UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 05-36
Statistics and Econometrics Series 05
May 2005

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

BAYESIAN ESTIMATION OF THE GAUSSIAN MIXTURE GARCH MODEL

María Concepción Ausín and Pedro Galeano*

Abstract

In this paper, we perform Bayesian inference and prediction for a GARCH model where the innovations are assumed to follow a mixture of two Gaussian distributions. This 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. The method is illustrated using the Swiss Market Index.

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

* Ausín, Departamento de Estadística, Universidad Carlos III de Madrid, c/Madrid, 126, 28903 Getafe (Madrid), e-mail:concepcion.ausin@uc3m.es. Galeano, Departamento de Estadística, Universidad Carlos III de Madrid, c/Madrid, 126, 28903 Getafe (Madrid), e-mail:pedro.galeano@uc3m.es. We wish to thank Michael P. Wiper for useful comments and suggestions. We acknowledge financial support by project SEJ2004-03303. The first author also acknowledges financial support by project 06/HSE/0181/2004. The second author also acknowledges financial support by project 06/HSE/0174/2004.

Bayesian estimation of the Gaussian mixture GARCH model

María Concepción Ausín*, Pedro Galeano

Departamento de Estadística, Universidad Carlos III de Madrid, Spain.

Abstract

In this paper, we perform Bayesian inference and prediction for a GARCH model where the innovations are assumed to follow a mixture of two Gaussian distributions. This 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. The method is illustrated using the Swiss Market Index.

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

[1]Corresponding author. Tel.: +34 91 624 98 59; fax: + 34 91 624 98 49
E-mail address: concepcion.ausin@uc3m.es (M. C. Ausín)

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, 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.

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 a predictive distribution of the one-step ahead

volatility can be obtained which is more informative than a simple point forecast as it provides the quantiles needed for Value at Risk calculation. Furthermore, 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), Nakatsuma (2000) and Vrontos, Dellaportas and Politis (2000).

In this paper, we present an exact 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 combine 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 estimating and predicting volatilities is also addressed. Section 4 presents a Monte Carlo simulation which illustrates the accuracy in the estimation of the parameters and volatilities. Section 5 illustrates our procedure for the 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

In this article, throughout, we consider a GARCH(1,1) model as this is a parsimonious model which is found to be appropriate in most applications. A similar analysis to that proposed here can be carried out for a general GARCH(p,q) model.

The GARCH(1,1) model for a series y_t is given by,

$$\begin{aligned} y_t &= \mu + \sqrt{h_t}\epsilon_t, \\ h_t &= \omega + \alpha\left(y_{t-1} - \mu\right)^2 + \beta h_{t-1}, \end{aligned} \tag{1}$$

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, α, β) follow the restrictions, $\omega \geq 0$, $\alpha > 0$ and $\beta \geq 0$, to ensure positivity of h_t, for all t. Other restrictions imposed are $0 < \alpha + \beta < 1$ and $\beta < 1$ to ensure covariance and strong stationarity, respectively.

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 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 larger than 0.5 to ensure that the component with largest number of elements is the one with smallest variance.

Some of the reasons of using this 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, that 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, and is

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

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,

$$K_g = \frac{6\alpha^2}{1 - 2\alpha^2 - (\alpha + \beta)^2}, \tag{5}$$

which depends only on the parameters of the volatility equation. Note that the fourth moment of y_t only exists if $1 - 2\alpha^2 - (\alpha + \beta)^2 > 0$ or, in other words, if K_g is positive (see Bollerslev, 1986). 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,

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

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).

Figure 1 about here

3 Bayesian Inference for the Gaussian mixture GARCH(1,1) 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. These authors compare this approach with other MCMC methods such as importance sampling and Metropolis-Hastings algorithms for GARCH models with t-distributed errors. Although Griddy-Gibbs sampling has a slightly larger computational cost than these other methods, it is easier to implement and avoids the difficulty of finding a good importance function or a proposal distribution.

Let $\boldsymbol{\theta} = (\rho, \lambda, \mu, \omega, \alpha, \beta)'$, be the parameter vector of model (1). Given a series, y_t, $t = 1, \ldots, T$,

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,

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

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 ρ, λ, μ, ω, α and β are uniformly distributed over their respective domains, i.e.,

$$\begin{aligned} &p\left(\rho\right) \sim U\left(0.5, 1\right), \qquad p\left(\lambda\right) \sim U\left(0, 1\right), \qquad p\left(\omega\right) \sim U\left(0, \infty\right), \\ &p\left(\alpha\right) \sim U\left(0, 1\right), \qquad p\left(\beta\right) \sim U\left(0, 1\right), \qquad p\left(\mu\right) \sim U\left(-\infty, \infty\right), \end{aligned} \tag{9}$$

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, ν, 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 ν 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 $\boldsymbol{\theta}$, $p(\boldsymbol{\theta} \mid \boldsymbol{y})$. 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 $\boldsymbol{\theta}^{(n)} = \left(\rho^{(n)}, \ldots, \beta^{(n)}\right)'$, which has equilibrium distribution $p(\boldsymbol{\theta} \mid \boldsymbol{y})$, the posterior distribution of the parameter vector. 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,

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

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

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

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

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 $(\omega, \alpha, \mu, \beta)$ as given in (1).

