit. Observe that these intervals are quite symmetric and their amplitudes are not very wide, meaning that the

Table 4: True, predictive mean and 95% predictive interval for future volatilities at times $T + 1, \ldots, T + 6$ for the three simulated series.

|		$T = 500$			$T = 1000$			$T = 2000$	
s	True	Mean	Interval	True	Mean	Interval	True	Mean	Interval
1	.0042	.0037	(.0026, .0051)	.0050	.0060	(.0042, .0086)	.0040	.0050	(.0034, .0070)
2	.0040	.0039	(.0025, .0073)	.0045	.0060	(.0040, .0108)	.0040	.0052	(.0033, .0091)
3	.0038	.0041	(.0024, .0085)	.0049	.0060	(.0038, .0126)	.0039	.0053	(.0032, .0107)
4	.0036	.0042	(.0024, .0092)	.0044	.0061	(.0036, .0139)	.0038	.0055	(.0031, .0119)
5	.0037	.0043	(.0023, .0098)	.0050	.0061	(.0034, .0148)	.0036	.0056	(.0031, .0128)
6	.0039	.0044	(.0023, .0103)	.0045	.0061	(.0033, .0155)	.0042	.0057	(.0030, .0134)
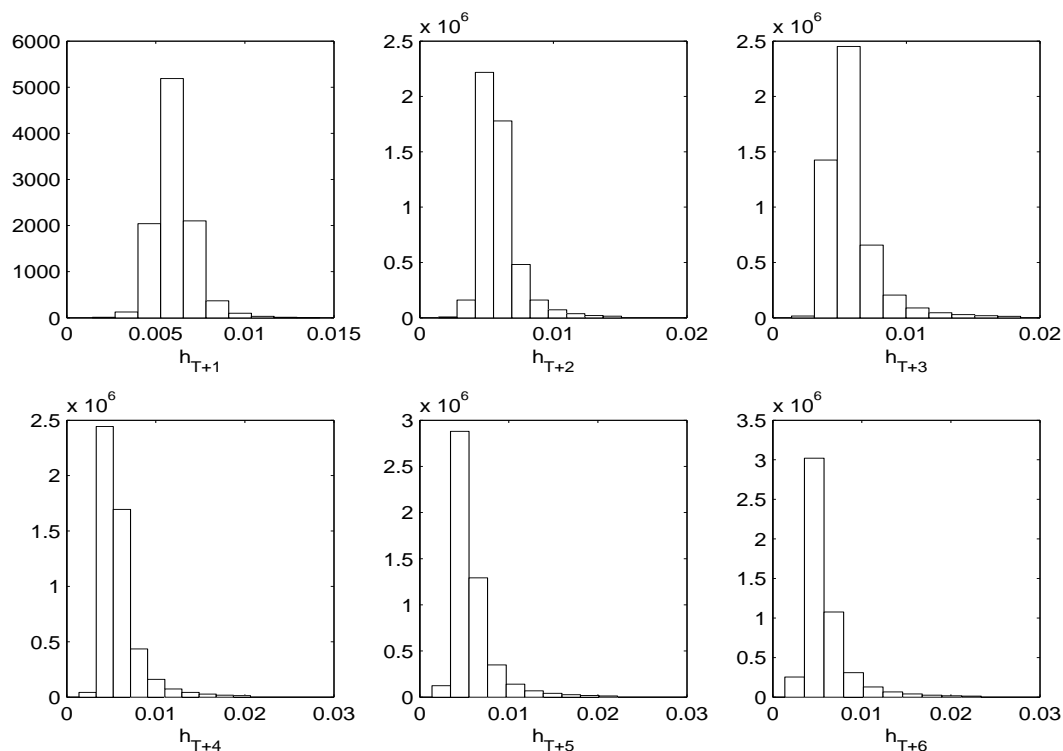
Figure 5: Histograms of the predictive distribution of the future volatilities, $h_{T+1}, \ldots, h_{T+6}$, for the simulated series with $T = 1000$.

Table 5: Predictive mean and 95% predictive interval of VaR for future times $T+1, \ldots, T+6$ with probability $\pi = 0.01$ and initial investment $A = 1$ unit.

