Assistant Professor Yakup ARI, PhD
Department of Economics and Finance
Alanya Alaaddin Keykubat University, Alanya, TURKEY
E-mail: yakup.ari@alanya.edu.tr
Professor Emeritus Alex PAPADOPOULOS, PhD
Department of Mathematics and Statistics
University of North Carolina at Charlotte, Charlotte NC, USA
E-mail: apapdpls@uncc.edu

BAYESIAN ESTIMATION OF STUDENT-T GARCH MODEL USING LINDLEY'S APPROXIMATION

Abstract. *The dependency of conditional second moments of financial time series is modelled by Generalized Autoregressive conditionally heteroscedastic (GARCH) processes. The maximum likelihood estimation (MLE) procedure is most commonly used for estimating the unknown parameters of a GARCH model. In this study, the parameters of the GARCH models with student-t innovations are discussed for estimations using the Bayesian approach. It is assumed that the parameters of the GARCH model are random variables having known prior probability density functions. Lindley's approximation will be used to estimate the Bayesian estimators since they are not in a closed form. The Bayesian estimators are derived under squared error loss function. Finally, a simulation study is performed in order to compare the ML estimates to the Bayesian ones and in addition to simulations an example is given in order to illustrate the findings. MLE's and Bayesian estimates are compared according to the expected risks in the simulation study which shows that as the sample size increases the expected risks decrease and also it is observed that Bayesian estimates have performed better than MLE's.*

Keyword:*GARCH, MLE, Lindley's Approximation, Bayesian Methods, Squared Error.*

JEL Classification:C11, C15, C22, C51

1. Introduction

The Generalized Autoregressive Conditionally Heteroscedastic (GARCH) model that is introduced by Bollerslev (1986) assumes that the conditional variance depends on its own p past values and q past values of the squared error terms. This model is represented as GARCH(p,q). Especially, the GARCH (1,1) model is very successful to capture the volatility of financial data in most applications which

DOI: 10.24818/18423264/53.1.19.05

have been utilized widely by practitioners and academicians. Therefore, it has been studied extensively in the literature, see for example Tsay (2013). The most common method to make inferences for the GARCH model is Maximum Likelihood Estimation (MLE) since it is easy to implement and is available in statistical software packages. Furthermore, the MLE's are asymptotically optimum that have been shown by Lee and Hansen (1994) and Bollereslev *et al.* (1994).

The Bayesian approach is another well-known approach to make inferences about the GARCH models. The parameters of the model that have known prior probability function are assumed to be random variables in the Bayesian approach. Then, the prior information about the distribution of the parameters along with the data is used to determine the posterior distribution which is a base for statistical inferences regarding the parameter.In Bayesian studies, the posterior distribution for the parameters of GARCH model is found to be analytically intractable. Hence, numerical or a proper approximation method is required to get inferences for the parameters.Markov Chain Monte Carlo (MCMC) techniques are the primary method that makes enable to draw samples from the posterior and predictive distributions using Markov chains and thus it satisfies that sample averages can be used to approximate expectations.There are various ways of generating the required Markov chain; particularly, Metropolis-Hastings (MH) algorithm, introduced by Metropolis *et al.* (1953), can be used to converge target distribution using an acceptance/rejection rule.And also, all other MCMC methods can be considered as special cases of MH algorithm. In the time series studies, MCMC procedures are applied to determine the joint posterior distribution of the GARCH (1,1) model by Müller and Pole (1998), Nakatsuma (2000) and Ardia (2010). In addition to these studies, the Griddy Gibbs sampler approach is proposed to estimate the GARCH models by Bauwens *et al.* (1998). Kim et al. (1998) utilized the adaptive rejection Metropolis sampling (ARMS) technique that is proposed by Gilks et al (1995) is used for estimating the parameters of GARCH-t models.ARMS is applied by developing an envelope function of the log of the target density, which is then utilized in rejection sampling.

Mitsui and Watanabe (2003) developed another Bayesian estimation method that is a Taylored approach based on the acceptance-rejection Metropolis-Hastings algorithm. This method can be applied for any kind of parametric ARCH-type models. Marín et al. (2015) use the data cloning methodology to estimate the parameters of the GARCH and Continuous GARCH model. In the data cloning methodology, a Bayesian approach is used to obtain approximate maximum likelihood estimators of GARCH and continuous GARCH models avoiding numerically maximization of the pseudo-likelihood function. One can find the review of the existing literature on the most relevant Bayesian inference methods for GARCH models in the papers Asai (2006) and Virbickaite et al. (2015) that contain the comparison of the Bayesian approach versus classical procedures.

The aim of the study is to derive Bayesian estimators for the Student-t GARCH model using Lindley's approximation. The symmetric SEL function is considered in this study since it is the most commonly used loss function for constructing a Bayes estimator of the parameter and, due to mathematical tractability that was pointed out by Moore and Papadopoulos (2000). The paper is organised as follows; general form of the GARCH Model, GARCH Models with Standardized Student-t distribution, Squared Error Loss Function and Expected Risk, Lindley's Approximation and Bayesian Estimation of the GARCH model are mentioned in the section under the methodology topic. Simulation study and illustration for the method are given in the separate sections. The last part of the study includes conclusions and discussions.

