Computational Tools for Comparing Asymmetric GARCH Models via Bayes Factors

Ricardo S. Ehlers*
ICMC-USP

Abstract

In this paper we use Markov chain Monte Carlo (MCMC) methods in order to estimate and compare GARCH models from a Bayesian perspective. We allow for possibly heavy tailed and asymmetric distributions in the error term. We use a general method proposed in the literature to introduce skewness into a continuous unimodal and symmetric distribution. For each model we compute an approximation to the marginal likelihood, based on the MCMC output. From these approximations we compute Bayes Factors and posterior model probabilities.

Key Words: GARCH, Markov chain Monte Carlo, Metropolis-Hastings, marginal likelihood.

1 Introduction

Autoregressive conditional heteroskedastic (ARCH) models of Engle (1982) and its generalization, the GARCH model of Bollerslev (1986) have been around for a long time now and a large amount of theoretical and empirical research has been produced in the past two decades or so. Most of the work was based on (quasi-) likelihood methods and the generalized method of moments (see for example Bollerslev et al. 1992) and much less attention was paid to inference procedures from a Bayesian perspective. More recently however Bayesian computational methods based on Markov chain Monte Carlo (MCMC) have been utilized to address the complexity of these models (see for example, Bauwens and Lubrano 1998 and Nakatsuma 2000).

*ehlers@icmc.usp.br
The GARCH\((p, q)\) model estimates the volatility of a return \(y_t\) as

\[
y_t = \epsilon_t \sqrt{h_t}, \quad \epsilon_t \sim D(0, 1)
\]

\[
h_t = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i}.
\]

where \(h_t\) is the (unobservable) conditional variance of \(y_t\) given previous information \(I_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}\), the \(\epsilon_t\) are i.i.d. and \(D(0, 1)\) denotes a distribution with mean zero and variance 1. Positivity and covariance stationarity constraints are \(\omega > 0, \quad \alpha_i \geq 0, \quad i = 1, \ldots, p, \quad \beta_i \geq 0, \quad i = 1, \ldots, q\) and \(\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i < 1\). The conditional likelihood function of the model is given by

\[
l(\theta) = \prod_{t=s+1}^{n} h_t^{-1/2} p_c(y_t/\sqrt{h_t})
\]

where \(s = \max(p, q)\), \(p_c\) is the density function for \(\epsilon_t\) and the set of all model parameters is represented by \(\theta = (\theta_0, \theta_1, \ldots, \theta_{p+q}) = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)\).

Of course, even if \(\epsilon_t \sim N(0, 1)\) the unconditional distribution of \(y_t\) is nonnormal, in particular having fatter tails than the normal distribution. However, there is evidence in the literature that many financial time series tend to have observed kurtosis even higher than that implied by a GARCH model with normal errors. This has led many authors to use fat-tailed distributions for \(\epsilon_t\), most commonly the Student-\(t\) (e.g. Baillie and Bollerslev 1989). So, as a first extension we let \(\epsilon_t\) still follow a symmetric distribution but with an additional shape parameter which models the kurtosis. Here we consider a standardized Student-\(t\) distribution with \(\nu\) degrees of freedom with p.d.f. given by

\[
f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)}} \left[1 + \frac{x^2}{\nu - 2}\right]^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}
\]

and a standardized generalized error distribution (GED) also known as exponential power distribution with p.d.f. given by

\[
f(x|\nu) = \frac{\nu}{\lambda 2^{1+1/\nu} \Gamma(1/\nu)} \exp\left(-\frac{|x/\lambda|^\nu}{2}\right), \quad x \in \mathbb{R}
\]