	$T = 500$		$T = 1000$		$T = 2000$	
s	Mean	Interval	Mean	Interval	Mean	Interval
1	$-.1459$	$(-.1563, -.1375)$	$-.1980$	$(-.2149, -.1832)$	$-.1768$	$(-.1931, -.1623)$
2	$-.2066$	$(-.2191, -.1946)$	$-.2734$	$(-.2908, -.2579)$	$-.2499$	$(-.2678, -.2334)$
3	$-.2499$	$(-.2650, -.2352)$	$-.3273$	$(-.3481, -.3094)$	$-.3001$	$(-.3197, -.2806)$
4	$-.2856$	$(-.3026, -.2705)$	$-.3719$	$(-.3957, -.3496)$	$-.3417$	$(-.3624, -.3200)$
5	$-.3166$	$(-.3343, -.3014)$	$-.4090$	$(-.4321, -.3870)$	$-.3777$	$(-.4006, -.3565)$
6	$-.3436$	$(-.3634, -.3258)$	$-.4430$	$(-.4678, -.4175)$	$-.4090$	$(-.4346, -.3837)$

VaR estimates are quite accurate. The proposed method for obtaining the replications needed for volatilities and VaR prediction is not very time consuming. In fact, only two minutes were enough for each series.

4.2 Sensitivity analysis and comparison with other approaches

In this subsection, we develop a sensitivity analysis to the choice of priors and compare the proposed Griddy-Gibbs sampler procedure with other alternative approaches.

Firstly, we use the three simulated series to explore the sensitivity to our election to prior distributions. We replace the uniform distributions of ρ and λ given in (9) for more informative beta distributions. We consider a priori that $\rho \sim Beta(2, 1)$ restricted to the interval $(0.5, 1)$, to increase the prior probability that an observation is generated by the smaller variance component, and $\lambda \sim Beta(1, 2)$, to let the variance of the second component be large enough compared with the variance of the first one. We also assume a non-informative prior on the unconditional variance,

$$f(\omega, \alpha, \beta) \propto \frac{1 - \alpha - \beta}{\omega}, \tag{20}$$

restricted to the stationary region, as considered in Bauwens et al. (1999). Note that the unconditional variance of our GARCH-mixture model, which is given by the inverse of (20), does not depend on (ρ, λ) as $var(\epsilon_t) = 1$. Columns 2, 5 and 8 of Table 6 show the posterior means and standard deviations of the model parameters for the three series simulated in the previous subsection using this prior specification. Observe that there is little sensitivity to this prior choice compared

Table 6: Posterior mean and standard deviations with the Griddy-Gibbs sampler with informative priors (GG2), Random Walk Metropolis-Hastings (RWMH) algorithm and Maximum Likelihood (MLE) estimates with standard errors.

	$T = 500$			$T = 1000$			$T = 2000$		
Param.	GG2	RWMH std	MLE std	GG2	RWMH std	MLE std	GG2	RWMH std	MLE std
$\rho = .9$	.8860 .0781	.8816 .0745	.9392 .0415	.8981 .0423	.8895 .0350	.9098 .0265	.9187 .0325	.9123 .0223	.9230 .0131
$\lambda = .15$	.1956 .0762	.2038 .0736	.1505 .0873	.1521 .0465	.1698 .0366	.1529 .0038	.1279 .0402	.1436 .0251	.1336 .0197
$\mu = .01$	.0093 .0026	.0088 .0026	.0091 .0026	.0084 .0019	.0082 .0019	.0084 .0019	.0092 .0014	.0093 .0013	.0092 .0013
$\omega = .001$	.0011 .0006	.0009 .0004	.0005 .0002	.0007 .0003	.0004 .0001	.0004 .0001	.0010 .0004	.0007 .0001	.0006 .0001
$\alpha = .15$	.1479 .0612	.1432 .0494	.1127 .0467	.1176 .0482	.1057 .0279	.0985 .0298	.1170 .0452	.0992 .0202	.0961 .0221
$\beta = .7$	.6038 .1788	.6492 .1246	.7684 .0957	.7578 .0920	.8102 .0375	.8161 .0528	.7188 .1033	.7749 .0334	.7813 .0047
Time	11.7 m.	19.9 m.	0.4 s.	19.1 m.	27.5 m.	0.7 s.	35.8 m.	45.3 m.	2.3 s.

with the results obtained in Table 2. Also, we have observed that there is also little sensitivity in the prediction of volatilities and VaR.