2. METHODOLOGY

2.1 GARCH Model

The variance equation of the GARCH (p,q) model can be expressed as

$$a_t = \sigma_t \varepsilon_t$$
$$\varepsilon_t \sim f_v(0,1)$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 \tag{1}$$

where $f_v(0,1)$ is the probability density function of the innovations or residuals with zero mean and unit variance. In non-normal case, v are used as additional distributional parameters for the scale and the shape of the distribution

Bolerslev (1986) has shown that the GARCH(p,q) process is covariance stationary with E(a_t) = 0, var(a_t) = α_0 / $(1 - \left(\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i\right))$ and cov(a_t, a_s) = 0 for t ≠ s if and only if $\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i < 1$. He used the MLE method by maximizing the given log-likelihood function

$$L(\vartheta) = ln \prod_t f_v(a_t, E(a_t|I_{t-1}), \sigma_t)$$

where f_v is the conditional distribution function. The second argument of $f_v$denotes the mean, and the third argument the standard deviation. The full set of parameters ϑ includes the parameters from the variance equation $\vartheta = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$and the distributional parameters (v) in the case of a non-normal distribution function.

2.2 GARCH Models with Standardized Student-t Distribution

In some applications it should be assumed that the errors are follow a heavy tailed distribution such as a standardized Student t-distribution. Bollerslev (1987) proposed the GARCH model with standardized Student t-distribution for the innovations which is called GARCH-t model. For example, in the financial markets asset returns often exhibit heavy tails, and as is pointed by Ardia (2010) a distribution with heavy tails makes extreme outcomes such as crashes more likely than does a normal distribution.

The likelihood function of the GARCH-t model is

$$f\ \left(a_{p+1}, a_{p+2}, \dots, a_T \middle| \alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q\right) = \prod_{t=p+q+1}^{T} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi}} \frac{1}{\sigma_t} \left(1 + \frac{a_t^2}{(\nu-2)}\right)^{-\frac{\nu+1}{2}} \quad (2)$$

and the Log-likelihood function is

$$L = \sum_{t=p+q+1}^{n} \left[ln\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{1}{2} ln\left((\nu-2)\pi\right) - \frac{1}{2} ln\sigma_t^2 - \frac{\nu+1}{2} ln\left(1 + \frac{a_t^2}{(\nu-2)\sigma_t^2}\right)\right] \quad (3)$$

2.3 Squared Error Loss Function and Expected Risk

Loss function defines the "penalty" that one pays when θ is estimated by $\hat{\theta}$. Bayesian estimates are based on minimization of the expected loss function. The expected loss is integrated over all possible settings of θ weighted by their relative probabilities and indicates how much loss can be expected when $\hat{\theta}$ is chosen as the estimate. The optimal decision procedure has to choose a $\hat{\theta}$ that minimizes this expected loss.

$$\hat{\theta}^* = E\left[L\left(\hat{\theta}, \theta\right)\right] = min_\theta \int L\left(\hat{\theta}, \theta\right) h(\theta | \underline{x}) d\theta \quad (4)$$

The Squared Error Loss Function $L\left(\hat{\theta}, \theta\right) = \left(\hat{\theta} - \theta\right)^2$ is a symmetrical loss function. The Squared Error Loss Functiongives equal losses to over estimation and underestimation. The Bayes estimator under squared error loss function is $\hat{\theta} = E\left[\theta | \underline{x}\right]$

The Bayesian and ML estimators of some distributions were compared by using expected risks (ERs) of Monte Carlo simulations (Nadar et.al, 2015). The expected risk (ER) of θ under the SEL function is

$$ER(\theta) = \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\theta}_n - \theta_n\right)^2$$

2.4 Lindley's Approximation

Lindley (1980) developed approximate procedures for the evaluation of the ratio of two integrals which are in the form of

$$\frac{\int v(\underline{\theta}) \exp\{L(\underline{\theta})\} d\underline{\theta}}{\int g(\underline{\theta}) \exp\{L(\underline{\theta})\} d\underline{\theta}} \quad (5)$$

where $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$, $L(\underline{\theta})$ is the logarithm of the likelihood function, and $g(\underline{\theta})$ and $v(\underline{\theta}) = u(\underline{\theta})g(\underline{\theta})$ are arbitrary functions of $\underline{\theta}$. The posterior expectation of the function $u(\underline{\theta})$, for given $\underline{x}$, is

$$E\left[u(\underline{\theta})\middle|\underline{x}\right] = \frac{\int u(\underline{\theta}) \exp\{L(\underline{\theta}) + \rho(\underline{\theta})\} \, d\underline{\theta}}{\int \exp\{L(\underline{\theta}) + \rho(\underline{\theta})\} \, d\underline{\theta}} \quad (6)$$

