COMPARISON BETWEEN THE BAYESIAN AND FREQUENTIST ESTIMATORS: UNIVARIATE GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY (GARCH) MODEL

SUBMITTED TO

DR. ASAD ZAMAN (SUPERVISOR)

DR. SAUD AHMAD KHAN (CO-SUPERVISOR)

SUBMITTED BY

FARRUKH MAHMOOD

06/PHD-ETS/2014/PIDE

DEPARTMENT OF ECONOMETRICS AND STATISTICS

PAKISTAN INSTITUTE OF DEVELOPMENT ECONOMICS

ISLAMABAD

Author's Declaration

I Mr. Farrukh Mahmood hereby state that my PhD thesis titled **"Comparison between the Bayesian and Frequentist Estimators: Univariate Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model"** is my own work and has not been submitted previously by me for taking any degree from **Pakistan Institute of Development Economics, Islamabad'** or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my Graduation the university has the right to withdraw my PhD degree.

Farrukh

(Mr. Farrukh Mahmood)

06/PhD-ETS/PIDE/2014

Plagiarism Undertaking

I solemnly declare that research work presented in the thesis titled **"Comparison between the Bayesian and Frequentist Estimators: Univariate Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model"** is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the **HEC** and **Pakistan Institute of Development Economics, Islamabad** towards plagiarism. Therefore I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of PhD degree, the University reserves the rights to withdraw/revoke my PhD degree and that HEC and the University has the right to publish my name on the HEC/University Website on which names of students are placed who submitted plagiarized thesis.

Students/Author Signature: ________________

Mr. Farrukh Mahmmod

06/PhD-ETS/PIDE/2014

Pakistan Institute of Development Economics

Certificate of Approval

This is to certify that the research work presented in this thesis, entitled: **"Comparison between the Bayesian and Frequentist Estimators: Univariate Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model"**was conducted by Mr. Farrukh Mahmood under the supervision of Dr. Asad Zaman and Dr. Saud Ahmed Khan. No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Econometrics from**Pakistan Institute of Development Economics, Islamabad.**

Student Name:**(Farrukh Mahmood)**
06/PhD-ETS/PIDE/2014

Examination Committee

Supervisor:

Dr. Asad Zaman
Ex-Vice Chancellor
PIDE, Islamabad.

Co-supervisor:

Dr. Saud Ahmed Khan
Assistant Professor
PIDE, Islamabad.

External Examiner:

20.6.19
Dr. Zahid Asghar
School of Economics
Quaid-i-Azam University, Islamabad

Internal Examiner:

Dr. Abdul Rashid
International Institute of Islamic Economics
International Islamic University,Islamabad

Head,
Department of Econometrics and Statistics:

Dr. Amena Urooj

ABSTRACT

The estimates of the Maximum Likelihood estimation method are the estimates of the global maximum likelihood function, by definition. However, the present study showed empirically that the likelihood function of the GARCH model is multimodal. Due to the presence of multimodality in the likelihood function leads to a difference in estimates at local and global maxima, and hence, Maximum Likelihood estimation methods can have unstable performance in such situations. Therefore, it will face the problem in inference and prediction, due to the difference in estimates at local and global maxima(s). Two estimation methods are chosen from the Frequentist and the Bayesian approach, respectively, to measure the significance of the difference in estimated parameters due to the presence of multimodality in the likelihood function. Besides, to calculate the level of difference, a standard method of Monte Carlo simulation method is used. The surface plot is constructed by changing the value of the Monte Carlo simulation method to evaluate their performance along the whole surface. these surfaces are then compared within each approach.

Subsequently, the preferable algorithms are compared across the Bayesian and Frequentist approaches. For comparison, the present study has calculated bias and variance around the true data generating process. Empirically it is found that in case of Frequentist approach Differential Evolution (DE) algorithm is preferable estimation method for GARCH type models, as compared to Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Because there is multimodality in the likelihood function of the GARCH model, and BFGS uses a single starting value to search maximum point in the likelihood function, and often this single starting value traps into local maxima. Therefore, the estimated

parameter at the local and global maxima vary, and hence, inferences and predictions. Conversely, DE uses multiple starting values with multiple chains, due to which it automatically avoid local maxima and converge**s** to global maxima.

In the case of the Bayesian approach, Robust Adaptive Metropolis (RAM) is a preferable estimation for GARCH type models as compared to Metropolis Hasting (MH). Because RAM is based on the strategy of adaptive mechanism, i.e., the Markov Chain of the RAM move to the next point, after taking information from the previous point, and finally converge to some particular value of the estimate. While MH use chain of independent nature, i.e., it does not take information while moving from one point to another point in the Markov Chain. After confirming the best estimator from frequentist and the Bayesian approach, this study compared these approaches with each other. Empirically, it is found that the Bayesian approach (RAM) is the preferable estimation method than the Frequentist approach (DE) because the level of bias and variance around the true parameter for RAM is lower than DE.

Pakistan Stock Exchange (PSX) is used as a real-world application. Empirically it is found that the Bayesian approach is preferable estimation method than the frequentist approach. Reasons are followed; first, in the frequentist approach estimated parameters are the point estimates, while in the case of the Bayesian approach, the complete distribution of the estimated parameter is obtained at the low cost of simulation. Second, the distribution of the point estimate is hypothetically assumed to be normal, while in case of Bayesian approach it is not valid, i.e., the distribution of the estimates could be skewed in either direction. Therefore, the frequentist approach either over or underestimate the true value of the parameter. Finally, the standard error of the estimates which are obtained through the DE algorithm is more precise as compared to the estimates of BFGS. Therefore, the forecasting based on DE is more accurate about risk and return.

Keywords: Multimodality, Likelihood Function, Estimation Methods, Bayesian Approach, Frequentist Approach, Monte Carlo Simulation, Surface Plot, Single Starting Value, Multiple Starting Value, Pakistan Stock Exchange.

DEDICATION

Grandmother Snobar Bibi, Grandfather; Mughfoor Elahi, Parents, Sibling

and

Shumaila Hashim, my wife.

Whose prayers for me, were what sustained me thus far.

ACKNOWLEDGEMENTS

I will never forget the day when Dr. Attiya Yasmin Javid enforced me to take admission in Ph.D. Econometrics. After a few days, a list of the twenty-selected candidate was posted by PIDE, for the Ph.D. Econometrics. Few students left in their starting days, few dropped-out through course-work, three students appeared in comprehensive exams, and I am the only one who got success as a pioneer student of Econometrics and Statistics in PIDE, and other students were also dropped-out. Allah blessed me that I have completed my Ph.D. Thesis. I am especially thankful to Dr. Attiya Yasmin Javid, Dr. Eatzaz Ahmed, Dr. Waseem Shahid, for their cooperation, teaching, and guidance. I am also thankful to Dr. Amna Urooj – Head of Department for their kindness, moral support and guidance.

Alhamdulliah, today, I am submitting my thesis, which could not be completed without the focus, attention, and keen observation of my research supervisor Dr. Asad Zaman. It is with pleasure and gratitude that I acknowledge his efforts. Besides Econometrics, I learned from him the true way of peaceful life, value, and respect to humanity, how to be a true Muslim and well-wisher of human beings. From his teachings now, I know what success is and how to achieve it. I want to thank him. He has been very patient, supportive, and providing high-quality supervision since the beginning of my thesis. He has been guiding me to develop my independent research ability and very detailed and insightful comments on my work. I forfeit my enormous gratitude to my co-supervisor Dr. Saud Ahmed Khan, for his generous support and precious time.

I am also thankful to Dr. Abdul Qayyum, Dr. Abdul Rasheed – Internal Examiners, and Dr. Sumru Altug, Dr. Giovanni Ricco – International Reviewers for reviewing this research document at different stages. They suggested valuable comments on this document, which have increased the worth of this document.

In the compilation of this document, the hardest part is computational analysis. In the earliest days, when I started simulation analysis; one-time iteration takes about 5 minutes. According to this calculated time, 1000 time iterations take about nine days. To complete this study, it takes about two years. It seems impossible to complete the analysis. Latterly, I learned R language with myself then profoundly with the help of Dr. Ijaz from Quaid e Azam University. I am also thankful to Dr. Zafar Mahmood for allowing me to use a computer lab at S3H department at NUST.

Furthermore, Time length issues did not solve; then I moved Peshawar to resolve these issues. I meet Dr. Roshan Ali from Khyber Medical University, Peshawar. Due to his teaching, the time length has shortened to about 20 hours on the CORE i3 system or 11 hours on CORE i7, for 1000 time iterations. At this stage, to complete analysis, about 1.5 months will be required.

Furthermore, to squeeze time length more for complete simulation analysis, one method is to run analysis on multiple computers at the same time. For this purpose, My supervisor has permitted me to use the computer lab of PIDE. The ICT department of PIDE facilitated me in this regards, and the ICT allowed me to engage the whole lab for 3 - 4 days with 20 computers. Fortunately, I did my entire work of simulation for approximately about two weeks. I am especially thanks to the team of ICT lab, for their kind cooperation. The period of learning this shortcut method is about 1.5 years.

I want to extend my sheer gratitude to my friends Saud Ahmed, Adeel Sheikh, Muhammad Zubair Chasti, Raja Mohsin, Mohsin Kayani, Mohsin Ali, Rubina Ilyas and

all others colleague from PIDE, who have directly or indirectly supported me to concentrate on my thesis and stay me on track. I also need to thank the faculty, staff, and students at PIDE. Indeed, an excellent academic environment encourages me to develop my research ideas and motivate me to pursue to the next level of research.

Lastly, I am thankful to the Higher Education Commission (HEC) of Pakistan, for their monetary support through Indigenous Scholarship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Volatility plays a crucial role in empirical finance and financial risk management and is the basis of any model for financial time series, such as; stock prices, exchange rate, interest rate. Since the seminal work of Engle (1982) on Auto-Regressive Conditional Heteroscedasticity (ARCH) model, extensive research has been conducted on changing volatility (i.e., conditional variance) using time series models. Bollerslev (1986) proposed a Generalized ARCH (GARCH) model, which extended the specification of the conditional variance. Since 1982, ARCH-type models overgrew into a family of empirical financial models for forecasting volatility. The primary purpose of ARCH-type modeling is to understand systematic changes in the volatility of financial time series to minimize the risk in the future. These GARCH-type models are now general and vital tools in financial econometrics.

Maximum Likelihood (ML) estimation method is usually used to estimate the GARCH-type models. The standard theory showed that these estimates follow ideal asymptotic properties (Bollerslev et al., 1994). Conversely, these asymptotic properties are slightly different when there are boundary constraints on the estimates (Zaman, 2002), as covariance stationarity condition and positive conditional variance conditions are applied to estimate GARCH models. Besides, the estimates of the ML estimation method are the estimates of the global maximum of the likelihood function, by definition. However, it has been shown theoretically by Jerrell and Campione (2001), and the present study also confirms it empirically that, the likelihood function may fail to have a global maximum; there are singularities at which the likelihood peaks to an infinite value. After excluding neighborhoods of singularities, the highest local maximum has the optimality properties of the ML, but numerical estimation methods may fail to find this and instead converge to some lesser local maximum.

