BAYESIAN ESTIMATION OF THE GJR-GARCH (*p, q*) MODEL

By: Resa Mae R. Sangco
Adviser: Arnulfo P. Supe, Ph.D.

Abstract

This study derives the posterior distribution of the GJR-GARCH (*p, q*) model in Bayesian approach, and used student-*t* distribution as prior error distribution. The Markov chain Monte Carlo (MCMC) method, particularly the Metropolis-Hastings Algorithm is used in estimating the parameters of the GJR-GARCH (*p, q*) model. Furthermore, the estimates of the classical Maximum Likelihood Estimation (MLE) is compared to the estimates of Bayesian Estimation in terms of Mean Squared Error (MSE). Simulation study shows that Bayesian estimates of the GJR-GARCH (*p, q*) model with student-t error distribution is more efficient than the classical Maximum Likelihood estimates.

Background of the Study

In many financial applications, forecasting the values of an economic commodity, like natural resources such as oil, basic foods and other classes of assets such as stocks, is very important information. Financial data is replete with the effect of volatility which causes instability in the prices of security. It is therefore important that financial time series be accurately modeled and predicted so that gains or profits can be optimized and economic decision will be more reliable.

Robert Engle [2] in 1982 developed the Autoregressive Conditional Heteroskedastic (ARCH) model which was the first model to model time-varying volatility. For this contribution, Engle won the 2003 Nobel Memorial Prize in Economics. ARCH model grew rapidly as a volatility forecasting technique during the last thirty years and have been applied to numerous economic and financial data series. However, in many applications with the ARCH model, a long lag length or a large number of parameters are required to approximately model the data. Thus, Engle's student, Tim Bollerslev [1], developed the Generalized Autoregressive Conditional Heteroskedastic (GARCH) model in which there is past conditional errors aside from conditional variance as part of the model. GARCH model had the same properties as the ARCH model but requires less parameters to precisely model heteroskedasticity. When using these models, there is an imposed restriction on the parameters to assure that the variance is positive. For this reason, Nelson [20] in 1991 presented an alternative way to the GARCH model, the Exponential GARCH (EGARCH) model, by modifying it to allow asymmetric effect of positive and negative stock return. Another model which allows the positive and negative shocks to have different impact in the volatilities is the GJR-GARCH model, which was introduced by Glosten, Jagannathan and Runkle (1993). Later, many models were developed and extended regarding volatility models.

In estimating the parameters of these financial time series model, one of the main objectives of many researcher is to find an estimate closer to the true value, most commonly used estimation procedure is the Maximum Likelihood Estimation (MLE). Maximum Likelihood estimates are known to be statistically efficient and its likelihood ratio test provide a powerful and general method of inference. However, the complexity of the computations of maximum likelihood estimates made it has become less practical in numerous situations.

Thus, this paper focuses on the Bayesian estimation of the GJR-GARCH (*p, q*) model, and consider student-*t* distribution as distribution of error of the model.

Objectives of the Study

This study aims to provide parameter estimates of the GJR-GARCH (*p, q*) model using the Bayesian approach. Specifically, this study aims:

1. To provide estimates of the GJR-GARCH (*p, q*) model with student-*t* distribution error and derive
   1.1 The posterior densities of γ;
   1.2 The posterior densities of parameters α and β;
   1.3 The full conditional posterior density of the latent variable ω; and
   1.4 The posterior densities of degrees of freedom parameter v.

2. To apply the Metropolis-Hastings algorithm using the R software computing the Bayesian estimates of the parameters of the GJR-GARCH (*p, q*) model with student-*t* distribution

3. To compare the classical Maximum Likelihood estimates to the estimates of Bayesian Estimation using GJR-GARCH (*p, q*) with student-*t* error distribution.

Methodology

GJR-GARCH model was named from the authors who introduced it, Glosten, Jagannathan and Runkle [3] as an alternative way to model asymmetric effects. Following Ardia (2008), the GJR-GARCH (*p, q*) model is defined as follows:

$$u_t = \varepsilon_t(\sigma h_t)^{1/2}$$
$$\varepsilon_t \sim iid\, S_v(0,1)$$
$$\sigma = \frac{v-2}{v}$$
$$h_t = \alpha_0 + \sum_{i=1}^{q}\left(\alpha_i \mathrm{I}_{\{u_{t-i}\geq 0\}} + \alpha_i{}^* \mathrm{I}_{\{u_{t-i}<0\}}\right)u_{t-i}{}^2 + \sum_{j=1}^{p}\beta_j h_{t-j}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0\ (i = 1, \dots, q)$, $\alpha_i{}^* \geq 0\ \ (i = 1, \dots, q)$ and $\beta_j \geq 0\ \ (j = 1, \dots, p)$ to guarantee that the conditional variance is positive. This model is equivalent to the model given by