where $exp\{L(\underline{\theta}) + \rho(\underline{\theta})\}$is the the posterior distribution of $\underline{\theta}$ except for the normalizing constant and $\rho(\underline{\theta}) = lng(\underline{\theta})$. Expanding $exp\{L(\underline{\theta}) + \rho(\underline{\theta})\}$in equation (6) into a Taylor series expansion about the ML estimates of $\underline{\theta}$ gives $E\left[u(\underline{\theta})\middle|\underline{x}\right]$. So, $E\left[u(\underline{\theta})\middle|\underline{x}\right]$ can be estimated asymptotically by

$$\hat{u}_B = u + \frac{1}{2}\sum_i\sum_j (u_{ij} + 2u_i\,\rho_j)\varphi_{ij} + \frac{1}{2}\sum_i\sum_j\sum_k\sum_l L_{ijk}\,\varphi_{ij}\varphi_{kl}u_l \quad (7)$$

where $i, j, k, l = 1,2, \ldots, n$, and

$$u = u(\underline{\theta}),\ \ u_i = \frac{\partial u}{\partial \theta_i},\ \ u_{ij} = \frac{\partial^2 u}{\partial \theta_i \partial \theta_j},\ \ L_{ijk} = \frac{\partial^3 L}{\partial \theta_i \partial \theta_j \partial \theta_k}, \rho_j = \frac{\partial \rho}{\partial \theta_j}, L_{ij} = \frac{\partial^2 L}{\partial \theta_i \partial \theta_j}$$

and φ_{ij} is the $(i,j)th$ element of the inverse matrix $\{-L_{ij}\}$ and all are evaluated at the MLE of the parameters.

2.5 Bayesian Estimation of the Parameters of GARCH(p,q)Model

Let $\{a_t\}$ where $t = 1,2, \ldots,$ n, denote the GARCH(p,q) process defined by equation (1) where the parameters $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i = 1,2, \ldots, p$ and $\beta_i \geq 0$ for $i = 1,2, \ldots, q$. In this study it will be assumed that the process is stationary and thus the coefficients $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ satisfy the condition $\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i < 1$.

Let $\vartheta = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ denote the parameters of the GARCH(p,q) model and $\underline{x} = (a_{p+1}, a_2, \ldots, a_T)$ the observed series. If a_t is distributed with mean zero and conditional variance σ_t^2, then the density function is

$$f(a_t|\vartheta, a_0, a_1, \ldots, a_{t-1}) = \prod_{t=p+1}^{T} f_v(a_t, E(a_t|I_{t-1}), \sigma_t) \quad (8)$$

The estimates obtained by maximizing eq (8) are known as the conditional maximum likelihood estimates. Usually, it is easier to maximize the log of the likelihood function, i.e.

$$L = ln \prod_{t=p+1}^{T} f_v(a_t, E(a_t|I_{t-1}), \sigma_t) \quad (9)$$

Bollerslev (1986) derived the MLE estimates of $\underline{\alpha}$ and $\underline{\beta}$, denoted by $\underline{\hat{\alpha}}$ and $\underline{\hat{\beta}}$ by maximizing eq(9)

It will be assumed that the parameters of the GARCH(p,q) model behave as random variables and thus they will be estimated using Bayes theorem. It will be assumed that α_0 has gamma prior ; $g_1(\alpha_0, r, \theta)$ with parameters (r, θ),

$$g_1(\alpha_0; r, \theta) = \frac{1}{\Gamma(r)\beta^r} \alpha_0^{r-1} e^{-\frac{\alpha_0}{\theta}} \quad \text{for} \quad \alpha_0 > 0 \text{ and } r, \theta > 0 \tag{10}$$

Furthermore, we will assume that the joint density function of $\tau = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$ with $p \geq 2$ is the Dirichlet probability function with parameters $\omega_1, \omega_2, \dots, \omega_{p+1}$>0 given as

$$g_2(\tau; \omega_1, \omega_2, \dots, \omega_{p+q+1}) = \frac{1}{B(\underline{\omega})} \prod_{i=1}^{p+q+1} \tau_i^{\omega_i - 1} \tag{11}$$

where $\tau_1 + \dots + \tau_{p+q+1} < 1$ and $\tau_{p+q+1} = 1 - \tau_1 - \dots - \tau_{p+q}$. The normalizing constant $B(\underline{\omega})$is the multinomial beta function given as

$$B(\underline{\omega}) = \frac{\prod_{i=1}^{p+q+1} \Gamma(\omega_i)}{\Gamma(\sum_{i=1}^{p+q+1} \omega_i)}$$

where $\underline{\omega} = (\omega_1, \omega_2, \dots, \omega_{p+q+1})$. Since α_0 and $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ are independent, their joint pdf is given by

$$g(\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q) = \frac{1}{\Gamma(r)\beta^r} \frac{1}{B(\underline{\omega})} \alpha_0^{r-1} e^{-\frac{\alpha_0}{\beta}} \prod_{i=1}^{p+q+1} \tau_i^{\omega_i - 1} \tag{12}$$