where \(\lambda = \sqrt{2^{-2/\nu}\Gamma(1/\nu)/\Gamma(3/\nu)}\) and \(\Gamma(\cdot)\) is the Gamma function. This error distribution was suggested in Nelson (1991). We note that both of the above distributions have unit variances. Important special cases of the standardized GED are, the Laplace (or double exponential) distribution for \(\nu = 1\), the standard normal distribution for \(\nu = 2\) and the uniform distribution in the interval \((-2\sqrt{3}, 2\sqrt{3})\) when \(\nu \to \infty\). The kurtosis of this GED distribution is given by \(\Gamma(1/\nu)\Gamma(5/\nu)/\Gamma(3/\nu)^2 - 3\) so that values \(\nu < 2\) lead to leptokurtic distributions.
Figure 1: Tail behaviour of the standard normal, standardized Student-$t$ with 5 degrees of freedom, standardized Laplace and standardized GED with $\nu = 1.5$.

Figure 1 shows the tails of the p.d.f. of the standard normal, standardized Student-$t$ with 5 degrees of freedom, Laplace and GED with $\nu = 1.5$.

Also, it is often the case that data contain asymmetries and many empirical studies found evidence that modelling asymmetry in the volatility process cannot capture skewness in the returns distribution. Therefore, estimation methods that accommodate skewness should be taken into account too. This can be dealt with by allowing a certain amount of skewness in the error distribution. Possibly asymmetric distributions have been proposed and applied in recent works (see for example, Bauwens and Lubrano 2002 and Bauwens and Laurent 2005). In terms of model comparison, the superiority of conditionally skewed models is usually reported on the basis of asymptotically based tests of skewness excess or informal likelihood inference (see for example Verhoeven and McAleer 2003). In this paper, we adopt a formal and fully Bayesian approach to investigating the explanatory power of each model. We compare the following different distributions for the error term: symmetric normal distribution, standardized Student-$t$ distribution, the generalized error distribution and their skewed versions.

1.1 Skewed Distributions

There are a number of proposals in the literature to introduce skewness in unimodal symmetric distributions (e.g. Azzalini 1985, Fernandez and Steel 1998, Azzalini and Capitanio 2003 and Jones and Faddy 2003). In particular, Fernandez and Steel (1998) presented a general method for transforming any continuous unimodal and
symmetric distribution into a skewed one by changing the scale at each side of
the mode. They proposed the following class of skewed distributions indexed by a
shape parameter \( \gamma \in (0, \infty) \), which describes the degree of asymmetry,
\[
s(x|\gamma) = \frac{2}{\gamma + 1/\gamma} \left\{ \frac{f\left(\frac{1}{\gamma}x\right)}{I_{[0, \infty)}(x)} + \frac{f(\gamma x)}{I_{(-\infty, 0)}(x)} \right\}.
\]

Note that \( \gamma = 1 \) yields the symmetric distribution as \( s(x|\gamma = 1) = f(x) \), and
values of \( \gamma > 1 \) (\(< 1 \)) indicate right (left) skewness. Also, the mode of this density
remains at zero irrespective of the particular value of \( \gamma \). Mean and variance of
\( s(x|\gamma) \) depend on \( \gamma \) and are given by
\[
\mu_\gamma = m_1(\gamma - 1/\gamma) \quad \text{and} \quad \sigma^2_\gamma = (m_2 - m_1^2)(\gamma^2 + 1/\gamma^2) + 2m_1^2 - m_2
\]
where
\[
m_r = 2 \int_0^\infty x^r f(x)dx,
\]
is the \( r \)-th absolute moment of \( f(x) \) on the positive real line. Note also that, if
\( f(x) \) is a standardized distribution then \( m_2 = 1 \).