Secondly, we compare our Griddy-Gibbs sampling approach with uniform priors with a frequently used MCMC method, the Random Walk Metropolis-Hastings (RWMH), see e.g. Robert and Casella (2004). In this method, for each iteration and parameter, a candidate value is generated from a symmetric distribution and accepted with an easily computable probability. For instance, for the parameter ρ, a candidate value, $\tilde{\rho}$, is sampled from a normal distribution, $N(\rho^{(n-1)}, V_\rho)$, where the variance, V_ρ, is a previously chosen constant, and $\tilde{\rho}$ is accepted with probability $\min\left\{1, \kappa(\tilde{\rho})/\kappa\left(\rho^{(n-1)}\right)\right\}$. Note that the main difficulty of this approach is the choice of V_ρ, which has to be tuned to obtain the best acceptance rate as possible.

Table 6 shows the posterior means and standard deviations of the model parameters for the three simulated series obtained with the RWMH algorithm using a 100000 MCMC sample (after discarding the first 100000 runs). We needed to run such a large number of iterations to achieve convergence based on the values of the statistic (19) and to obtain reasonable convergence diagrams. In fact, the best acceptance rates we could get was quite low (around 10% for each parameter). Note that the posterior means are quite close to those obtained previously using the Griddy-Gibbs sampling (see Table 2) and the computational cost is even larger with RWMH algorithm. Moreover, we have observed in the histograms (not reported) that the RWMH approach did not explore

completely the tails of the marginal distributions. Observe that the posterior standard deviations are smaller than those obtained with the Griddy-Gibbs sampling. Although other alternative Metropolis-Hastings algorithms could improve the efficiency of the method, considering the effort required to calibrate the candidate distributions, we believe that the Griddy-Gibbs sampling is a simple method to implement which seems to work well in this type of models.

Finally, we consider Maximum Likelihood Estimation (MLE), which is the usual non-Bayesian procedure for model parameter estimation. Table 6 also shows the maximum likelihood estimates and standard errors of the parameters of model (18) for the three series. As expected, MLE estimates differ clearly from the posterior means when the posterior distributions are asymmetric, in particular for ρ, ω and β. Moreover, the MLE method for GARCH models only provides point predictions of future observations and volatilities. Furthermore, it is difficult to obtain VaR estimates and intervals for the mixture model. It can be shown that, given the parameters, the distribution of $y_{T+1}+\ldots+y_{T+s}$ is a mixture of 2^s components and the percentiles of this distribution are not easily available, even if s is small.

5   A Gaussian mixture GARCH(1,1) model for the SMI index

For illustration, we apply our Bayesian procedure described in Section 3 to the daily closing prices of the stock SMI index, for the period 1/Jul/1991-14/Aug/1998. The log return series, which contain $T = 1859$ data points, is plotted in Figure 6. Note that the series includes several extreme returns. For instance, the large negative return in August, 1991, corresponds to the fall of the communist regime in the USSR. The sample mean, variance and kurtosis coefficient of the log return series are 8.17×10^{-4}, 8.55×10^{-5} and 8.73 respectively. Note the large sample kurtosis of the returns. The autocorrelation function of the returns does not show any significant autocorrelation. The Ljung-Box statistics for the log returns for lags 20 and 25 are 26.01 and 29.36 with $p-$values 0.16 and 0.24, respectively.