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

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

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

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

The kernels of the conditional posterior densities $p(\omega \mid \boldsymbol{\theta}_{-\omega}, \mathbf{y}, \mathbf{z})$, $p(\alpha \mid \boldsymbol{\theta}_{-\alpha}, \mathbf{y}, \mathbf{z})$ and $p(\beta \mid \boldsymbol{\theta}_{-\beta}, \mathbf{y}, \mathbf{z})$ have the same expression (13), but for given w, α and β, respectively.

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 this 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,

   $$\Phi_i \simeq \int_{\rho_1}^{\rho_i} \kappa\left(\rho \mid \theta_{-\rho}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}\right) d\rho, \qquad i = 2, ..., G, \quad (14)$$

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

   (b) Generate $u \sim U(0, \Phi_G)$ 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 $\lambda^{(n+1)} \sim \lambda|\rho^{(n+1)}, \mu^{(n)}, \omega^{(n)}, \alpha^{(n)}, \beta^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}$, as in 3.

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

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

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

8. Update β by sampling from $\beta^{(n+1)} \sim \beta|\rho^{(n+1)}, \lambda^{(n+1)}, \mu^{(n+1)}, \omega^{(n+1)}, \alpha^{(n+1)}, \mathbf{y}, \mathbf{z}^{(n+1)}$, 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 Lumbrano (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, α and β, 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 looks truncated. Finally, we use linear interpolation between adjacent points in point 3.b. and to ensure the assumed stationarity, we simply reject the draws with $\alpha^{(n)} + \beta^{(n)} \geq 1$.

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_i, of the grid using Rao-Blackwellization, see Casella and Robert (1996), as follows,

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

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, $E[\rho \mid \mathbf{y}]$ with $\sum_{t=s+1}^{N} E\left[\rho \mid \theta_{-\rho}^{(n)}, \mathbf{y}, \mathbf{z}^{(n+1)}\right]/(N-s)$ instead of considering the usual sample mean, $\sum_{t=s+1}^{N} \rho^{(n)}/(N-s)$. 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 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,

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

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.

Suppose now that we are interested in the prediction of $h_{T+2}, \ldots, h_{T+r}$. We cannot obtain samples from their predictive distributions using the same procedure as for h_{T+1} because the values of y_t are unknown for $t \geq T+1$. However, we can obtain an estimation of their predictive means as follows. Firstly, we derive the expression for the conditional predictive expectation of h_{T+i}, for $i = 2, \ldots, r$,

$$\begin{aligned} E[h_{T+i}|\mathbf{y}, \boldsymbol{\theta}] &= E\left[\omega + \alpha\left(y_{T+i-1} - \mu\right)^2 + \beta h_{T+i-1}|\mathbf{y}, \boldsymbol{\theta}\right] = \omega + (\alpha+\beta) E[h_{T+i-1}|\mathbf{y}, \boldsymbol{\theta}] = \\ &= \omega\left[\sum_{j=0}^{i-2} (\alpha+\beta)^j\right] + (\alpha+\beta)^{i-1} E[h_{T+1}|\mathbf{y}, \boldsymbol{\theta}] = \omega\left[\sum_{j=0}^{i-2} (\alpha+\beta)^j\right] + (\alpha+\beta)^{i-1} h_{T+1}, \end{aligned} \tag{17}$$

which only depends on the parameters, $\boldsymbol{\theta}$, and the one-step ahead volatility, h_{T+1}. Note that when the forecast horizon increases, $i \to \infty$, the conditional mean (17) tends monotonically to the unconditional variance of the series, $\omega/(1-\alpha-\beta)$. Now, the predictive mean of h_{T+i} can be

estimated as the sample mean of (17) for all the draws of the MCMC sample,

$$E\left[h_{T+i}|\mathbf{y}\right] = E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[E\left[h_{T+i}|\mathbf{y},\boldsymbol{\theta}\right]\right] \simeq \frac{1}{N-s}\sum_{n=s+1}^{N} E\left[h_{T+i}|\mathbf{y},\boldsymbol{\theta}^{(n)}\right]. \tag{18}$$

Note that (18) is also an increasing function of i that will converge to the posterior unconditional variance of the series.