ML estimation method is easy to understand and implement and is readily available in statistical software. Most commonly used estimation method is Broyden-Fletcher-Goldfarb-Shanno (BFGS). BFGS is based on a "hill-climbing" strategy, i.e., it starts from an initial value and searches for the maxima in the likelihood function. Besides BFGS, Stron and Price (1997) developed a Differential Evolution (DE) estimation method, and was implemented by Mullen et al., (2011). DE uses multiple chains with multiple starting values to search global maxima in the likelihood function. Due to multiple chains and multiple starting values, it avoids local maxima automatically.

Moreover, the estimation of GARCH type models is achieved after imposing two types of constraints, i.e., positive conditional variance and covariance stationary condition. If an inequality constraint, then the procedure of optimization can be cumbersome. Additionally, if the value of the actual parameter is close to boundary space, then the numerical method may fail to converge (Ardia, 2008). These are theoretical problems of ML because the asymptotic theory for constrained estimators is different from the standard ML asymptotic theory (Zaman, 2002).

Contrary to the frequentist approach, there is a Bayesian approach and all the difficulties which are found in the frequentist approach – as discussed above, disappear when the Bayesian approach is used to estimate GARCH type models (Ardia, 2008). First, the Bayesian method estimates using averages over the posterior distribution and is not directly affected by the presence of multiple local maxima. Second, any constraint on the model parameter can be incorporated in the modeling through the appropriate prior specification. Third, the exact distribution of the parameter is obtained at the low cost of simulation, which resolves the problem created by multiple maxima and constraints, and permits direct inference. Therefore, the present study also uses a Bayesian approach for the estimation of GARCH type models.

The main objective of this study is to select different estimation techniques from the literature, which are commonly used for the estimation of GARCH type models and assess their ability to estimate volatility accurately. It would allow ascertaining whether these estimation techniques have practical value in the context of volatility models. The present study uses simulation to evaluate the power, and afterward, algorithms will be compared within each approach. Second, to increase the reliability of the first objective present study repeat the experiment of simulation with multiple data generating process. Repeating the experiment with multiple data generating process will give a clue about the estimation methods that either the result of each algorithm is consistent across different data generating process or not. Because, in the real world application, the true data generating process is unknown; therefore, in a real-world application, econometrician is interested in approximation.

Thirdly, to increase the enforceability of comparison, present study repeat**s** experiment for a large number of combination, to construct the surface plot analysis, to analyze the difference along the whole surface. Fourth, after comparing algorithm within each approach, the selected algorithm will be compared across approaches. Fifth, the Pakistan Stock Exchange (PSX) will be used as a real-world application and apply GARCH (1, 1) model, to analyze the significance of the difference in inference and prediction, in the presence of multimodality. In the frequentist approach, estimated parameters are the point estimates, and the distribution of these are assumed to be normal, hypothetically. To analyze this issue through posterior distribution is the sixth objective of the present study.

After achieving six objectives of this study, mentioned above, will prove the hypothesis, i.e., the choice of the algorithm in the presence of multimodality for the estimation of GARCH type models, both in Bayesian and frequentist approach. The present study will now move towards the estimation of asymmetric GARCH type model for the window of PSX. This step will be done after confirming the presence of Leverage effect; this is the seventh objective of the study. Consequently, the good fitted asymmetric empirical model will be extended in the presence of skewed student t distribution. It is the eighth objectives of the study. As already discussed that the main objective of the GARCH type model is forecasting; therefore, this study will also discuss forecasting based on Value at Risk (VaR), Expected Short Fall and Predictive density measure. It is the ninth and final objective of the present study.

Most of the applications of GARCH type models are found in financial economics, where researchers are focusing on two types of issues; best specification of error and the choice of most efficient approach for inference (Virbickaite & Ausin, 2015). The present study focuses on both issues, i.e., the performance of different estimation techniques are evaluated using simulation with different error specification. By doing so, it will give a clue about the choice of estimation technique, which is consistent across the various distribution. Also, it removes the ambiguity among the selection of estimation approach as well, i.e., Bayesian and Frequentist. Consequently, it will lessen the risk due to the choice of inappropriate estimation technique for the inference and prediction.

1.1. Illustration of Multimodality:

Two illustrative examples of multimodality in the likelihood function are taken from literature. The choice of these examples is made on the bases of the empirical studies of the GARCH $(1, 1)$ model. The first example is based on the exchange rate, which is used in the seminal paper of Bollerslev (1987). The second example is also based on the exchange rate and is taken from Ardia (2008). While the third illustration is from the daily Pakistan Stock Exchange (PSX).

1.1.1 Illustration from Bollerslev (1987)

The series of US dollar versus the British pound exchange rate is chosen from the study of Bollerslev (1987) to illustrate the multimodality in the likelihood function. The frequency of the exchange rate is daily, from March 1, 1980, to January 28, 1985. The likelihood function is constructed by using the same window of the exchange rate and is presented in Fig. 1. Contour plots are used to analyze this issue, and it represents in different circles or layers. These circles are centered in a different region of the plot, and each of these centered represented unique modality. For example, the value of ML 11640 represents three different circles, which implies that there are three peaks in ML function. Moreover, 11460 and 11400 are other peaks in the same ML function. Hence, there is multimodality in the ML function of exchange rate US dollar vs. British pound.

1.1.2 Illustration from Ardia (2008)

The daily data of Deutschmark vs. British Pound exchange rate is taken from Ardia (2008). The lower plot of Fig. 2 represents multi-modality for this data set. It is observed that the maximum peak with the value of 210, with this same value, apparently there are four peaks. On the other hand, there are also other local peaks with a value of 200, 180, and 190. Hence, similar to the previous example, this series also highlighted multimodality in the ML function of GARCH (1, 1) model.

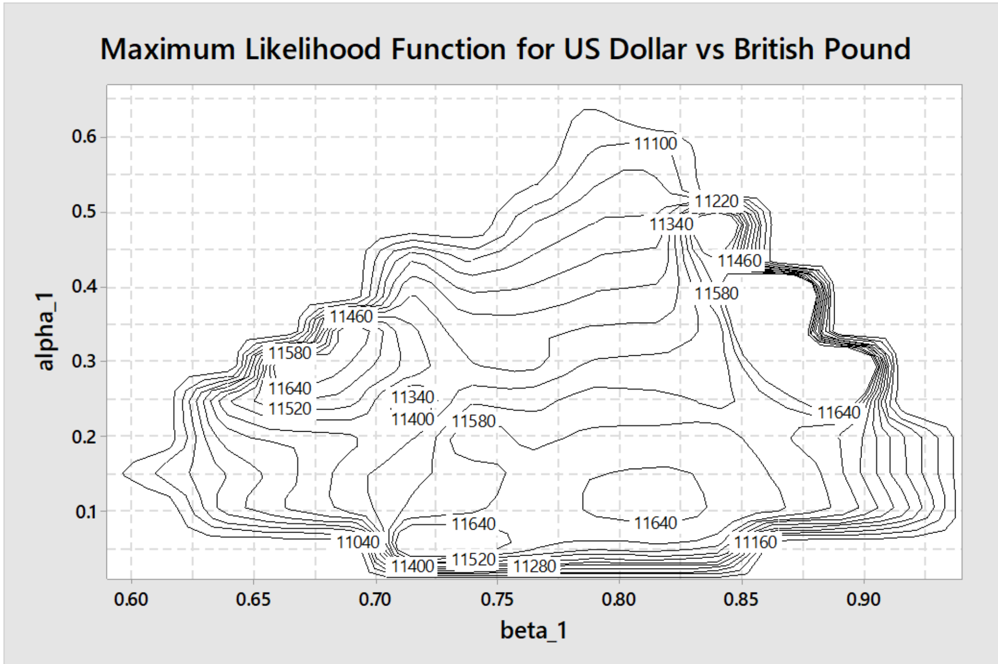


Figure 1. represents the ML Function of GARCH (1,1) for the exchange rate of the US Dollar vs British Pound. The x-axis represents different values for β, and the y-axis represents different values for α_1. Layer in graphs represent the level of ML, and different circles represent peaks in ML, implies multimodality.

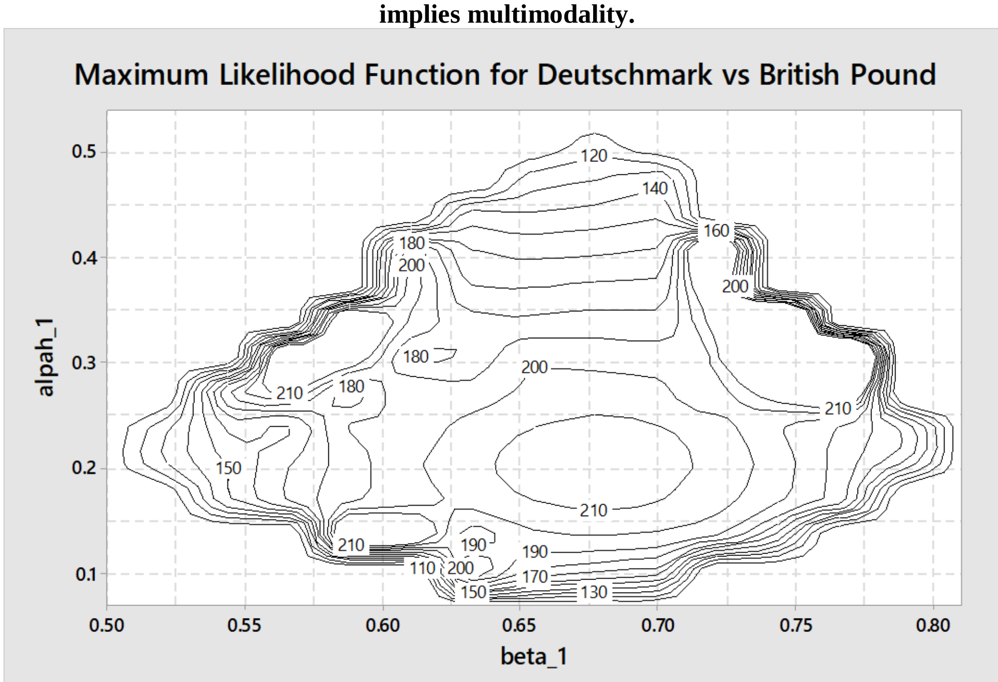


Figure 2. represents the ML Function of GARCH (1,1) for the exchange rate of the Deutschmark vs British Pound. The x-axis represents different values for β, and the y-axis represents different values for α_1. Layer in graphs represent the level of ML, and different circles represent peaks in ML, implies multimodality.

1.1.3 Illustration of Pakistan Stock Exchange

The daily data of the Pakistan Stock Exchange index is used as the third illustration to present the multimodality in the likelihood function. The sample period is from March 20, 2002, to January 11, 2005, with 696 number of observations, excluding weekdays and holidays. The data span covers about three financial years and is long enough to apply ML Estimation Method (Arida, 2008). Also, this sample period satisfies the condition of asymptotic normality condition. Hence, the data is large enough to represent the ML function. The nominal returns are expressed in percent as in Bollerslev and Ghysels (1996). The data set is readily available on the open source of yahoo finance. Fig. 3 represents the ML function of PSX. Similar to the previous example, this graph also represents multimodality in the ML function of GARCH (1, 1).