In order that the innovation process has again zero mean and unit variance we
use a reparameterization introduced in Wurtz et al. (2009). After computing \( \mu_\gamma \)
and \( \sigma^2_\gamma \) we set \( x^* = \sigma_\gamma x + \mu_\gamma \) and the p.d.f. of a standardized skewed distribution
can be expressed as,
\[
f(x|\gamma) = \frac{2\sigma_\gamma}{\gamma + 1/\gamma} \left\{ \frac{f^*(\frac{x^*}{\gamma})}{I_{[0, \infty)}(x^*)} + \frac{f^*(\gamma x^*)}{I_{(-\infty, 0)}(x^*)} \right\}.
\]

Table 1 shows the expressions for the first absolute moment \( m_1 \) of the skewed
normal, skewed Student-\( t \) and skewed GED distributions.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( m_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>( 2/\sqrt{2\pi} )</td>
</tr>
<tr>
<td>Student-( t )</td>
<td>( 2\sqrt{\nu} - 2/(\nu - 1)/\text{B}(1/2, \nu/2) )</td>
</tr>
<tr>
<td>GED</td>
<td>( 2^{1/\nu} \chi \Gamma(2/\nu)/\Gamma(1/\nu) )</td>
</tr>
</tbody>
</table>

We note that this approach entirely separates the effects of the skewness and
tail parameters thus making prior independence between the two a plausible
assumption, and hence facilitates the choice of their prior distributions.
1.2 Prior Distributions

Following the Bayesian paradigm, we need to complete the model specification by specifying the prior distributions of all parameters of interest. By Bayes Theorem, the posterior distribution is proportional to the product of the likelihood function (1) by the prior.

For the GARCH coefficients, we define \( \phi_0 = \log(\theta_0) \) and \( \phi_j = \log(\theta_j/(1 - \theta_j)) \), with reverse map given by \( \theta_j = e^{\phi_j}/(1 + e^{\phi_j}) \in (0, 1) \), \( j = 1, \ldots, p + q \). Prior distributions are then assigned as \( \phi_j \sim N(0, \sigma^2_\phi) \), \( j = 0, \ldots, p + q \) where the value of \( \sigma^2_\phi \) is kept fixed. So, the parameter vector \( \phi = (\phi_0, \phi_1, \ldots, \phi_{p+q})' \) is generated without restriction, then we map back to \( \theta \) and can check for stationarity. It is worth noting that this reparameterization for the GARCH coefficients makes the MCMC algorithm more efficient.

We also need to specify prior distributions for the parameters \( \nu \) and \( \gamma \). For the GED recall that we are interested in values \( \nu < 2 \), so we reparameterize as \( \psi = \log(\nu/(2 - \nu)) \in \mathbb{R} \), so that \( \nu = 2e^\psi/(1 + e^\psi) \in (0, 2) \), and assign a prior as \( \psi \sim N(0, \sigma^2_\psi) \). For the Student-t degrees of freedom parameter \( \nu \) we need to be more careful. As shown in Bauwens and Lubrano (1998) the posterior density is not proper when one chooses a flat prior on \((0, \infty)\) which would make model comparison based on marginal likelihoods rather questionable. They suggest other possible priors on \( \nu \) that do not suffer from this problem and here we adopt an exponential distribution as in Geweke (1993). Prior mean and standard deviation of this exponential equal to 10 has been reported in the literature as a reasonable choice (see Fernandez and Steel 1998) so that an exponential distribution with parameter 0.1 is our default choice. This represents sufficient prior information to force the posterior to tend to zero quickly enough at the tail but also avoids introducing strong prior information on \( \nu \) essentially allocating substantial prior mass to both very thick tails and almost normal tails.

As for the skewness parameter in the skew normal, skew Student-t and skew GED, we find it reasonable to choose a prior that is centered around the symmetric version of the skewed distribution and gives approximately equal weights to left and right skewness. Following Fernandez and Steel (1998), we shall use a Gamma\((a, b)\) prior on \( \gamma^2 \) which is the ratio of probability masses above and below the mode, i.e. \( \gamma^2 = Pr(X \geq 0)/Pr(X < 0) \). If we choose the hyperparameters \( a \) and \( b \) such that \( E(\gamma) = 1 \) then it is not difficult to see that this implies choosing \( b = [\Gamma(a + 1/2)/\Gamma(a)]^2 \) and \( a \) can be elicited by controlling the prior variance and prior mass of \( \gamma \) on the interval \((0, 1)\). Fernandez and Steel (1998) found that \( a = 1/2 \), which leads to \( Var(\gamma) \approx 0.57 \) and \( P(0 < \gamma < 1) \approx 0.58 \), is a reasonable choice. This is our default choice here too. We note that this particular choice is equivalent to setting \( \gamma \sim N(0, 0.64^{-1}) \) truncated to \( \gamma > 0 \) (i.e. a half-normal prior).
Finally, we note that since we are using proper priors on model specific parameters, model comparison may formally be done through the use of Bayes factors as these can meaningfully be computed.