Finally, consider the prediction of $y_{T+1},\ldots,y_{T+r}$. The predictive density of y_{T+i}, $i=1,\ldots,r$, is given by,

$$p\left(y_{T+i}|\mathbf{y}\right) = \int_{\boldsymbol{\theta}} p\left(y_{T+i}|\mathbf{y},\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}|\boldsymbol{y}\right) d\boldsymbol{\theta}. \tag{19}$$

The density $p\left(y_{T+i}|\mathbf{y},\boldsymbol{\theta}\right)$ is of unknown form except for the case $i=1$, when it is a Gaussian mixture with mean μ and variance h_{T+1}. Thus, for $i=1$, the predictive density (19) can be estimated as the mean of the density functions obtained for all the draws of the MCMC sample. The predictive mean and variance of (19) are given by,

$$E\left[y_{T+i}|\mathbf{y}\right] = E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[E\left[y_{T+i}|\mathbf{y},\boldsymbol{\theta}\right]\right] = E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[\mu\right], \tag{20}$$

$$\begin{aligned} Var\left[y_{T+i}|\mathbf{y}\right] &= E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[Var\left[y_{T+i}|\mathbf{y},\boldsymbol{\theta}\right]\right] + Var_{\boldsymbol{\theta}|\boldsymbol{y}}\left[E\left[y_{T+i}|\mathbf{y},\boldsymbol{\theta}\right]\right] = \\ &= E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[E\left[h_{T+i}|\mathbf{y},\boldsymbol{\theta}\right]\right] + Var_{\boldsymbol{\theta}|\boldsymbol{y}}\left[\mu\right], \end{aligned} \tag{21}$$

which are straightforward to estimate because the predictive mean (20) is the posterior mean of μ and the predictive variance (21) is the sum of the predictive mean of h_{T+i}, given in (18), and the posterior variance of μ.

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. We illustrate our MCMC procedure with an artificial series simulated from the following

GARCH(1,1) model,

$$\begin{aligned} y_t &= 0.5 + \sqrt{h_t}\epsilon_t, \\ h_t &= 0.1 + 0.15\,(y_{t-1} - 0.5)^2 + 0.7h_{t-1}, \end{aligned} \tag{22}$$

where ϵ_t follows a Gaussian mixture as in (2) with parameters $\rho = 0.9$ and $\lambda = 0.15$. Thus, the excess kurtosis of y_t and ϵ_t are given by 8.84 and 3.53, respectively. We generated a series with 1005 observations from this model and estimated the model (22) using the first $T = 1000$ observations whose sample mean, variance and kurtosis are 0.5013, 0.5354 and 6.069, respectively. We generated $N = 30000$ 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 gives similar results) and discarded the initial 15000 runs. To asses 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. The convergence diagnostic is given by,

$$CD = \frac{(\overline{\rho}_1 - \overline{\rho}_2)}{\left[\frac{\widehat{S}_1^{\rho}(0)}{n_1} + \frac{\widehat{S}_2^{\rho}(0)}{n_2}\right]^{\frac{1}{2}}}, \tag{23}$$

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 (23) has asymptotically the standard Gaussian distribution. The values of the statistic (23) for each element of the chain are -1.088, 0.4743, -0.0646, 0.1163, -0.1661 and -0.0598, respectively for each parameter in $\boldsymbol{\theta}$, and we conclude that the chain has converged. Figure 2 shows the convergence diagrams of the posterior sample for each parameter. Note that only the draws in equilibrium are plotted. Figure 3 shows the histograms of the MCMC output. 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. In Table 1, we report the estimation results. Columns 2, 3 and 4 in Table 1 show the mean and standard deviation, median and mean absolute deviation and mode of the posterior distribution of each parameter, respectively. Note that these estimates are close to the true parameters in all cases, especially

for the mode. In particular, when the posterior distribution is asymmetric, the median and mode seems to be better estimates than the mean as should be expected. Note also the accuracy of the estimation of the parameters of the mixture model, ρ and λ.

Table 1 about here

Figure 2 about here

Figure 3 about here

Table 2 shows the correlations between parameters. 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 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.

Table 2 about here

Figure 4 shows the true and posterior mean (16) of the last 100 conditional volatilities 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 conditional volatility for the observation $T+1 = 1001$ which is equal to 0.4509 with predictive interval (0.3467,0.5708). Figure 5 shows the histogram of the distribution of the conditional volatility at time $T+1 = 1001$. Note that the distribution is apparently symmetric. In fact, the predicted median coincides with the predictive mean. Finally, Table 3 shows the true and estimated predictive conditional volatilities for times $1001, \ldots, 1005$ obtained with (18). Note that the predictive volatilities increase as pointed out in section 3 because (17) is an increasing function of i.

Figure 4 about here

Figure 5 about here

Table 3 about here

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, which contain 1860 data points. The index and return series are 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 return series are 8.15×10^{-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.