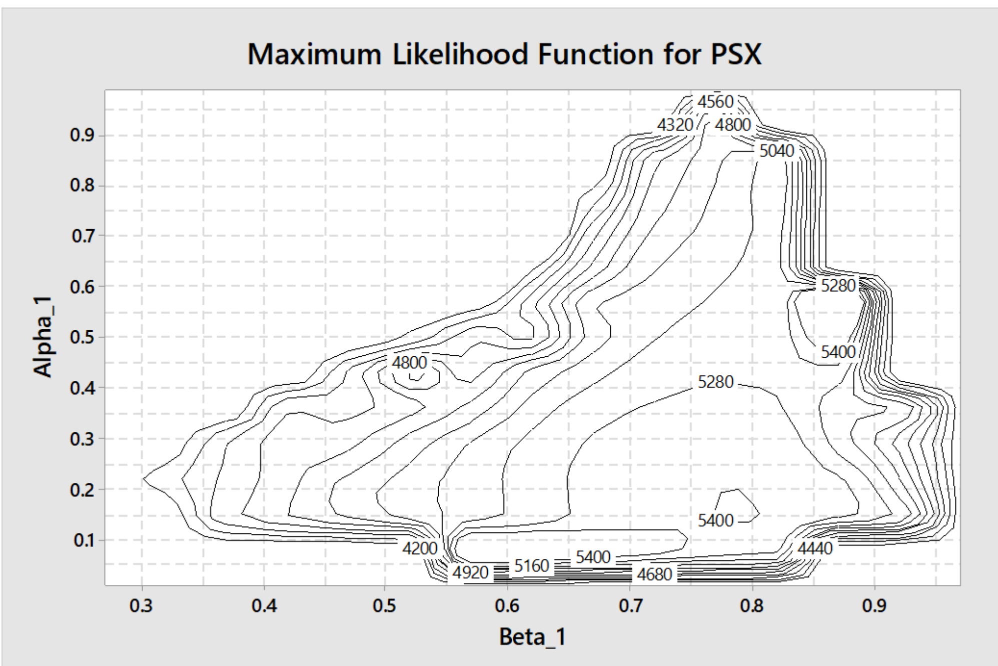


Figure 3. represents the ML Function of GARCH (1,1) for PSX. The x-axis represents different values for β, and the y-axis represents different values for α_1. Layer in graphs represent the level of ML, and different circles represent peaks in ML, implies multimodality.

From these illustrations, it can be inferred that there is multimodality in the likelihood function of the GARCH model. Hence inference of each empirical model at each mode in these contour surface plots is different, consequently predictions. Most commonly, the assumption of covariance stationarity is violated in all cases. Furthermore, there are two types of models for each of the data set presented above depending on the modes, i.e., standard GARCH with $(alpha + beta < 1)$ and Integrated GARCH with $(alpha + beta > 1)$.

This study comprises of six chapters. Chapter 1 introduces the purpose of this study. Chapter 2 is the literature review of GARCH type models and then followed by brief detail about the choice of distribution in the GARCH type models, and contains literature on a different type of algorithms in Bayesian and Frequentist approaches. Chapter 3 discusses the basic model for the simulation-based empirics to analyze and compare the performance of different algorithms and samplers. Chapter 4 is divided into two major sections. The first section of Chapter 4 will discuss the empirical results of GARCH (1, 1) model with a normal distribution, while the second section of Chapter 4 will discuss the empirical results of GARCH (1, 1) for PSX with the student-t distribution.

Similarly, Chapter 5 is divided into two major sections. The first section of Chapter 5 will discuss the choice of most appropriate asymmetric model for the same window of PSX, and the second section the same chapter will extend the selected asymmetric model in the presence of skewed distribution, subsequently, value at risk, expected shortfall, and predictive density. Finally, Chapter 6 includes the conclusion for the study and recommendations for future research based on this dissertation.

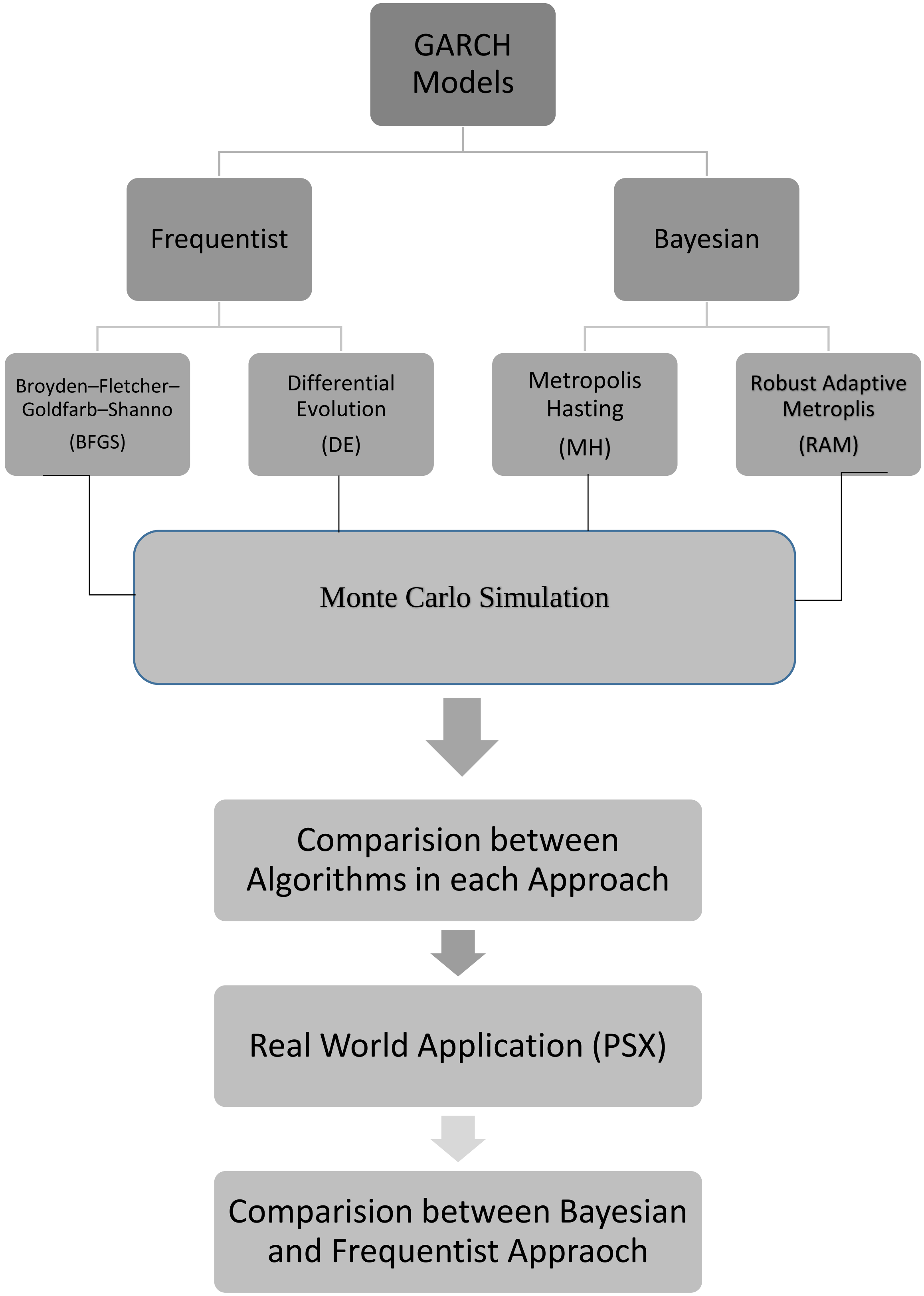


Figure 1.4: Flowchart of Empirical Study

CHAPTER 2

LITERATURE REVIEW

Engle (1982) introduced the ARCH model to estimate the volatility of financial time series, which was extended by Bollerslev (1986), as a generalized ARCH model (GARCH). These models assume a normal distribution for the error, but Bollerslev (1987) observed that the financial time series have more massive tail compared to the normal distribution; therefore, he suggested t-distribution for the error in GARCH models. These models were updated into different dimensions, based on error specifications and choice of the most efficient approach for inference (Virbickaite et al., 2015).

Usually, Maximum likelihood estimation method is used to estimate the GARCH type models, although some authors (Gallant et al., 1989, Pagan et al., 1990) have also applied semi and non-parametric techniques. The primary appeal of the Maximum likelihood technique stems from the well-known asymptotic optimality conditions of the resulting estimators under ideal conditions (Bollerslev et al., 1994). However, asymptotic theory for constrained parameters; as in case GARCH models, is different from the standard Maximum likelihood asymptotic theory (Zaman, 2002). The standard theory of Maximum likelihood is incorrectly applied to GARCH models since the GARCH model has multi-modality and boundary constraint. Doornik and Ooms (2000, 2008) first noted multimodality. However, according to them, multi-modality arises when the dummy variable is added in the mean equation to correct the effect of an additive outlier.

On the other hand, Jerrell and Campione (2001) applied two different algorithms to estimate the GARCH type model and found different estimated results. The difference in the estimated parameter directly implies that multiple maxima exist in the likelihood function. Nevertheless, the authors did not explore this further.

Moreover, Maximum Likelihood estimation method is simple to understand and easy to implement, and readily available in the econometric or statistical software. Most of this software, like Ox-Metrix use BFGS estimation method to estimate GARCH type models. It is a numerical method of maximization. Numerical methods are often susceptible to the starting value of algorithms, since they converge to the nearest local maximum, instead of the global maximum. BFGS is a hill-climbing strategy where it starts searching for a maximum point in the likelihood function, from the initial value. In the case of BFGS, this initial value is specified by using the method of Fiorentini et al. (1996).

In the GARCH models, there are multiple maxima, and therefore, the numerical maximization algorithms such as BFGS are sensitive to the starting value. Due to this sensitivity, the system might not converge to the global maximum. Moreover, these numerical algorithms use the line search criterion to make the decision where a new value of the parameter is accepted if it maximizes the value of likelihood function. Although with the line search criterion, the system converges very fast, yet there is a risk of getting trapped by local maxima. On the other hand, there is a parallel search method similar to DE strategies. In a parallel search, there is a built-in safeguard to prevent misconvergence into local maxima.

Furthermore, these parallel search techniques satisfy all four requirements, (i) Ability to handle non-differentiable, nonlinear and multimodal cost functions (ii) Computational efficiency (iii) Ease of use (iv) Good convergence. One such parallel search technique is DE – (Stron & Price, 1997). Price et al., (2006) developed a practical procedure to find a good starting point, which was implemented by Ardia et al., (2015). The problem of misconvergence can easily be overcome by combining starting value procedure from Price et al., (2006) with DE from Stron and Price (1997).

For the Bayesian approach, several Monte Carlo Markov chain methods are available to form the posterior distribution. These Markov chains are used to compute the information about different parameters for a particular posterior distribution. Tierney (1994) mentioned several well-known Markov chain to compute posterior density, but few of them can be used to estimate GARCH type models. More specifically, Tierney (1994) suggested that the Markov chains in hybrid strategies are more appropriate for the estimation of GARCH type models. Ardia and Hoogerheide (2010) presented Markov chains, which are relevant to the GARCH-type models, i.e. Griddy Gibbs Sampler and Metropolis Hasting (MH).