We estimate the model (1) with mixture innovations in (2) using the whole sample so the model estimated uses 1858 observations. We generate $N = 20000$ runs of the Markov chain and discard the initial 10000 runs. The values of statistic (19) for each parameter in $\boldsymbol{\theta}$ for the chain are 0.2789, -0.5647, 0.9257, 0.1037, 0.4865 and -0.0231, respectively, indicating that the convergence
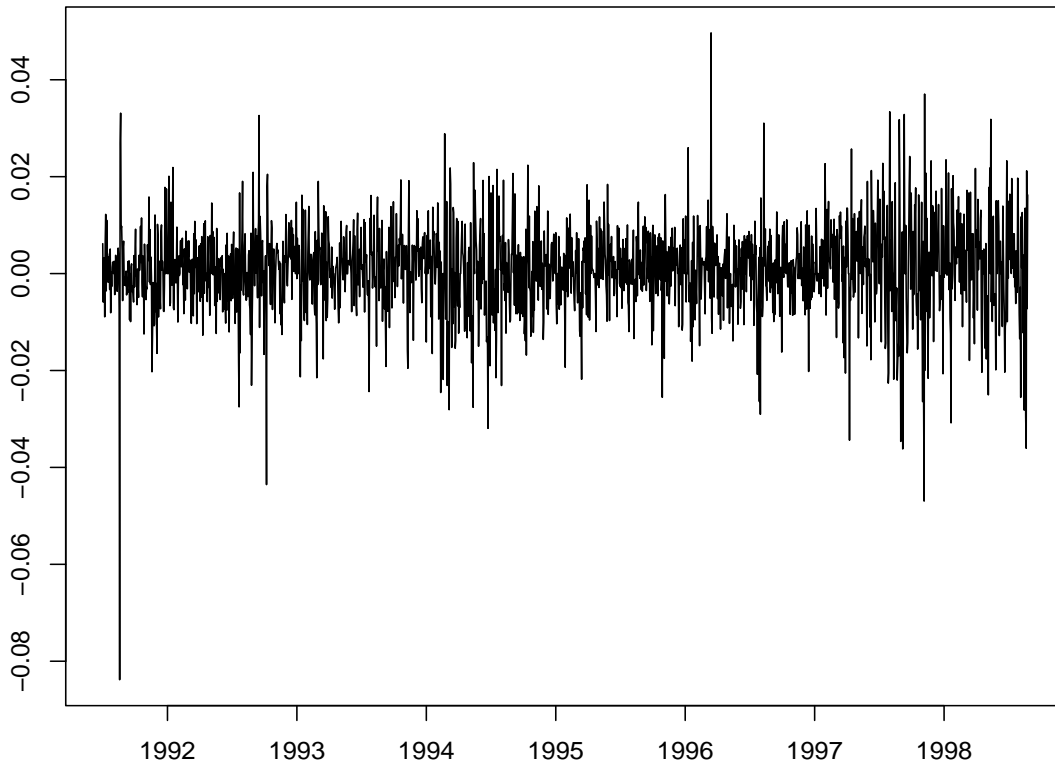
Figure 6: Return series of the MSI stock index

has been achieved. Table 7 reports the estimation results. Our Bayesian procedure estimates that the variance of the second component is approximately seven times larger than the variance of the first component. The posterior probability that an observation belongs to the component with larger variance is 0.077. Thus, the estimated model suggests that most of the innovations (92.3%) are generated by the first component, while a small number (7.7%), presumably including the extreme events, are generated by the second component. The estimation indicates the existence of a significant positive mean so that the SMI stock index has an overall upward trend on the observed period. The posterior probability that the extreme observation commented above has been generated by the largest variance component of the mixture is 0.9999, i.e., 99.99% of the MCMC draws assign this observation to the second component. Last column of Table 7 shows the maximum likelihood estimates of the model parameters. Note the presence of small differences between both kind of estimates mainly due to the asymmetry of the posterior distributions as in the previous section.

Figure 7 shows the histograms of the predictive distribution of the first six-step ahead forecasts for the conditional volatility. Note that the distribution for the one-step ahead forecast is quite

Table 7: Estimation results for the stock SMI index.

Parameter	Mean std	Median mad	MLE
ρ	0.923 0.047	0.932 0.038	0.973 0.004
λ	0.135 0.050	0.137 0.041	0.086 0.012
μ	1.113×10^{-3} 1.88×10^{-4}	1.114×10^{-3} 1.51×10^{-4}	1.066×10^{-3} 1.867×10^{-4}
ω	1.130×10^{-5} 5.40×10^{-6}	1.080×10^{-5} 4.30×10^{-6}	5.610×10^{-5} 1.34×10^{-6}
α	0.151 0.051	0.152 0.041	0.121 0.019
β	0.741 0.084	0.746 0.067	0.826 0.027
Time	36.33 min.		4.37 sec.

Table 8: First Part: Predictive means of future volatilities at times $T+1,\ldots,T+6$ joint with 95% predictive intervals. Second Part: Value at Risk (VaR) estimates and 95% predictive intervals.