Figure 6 about here

We estimate the model (1) with mixture errors in (2) using the whole sample so the model estimated uses 1858 observations. We generate $N = 30000$ runs of the Markov chain and discard the initial 15000 runs. The values of statistic (23) for each parameter in $\boldsymbol{\theta}$ for the chain are 0.7544, -1.009, 0.7676, 0.3355, 0.5506 and -0.4604, respectively, indicating that the convergence has been achieved. Table 4 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 0.096. Thus, the estimated model suggests that most of the innovations (90.4%) are generated by the first component, while a small number (9.6%), 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 probability of stationarity is 0.9306. Also the posterior probability that the extreme observation commented above has been generated by the largest variance component of the mixture is 0.9998, i.e., 99.98% of the MCMC draws assign this observation to this component.

Table 4 about here

Figure 7 shows the histogram of the predictive distribution of the one-step ahead forecast for the conditional volatility. Note that the distribution is quite symmetric. The predictive mean and median are 2.654×10^{-4} and 2.610×10^{-4}, respectively, and the 95% predictive interval is $(1.6 \times 10^{-4}, 3.9 \times 10^{-4})$.

Figure 7 about here

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 this model for the innovations accounts for large kurtosis and extreme events better than the normal and t distributions. A Griddy-Gibbs sampler has been constructed which is straightforward to implement and has been shown to work well with simulated and real data. Moreover, it is much simpler to implement than other Bayesian procedures. The mixture model also avoids the use of informative priors for the degrees of freedom parameter of the $t-$distribution.

This approach can be straightforwardly extended to the general GARCH(p, q) model with mixture innovations. Furthermore, it can be 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

We wish to thank Michael P. Wiper for useful comments and suggestions. We acknowledge financial support by project SEJ2004-03303. The first author also acknowledges financial support by project 06/HSE/0181/2004. The second author also acknowledges financial support by project 06/HSE/0174/2004.

References

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

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.

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. Semiparametric arch models. J. Bus. Econom. Statist., 9, 345-359.

Geweke, J., 1989. Exact predictive densities in linear models with ARCH disturbances. J. Econometrics, 40, 63-86.

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.

Kleibergen, F., van Dijk, H.K., 1993. Non-stationarity in GARCH models: a Bayesian analysis. J. Appl. Econom., 8, 41-61.

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. Econo-

metrica, 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.

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.

Tanner, M., Wong, W., 1987. The calculation of posterior distributions by data augmentation. J. Amer. Statist. Assoc., 82, 528-550.

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.

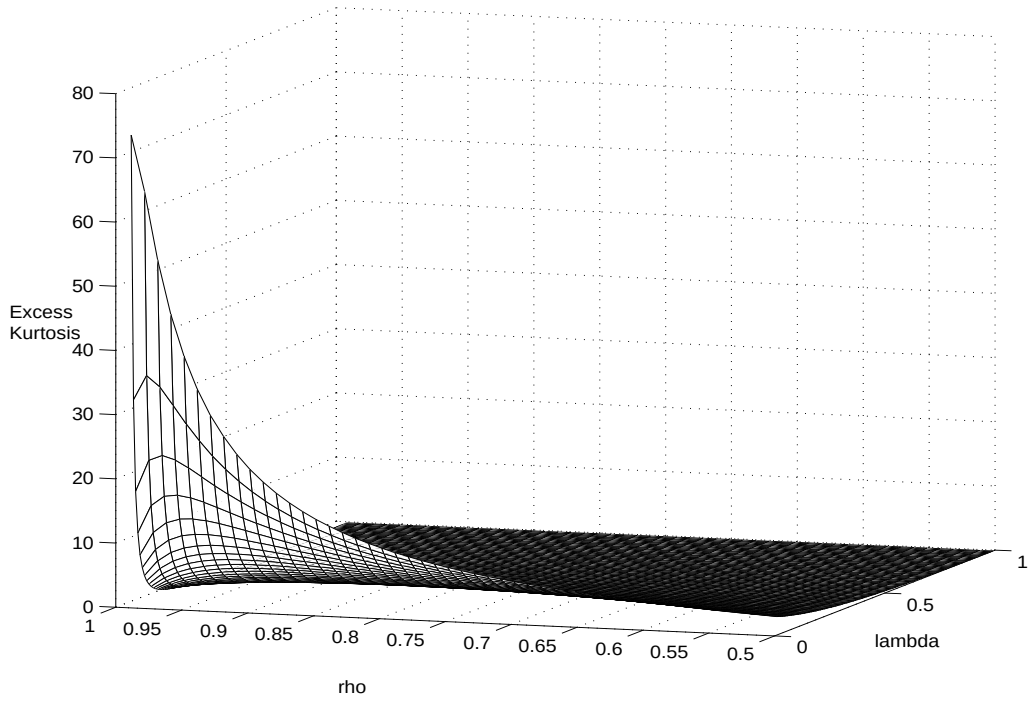

Figure 1: Excess kurtosis of the innovations as a function of the parameters λ and ρ