The simple Gibbs sampler is used only when there is full conditional density, i.e. conjugacy property (Ardia, 2008). Some complicated Bayesian problems cannot be solved by using the Gibbs sampler. For example, when it is not easy to break down the joint density into full conditional density or when the full conditional densities are in the unknown form (Ardia, 2008). Ritter and Tanner (1992) overcame this problem by suggesting the Griddy Gibbs sampler for the non-conjugacy situations. The basic idea is to approximate the true distribution by a piecewise linear function. This sampler is easy to understand and implement, but it is time-consuming. This procedure was extended by using student t-distribution by Bauwens and Lubrano (1998), and practically, Ausin and Galeano (2007) used it. However, Arida (2008) suggested that the Gibbs sampler is not an appropriate choice of estimation, due to issues explained above, while MH is the most preferable choice of estimation for GARCH type models.

Metropolis algorithm was introduced by Metropolis *et al*., (1953) and generalized by Hastings (1970). In the Markov chain, candidate draws are generated from conditional density. The candidate is then accepted or rejected based on acceptance probability. If the candidate's density is accepted, then the chain will move to the next step. Otherwise, it will stay on the current stage. Two varieties of Metropolis-Hastings chain are most common. One is *independence chain*, in which the candidate draws are generated from the unconditional candidate distribution. The other is a *random walk chain;* in which current draw is a condition on the current value of the chain. In both methods, the chain distribution must be tuned to achieve a reasonable acceptance rate and to explore sufficiently the domain of the posterior distribution. The values of the chains are correlated,[1] so, it takes some time to control the effect of the arbitrarily chosen initial value. To produce the *iid* sample from the posterior distribution, samples are taken at a long distance from each other within the chain. In random walk and independence chains, preliminary runs and tuning are necessary. Therefore, the method is not entirely automatic, which is not desirable.

Ardia (2008) used the Metropolis-Hasting sampler independent chains.[2] Independence means that the current proposal draw is independent of the previous draws of the chain (Sarkka, 2013). The chain approaches its equilibrium distribution as the

[1]Bauwens and Lubrano (1998), explains that if the parameter have correlation in the posterior distribution then the convergence is slower. As the level of autocorrelation increases the level of convergence become slower. In addition when correlation increases in the posterior distribution, the draws then to cluster, so that it takes a lot of draws to explore the posterior, and convergence is slower. This phenomenon can also be explained by the correlogram of the Markov chain.

[2]Metropolis-hasting samplers also follow the procedure of random walk chains – for detail see Tierney 1994.

number of iterations increases (Tierney, 1994). On the other hand, Vihola (2012) suggested that the dependence of the chain converges to the excellent choice for approaching the equilibrium distribution. Vihola (2012) developed a new sampler, i.e. RAM. It was practically used by Ardia et al., (2016) for the estimation of the GARCH type models. This latest algorithm has a speciality that it estimates the shape of the target distribution and simultaneously forces the acceptance rate. Furthermore, adaptive algorithms require burn-in of some iteration to obtain independent chain; this algorithm does not require any burn-in iteration. Lastly, as it is shown that the likelihood function of the GARCH type model is multimodal, and most of the algorithm fails to converge global maxima, while this latest algorithm also works well in this situation.

CHAPTER 3

THE MODEL AND ESTIMATION APPROACHES

Volatility is a statistical measure of the dispersion from the mean. It can be measured by using variance (or standard deviation); the higher value of variance means more uncertainty about the next event and vice versa. The conventional time series econometrics assumes that the variance of the disturbance term is constant. However, time series data exhibit unusually large volatility followed by periods of relative tranquillity. It is one of the stylized properties.[3] In such circumstances, the assumption of constant variance is violated. Therefore, it is needed to specify this non-constant variance (conditional) along with the mean (conditional) equation. These type of models are also called volatility models and are extensively studied in financial econometrics. The most well-known volatility models are the GARCH type models. The purpose of these models is to analyze the historical pattern of volatility and predict future volatility by incorporating stylized properties.

Extensive literature is available on the GARCH type models, which either discuss different extensions of the GARCH models or focus on the specification of the error distribution. There is no previous study, which gives attention to multimodality in the likelihood function of the GARCH type models. However, theoretically, it is discussed by Zaman (2002) in the context of random coefficient models. The present thesis shows empirically that the likelihood function of the GARCH model is multimodal. When the

[3] Other are; non-stationarity of the price series, absence of autocorrelation for the prices variations, autocorrelation of the squared prices returns, fat-tail distribution, leverage effect and seasonality, for detail see, Francq and Zakoian (2010), pp. 7-10.

likelihood function is multimodal, then the starting value should be taken with much consideration. Otherwise, inconsistent starting value converges to local maxima, and hence, the estimated parameter will not maximize the value of the likelihood function. In the previously available literature, different algorithms have been used to estimate the likelihood function of the GARCH type models. The most well-known and commonly used algorithm is BFGS, that is a numerical method, and the most recent algorithm used is DE. These algorithms are discussed in detail in the following sections. This study compared the performance of these algorithms for the estimation of the GARCH type models while keeping in mind the concept of multimodality in the likelihood function.

The Maximum Likelihood estimation method is generally preferred to estimate the GARCH type models. It is simple to understand and implement. However, there are some practical difficulties. First, the estimation of the GARCH type models require some constraints on parameters, i.e., positive conditional variance, and covariance stationary condition. These constraints are of inequality, which makes optimization difficult. Also, the optimization procedure might not converge if the value of the true parameter is close to the boundary of parameter space or if the process is a non-stationary process. Second, numerical algorithms are used to estimate the likelihood function of the GARCH type models. These algorithms are acceptable if the likelihood function of the GARCH type model is unimodal. However, empirically, it has been shown that the likelihood function of the GARCH type model is multimodal. In this case, the estimates of the Maximum

Likelihood method are sensitive to the starting values.[4] These difficulties disappear when a Bayesian approach is used.

The plan of the study is as follow: Section 1 explains the model. Section 2 explains different approaches for estimation of GARCH type models. Section 3 explains the different steps for the comparison of different algorithms. The last section of this chapter explains the comparison with the help of visualization.

3.1. The Model

The standard approach of modeling volatility is to introduce an exogenous variablex_t, which helps to predict the variance of y_t which is the dependent variable. The expected mean of x_t is zero, and the model can be written as,

$$y_t = \varepsilon_t x_{t-1} \qquad 1$$

where, y_t is a variable of interest, $\varepsilon_t$is white noise disturbance term with constant variance δ^2, x_{t-1} is exogenous variable, $t = 1,2,3 \dots ,T$. The variance of y_t is simply $\delta^2 x_{t-1}^2$, i.e., the variance of y_t depends upon the variance of x_t. This specification does not seem convincing because this model assumes a specific cause, i.e., x_t, for changing variance in y_t. Also, often the choice of x_t is theoretical, i.e., it could be oil price shock, monetary policy shock, or any other factor. Granger and Andersen (1978) modified it by replacing the exogenous variable with the past realized values of y_t. A simple case is

$$y_t = \varepsilon_t y_{t-1} \qquad 2$$

[4] For other issues with Maximum Likelihood methods, see Ardia 2008, pp 2-4.

Now the variance (conditional) of y_t is $\delta^2 y_{t-1}^2$, while the unconditional variance is either zero or undefined (Engle, 1982) which makes this type of modeling odd for practical use. Engle (1982) introduced conditional variance h_t at time t and postulated as a linear function of the square of past observations

$$y_t = \varepsilon_t h_t^{1/2} \qquad 3$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 \qquad 4$$

where, $\alpha_0 > 0$ and$\alpha_1 \geq 0$, in order to ensure that the conditional variance is positive. Variance of ε_t is equal to 1. This model is known as the ARCH (1) model. The general form of variance function can be expressed as:

$$h_t = \alpha_0 + \sum_1^q \alpha_i y_{t-i}^2 \qquad 5$$

where q is the order of the ARCH process and α is a vector of unknown parameters. Creating good fits to data would often require large numbers of lags and hence ARCH model with a large number of parameters; this creates difficulties in finding reasonable estimates, especially with smaller samples. To bypass this problem, Bollerslev (1986) proposed the Generalized ARCH or GARCH (p, q) model, which can be written as follows:

$$h_t = \alpha_0 + \sum_1^q \alpha_i y_{t-i}^2 + \sum_1^p \beta_j h_{t-i} \qquad 6$$

Where, $\alpha_0 > 0$, $\alpha_i \geq 0$ $(i = 1,2,3, \ldots, q)$ and $\beta_j \geq 0$ $(i = 1,2,3, \ldots, p)$. Now in Eq. 6, conditional variance depends upon its past values, which make the model more parsimonious. In most of the empirical applications, the simple specification $p = q = 1$has been found to work well. This specification led to the GARCH (1, 1) model. Because it has become the most frequently used, default model of choice, it is often known as the workhorse model.

3.2. Estimation Approach

There are two primary approaches used to estimate the GARCH type models, i.e., the frequentist approach and the Bayesian approach. These approaches use different algorithms to estimate models, consequently, have a difference in inference. However, the present study only used two algorithms from each approach. Each of these approaches along with algorithms is discussed in the following sub-sections;

3.2.1. Frequentist Approach

Most of the available statistical software has a built-in package to estimate the GARCH type models. In these packages, the researcher specifies the order of process, and the computation is performed. However, there are also some softwares which requires writing a program for estimation. This section explains the Maximum Likelihood method required to understand and write a program for the GARCH type models.[5]

The log-likelihood function for the GARCH (1, 1) model is as follow:

$$lnL = -\frac{T-1}{2}\ln(2\pi) - 0.5\sum_{t=2}^{T}\ln(\alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}) - 0.5\sum_{t=2}^{T}\left(\frac{y_t^2}{\alpha_0+\alpha_1 y_{t-1}^2+\beta_1 h_{t-1}}\right)$$

7

Maximizing the log-likelihood function, with respect to $w = (\alpha_0, \alpha_1\ and\ \beta_1)$ yields:

[5] For detail, see Enders (2015), pp. 152-154, and Tsay (2010), pp. 120-122.

$$\frac{\partial lnL}{\partial w} = -0.5\sum_{t=2}^{T}\frac{1}{(\alpha_0+\alpha_1 y_{t-1}^2+\beta_1 h_{t-1})} * \frac{\partial(\alpha_0+\alpha_1 y_{t-1}^2+\beta_1 h_{t-1})}{\partial w}$$

$$-0.5\sum_{t=2}^{T}\left[\frac{y_t^2}{(\alpha_0+\alpha_1 y_{t-1}^2+\beta_1 h_{t-1})^2}\right] * \frac{\partial(\alpha_0+\alpha_1 y_{t-1}^2+\beta_1 h_{t-1})}{\partial w}$$

$$= -0.5\sum_{t=2}^{T}\frac{1}{(\alpha_0+\alpha_1 y_{t-1}^2+\beta_1 h_{t-1})} * \frac{\partial(\alpha_0+\alpha_1 y_{t-1}^2+\beta_1 h_{t-1})}{\partial w}$$

$$+0.5\sum_{t=2}^{T}\left[\frac{y_t^2}{(\alpha_0+\alpha_1 y_{t-1}^2+\beta_1 h_{t-1})^2}\right] * \frac{\partial(\alpha_0+\alpha_1 y_{t-1}^2+\beta_1 h_{t-1})}{\partial w}$$

$$\ldots 8$$

It is possible to maximize the lnL concerningα_0, α_1 and β_1. However, first order conditions are the non-linear function as shown in Eq. 8, because h_t is a function of β_1; therefore, an analytical solution is not possible. Instead, the solution requires the search algorithm. Statistical software uses these algorithms and selects the value of parameters at which log-likelihood function is maximized. One such algorithm that is used to estimate the GARCH type model is BFGS.