The posterior function for GARCH model in general form is given as

$$h_1(\underline{\vartheta}|\underline{x}) = \frac{\{ln \prod_{t=p+q+1}^{T} f_v(a_t, E(a_t|I_{t-1}), \sigma_t)\} \frac{\alpha_0^{r-1} e^{-\frac{\alpha_0}{\theta}}}{\theta^r \Gamma(r)} \frac{1}{B(\omega)} \prod_{i=1}^{p+q+1} \tau_i^{\omega_i - 1}}{\iint \dots \int \{ln \prod_{t=p+q+1}^{T} f_v(a_t, E(a_t|I_{t-1}), \sigma_t)\} \frac{\alpha_0^{r-1} e^{-\frac{\alpha_0}{\theta}}}{\theta^r \Gamma(r)} \frac{1}{B(\omega)} \prod_{i=1}^{p+q+1} \tau_i^{\omega_i - 1} d\alpha_0 d\tau_1 \dots d\tau_{p+q+1}} \tag{13}$$

which can not be expressed in a closed form.

The estimation of GARCH parameters under SE loss function is

$$\vartheta_{SEL}^{*} = E[(\underline{\vartheta}|\underline{x})] = \iint \dots \int u(\underline{\vartheta}) h_1(\underline{\vartheta}|\underline{x}) d\vartheta$$

$$= \frac{\iint \dots \int u(\underline{\vartheta})\{ln \prod_{t=p+q+1}^{T} f_v(a_t, E(a_t|I_{t-1}), \sigma_t)\} \frac{\alpha_0^{r-1} e^{-\frac{\alpha_0}{\theta}}}{\theta^r \Gamma(r)} \frac{1}{B(\omega)} \prod_{i=1}^{p+q+1} \tau_i^{\omega_i - 1} d\alpha_0 d\tau_1 \dots d\tau_{p+q+1}}{\iint \dots \int \{ln \prod_{t=p+q+1}^{T} f_v(a_t, E(a_t|I_{t-1}), \sigma_t)\} \frac{\alpha_0^{r-1} e^{-\frac{\alpha_0}{\theta}}}{\theta^r \Gamma(r)} \frac{1}{B(\omega)} \prod_{i=1}^{p+q+1} \tau_i^{\omega_i - 1} d\alpha_0 d\tau_1 \dots d\tau_{p+q+1}} \quad (14)$$

Lindley's approximation will be applied to find the parameters of GARCH model with Student-t distributed innovations in the following sections since above Bayesian estimate under SE loss function has no closed form.

2.6 Bayesian Estimation of the Parameters of GARCH (1,1) Model with Student-t Innovations

Then under the assumption of Student-t innovations, the conditional likelihood function of an GARCH(1,1) model is

$$f\left(a_{p+q+1}, a_{p+q+2}, \dots, a_T \middle| \underline{\vartheta}\right) = \prod_{t=2}^{T} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi}} \frac{1}{\sigma_t} \left(1 + \frac{a_t^2}{(\nu-2)\sigma_t^2}\right)^{-\frac{\nu+1}{2}}$$

Log-likelihood function is

$$L = \sum_{t=2}^{n} [\ \ln\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{1}{2}\ln\left((\nu-2)\pi\right) - \frac{1}{2} ln\sigma_t^2 + \frac{\nu+1}{2}\ln\left(1 + \frac{a_t^2}{(\nu-2)\sigma_t^2}\right)]$$

After omitting the constant terms and plugging the $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, the log-likelihood function is

$$L = \ln\left(\underline{x}\middle|\underline{\vartheta}\right) = -\sum_{t=2}^{n} \left[\frac{1}{2}\ln(\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2) + \left(\frac{\nu+1}{2}\right)\ln\left(1 + \frac{a_t^2}{(\nu-2)(\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2)}\right)\right]$$

Let $m_1 = \frac{\nu+1}{2}$, $m_t = \frac{a_t^2}{(\nu-2)}$ and $c_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$. So, log-likelihood function becomes as

$$L = -\sum_{t=2}^{n} \left(\left(\frac{1}{2} - m_1\right) lnc_t + m_1 \ln(c_t + m_t) \right)$$

It will be assumed that α_0 has gamma prior and the joint density function of α_1 and β_1 is the Dirichlet probability function that are given in the equations (10) and (11) respectively. Since α_0,α_1 and $\beta_1$are independent, their joint pdf is given by in eq(12)