$$u_t = \varepsilon_t(\omega_t \sigma h_t)^{1/2}$$
$$\varepsilon_t \sim iid\, N(0,1)$$
$$\omega_t \sim iid\, IG\left(\frac{v}{2}, \frac{v}{2}\right)$$

where *IG* denotes the Inverted Gamma density with parameters $\frac{v}{2}$ and $\frac{v}{2}$. ε_t is a sequence of independent and identically distributed random variables with $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = 1$.

In Bayesian inference, MCMC allows one to generate samples of parameters in the model from joint posterior distribution. One of the MCMC technique is the Metropolis-Hastings algorithm.

Metropolis-Hastings (MH)

The Metropolis-Hastings algorithm is one of the Markov Chain Monte Carlo (MCMC) sampling algorithm introduced by Metropolis, et al (1953) which allows to generate draws from any density of interest. Some complicated Bayesian Problem is hard to solve using the Gibbs sampling and this method is similar to Gibbs sampler that is for *k*th iteration, the distribution of $\theta^{(k)}$ will converge to $\pi(\theta)$ for large enough *k* but this is the case when the full conditional densities are of unknown form or when it is not easy to break down the join density into full conditionals.

Suppose we have a density that can generate candidates, so since we are dealing with Markov Chains this implies that the density will depend on the current state of the process say. Consider a *candidate- generating density or a proposal density* denoted by $q(x,y)$ where $\int q(x,y)dy = 1$. This density means that when a process is at point *x* and the density generates a value *y* for $q(x,y)$. Then we can say that

$$\pi(x)q(x,y) = \pi(y)q(y,x)\ ,$$

for all *x, y*. But if the process moves from *x* to *y* too often, that is

$$\pi(x)q(x,y) > \pi(y)q(y,x),$$

then the convenient way to correct this condition is to reduce the number of moves from *x* to *y* by introducing *acceptance probability* $\alpha(x,y) < 1$ where $\alpha(x,y)$ is called the probability of move. If the move is not made, then the process will return to *x* as the value from the proposal density. Thus, transitions kernel from *x* to *y* are made according to

$$p(x,y) \equiv q(x,y)\alpha(x,y) \quad x \neq y\ ,$$

where $\alpha(x,y)$ must be determined. From equation (4.7), since the movement from *x* to *y* is not made enough, we should have defined $\alpha(x,y)$ to be as large as possible. The probability of move $\alpha(x,y)$ is determined by requiring $p(x,y)$, then

$$\begin{aligned}\pi(x)q(x,y)\alpha(x,y) &= \pi(y)q(y,x)\alpha(y,x) \\ &= \pi(y)q(y,x)\ .\end{aligned}$$

We now see that

$$\alpha(x,y) = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}.$$

Thus, the probability of move must be

$$\alpha(x,y) = min\left\{\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right\},$$

and since $\alpha(x,y)$ is a probability, that's why its upper limit is 1.

There are two-stage process to generate $\theta^{(k+1)}$ in M-H algorithm. The first stage is to generate a candidate θ^* from a proposal density $q\left(\bullet|\theta^{(k-1)}\right)$ which depends on the current state, $\theta^{(k)}$. And the second stage is to accept or reject θ^* with the acceptance probability $\alpha(x,y)$.

The algorithm of Metropolis-Hasting consists of the following steps:

1. Initialize the iteration counter to $j = 1$ and set an initial value $\theta^{(0)}$;
2. Move the chain to a new value θ^* generated from the proposal density $q\left(\bullet|\theta^{(j-1)}\right)$;
3. Evaluate the acceptance probability of the move from $\theta^{(j-1)}$ to θ^* given by:

$$\min\left\{\frac{p(\theta^*|y)}{p(\theta^{(j-1)}|y)}\frac{q\left(\theta^{(j-1)}|\theta^*\right)}{q(\theta^*|\theta^{(j-1)})},1\right\}.$$

   If the move is accepted, set $\theta^{(j)} = \theta^*$, if not, set $\theta^{(j)} = \theta^{(j-1)}$ so that the chain does not move;
4. Change the counter from *j* to *j+1* and go back to step 2 until convergence is reached.