BFGS is a numerical algorithm and is based on a “hill-climbing” strategy. It starts from an initial value and then search for the maxima in the likelihood function, or choose the value of the parameter at which value of likelihood function is maximized. This search for maxima done in a specific line (one-dimensional) that is why it is called a line search method. Therefore, it is essential to initialize the algorithm with a consistent starting point. According to Fiorentini et al., (1996), the unconditional expectation can be employed as an estimate of the initial value. Mathematically:

$$h_t = y_t^2 = \frac{1}{T}\sum_{s=1}^{T} y_s^2 \qquad 9$$

where y_s is the return series.[6] h_t is used in Eq. 8, and derivative of a log-likelihood function is computed. These derivative values are used as initial values of the algorithm. This Eq. 8 permits the recursive calculation for the derivatives.

This procedure is valid only if the likelihood function of the GARCH model is unimodal. However, empirically has been shown that the likelihood function of the GARCH model is multimodal (see section 1). When the likelihood function is multimodal, the sensitivity in the choice of starting value increases, i.e., the choice of starting value, and the direction of the algorithm in which search for maxima. Therefore, BFGS might not converge to global maxima. Contrary, there is a strategy, i.e., DE, which use multiple starting values with parallel chains method. This procedure is based on four steps. First, different mods are identified in the likelihood function using the Expectation-Maximization algorithm. Second, treat each observation as a regime of Markov chain, using a Viterbi algorithm (Viterbi, 1967). Stack all these Markov chains in K vectors, one for each regime. Third, estimate volatility model via Quasi-Maximum Likelihood for each vector. Finally, estimate the shape parameter of conditional distribution via the Maximum Likelihood method. During these steps, positivity and covariance-stationarity constraints are guaranteed through specific parameter-mapping functions.

[6] In case of present study, this is return series, because the model is GARCH (1, 1). However, if the model is ARMA-GARCH, then estimate ARMA model by using OLS at first step, and then use consistent estimated residual ($\varepsilon_s = y_s - x_s'\hat{b}$) instead of return series, for detail see Fiorentini et al. (1996).

For the comparison purpose, this study used both of these algorithms, i.e., BFGS and DE, to estimate the GARCH (1, 1) model. The performance of these algorithms was evaluated by comparing the estimates and the true parameter values in repeated simulations

3.2.2. Bayesian Approach

Unlike the Frequentist approach to estimation, the Bayesian approach assumes a vector of $y = (y_1, \dots, y_T)'$ of observations defined through a probability density$p(y|\theta)$. The parameter θ serves as an index of the family of the possible distribution for the observations. The significant difference between the Frequentist and Bayesian approach is the interpretation of θ. Frequentist approach assumes that there exists a true value of parameterθ, while in the Bayesian approach, it is assumed as a random variable. A prior density denoted by $p(\theta)$ characterizes this random variable. Classical Bayesian approaches rely on natural conjugate priors which were too restrictive for applications. More recent work uses much more flexible priors, using the technological device of hyper-parameter[7], to be discussed in greater detail later.

Moreover, depending on the researcher's prior information, this density can be more or less informative. By combining the likelihood function of the model and parameter with the prior density, can construct probability density using Bayes' rule to get the posterior density $p(\theta|y)$.[8] This posterior density is quantitative; a probabilistic description of the knowledge about the parameter θ after observing the data.

---

[7]In Bayesian statistics, a hyper-parameter is a parameter of a prior distribution; the term is used to distinguish them from parameters of the model for the underlying system under analysis.
[8] For basic understanding, see Kennedy (2008), pp 213-217.

It is often convenient to choose a prior density which is conjugate to the likelihood, i.e., the same distribution of prior and likelihood. In effect, conjugate priors permit posterior densities to emerge without numerical integration. However, easy calculations of this specification come with the restrictions they impose in the form of the prior. In many cases, it is unlikely that the conjugate prior is an adequate representation of the prior state of knowledge. If a conjugate prior is replaced by more flexible priors which allow the adequate representation of prior information, then the evaluation of posterior density formula is analytically intractable. Therefore, asymptotic approximation or Monte Carlo methods are required. Deterministic techniques can provide excellent results for low dimensional models. However, when the dimension of the model becomes large, simulation is the only way to approximate the posterior density.

Metropolis et al., (1953)first introduced the idea of Monte Carlo Markov Chain sampling., and was consequently generalized by Hastings (1970). This study will focus on Bayesian inference. A general and detailed statistical theory of Monte Carlo Markov Chain methods can be found in Tierney [1994].

The idea of the technique is straightforward and is explained in Zaman (1996). Let $\theta = (\theta_1, \dots, \theta_p)$ be the parameters to be estimated. In order to use the usual Bayesian technique, the posterior density of θ given the data y is computed. Most of the methods known as "samplers" bypass numerical integrations required to compute the posterior, obtaining a sample or a random vector instead $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_p)$ from the posterior distribution of $\theta \mid y$. The estimate of the posterior density can be obtained through extensive sampling of the posterior distribution. Usually, the full posterior density is not needed; the posterior mean and covariance suffice. These are also easily estimated based on a random sample from the posterior density. However, due to the increase in computational power, the posterior density can be easily be obtained at a low cost nowadays. The distribution of the posterior density helps to study either the distribution is symmetric or asymmetric, because the Frequentist approach assumes a normal distribution, which is not true in the case of Bayesian (Ardia, 2008). The well-known Monte Carlo Markov Chain methods in the literature are Gibbs sampler and MH. Sometime situation gets complicated and cannot be easily handled by the Gibbs sampler, such as; when it is not easy to breakdown the joint density into the full conditional densities are of unknown form or when the joint density is of nonstandard form. Metropolis-Hastings algorithm is a simulation scheme which allows generating draws form any density of interest whose normalizing constant is unknown.

One algorithm is MH. This algorithm with independence chain is often used with normal and student t distribution, for the estimation of the GARCH models. Practically, Ardia (2008) and Ardia et al., (2009) used MH with independence chain for the estimation of the GARCH model. On the other hand, there is RAM, developed by Vihola (2012), which worked well as compared to other Adaptive chains, for the student t distribution.[9] For the comparison purpose, the present study used both of these algorithms to estimate the GARCH model.

3.3. Evaluation of Algorithms and Approaches: Monte Carlo Simulation

For the comparison purpose, this study has selected four algorithms for the estimation of the GARCH type models. Two of them are from the Frequentist approach,

[9]Andreas (2012) develop package for RAM, with the name of “adapt MCMC” in R language, URL https://cran.r-project.org/package=adaptMCMC.

i.e., BFGS and DE, and two are from the Bayesian approach, i.e., MH and RAM. The assessment is comprised of two steps. In the first step, algorithms will be compared within an approach, and in the second step, the comparison will be made across approaches. Hence, either of these algorithms will be generalized for estimation of the GARCH-type models.

Furthermore, for generalization, two things are imperative. First, which algorithm is best suited for the estimation of the GARCH type models. Monte Carlo simulation is a technique through which can generate many series with known parameters. It is assumed that the data generating process is unknown, and then estimation is conducted with selected algorithms. The algorithm that closely estimates the true data generating process is selected as the best algorithm. Second, in the real world, the data generating process is unknown. In the real world, econometrician and statistician are not interested in the mechanism through which the series is generated; instead, they are interested in appropriate approximation. It is, therefore, this study will generate data with different distributions, and then estimated through selected algorithms. The algorithm that remains consistent across different data generating process is selected as preferable for the estimation of the GARCH type models (James et al., 2013).

The Monte Carlo Simulation technique has been used for the selection of a robust algorithm for the estimation of the GARCH-type models. The present study has also analyzed the robustness of this algorithm along the whole surface. The methodology of evaluating the estimation power of each algorithm consists of the following steps to compare the surface of different algorithms:

1. In the first step, the GARCH (1, 1) is generated by setting $\alpha_0 = 0.01, \alpha_1 = 0.1, \beta_1 = 0.1$. with a normal distribution.

$$y_t = \varepsilon_t h_t^{1/2}$$

$$\varepsilon_t \sim i.i.d\ \mathcal{N}(0,1)$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} \qquad 10$$

   A total of 1500 observations were generated, out of which initial 500 were discarded to control for the initial value effect. "Rugarch" package is used to generate the simulated sample. It is available in R.

2. Estimate the volatility of simulated samples obtained in step 1 with the GARCH (1, 1) model. For the estimation of this GARCH model, Maximum Likelihood estimation method is used. To search for the maximum point in the likelihood function, this study used the following four different algorithms separately:
   a. First, the BFGS algorithm was used to search for the parameters at which the value of the likelihood function is maximized. BFGS algorithm is built-in and available in Ox-Metrix package for the estimation of the GARCH type models. This study used the same package for the estimation, and then noted the value of the estimated parameters.
   b. GARCH (1, 1) model was estimated with the Metropolis-Hasting algorithm. The same variable was used for an estimation, which was obtained in step 1. This algorithm is available in "bayesGARCH" package in R-languages. The estimated parameters were then noted.
   c. Similarly, DE and RAM algorithms were used to search for the parameters at which the value of likelihood function is maximized for the GARCH (1,

1) model. The estimated parameters were then noted, respectively. Both of these algorithms are available in "MSGARCH" package of R-Language.

3. Calculate the bias, i.e., calculate the absolute difference between the estimated parameters in step 2 and the true parameters in step 1, respectively. Note this bias for each algorithm, separately.
4. One time estimation of the true data generating process could be misleading. To overcome this bias, the experiment of simulation is repeated the process from step 1 to step 3, 1000 times.
5. Calculate the variance of these 1000 points obtained in step 4, for each parameter separately. Also, note this variance separately according to each algorithm.
6. From step 1 to step 5, the value of true parameter was the same. Next, this study then executed the surface analysis, and the values are presented in Table 3.1. This table gives information about the whole surface of each algorithm separately. The values assume the shape of a triangle because, for all other values of the table, the assumption of covariance stationary does not hold.[10]

Table 3.1: α_1 and β_1 Value Changes

		β_1							
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
α_1	0.1	(0.1, 0.1)	(0.1, 0.2)	(0.1, 0.3)	(0.1, 0.4)	(0.1, 0.5)	(0.1, 0.6)	(0.1, 0.7)	(0.1, 0.8)
	0.2	(0.2, 0.1)	(0.2, 0.2)	(0.2, 0.3)	(0.2, 0.4)	(0.2, 0.5)	(0.2, 0.6)	(0.2, 0.7)	
	0.3	(0.3, 0.1)	(0.3, 0.2)	(0.3, 0.3)	(0.3, 0.4)	(0.3, 0.5)	(0.3, 0.6)		
	0.4	(0.4, 0.1)	(0.4, 0.2)	(0.4, 0.3)	(0.4, 0.4)	(0.4, 0.5)			
	0.5	(0.5, 0.1)	(0.5, 0.2)	(0.5, 0.3)	(0.5, 0.4)				
	0.6	(0.6, 0.1)	(0.6, 0.2)	(0.6, 0.3)					
	0.7	(0.7, 0.1)	(0.7, 0.2)						
	0.8	(0.8, 0.1)							

[10] Intercept is fixed throughout the experiment, because in real-world problem, it is possible that relevant variable are omitted from the econometric model, creating bias. This bias can be lessened (but not eliminated) by including an intercept term. On the other hand, no bias is created by including an unnecessary intercept (for detail see, Kennedy 2008, pp 109-110).