The joint posterior function of α_0, α_1 and β_1

$$h_2(\underline{\vartheta}|\underline{x}) = \frac{\{\prod_{t=2}^{T}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi}}\frac{1}{\sigma_t}(1+\frac{a_t^2}{(\nu-2)})^{-\frac{\nu+1}{2}}\}\frac{\alpha_0^{r-1}e^{-\frac{\alpha_0}{\beta}}}{\beta^r\Gamma(r)}\frac{1}{B(\omega)}\prod_{i=1}^{3}\tau_i^{\omega_{i-1}}}{\iiint\{\prod_{t=2}^{T}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi}}\frac{1}{\sigma_t}(1+\frac{a_t^2}{(\nu-2)})^{-\frac{\nu+1}{2}}\}\frac{\alpha_0^{r-1}e^{-\frac{\alpha_0}{\beta}}}{\beta^r\Gamma(r)}\frac{1}{B(\omega)}\prod_{i=1}^{3}\tau_i^{\omega_{i-1}}\,d\alpha_0 d\tau_1 d\tau_2}$$

The estimation of Student-t GARCH(1,1) parameters under SEL function is:

$$\vartheta_{SEL}^{*} = E[(\underline{\vartheta}|\underline{x})] = \iiint u(\underline{\vartheta})h_2(\underline{\vartheta}|\underline{x})d\vartheta$$

$$= \frac{\iiint u(\underline{\vartheta})ln\{\prod_{t=2}^{T}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi}}\frac{1}{\sigma_t}(1+\frac{a_t^2}{(\nu-2)})^{-\frac{\nu+1}{2}}\}\frac{\alpha_0^{r-1}e^{-\frac{\alpha_0}{\beta}}}{\beta^r\Gamma(r)}\frac{1}{B(\omega)}\prod_{i=1}^{3}\tau_i^{\omega_{i-1}}\,d\alpha_0 d\tau_1 d\tau_2}{\iiint ln\{\prod_{t=2}^{T}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi}}\frac{1}{\sigma_t}(1+\frac{a_t^2}{(\nu-2)})^{-\frac{\nu+1}{2}}\}\frac{\alpha_0^{r-1}e^{-\frac{\alpha_0}{\beta}}}{\beta^r\Gamma(r)}\frac{1}{B(\omega)}\prod_{i=1}^{3}\tau_i^{\omega_{i-1}}\,d\alpha_0 d\tau_1 d\tau_2}$$

α_0, α_1 and β_1 can be estimated by using Lindley's equation which are given in Appendix A under SE loss function.

3. SIMULATIONS

The Expected Risk of Monte Carlo simulations is used to compare the Bayesian estimators with MLEs. The simulation study is done using a Student-t distributed innovations and for different sample sizes which are 200, 400, 600, 800 and 1000.

A gamma or an improper (vague) prior and Dirichlet prior are assumed as priors for α_0 and the set of parameters (α_1, β_1) respectively.

The ML and Bayes estimates of the parameters under an SE loss function are obtained using the above-mentioned innovations, sample sizes, and priors.

Table 1 presents the mean true values of each parameter that are randomly generated using the prior distributions. The average values of the ML and Bayesian estimates are given in Table 2 through Table 4 with the expected risks.

All the results are based on 1000 repetitions. The degrees of freedom parameter of the Student-t distribution is assumed as a fixed parameter that is equal to 4.

The prior of α_0 is either a gamma distribution with the parameters $r = 3$ and $\beta = 1$ or a vague prior. Dirichlet distribution is assumed as prior for α_1 and $\beta_1$with parameters $\omega_1 = 1$,$\omega_2 = 2$ and $\omega_3 = 3$. Dirichlet prior is chosen for α_1 and β_1 since it satisfies the stationarity condition for the model. All simulations are done in R 3.2.3 (R metrics)

It is observed that as the sample size increases the expected risks decrease which should be expected since the MLEs are consistent. It is also observed for the Bayes estimates when the sample sizes increase the expected risks decrease.

The expected risks, when proper priors are used for the Bayes estimates, are smaller than the MLEs and this situation is also valid when a vague prior is used for α_0. The MLEs and the Bayes' estimates when a vague prior is utilized have a little difference according to the expected risks.

Table 1. Student-t GARCH (1,1) Model Simulation: Average Values of Generated Parameters

Priors	Average Values				
α_0 Gamma(3,1)	2.400738	2.738164	2.789534	3.008514	2.925643
α_1 Dirichlet$(1,2,3)$	0.3251807	0.3129626	0.3195718	0.3367621	0.3329756
β_1 Dirichlet$(1,2,3)$	0.3541763	0.3470459	0.336607	0.3348294	0.3277689

Table 2. Student-t GARCH (1,1) Model Simulation: Results of α_0

Sample Size	*MLEs for α_0*	*ER of MLEs for α_0*	*Bayes for α_0 with vague prior*	*ER of Bayes for α_0 with vague prior*	*Bayes for α_0 with Gamma prior*	*ER of Bayes for α_0 with Gamma prior*
200	2.912492	0.5864247	2.869875	0.5482291	2.87149	0.4948494
400	2.996625	0.2808601	2.975769	0.2622993	2.97615	0.2413418
600	3.082227	0.2231313	3.069906	0.2144665	3.069987	0.203641
800	3.023868	0.130692	3.015687	0.1290079	3.015817	0.1288259
1000	2.927911	0.0956071	2.921488	0.0949001	2.921669	0.0947229