As the number of iteration increases, the chain approaches its equilibrium distribution same in the Gibbs sampler. The power of the M-H algorithm is from the fact that for any proposal *q* who's supported by the joint posterior distribution is obtained. If the proposal density is symmetric, then $q\left(\theta^{(j)}|\theta^*\right) = q\left(\theta^*|\theta^{(j)}\right)$, the acceptance probability of the M-H algorithm will become

$$min\left\{\frac{p(\theta^*|y)}{p(\theta^{(j)}|y)},1\right\}.$$

So that the proposal density does not need to be evaluated.

Result

A linear regression model can be written as

$$y_t = \boldsymbol{x}_t{}'\boldsymbol{\gamma} + \boldsymbol{u}_t \quad \text{for } t = 1, \ldots, T,$$

where y_t is a scalar dependent variable, $\boldsymbol{x}_t$ is a *m* x *1* vector of exogenous independent variable, $\boldsymbol{\gamma}$ is a *m* x *1* vector of regression coefficients.

Theoretical Results

Using the following priors,

$$p(\alpha) \propto N_{2q+1}(\alpha|\mu_\alpha, \Sigma_\alpha) I_{\{\alpha>0\}}$$
$$p(\beta) \propto N_p\left(\beta|\mu_\beta, \Sigma_\beta\right) I_{\{\beta>0\}}$$

The Posterior Density of $\boldsymbol{\alpha}$

The approximate likelihood function of α is

$$L(\alpha|\gamma, \beta, y, X) \propto \exp\left\{-\frac{1}{2}(v - C'\alpha)'\Lambda^{-1}(v - C'\alpha)\right\}$$

Then the posterior density of α is

$$\pi(\alpha|y) \propto \exp\left\{-\frac{1}{2}(\alpha - \hat{\mu}_\alpha)'\hat{\Sigma}_\alpha{}^{-1}(\alpha - \hat{\mu}_\alpha)\right\}$$

with

$$\hat{\Sigma}_{\alpha}^{-1} = C'\tilde{\Lambda}^{-1}C + \Sigma_{\alpha}^{-1} \text{ and}$$
$$\hat{\mu}_{\alpha} = \hat{\Sigma}_{\alpha}^{-1}(C'\tilde{\Lambda}^{-1}v + \Sigma_{\alpha}^{-1}\mu_{\alpha})$$

where the T x T diagonal matrix $\tilde{\boldsymbol{\Lambda}} = diag(\{2{h_t}^2(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta})\}_{t=1}^{T})$ and $\tilde{\boldsymbol{\alpha}}$ is the previous draw of $\boldsymbol{\alpha}$ in the M-H sampler.

The Posterior Density of β

The approximate likelihood function of parameter β is,

$$L(\beta|\alpha, \gamma, y, X) \propto \exp\left\{-\frac{1}{2}(b - \varphi\,\beta)'\Lambda^{-1}(b - \varphi\,\beta)\right\}$$

And the prior density of β is given above, then the posterior density is

$$\pi(\beta|\alpha, \gamma, y, X) \propto \exp\left\{-\frac{1}{2}(\beta - \hat{\mu}_{\beta})'\hat{\Sigma}_{\beta}^{-1}(\beta - \hat{\mu}_{\beta})\right\}$$

with

$$\hat{\Sigma}_{\beta}^{-1} = \varphi'\tilde{\Lambda}^{-1}\varphi + \Sigma_{\beta}^{-1} \text{ and}$$
$$\hat{\mu}_{\beta} = \hat{\Sigma}_{\beta}^{-1}(\varphi'\tilde{\Lambda}^{-1}b + \Sigma_{\beta}^{-1}\mu_{\beta})$$

where the T x T diagonal matrix $\tilde{\Lambda} = diag(\{2{h_t}^2(\alpha, \tilde{\beta})\}_{t=1}^{T})$ and $\tilde{\beta}$ is the previous draw of β in the M-H sampler.