2 Simulating the Joint Posterior

In this paper we use Markov chain Monte Carlo (MCMC) methods to obtain a sample from the posterior distribution of the parameters for each model. In particular, we adopt a Metropolis-Hastings algorithm where the parameters are updated as a block. In the steps below, describing the simulation scheme, assume that the vector $\phi$ includes transformations of the extra parameters $\nu$ and/or $\gamma$ depending on which distribution is assigned for the error term.

1. Set initial values for the GARCH parameters $\theta^{(0)}$ and transform to $\phi^{(0)}$.
2. At iteration $j$, generate a vector $\phi'$ from the random walk kernel, $\phi' = \phi^{(j-1)} + \epsilon, \epsilon \sim N(0, \tau \Sigma)$.
3. If $\sum_{j=1}^{p+q} e^{\phi'}/(1 + e^{\phi'}) < 1$ accept the move and set $\phi^{(j)} = \phi'$ with probability
   $$\alpha(\phi, \phi') = \min\left\{1, \frac{l(\phi')p(\phi')}{l(\phi)p(\phi)}\right\},$$
   otherwise reject the move and set $\phi^{(j)} = \phi^{(j-1)}$.
4. Repeat until convergence.

In step 2 above, $\tau$ is a constant to tune the acceptance rate and $\Sigma$ is estimated from the approximate Hessian matrix of the target density evaluated at its mode.

After convergence a sample $\theta^{(1)}, \ldots, \theta^{(N)}$ is available from the sample $\phi^{(1)}, \ldots, \phi^{(N)}$. Using this MCMC output, it is easy to obtain a sample from the posterior distribution of each conditional variance, $h^{(1)}_t, \ldots, h^{(N)}_t$, $t = 1, \ldots, n$ and the posterior mean, median and credible intervals can be calculated. In the next section we will see how to use this sample to compute an approximation for the marginal likelihood for each model.

2.1 Model Comparison

Denoting the competing models by $M_1, M_2, \ldots$ indexed by a model indicator $k$ then, associated with each model there is a likelihood function $p(y|\theta_k, M_k)$ and a prior $p(\theta_k|M_k)$. From a Bayesian viewpoint, we would like to compute the posterior model probabilities which are given by $p(M_k|y) \propto p(y|M_k)p(M_k)$ where

$$p(y|M_k) = \int p(y|\theta_k, M_k)p(\theta_k|M_k)d\theta_k \quad (2)$$
is the marginal likelihood of $M_k$. If the number $K$ of competing models is not too large, pairwise comparison of the models is done through their posterior odds, which for any two models $M_i$ and $M_j$ is given by

$$
\frac{p(M_i | y)}{p(M_j | y)} = \frac{p(y | M_i)}{p(y | M_j)} \times \frac{p(M_i)}{p(M_j)}
$$

where the ratio of marginal likelihoods is the Bayes factor and the second fraction is known as the prior odds. Posterior model probabilities are given by,

$$
p(M_i | y) = \frac{p(y | M_i)p(M_i)}{\sum_{j=1}^{K} p(y | M_j)p(M_j)} = \left[ \sum_{j=1}^{K} B_{ji} \frac{p(M_j)}{p(M_i)} \right]^{-1}
$$

where $B_{ji} = p(y | M_j)/p(y | M_i)$. Computation of the marginal likelihood requires a proper prior and the integral (2) is in general difficult to calculate (see Kass and Raftery 1995 for an extensive description and comparison of available numerical strategies). So, we need a method which is straightforward to implement and can be used for complex models.