Table 3. Student-t GARCH (1,1) Model Simulation: Results of α_1

Sample Size	*MLEs for α_1*	*ER of MLEs for α_1*	*Bayes for α_1 with Dirichlet prior and α_0 with vague prior*	*ER of Bayes for α_1 with Dirichlet prior and α_0 with vague prior*	*Bayes for α_1 with Dirichlet prior*	*ER of Bayes for α_1 with Dirichlet prior*
200	0.2412551	0.0186213	0.2498856	0.0163551	0.2493107	0.0143522
400	0.2695071	0.0096773	0.2710411	0.0091329	0.2709386	0.0086339
600	0.2753703	0.0081199	0.2783593	0.0077794	0.2781737	0.0076873
800	0.3360034	0.0038143	0.3368273	0.0037795	0.3367491	0.0037791
1000	0.331809	0.0028863	0.3322847	0.0028718	0.3322128	0.002835

Table 4. Student-t GARCH (1,1) Model Simulation: Results of β_1

Sample Size	*MLEs for β_1*	*ER of MLEs for β_1*	*Bayes for β_1 with Dirichlet prior and α_0 with vague prior*	*ER of Bayes for β_1 with Dirichlet prior and α_0 with vague prior*	*Bayes for β_1 with Dirichlet prior*	*ER of Bayes for β_1 with Dirichlet prior*
200	0.2499502	0.0190152	0.2633075	0.0179713	0.2627652	0.0158404
400	0.2960466	0.0078536	0.3007127	0.0075232	0.3005076	0.0073171
600	0.2822612	0.0060151	0.2852617	0.0054279	0.2851629	0.0054135
800	0.3343117	0.0039281	0.3358129	0.0038931	0.3357285	0.003893
1000	0.3249313	0.0033352	0.3266395	0.0031782	0.3265664	0.0031269

4. ILLUSTRATION

The daily observations of the Deutschmark vs British Pound (DEM/GBP) foreign exchange log-returns are used to apply Bayesian estimation methods. The period of data is from January 3, 1985, to December 31, 1991. This dataset has been promoted as an informal benchmark for GARCH time series software validation. The first 750 observations are used to illustrate the estimation method (Ardia, 2010). The plot of the dataset is shown in Figure 1.

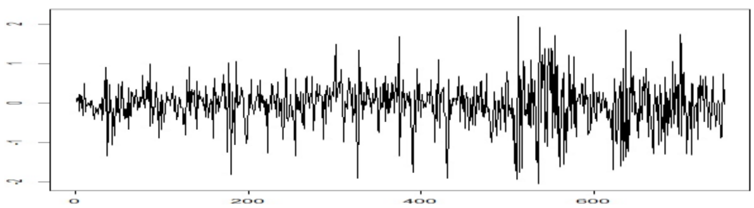

Figure 1. DEM/GBP FOREX log-returns

Table 5. Parameter estimates for the Student-t GARCH (1,1)

Coefficients	MLEs	Bayes SE
$\boldsymbol{\alpha_0}$	0.0359309	0.0436224
$\boldsymbol{\alpha_1}$	0.2668964	0.2675840
$\boldsymbol{\beta_1}$	0.6942793	0.6654374

For the parameter α_0 a vague prior is assumed. Dirichlet prior is assumed for α_1 and $\beta_1$where the hyperparameters of Dirichlet distribution are assumed equal and unknown and are estimated using the method of moments. So, $\omega_1 = \omega_2 = \omega_3$ =0.01094683.The log-returns is used to estimate the parameters which are given in Table 5 and then using the student-t GARCH(1,1) model the next 10 values will be predicted.

Table 6. Out-of-Sample Forecast Error Statistics

Method	ER	RMSE	MAE
MLE	0.1505108	0.38796	0.31099
BAYES	0.1444862	0.38011	0.30821

The predicted values that are compared to the real ones by computing the expected risks (ER), root mean square error (RMSE) and mean absolute error (MAE) are shown in Table 6.The Bayesian estimators' forecasting errors seem to be better than the MLE estimators, even with very small differences only for this period of data.

5. CONCLUSION

In this study, the parameters of GARCH model are assumed as random variables and are estimated using Bayes theory. The square error loss function is considered. The Bayes estimators are not in a closed form and thus Lindley's approximation is utilized. The error terms or financial asset returns are assumed to follow the Student-t distribution in estimating the parameters of GARCH model. Moreover, The gamma and vague priors are assumed for the constant coefficient of the GARCH model and the Dirichlet prior is assumed for the GARCH coefficients. The simulation studies are performed in order to compare the Bayesian estimates with the ML estimates. As expected, the Bayes estimates have lower expected risks than the MLE's.