	Volatilities		VaR	
Time Horizon	Mean	Pred. Interval	Mean	Pred. Interval
1	2.77×10^{-4}	$\left(1.59 \times 10^{-4}, 4.08 \times 10^{-4}\right)$	-0.040	$(-0.043, -0.038)$
2	2.62×10^{-4}	$\left(1.30 \times 10^{-4}, 5.06 \times 10^{-4}\right)$	-0.058	$(-0.061, -0.054)$
3	2.49×10^{-4}	$\left(1.08 \times 10^{-4}, 5.84 \times 10^{-4}\right)$	-0.069	$(-0.073, -0.065)$
4	2.38×10^{-4}	$\left(0.93 \times 10^{-4}, 6.36 \times 10^{-4}\right)$	-0.078	$(-0.083, -0.074)$
5	2.28×10^{-4}	$\left(0.82 \times 10^{-4}, 6.67 \times 10^{-4}\right)$	-0.086	$(-0.092, -0.081)$
6	2.20×10^{-4}	$\left(0.75 \times 10^{-4}, 6.87 \times 10^{-4}\right)$	-0.093	$(-0.098, -0.087)$

symmetric, while the remaining distributions are right-skewed. Table 8 shows the estimated future volatilities for times $T + 1,\ldots,T + 6$ obtained with the proposed method, joint with predictive intervals. Finally, second part of Table 8 shows point estimates and predictive intervals for the $s-$period 1% VaR for an amount of investment $A = \$1$, for $s = 1,\ldots,6$.

6   Conclusions

In this article we have described how to carry out Bayesian inference and prediction for a GARCH model with Gaussian mixture innovations. We have illustrated that the aforementioned model accounts for large kurtosis and extreme events better than the normal and t distributions. The mixture model also avoids the use of informative priors for the degrees of freedom parameter of the $t-$distribution. A Griddy-Gibbs sampler has been constructed which is straightforward to
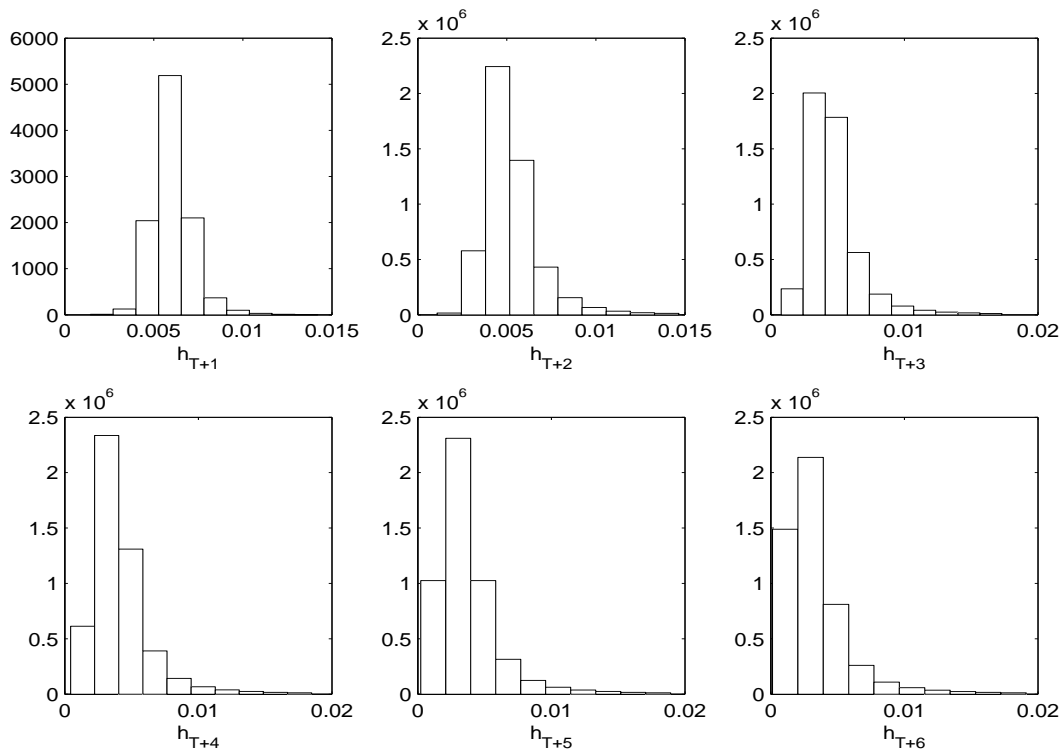
Figure 7: Histogram of the predictive distributions of the conditional volatility at times $T + 1, \ldots, T + 6$ for the stock SMI index.

implement and has been shown to work well with simulated and real data. Moreover, Bayesian prediction of volatilities and VaR has been also developed providing confidence intervals which give information about the precision in the estimates.