7. Table 3.1 represents the table of values for which Alpha and Beta changes. For 36 combinations, the estimation will be valid. The table seems triangle because if the value of the parameter increased in either direction, it violates the assumption of covariance stationary condition. Therefore, the table is half surface. To analyze the bias and variance along the whole surface, this study used the contour plots. These contour plots are constructed after completing the table of values. These contour plots show abrupt changes along the surface. This is because the small sample size of the simulation leads to high random errors of estimation in the bias and variance. Smoothing to create a response surface is a crucial way to avoid this problem and get a better global picture of results from simulations carried out at a small number of points. Exponential smoothing procedure is done to construct the surface of the contour plot smooth surface, and then the procedure "Akima's Polynomial Method" (Akima, 1969) is used to construct a contour plot.
8. The main objective of the study is to compare the different algorithms within the Frequentist and Bayesian approach. To keep the analysis simple, present study subtracted the surface of bias of DE from BFGS, from step 6. Then for this difference surface, the contour plot is constructed. This contour plot has a direct and straightforward interpretation, i.e., if the surface is positive, then it implies that the bias of BFGS is higher than DE and vice versa.
9. Repeat step 8 for the variance.
10. Repeat step 8 and 9 for MH and RAM by subtracting the RAM from MH.
11. A normal distribution of the true data generating process was assumed. The analysis based on the normal distribution gives a clue about the first step of generalization, i.e., which algorithm accurately estimates the true data generating process — now

repeat the same procedure from step 1 to 10 by assuming student t distribution for the error term. Again, it will give a clue about the performance of the different algorithm with a different distribution. By combining the results of both the steps based on the normal and student t distribution, will give a broader picture of the performance of different algorithms furthermore, as it is already explained that for the generalization of the algorithm, it is essential to test performance against different data generating processes. Hence, it will give evidence either any one of them is consistent across different data generating processes or not.

3.4. Application with Monte Carlo Simulation, for GARCH (1, 1)

"The more adequately a model fit whatever it stands for without being needlessly complex, and the easier it is for the intended audience to interpret it correctly, the better it will be."

Alberto Cairo (2016)

Any visualization is a model (Cairo, 2016) and is a powerful technique to summarize a large amount of data. Visualization not only minimizes the time to understand the information but an excellent way to present information more effectively. The same technique is used to summarize the empirical results of the Monte Carlo Simulation technique. It is not only an effective way of communication but also minimizes the time to understand the enormous difference between the different estimation methods. Moreover, contour plots are a more appropriate way to represent the response surface methodology and are used in the present study.

Fig. 3.1 represents a contour plot[11] for bias (upper plot) and variance (lower plot) around the true parameter of BFGS. This plot is construed by taking the difference of each estimated value of BFGS from the true parameter, for GARCH (1, 1) model. After taking the difference of the entire surface of BFGS from the true parameter. If the resultant of difference is positive, then the layer of contour plot will be positive, vice versa. The layers in each plot represent the variation along the surface, i.e., the minimum value for the layer is 0.56 and increase slightly increase toward the origin of plot, and reach to the maximum

[11] How to read or interpet contour plots, a short description for understanding is mention at the following link; https://www.youtube.com/watch?v=WsZj5Rb6do8&t=128s

level of 69. It can be interpreted as if the model moves from high persistence to low persistence region in the surface plot the level of bias increase.

Furthermore, there is, on an average 56 percent difference between BFGS and true data generating process, which reaches to the maximum level of 69 percent. It can also be interpreted as BFGS estimation method approximately 56 will overestimate the true parameters. A specific value represents each layer. Similarly, in case of variance around the true parameter is entirely positive, implies the difference remains positive through the entire surface of the contour plot.

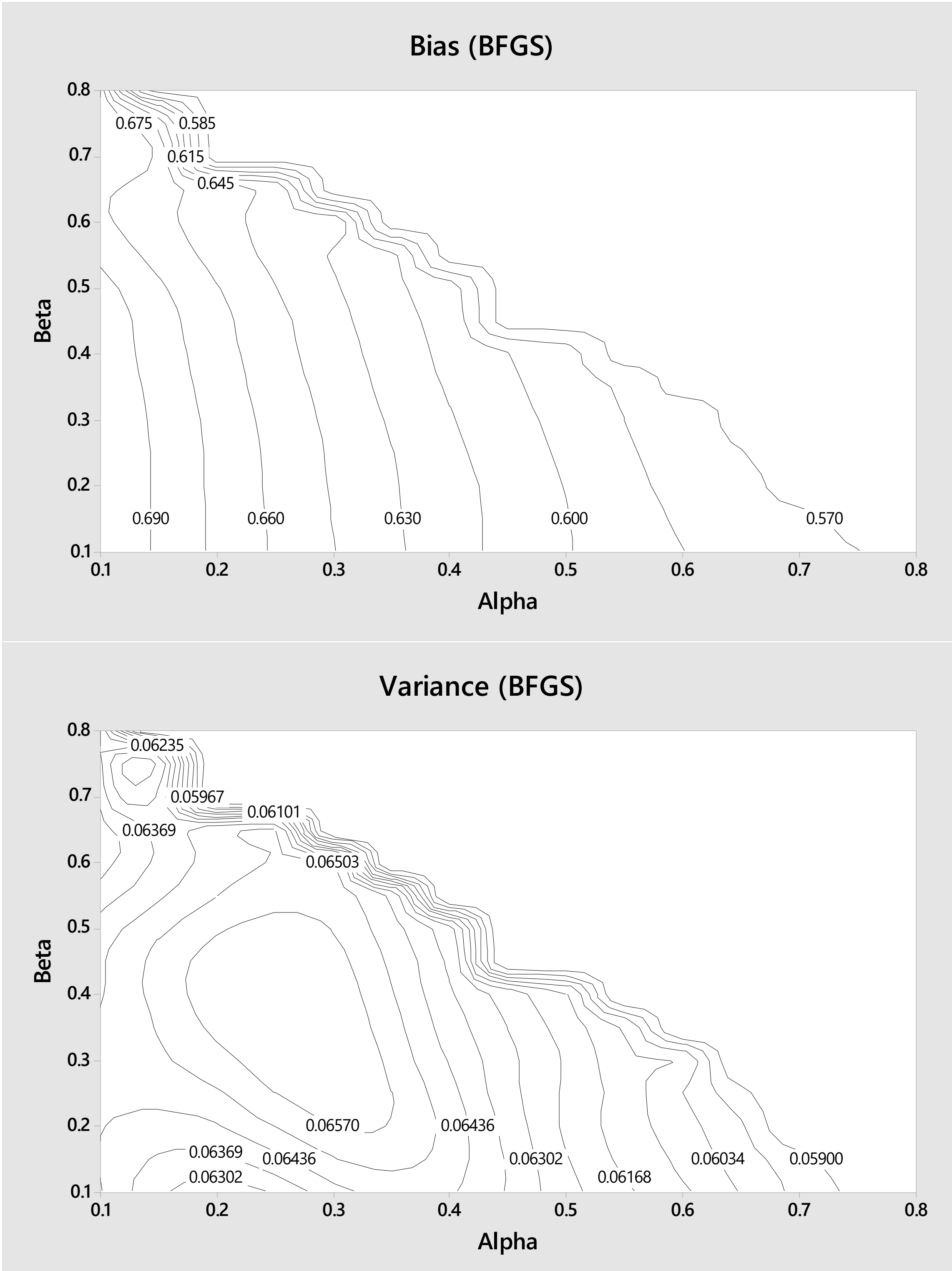


Figure 3.1: Contour plots represent the level of biases for the parameters of the GARCH (1, 1) model (Normal Distribution). The x-axis represents the change in the value of the Alpha parameter, from 0.1 to 0.9, while the y-axis represents the change in the value of the Beta parameter. The upper plot represents the level of bias for BFGS, while the lower plot represents the variance of the BFGS. The arrow points from low to high.

3.5. Comparison of Estimation Method within the Frequentist Approach

One objective of this study is to compare the BFGS with DE, in the Frequentist approach. For comparison purpose, the difference is calculated between BFGS and DE. Contour plots are then constructed for these difference values, the after smoothing the surface the resented in the upper plot of Fig. 3.2. The surface of the contour plot is entirely positive, and these values vary from 40 percent to 49 percent on average. It implies that BFGS has 40 percent more bias and therefore overestimate the GARCH (1, 1) model as compare to DE.

Similarly, the lower plot of Fig. 3.2 shows the difference in the variance of BFGS and DE. Layers for this contour plot are positive as well. By combining both bias and variance around the true parameter implies that the estimation power of BFGS estimation method is weaker than the DE.

In reality, most of the financial time series does not follow the normal distribution. Therefore, in the next stage of the experiment of the Monte Carlo simulation, the same step is repeated by assuming student t distribution for the data generating process. Results are presented in Fig. 3.3. The layers of the contour plots are positive; it implies that the BFGS estimation method has more bias as compared to the DE. Moreover, These results are consistent with normal distribution data generating process.

Furthermore, the level of difference increases if the contour plots are compared with the across normal and student t distribution. In the case of student t distribution, the minimum difference is about 57 percent on average, which increases to the maximum level of 64 percent. It implies, BFGS estimation method on average gives biased results about 57 percent of the times, as compared to DE estimation method.