Instead of using Lindley's approximation one could have used Tierney's and Kadane's approximation. Lindley's method requires the third order partial differentiation of log-likelihood function and one maximization whereas the Tierney and Kadane approximation requires the second order partial differentiation of the likelihood function and two maximizations. Singh et al. (2014) reported that Bayesian estimation of the parameters of the Marshall-Olkin extended exponential distribution using Lindley's approximation was better than using Tierney's and Kadane's approximation under informative setup. It is of interest to compare these two approximations for GARCH models.

Appendix A

The log likelihood for the Student-t GARCH (1,1) model reduces to

$$L = -\sum_{t=p+1}^{n} \left(\left(\frac{1}{2} - m_1\right) lnc_t + m_1 \ln(c_t + m_t) \right)$$

where $m_1 = \frac{\nu+1}{2}$, $m_t = \frac{a_t^2}{(\nu-2)}$ and $c_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

For the three parameter case when u is a function of only of one of the three parameters $\vartheta = (\alpha_0, \alpha_1, \beta_1)$ Lindley's approximation simplifies, when a SEL function is assumed, as

$$\alpha^{**}_{j-1\_SEL} = \hat{\alpha}_{j-1} + u_i\rho_i\varphi_{ii} + \frac{1}{2}\{Au_i\varphi_{1i} + Bu_i\varphi_{2i} + Cu_i\varphi_{3i}\} \qquad i = 1,2,3 \ j = 1,2$$

$$\beta^{**}_{1\_SEL} = \hat{\beta}_1 + u_i\rho_i\varphi_{ii} + \frac{1}{2}\{Au_i\varphi_{1i} + Bu_i\varphi_{2i} + Cu_i\varphi_{3i}\} \qquad i = 1,2,3$$

where $u_i = 1$ for $i = 1,2,3$ $\rho_1 = \frac{r-1}{\alpha_0} - \frac{1}{\theta}$, $\rho_2 = \frac{\omega_1-1}{\alpha_1} - \frac{\omega_3-1}{1-\alpha_1-\beta_1}$ and

$$\rho_3 = \frac{\omega_2 - 1}{\beta_1} - \frac{\omega_3 - 1}{1 - \alpha_1 - \beta_1}$$

$$A = \varphi_{11}L_{111} + 2\varphi_{12}L_{121} + 2\varphi_{13}L_{131} + 2\varphi_{23}L_{231} + \varphi_{22}L_{221} + \varphi_{33}L_{331}$$
$$B = \varphi_{11}L_{112} + 2\varphi_{12}L_{122} + 2\varphi_{13}L_{132} + 2\varphi_{23}L_{232} + \varphi_{22}L_{222} + \varphi_{33}L_{332}$$
$$C = \varphi_{11}L_{113} + 2\varphi_{12}L_{123} + 2\varphi_{13}L_{133} + 2\varphi_{23}L_{233} + \varphi_{22}L_{223} + \varphi_{33}L_{333}$$

The derived L_{ij} $i,j = 1,2,3$ and L_{ijk} $i,j = 1,2,3$ and the estimated variances and covariances of the MLE are

$$L_{11} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{1}{c_t^2} + \frac{m_1}{(c_t + m_t)^2}\right)L_{111} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{2}{c_t{}^3} + \frac{2m_1}{(c_t + m_t)^3}\right)$$

$$L_{22} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{a_{t-1}^4}{c_t^2} + \frac{m_1 a_{t-1}^4}{(c_t + m_t)^2}\right)L_{222} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{2a_{t-1}^6}{c_t{}^3} + \frac{2m_1 a_{t-1}^6}{(c_t + m_t)^3}\right)$$

$$L_{33} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{\sigma_{t-1}^4}{c_t^2} + \frac{m_1 \sigma_{t-1}^4}{(c_t + m_t)^2}\right)L_{333} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{2\sigma_{t-1}^6}{c_t{}^3} + \frac{2m_1 \sigma_{t-1}^6}{(c_t + m_t)^3}\right)$$

$$L_{12} = L_{21} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{a_{t-1}^2}{c_t^2} + \frac{c_1 a_{t-1}^2}{(c_t + m_t)^2}\right)$$

$$L_{13} = L_{31} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{\sigma_{t-1}^2}{c_t^2} + \frac{m_1 \sigma_{t-1}^2}{(c_t + m_t)^2}\right)$$

$$L_{23} = L_{32} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{a_{t-1}^2\sigma_{t-1}^2}{c_t^2} + \frac{m_1 a_{t-1}^2\sigma_{t-1}^2}{(c_t + m_t)^2}\right)$$

$$L_{122} = L_{221} = L_{212} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{2a_{t-1}^4}{c_t^3} + \frac{2m_1 a_{t-1}^4}{(c_t + m_t)^3}\right)$$

$$L_{112} = L_{121} = L_{211} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{2a_{t-1}^2}{c_t^3} + \frac{2m_1 a_{t-1}^2}{(c_t + m_t)^3}\right)$$

$$L_{113} = L_{131} = L_{311} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{2\sigma_{t-1}^2}{c_t^3} + \frac{2m_1 \sigma_{t-1}^2}{(c_t + m_t)^3}\right)$$