It is worth noting that computing Bayes factors from a MCMC output is not a trivial task. Because the marginal likelihood is obtained by integrating $p(y | \theta_k, M_k)$ with respect to the prior distribution of the parameters, the MCMC output can not be used directly and in particular is not easily implemented in all purpose packages. In this paper, we compute an approximation to the marginal likelihood, based on the MCMC output, using the methods described in Chib and Jeliazkov (2001). These methods were developed for MCMC chains produced by the Metropolis-Hastings algorithm. From these approximations we compute Bayes Factors and posterior model probabilities as a formal model comparison criterion. To our knowledge, this constitutes an advance on what is currently available in the ARCH/GARCH literature.

The method is based on the approach in Chib (1995) where the marginal likelihood is represented as

$$
p(y | M_i) = \frac{p(y | \theta_i, M_i)p(\theta_i | M_i)}{\pi(\theta_i | y, M_i)}
$$

since $p(y | M_i)$ is the normalizing constant in the posterior density of $\theta_i$. Evaluating the right hand side of (3) at some appropriate point $\theta_i^*$ and taking logarithms it follows that,

$$
\log p(y | M_i) = \log p(y | \theta_i^*, M_i) + \log p(\theta_i^* | M_i) - \log \pi(\theta_i^* | y, M_i).
$$

So, if we can find an estimate of the posterior ordinate $\pi(\theta_i^* | y, M_i)$ then the marginal likelihood can be calculated. For estimation efficiency we take the point $\theta_i^*$ as the posterior mode given model $M_i$. 7
We now drop the dependence on model $M_i$ to simplify the notation. If we let $q(\theta, \theta')$ denote the proposal density and $p(\theta, \theta') = \alpha(\theta, \theta')q(\theta, \theta')$ denote the sub-kernel of the Metropolis-Hastings algorithm then, by reversibility of this subkernel we can write

\[ p(\theta, \theta^*) = p(\theta^*, \theta)\pi(\theta^*|y) \]

for any point $\theta^*$. By integrating both sides with respect to $\theta$ we obtain that

\[ \pi(\theta^*|y) = \frac{\int \alpha(\theta, \theta^*)q(\theta, \theta^*)\pi(\theta|y)d\theta}{\int \alpha(\theta^*, \theta)q(\theta^*, \theta)d\theta} \]

which is estimated as

\[ \hat{\pi}(\theta^*|y) = \frac{N^{-1}\sum_{g=1}^{N}\alpha(\theta^{(g)}, \theta^*)q(\theta^{(g)}, \theta^*)}{J^{-1}\sum_{j=1}^{J}\alpha(\theta^*, \theta^{(j)})} \]

where $\{\theta^{(g)}\}$ are the sampled values from the posterior distribution and $\{\theta^{(j)}\}$ are draws from $q(\theta^*, \theta)$. In our application of this method with the random walk Metropolis we need to sample $\theta^{(j)} \sim N(\theta^*, \tau \Sigma)$, $j = 1, \ldots, J$.

A suite of R and Fortran functions were written by the author to run the Markov chains, compute the marginal likelihood for each model, Bayes factors, posterior model probabilities and to process the MCMC output.

Of course, in GARCH models one is interested in the estimation of in-sample volatilities and the prediction of future volatilities. These are easily obtained using the MCMC output by calculating the value of $h_t^{(i)}$ for each draw $\theta^{(i)}$ and the posterior mean of each conditional variance can be approximated by the sample average of these values.

## 3 Application

For an empirical example we consider the Daily Exchange Rates of several currencies relative to the US dollar. This data set is available from the URL http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/. The data were transformed in the standard way by taking logarithms of the ratio of consecutive daily exchange rates. Exchange rate returns have been analysed by several authors using GARCH models, mainly with (symmetric) normal and Student-\(t\) distributions for the errors. We analysed the Australian Dollar, British Pound, Canadian Dollar, Dutch Guilder, French Franc, German Mark and Japanese Yen exchanges rates from 3/1/1990 to 31/12/1998.