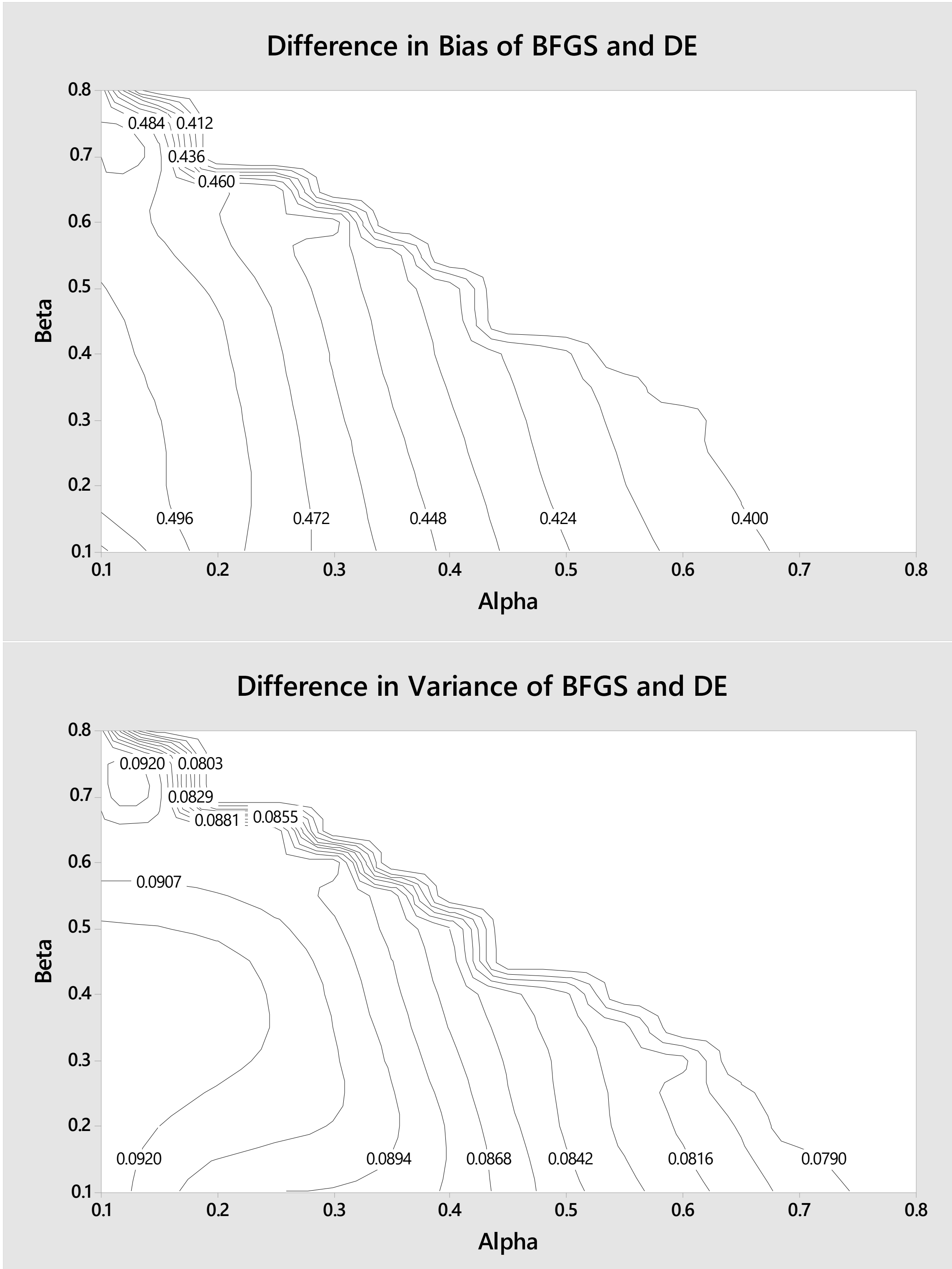


Figure 3.2: The difference in bias (upper graph) and variance (lower graph), for BFGS and DE, along with the surface. In this case, the data generating process in a normal distribution.

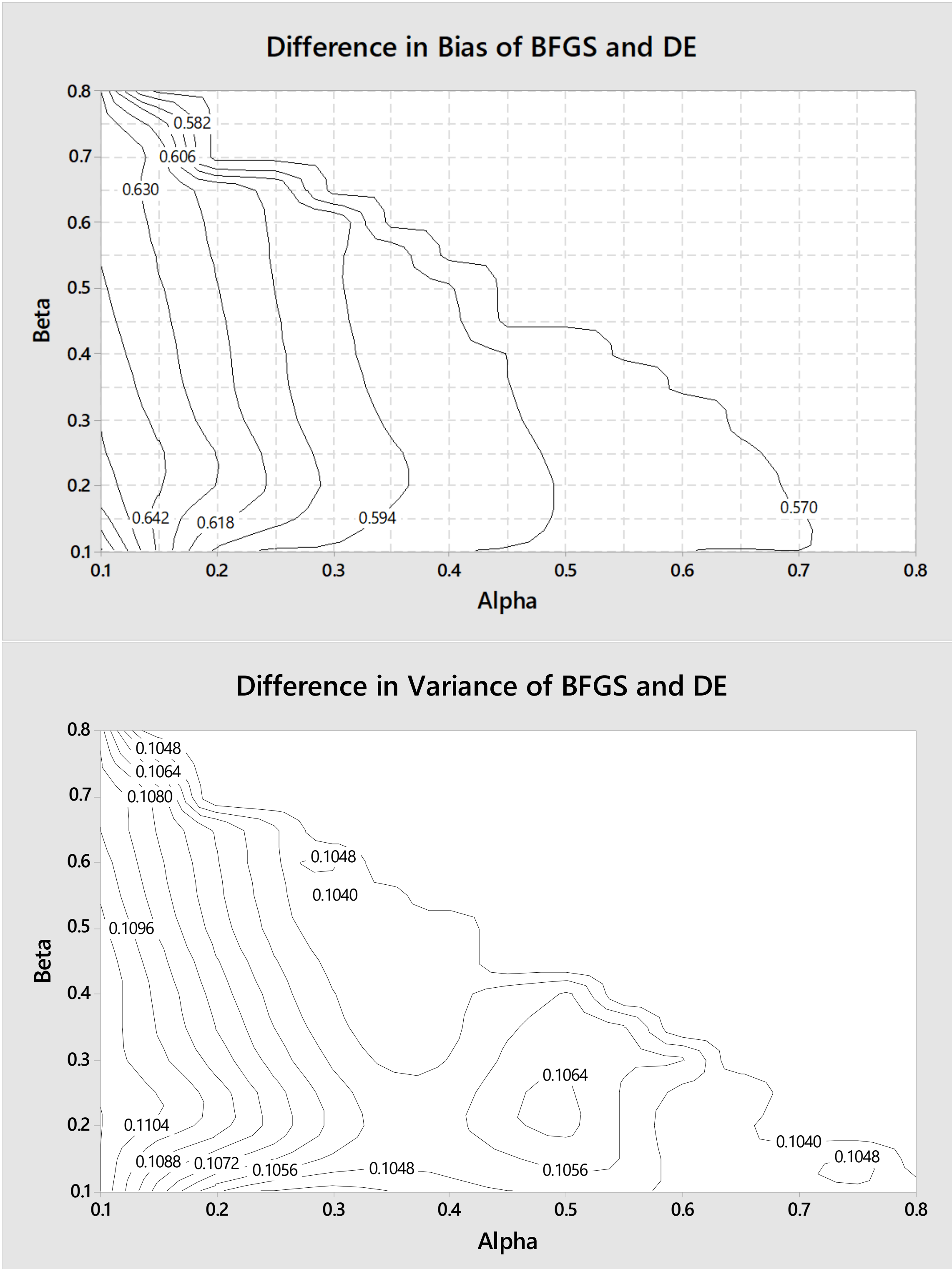


Figure 3.3: The difference in bias (upper graph) and variance (lower graph), for BFGS and DE, along with the surface. In this case, the data generating process in student-t distribution.

3.6. Comparison of the Estimation Methods within the Bayesian approach

A similar experiment is applied to the Bayesian approach as in the frequentist approach, i.e., repetition of experiment with normal and student t distribution, for the entire surface. In the case of the Bayesian approach, two estimation methods are selected, which are MH, and RAM. First, the contour plot is constructed by taking the difference of MH and RAM for the normal distribution, and results are presented in Fig. 3.4. The entire surface is positive with the minimum value of 0.28 and reaches to the maximum value of 0.33. At first, it implies that the surface is positive, i.e., MH has a higher level of bias around the true parameter as compared to the RAM. Second, it implies that on average, MH will give bias estimates about 28 percent, if the model is high persistence and the level of bias increase as the move from high persistence to the low persistent region of the contour plot. Similarly, in case of difference in variance around the true shown in the lower plot of Fig. 3.4.

Furthermore, if the data generating process changes from normal to the student t distribution two things are inferred, i.e., the MH is weaker estimation method as compared to the RAM and the level of difference has increased. Hence, it can be concluded that RAM estimation method is the more appropriate choice of estimation of the GARCH (1, 1) model as compared to the MH.

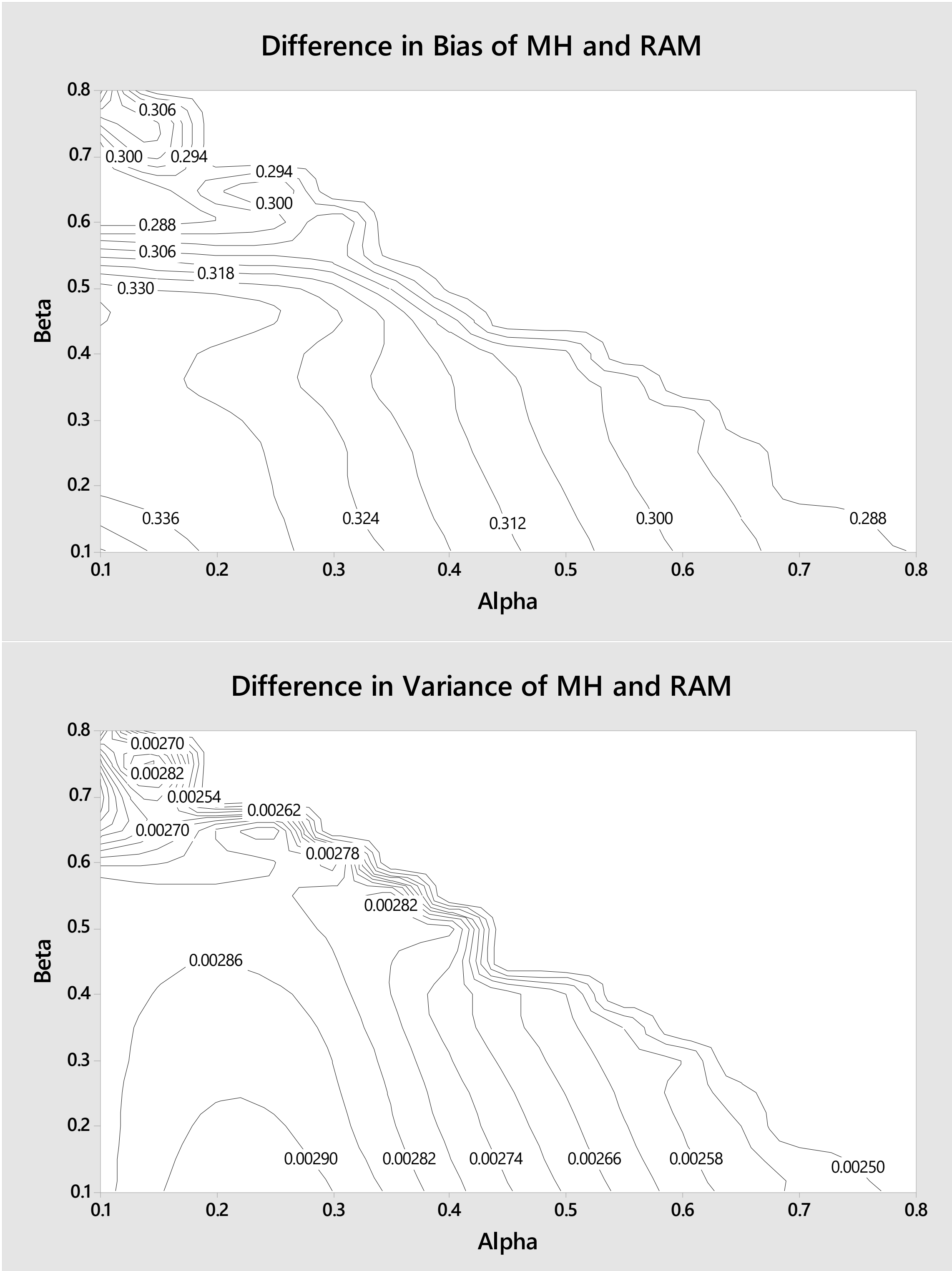


Figure 3.4: The difference in bias (upper graph) and variance (lower graph), for MH and RAM, along with the surface. In this case, the data generating process in a normal distribution.