$$L_{322} = L_{223} = L_{232} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{2a_{t-1}^4\sigma_{t-1}^2}{c_t^3} + \frac{2m_1 a_{t-1}^4\sigma_{t-1}^2}{(c_t + m_t)^3}\right)$$

$$L_{331} = L_{133} = L_{313} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{2\sigma_{t-1}^4}{c_t^3} + \frac{2m_1 \sigma_{t-1}^4}{(c_t + m_t)^3}\right)$$

$$L_{332} = L_{233} = L_{323} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{2a_{t-1}^2\sigma_{t-1}^4}{c_t^3} + \frac{2m_1 a_{t-1}^2\sigma_{t-1}^4}{(c_t + m_t)^3}\right)$$

$$L_{123} = L_{321} = L_{231} = L_{213} = L_{312} = L_{132} = -\sum_{t=p+1}^{n}(\frac{1}{2} - m_1)\left(\frac{2a_{t-1}^2\sigma_{t-1}^2}{c_t^3} + \frac{2m_1 a_{t-1}^2\sigma_{t-1}^2}{(c_t + m_t)^3}\right)$$

REFERENCES

[1] **Ardia D. and Hoogerheide L.F. (2010),*Bayesian Estimation of the GARCH (1,1) Model with Student-t Innovations in R;*** *The R Journal*, 2(2), pp. 41-47;
[2] **Asai M.(2006),*Comparison of MCMC Methods for Estimating GARCH Models;*** *Journal of the Japan Statistical Society,* 36: 199–212;
[3] **Bauwens L. and Lubrano, M.(1998*), Bayesian Inference on GARCH Models using the Gibbs Sampler;*** *Econometrics Journal* 1: 23–46;
[4] **Bollerslev T. (1986), *Generalized Autoregressive Conditional Heteroskedasticity;*** *Journal of Econometrics,* 31(3), pp. 307–327;
[5] **Bollerslev T. (1987),*A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return*;** *Review of Economics and Statistics*, 69, 542-547;
[6] **Gilks, W.R.,Best N.G. and Tan, K.K.C.(1995), *Adaptive Rejection Metropolis Sampling within Gibbs Sampling***; *Journal of the Royal Statistical Society Series*, C 44: 455–472;

[7] **Kim, S., Shephard, N. and Chib, S.(1998),*Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models*;** *Review of Economic Studies,* 65: 361–393;
[8] **Lee, S.W. and Hansen, B. E.(1994),*Asymptotic Theory for the GARCH(1, 1) Quasi-Maximum Likelihood Estimator*;***Econometric Theory,* 10, pp. 29–52;
[9] **Lindley, D.V. (1980), *Approximate Bayes Methods;***Trabajos de Estadistica,* 3 pp. 281–288;
[10] **Marín, J.M., Rodríguez-Bernal, M.T. and Romero, E.(2015),*Data Cloning Estimation of GARCH and COGARCH Models;***Journal of Statistical Computation and Simulation*, 85(9), pp. 1818-1831;
[11] **Metropolis, N., Rosenbluthv, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953),*Equation of State Calculations by Fast Computing Machines*;** *J. Chem. Phys*., 21, pp. 1087-1092;
[12] **Mitsui, H. and Watanabe, T.(2003),*Bayesian Analysis of GARCH Option Pricing Models***;*The Journal of the Japan Statistical Society (Japanese Issue),* 33, 307–324 (in Japanese);
[13] **Moore, D. and Papadopoulos, A.S.(2000),*The Burr Type XII Distribution as a Failure Model Under Various Loss Functions;***Microelectronics Reliability,* 40, pp. 2117-2122;
[14] **Müller, P. and Pole, A. (1998), *Monte Carlo Posterior Integration in GARCH Models*;***Sankhya, Series,* B 60: 127–14;
[15] **Nadar, M. & Kızılaslan, F. (2015),*Estimation and Prediction of the Burr Type XII Distribution Based on Record Values and Inter-Record Time;***.Journal of Statistical Computation and Simulation,*85:16, 3297-3321;
[16] **Nakatsuma, T. (2000),*Bayesian Analysis of ARMA-GARCH Models: A Markov Chain Sampling Approach;***Journal of Econometrics,* Vol. 95, pp. 57-69;
[17] **Tsay R. S. (2012), Ana*lysis of Financial Time Series*;** *Wiley and Sons*;
[18] **Singh, S. K., Singh, U. and Yadav, A. S. (2014),*Bayesian Estimation of Marshall-Olkin Extended Exponential Parameters Under Various Approximation Techniques;*** *Hacettepe Journal of Mathematics and Statistics,* Vol. 43 (2) 341-354;
[19] **Virbickaite, A., Ausin, M. C. and Galeano, P. (2015),*Bayesian Inference Methods or Univariate and Multivariate GARCH Models: A Survey*;***Journal of Economic Surveys*, Vol. 29, No. 1, pp. 76–9.