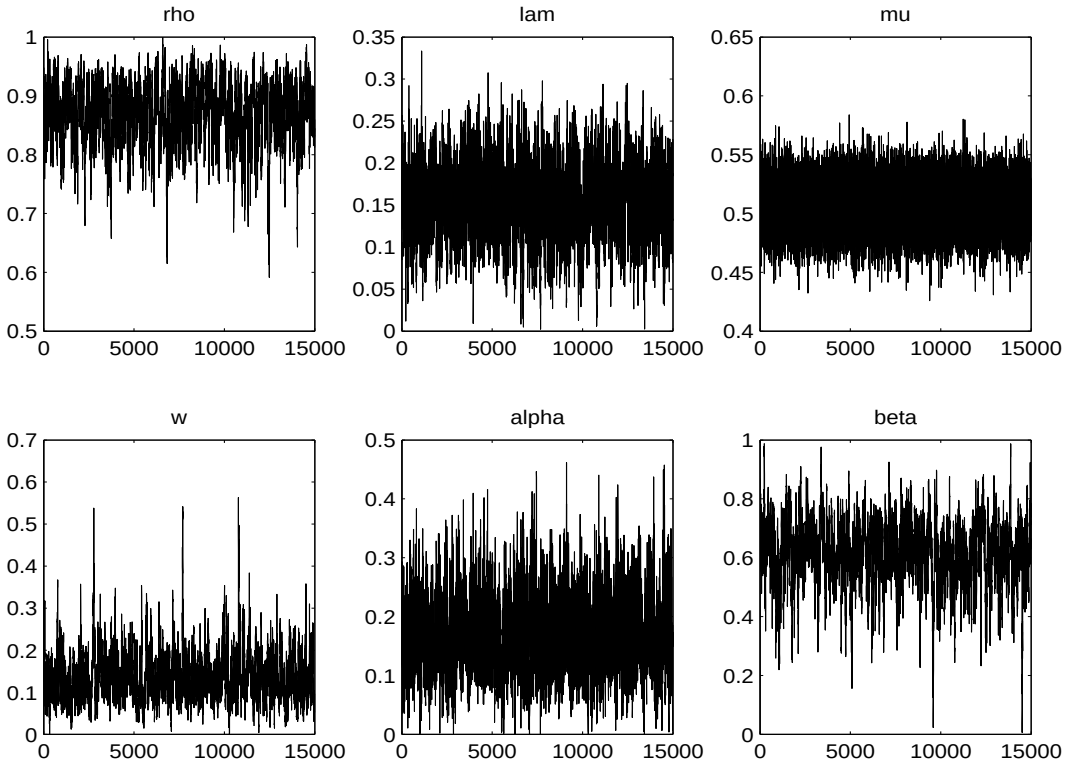

Figure 2: Convergence diagrams of the posterior sample for each parameter

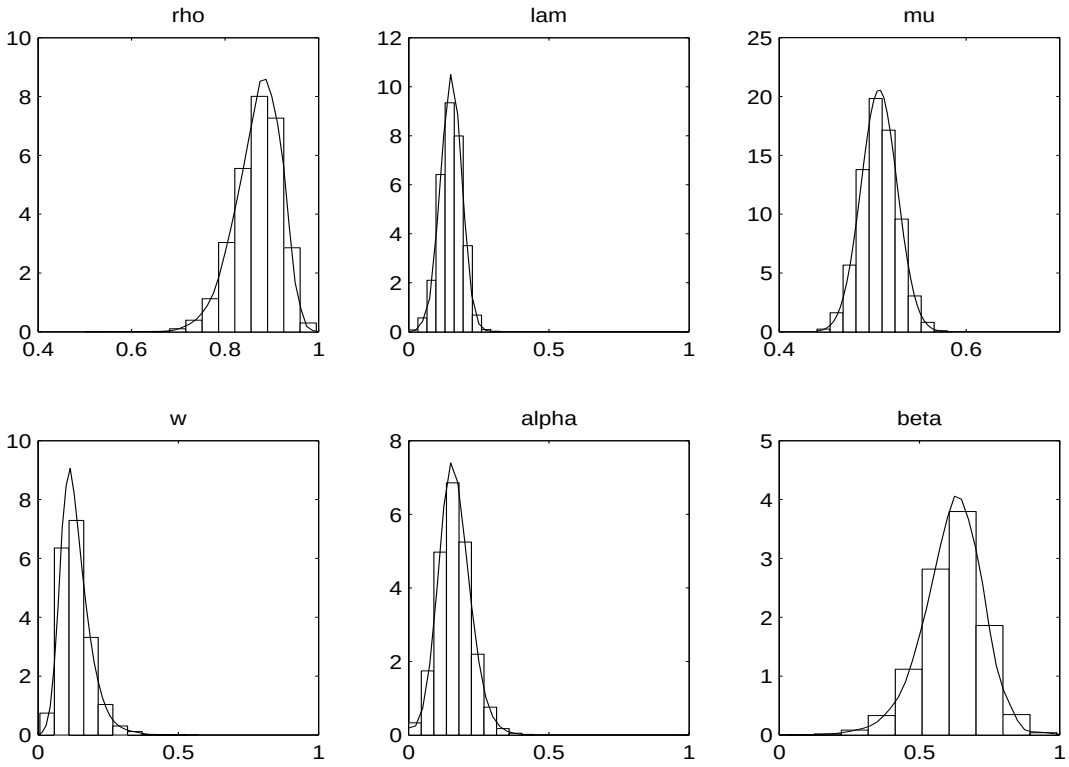

Figure 3: Histograms and marginal densities of the MCMC output for each parameter