Posterior Density of ω

The full conditional density of $\boldsymbol{\omega}$ is straightforward to derive. Note that $\omega_1, \omega_2, \dots, \omega_T$ are independent and identically distributed random variables from an Inverted Gamma density given by,

$$p(\omega_t|v) \propto \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\tau\left(\frac{v}{2}\right)}{\omega_t}^{-\frac{v}{2}-1}\exp\left\{-\frac{v}{2\omega_t}\right\}$$
$$\propto {\omega_t}^{-\frac{v}{2}-1}\exp\left\{-\frac{v}{2\omega_t}\right\}.$$

Then the joint density of T x 1 vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_T)'$ is

$$p(\boldsymbol{\omega}|v) \propto \prod_{t=1}^{T}{\omega_t}^{-\frac{v}{2}-1}\exp\left\{-\frac{v}{2\omega_t}\right\}.$$

The likelihood function of $\boldsymbol{\omega}$ is

$$L(\boldsymbol{\omega}|\boldsymbol{y}, \boldsymbol{X}) \propto \prod_{t=1}^{T}(\omega_t)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\frac{(y_t - {x_t}'\gamma)^2}{\omega_t\sigma h_t}\right]\right\}.$$

Then using Bayes Theorem, we obtain the joint posterior as

$$\pi(\boldsymbol{\omega}|\boldsymbol{u}) \propto \prod_{t=1}^{T}(\omega_t)^{-\frac{v+1}{2}-1} \exp\left\{-\frac{1}{2\omega_t}\left[\frac{(y_t - x_t'\gamma)^2}{\sigma h_t} + v\right]\right\},$$

which is the kernel of an Inverted Gamma Density with parameters $\frac{v+1}{2}$ and $\frac{1}{2}\left[\frac{(y_t-x_t'\gamma)^2}{\sigma h_t} + v\right]$.

The Posterior Density of v

The translated exponential density with parameters $\lambda > 0$ and $\delta \geq 2$ is given by

$$p(v) = \lambda \exp[-\lambda(v - \delta)] .$$

The prior density of vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_T)'$ conditional on v given in is

$$p(\boldsymbol{\omega}|v) = \left(\frac{v}{2}\right)^{\frac{Tv}{2}} \left[\tau\left(\frac{v}{2}\right)\right]^{-T} \prod_{t=1}^{T} \omega_t^{-\frac{v}{2}-1} \exp\left\{-\frac{v}{2\omega_t}\right\}.$$

Thus, the posterior density is

$$p(v|\omega) \propto \left(\frac{v}{2}\right)^{\frac{Tv}{2}} \left[\tau\left(\frac{v}{2}\right)\right]^{-T} \exp\{-\Delta v\} ,$$

where $\Delta = \lambda + \frac{1}{2}\left[\sum_{t=1}^{T}\left(\frac{1}{\omega_t} + \ln \omega_t\right)\right]$.

Simulation Results

The parameters $\alpha_i(i = 0,1, \ldots, q)$, $\alpha_i{}^*(i = 1, \ldots, q)$, $\beta_j(j = 1, \ldots, p)$ and v was preset to obtain the simulated data with sample sizes 100, 500, 2500, 1000 and 5000. From this data, estimates of the parameters where computed using Bayesian analysis through MCMC simulation with Metropolis-Hastings algorithm. The MCMC technique particularly the Metropolis- Hastings algorithm was used to draw a sample values from the posterior density of the parameters. The first step in MCMC procedure is to set the following initial values of the parameters

$$\theta^{(0)} = \left(\alpha^{(0)}, \beta^{(0)}, v^{(0)}\right)$$

then we generate iteratively a new value $\theta^{(j)} = \left(\alpha^{(j)}, \beta^{(j)}, v^{(j)}\right)'$from the posterior density of the parameters. For each MCMC simulation, two chains were run with 5,000 iterations for each chain. The burn-in period was set to 2,500 for each chain. This implies that the first 2,500 iterations from each chain were disregarded. Hence, a total of 5,000 values of the chain were considered as samples from the true posterior density of the parameters. The mean of the sample values of each parameter is the estimate of the parameter. These samples values were also used to find the summary statistics and to make inferences about the joint posterior.

The mean of the sample values of each parameter is the estimate of the parameter. The Bayesian and Maximum likelihood parameter estimate and their Mean Squared Errors (MSE) were calculated using 100 replications. The result for *n* = 500 and *n* = 5000 is given below.

Comparing the performance of the classical Maximum Likelihood Estimation to the Bayesian estimation of the GJR-GARCH (2, 2) model with student-*t* error distribution. For *n*=500 in Table 1, we observed that the parameters of the Bayesian estimates are closer to the true values except on the parameters α_1, α_2 and β_2. Examining the MSE, it shows that the Bayesian estimates has a smaller MSE value except on parameters α_1 and α_2. For *n* = 5000, the Bayesian estimation still perform better that the classical MLE estimation.