This approach can be straightforwardly generalized to capture other effects frequently observed in financial time series such as asymmetry. For instance, this can be done by increasing the number of components of the mixture and allowing the mean of these components to be different from zero. The theory and implementation of these issues are currently under research.

Acknowledgment

References

Alexander, C., Lazar, E., 2004. Normal Mixture GARCH(1,1): Applications to Exchange Rate Modelling. Working paper, ISMA Center, University of Reading.

Bai, X., Rusell, J., Tiao, G., 2001. Beyond Merton's utopia (i): effects of non-normality and dependence on the precision of the variance estimates using high-frequency financial data. Working paper, GSB, University of Chicago.

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

Bauwens, L., Bos, C.S., van Dijk, H.K., 1999. Adaptative polar sampling with an application to a Bayes measure of a value-at-risk. Working Paper, CORE, Université Catholique de Louvain.

Bauwens, L., Lubrano, M., 1998. Bayesian inference on GARCH models using Gibbs sampler. Econometrics J., 1, 23-46.

Bauwens, L., Lubrano, M., 2002. Bayesian option pricing using asymmetric GARCH models. J. Finan. Econom., 9, 321-342.

Bollerslev, T., 1986. Generalised autoregressive conditional heteroscedasticity. J. Econometrics, 51, 307-327.

Bollerslev, T., 1987. A conditional heteroscedastic time series model for speculative prices and rates of return. Rev. Econ. Stat., 69, 542-547.

Bollerslev, T., Chou, R. Y., Kroner, K. F., 1992 ARCH modeling in Finance-A review of the theory and empirical evidence. J. Econometrics, 52, 5-59.

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

Casella, G., Robert, C.P., 1996. Rao-Blackwellization of Sampling Schemes. Biometrika, 83, 81-94.

Davis, P.J., Rabinowitz, P., 1975. Methods of numerical integration. New York: Academic Press.

Diebolt, J., Robert, C.P., 1994. Estimation of finite mixture distributions through Bayesian Sampling. J. Roy. Statist. Soc. Ser. B, 56, 363-375.

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

Engel, R., González-Rivera, G., 1991. Semi-parametric ARCH models. J. Bus. Econom. Statist., 9, 345-359.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bayesian Statistics 4 (Eds. Bernado, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M.). Clarendon Press, Oxford.

Geweke, J., 1994. Bayesian comparison of econometric models. Working paper, Federal Reserve Bank of Minneapolis.

Haas, M., Mittnik, S., Paolella, M. S., 2004. Mixed Normal Conditional Heteroscedasticity. J. Finan. Econom., 2, 211-250.

Jorion, P., 2000. Value at Risk: The new benchmark for managing financial risk (2nd edition). McGraw-Hill, New York.

Müller, P., Pole, A., 1998. Monte Carlo posterior integration in GARCH models. Sankhya, 60, 127-144.

Nakatsuma, T., 2000. Bayesian analysis of ARMA-GARCH models: A Markov chain sampling approach. J. Econometrics, 95, 57-69.

Nelson, D.B., 1991. Conditional heteroscedasticity in asset pricing: a new approach. Econometrica, 59, 347-370.

Peña, D., Guttman, I., 1993. Comparing probabilistic methods for outlier detection in linear models. Biometrika, 80, 603-610.

Ritter, C., Tanner, M.A., 1992. Facilitating the Gibbs sampler: The Gibss stopper and the Griddy-Gibbs sampler. J. Amer. Statist. Assoc., 87, 861-868.

Robert, C.P., Casella, G., 2004. Monte Carlo Statistical Methods (2nd edition). Springer, New York.

Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In Cox, D. R., Hinkley, D. V., Barndorff-Nielson, O. E. (Eds.), Time series models with Econometric, Finance and other applications. Chapman & Hall, London, 1-67.

Taylor, S.J., 1986. Modelling financial time series. John Wiley, Chichester.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Ann. Stat., 22, 1701-1762.

Vrontos, I.D., Dellaportas, P., Politis, D.N., 2000. Full Bayesian inference for GARCH and

EGARCH models. J. Bus. Econom. Statist., 18, 187-198.

Wong, C.S., Li, W.K., 2001. On a Mixture Autoregressive Conditional Heteroscedastic Model. J. Amer. Statist. Assoc., 96, 982-995.