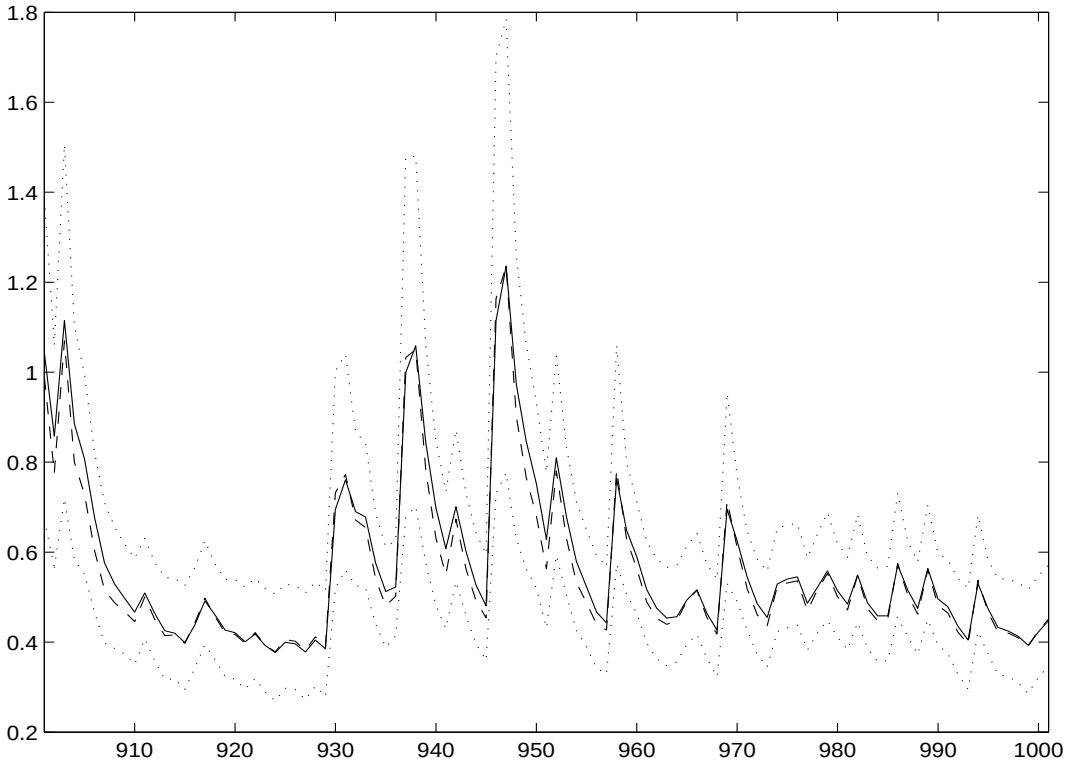

Figure 4: True (solid line) and predictive mean (dashed line) of the last 100 conditional volatilities with 95% predictive intervals (dotted lines)