Table 1 Bayesian and Classical Estimates of Student-*t* Distribution *n* = 500

	Parameter	True Value	Mean Estimate		MSE	
			Bayesian	MLE	Bayesian	MLE
GJR-GARCH(2,2)	α_0	0.02	0.0192427	0.1363798	**0.00000058**	0.063987
	α_1	0.06	0.1173753	0.0491917	0.00329339	**0.002394**
	α_2	0.04	0.1045225	0.0420948	0.00416427	**0.001935**
	$\alpha_1{}^*$	0.04	0.1044203	0.1132863	**0.00415148**	0.315362
	$\alpha_2{}^*$	0.05	0.1002810	0.1467286	**0.00252925**	0.426304
	β_1	0.1	0.1631649	0.1981398	**0.00399112**	0.081455
	β_2	0.2	0.2786250	0.2382006	**0.00618362**	0.083732
	v	7	7.0284896	7.5999792	**0.00084780**	3.481917

Table 2 Bayesian and Classical Estimates of Student-*t* Distribution *n* = 5000

	Parameter	True Value	Mean Estimate		MSE	
			Bayesian	MLE	Bayesian	MLE
GJR-GARCH(2,2)	α_0	0.02	0.01815224	0.05017454	**0.00000341**	0.0073761
	α_1	0.06	0.04888406	0.06347442	0.00012387	**0.0008280**
	α_2	0.04	0.04999618	0.03612814	**0.00010027**	0.0009979
	$\alpha_1{}^*$	0.04	0.05464162	0.07634240	**0.00021465**	0.1505255
	$\alpha_2{}^*$	0.05	0.04717600	0.17983332	**0.00000824**	0.2774307
	β_1	0.1	0.13912346	0.20697849	**0.00153212**	0.0555703
	β_2	0.2	0.24859422	0.16277576	**0.00236284**	0.0440146
	v	7	7.00689462	7.22856360	**0.00005341**	1.1063600

Figure 1 shows the graph of the true value, the MCMC Bayesian estimates and the MLE estimates for parameter α_1. For *n* = 50, some of the values in classical estimates are far from the true value. As *n* increases, the values tend to get closer to the true values. The behavior of the Bayesian method on the other provide estimations with small fluctuation for all the sample sizes. Hence, Bayesian estimation gives a smaller error for all sample sizes.

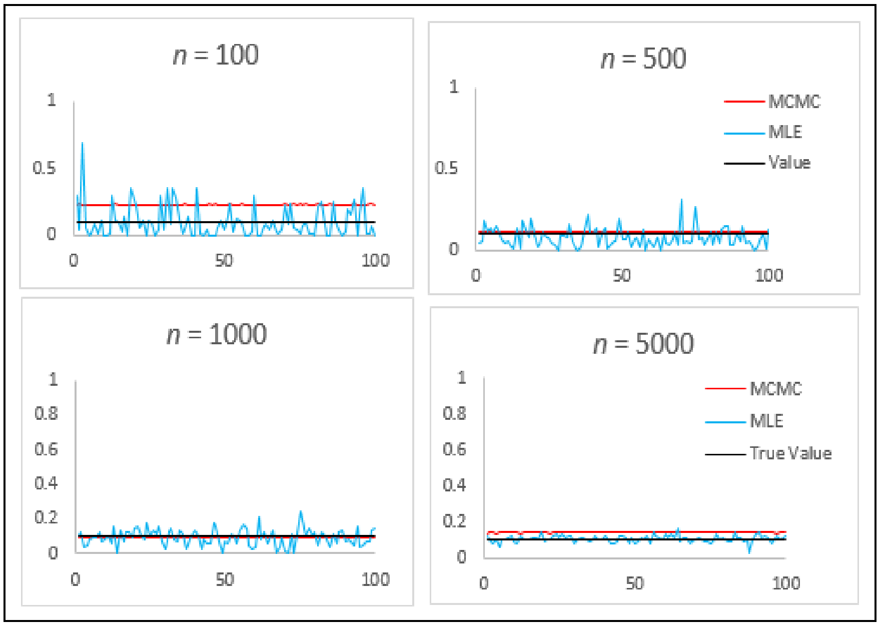

Figure 1 GJR-GARCH(1,1) Bayesian and Classical Estimates for α_1

Conclusion

The study shows that the Bayesian estimation of the GJR-GARCH (*p, q*) model with student-*t* distribution provides a better estimator than the classical Maximum Likelihood estimation.

References

[1] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307-327.

[2] Engle, R.F., (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica 50, 987-1007.

[3] Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993). On the Relation between the Expected Value and the Volatility of Nominal Excess Return on Stocks. Journal of Finance 48, 1779-1801.