For each of the six competing models, we ran a total of 20,000 iteration discarding the first half as burn-in. For each model, the simulated Markov chains were checked for convergence and good mixing by visual inspection of the marginal
traces, density estimates, autocorrelations and formal tests. In all models, these convergence diagnostics did not indicate lack of convergence. This is important to be checked in any application since in order to be efficient for estimating the marginal likelihood the sampling scheme should be efficient for sampling the posterior distribution. We then simulated another 10,000 iterations from the proposal distribution to obtain an approximation for the marginal likelihood. Also, when comparing models we considered equal prior model probabilities.

Table 2 shows MCMC estimates of the posterior model probabilities for each data set. Note the overwhelming superiority of the fat-tailed skew conditional distribution for all data considered.

Table 2: Estimates of posterior model probabilities for Daily Exchange Rates of currencies relative to the US dollar.

<table>
<thead>
<tr>
<th></th>
<th>normal</th>
<th>t</th>
<th>GED</th>
<th>skew normal</th>
<th>skew t</th>
<th>skew GED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian Dollar</td>
<td>0.0000</td>
<td>0.0137</td>
<td>0.00080</td>
<td>0.0000</td>
<td>0.5861</td>
<td>0.3994</td>
</tr>
<tr>
<td>British Pound</td>
<td>0.0000</td>
<td>0.1636</td>
<td>0.00001</td>
<td>0.0000</td>
<td>0.8364</td>
<td>0.0000</td>
</tr>
<tr>
<td>Canadian Dollar</td>
<td>0.0000</td>
<td>0.2787</td>
<td>0.00011</td>
<td>0.0000</td>
<td>0.7209</td>
<td>0.0004</td>
</tr>
<tr>
<td>Dutch Guilder</td>
<td>0.0000</td>
<td>0.2196</td>
<td>0.00000</td>
<td>0.0000</td>
<td>0.7804</td>
<td>0.0001</td>
</tr>
<tr>
<td>French Franc</td>
<td>0.0000</td>
<td>0.1757</td>
<td>0.00001</td>
<td>0.0000</td>
<td>0.8241</td>
<td>0.0001</td>
</tr>
<tr>
<td>German Marc</td>
<td>0.0000</td>
<td>0.2976</td>
<td>0.00001</td>
<td>0.0000</td>
<td>0.7024</td>
<td>0.0001</td>
</tr>
<tr>
<td>Japanese Yen</td>
<td>0.0000</td>
<td>0.0026</td>
<td>0.00001</td>
<td>0.0000</td>
<td>0.9974</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

4 Concluding Remarks

The main contribution of this paper was to provide computational tools for estimating and comparing GARCH models using MCMC methods and an approximation to the Bayes factor. We illustrated by comparing the usual symmetric normal error model with other error distributions that account for large kurtosis and asymmetry, which are frequently observed in financial time series. When applying the algorithms to real time series we found evidence in favour of skewed distributions for the error term. Similar empirical findings are reported by other authors (see for example Cappuccio et al. 2004).

Other recent approaches to model comparison via marginal likelihood estimation are currently under investigation by the author. In particular the methods developed in Friel and Pettit (2008), Chen (2005), Chib and Jeliazvok (2005) are very promising and we seek to adapt these methods to the context of models in the GARCH family.
Finally, the MCMC scheme adopted here is easy to implement and was used mainly for illustration purposes. We do not claim that this is the most efficient way of simulating the Markov chains. Of course we would need to adapt the computation of the marginal likelihood if other Metropolis-Hastings algorithms were used. Also, the analysis can be extended to variants of the GARCH model.

References


Verhoeven, P. and M. McAleer (2003). Fat tails and asymmetry in financial volatility models. Technical report, Faculty of Economics, University of Tokyo, CIRJE-F-211.