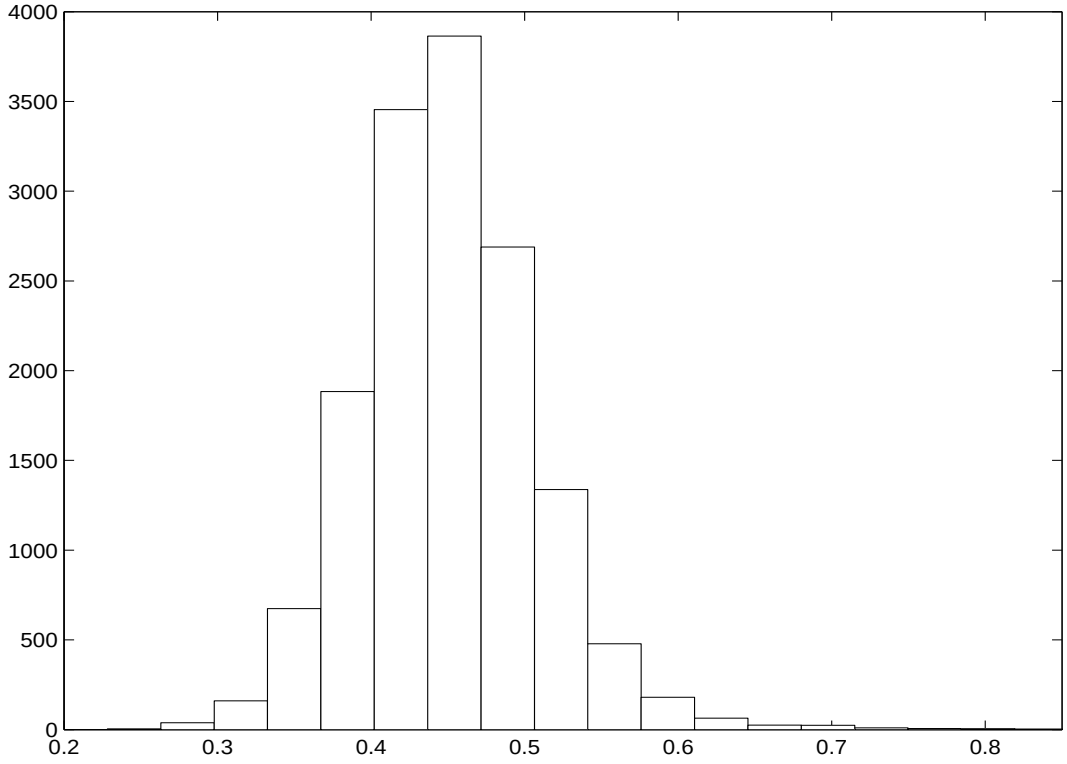

Figure 5: Histogram of the predictive distribution of the conditional volatility at time $T+1 = 1001$

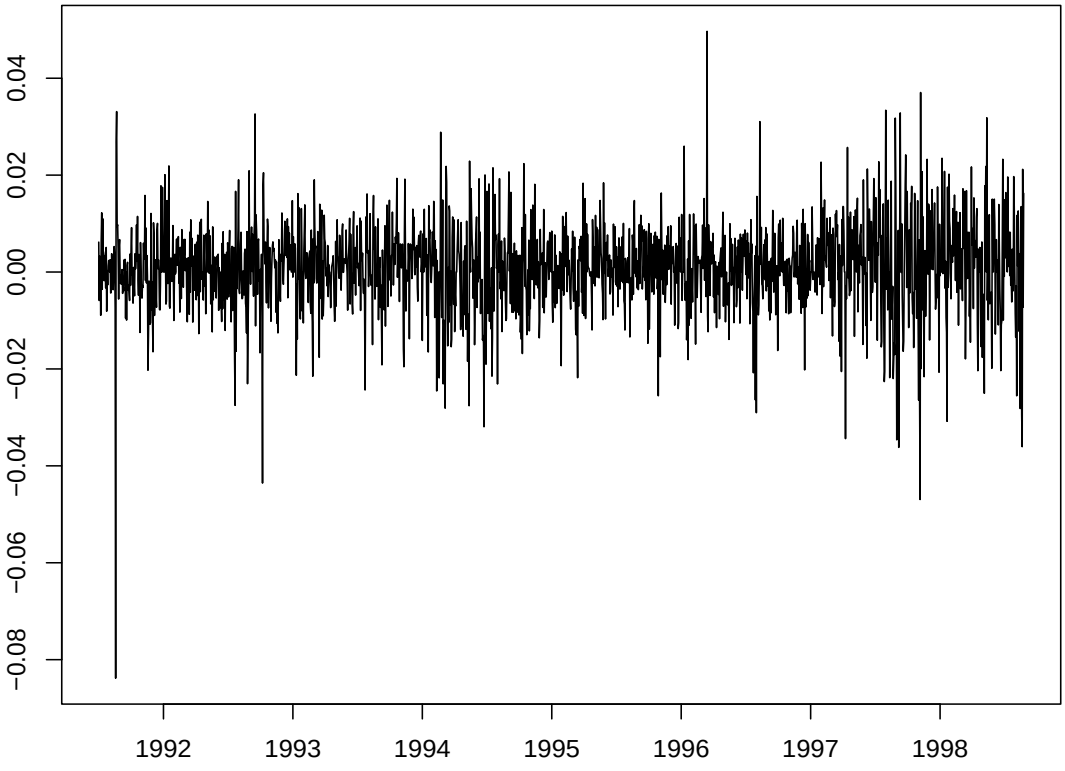

Figure 6: Return series of the MSI stock index

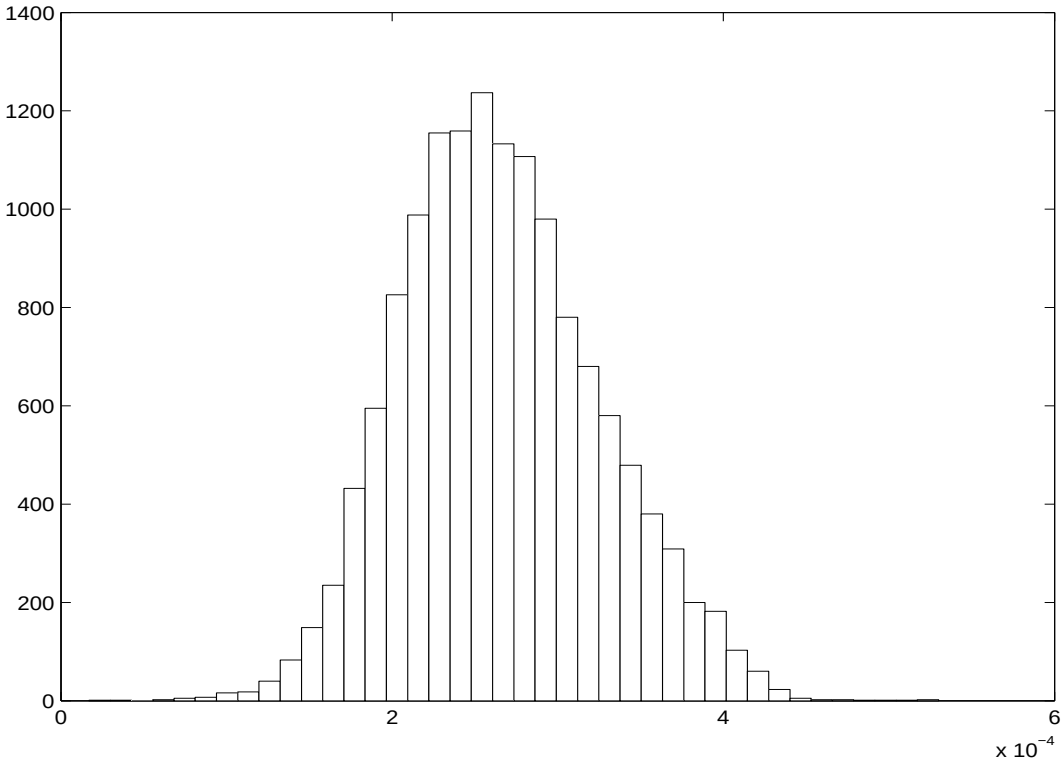

Figure 7: Histogram of the predictive distribution of the conditional volatility at time $T+1 = 1001$ for the stock SMI index.

Table 1: Estimation results.

Parameter	Mean std	Median mad	Mode
$\rho = 0.9$	0.8694 0.0487	0.8746 0.0389	0.8875
$\lambda = 0.15$	0.1513 0.0407	0.1522 0.0325	0.1501
$\mu = 0.5$	0.5067 0.0195	0.5065 0.0155	0.5083
$\omega = 0.1$	0.1337 0.0543	0.1250 0.0415	0.1144
$\alpha = 0.15$	0.1640 0.0584	0.1616 0.0460	0.1500
$\beta = 0.7$	0.6185 0.1092	0.6255 0.0847	0.6250

Table 2: Correlations between parameters.

	ρ	λ	μ	ω	α	β
ρ	1	−0.6374	−0.0459	−0.1305	0.2029	0.0948
λ		1	0.0247	0.0016	−0.1329	−0.0687
μ			1	−0.0350	−0.0154	0.0402
ω				1	0.1726	−0.8966
α					1	−0.4266
β						1

Table 3: True and estimated predictive conditional volatility for times $1001, \ldots, 1005$.

Volatilities	h_{T+1}	h_{T+2}	h_{T+3}	h_{T+4}	h_{T+5}
True	0.4473	0.4211	0.4223	2.9236	4.8878
Estimated	0.4509	0.4850	0.5119	0.5334	0.5507

Table 4: Estimation results for the stock SMI index.

Parameter	Mean std	Median mad	Mode
ρ	0.9038 0.0609	0.9161 0.0469	0.9250
λ	0.1454 0.0527	0.1474 0.0430	0.1501
μ	1.12×10^{-3} 1.86×10^{-4}	1.121×10^{-3} 1.48×10^{-4}	1.118×10^{-3}
ω	1.2×10^{-5} 6×10^{-6}	1.1×10^{-5} 4×10^{-6}	9×10^{-6}
α	0.14832 0.0518	0.1482 0.0411	0.1500
β	0.7331 0.0920	0.7363 0.0734	0.725