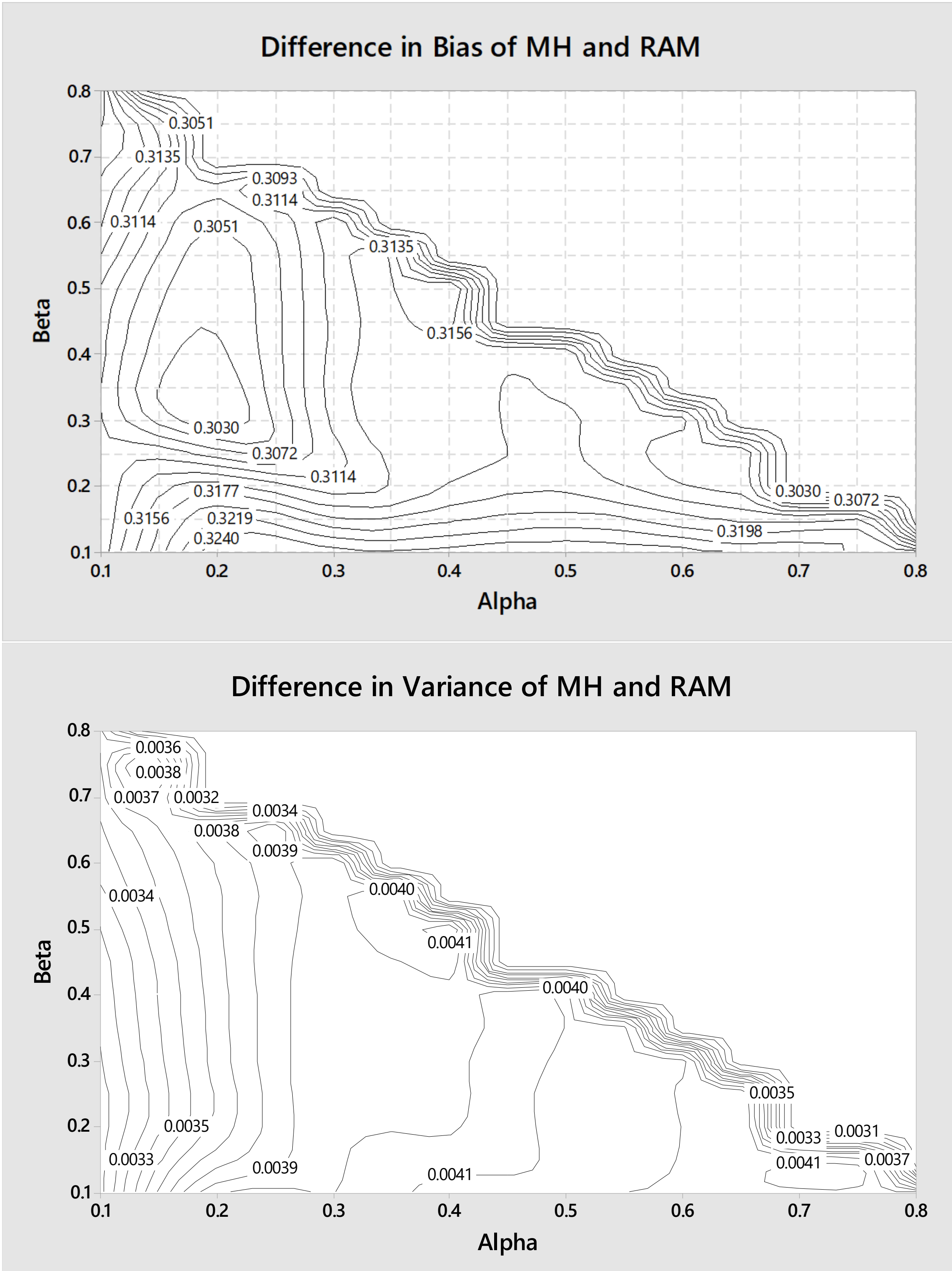


Figure 3.5: The difference in bias (upper graph) and variance (lower graph), for MH and RAM, along with the surface. In this case, the data generating process in student-t distribution.

3.7. Chapter Summary

The fundamental objective of the study is to compare the Frequentist and Bayesian approaches. To evaluate the estimation power of each algorithm around the true parameter, this study is used following steps; (i) true data generating process, (ii) true data generating process with different distributions, (iii) surface comparison. However, before comparing approaches, it is essential to compare algorithms within each approach. Simulation is used to evaluate the performance of each algorithm around the true data generating process. The experimental result of simulation implies that there is on average, about 40 percent difference between BFGS and DE across the surface. Moreover, the level of difference increase if the movement is done towards the low persistence region of the surface plot. The main reasons for this difference between BFGS and DE are just because of the multimodality in the likelihood function.

Similarly, the level of variance around the true parameter is on average, about 7 percent higher in case of BFGS as compare to DE. by combining both the bias and the variation implies that Mean Square Error of BFGS is higher than DE, hence, BFGS is inconsistent estimation method as compare to DE, for the estimation of GARCH type models.

Furthermore, BFGS use single starting value with hill climbing strategy, due to which it often converge to local maxima. Hence, due to converging at local maxima, the estimated parameters will be different from global maxima. On the other hand, DE uses multiple starting values with various chains, which automatically avoid local maxima, and converge to global maxima. Therefore, BFGS has a higher level of bias as compared to DE for the estimation of GARCH type models.

Contrary, in the Bayesian approach, the level of bias is, on average, about 28 percent, between MH and RAM. This is because MH estimation method uses the method independent MCMC, where the current draw of the posterior distribution does not take any information from the previous draw. While RAM uses the technique of Adaptive MCMC, where it receives data from the earlier draws of the posterior distribution, and ultimately converge to a stationary point. Therefore, RAM has a lower level of bias as compared to the MH estimation method.

To accomplish the second objective of the study, the same experiment of Monte Carlo simulation is repeated with a different distribution. In the real world problem, the value of the true data generating process is unknown, therefore to increase the reliability of experimental results, the present study repeats the same experiment with the student t distribution. Fortunately, the experimental results not only remain consistent but also, the level of bias and variance increase between BFGS – DE, and MH – RAM.

CHAPTER 4

REAL WORLD PROBLEM AND AN APPLICATION

"We refer here to emphasize the construction of approximations, not the mechanism by which they are constructed."

M.W. Trosset

The comparison of two algorithms in the frequentist approach has started in the previous chapter, i.e., BFGS and Difference Evolution. These algorithms have a different mechanism to choose the starting value. BFGS initiates with a single staring value, while the DE initiates with multiple starting values along with multiple chains. Hence there is a built-in safeguard that avoids local maxima and converges to global maxima. Because of this difference, BFGS often converges to local maxima while the DE converges to global maxima. By comparing on simulated sample data, it is proved that DE is preferable for the estimation of GARCH (1, 1) model as compared to BFGS. The estimation power of DE is valid for the whole surface, i.e., either the GARCH (1, 1) model is low, moderate, or of high persistence. Moreover, the estimation power of DE is consistent across different distributions as well, i.e., normal and student t distribution. Similarly, in the Bayesian approach, RAM and MH have been used to compare the estimation power for GARCH (1, 1) model, and it is also confirmed that the estimation power of RAM is better than MH.

In the case of simulation, the value of true parameters as well as the distribution is known. The problem arises when working with real data sets; where the value of the true parameter and distribution is unknown. In a real-world problem, statisticians construct approximates based on properties of data set in hand. Furthermore, if BFGS and MH are not working correctly in case of true data generating process, as compared to others, then they will also not work well in a real-world problem. Due to these issues, the Metropolis-Hasting and BFGS will not be discussed further.

In this chapter, the window of PSX is used as a real-world application to compare the estimation power of the Frequentist and the Bayesian approach for GARCH (1, 1) model. This comparison consists of two steps. First, empirical results will be compared, and in the second step, sensitivity about the choice of prior and residual testing will be analyzed to assess the misspecifications. Furthermore, this chapter is divided into two sections according to the distribution assumed for the GARCH (1, 1) model, i.e., normal and student t distribution. Finally, the present study illustrated some exciting aspects of the Bayesian approach through a probabilistic statement based on the parameters.

4.1. Estimations of GARCH (1, 1) – Normal distribution

Conventional econometrics assumes constant variance of time series data. However, financial time series violates this assumption such that the time series exhibit spans of unusually high and low volatility, i.e., volatility clustering. Therefore, in such cases, it is clear that the assumption of constant variance is very restrictive, and may lead to misleading inference and prediction. For an illustration purpose, consider an individual who is planning to invest in an asset at time t and sell at time $t+1$. For this investor, the forecast of the rate of return on this asset alone will not be enough. The investor would be interested in what the variance of the return over the holding period would be. Hence, the unconditional variance is of no use either. The investor would want to examine the behavior of the conditional variance of the series to estimate the risk associated with the asset at a specified period.

4.2. Methodology

4.2.1. Frequentist Approach

Suppose that values of ε_t is normally distributed, with mean zero and constant variance, σ^2. To write the ML function for ε_t as described in Standard distribution theory;

$$L_t = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right) \qquad 4.1$$

where L_t is the likelihood for ε_t. Assume that each ε_t is independent of other ε_{t-i}, then the joint likelihood is just a product of each realization, ε_1, ε_2, …, ε_T. Also, assume the homoscedasticity for these realizations, and then the joint likelihood of realization is as follows;

$$L_t = \prod_{t=1}^{T} \left(\frac{1}{\sqrt{2\pi}\sigma^2}\right) \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right) \qquad 4.2$$

By using natural log, the product can be converted to the sum or difference, which makes the system to handle the process easily;

$$lnL = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(\varepsilon_t)^2 \qquad 4.3$$

ML estimation method maximizes the likelihood or probability of drawing the observed sample. In the classical approach, the expected value of ε_t is zero, with constant variance (σ^2) having no serial correlation. In the case of the GARCH model, since the first order equation is nonlinear, the solution, therefore, requires a search algorithm. The simplest way to illustrate the issue is to introduce a GARCH (1, 1) error process into the regression model. Where the error term ε_t has a non-constant conditional variance. Since each realization of ε_t has the conditional variance h_t, the joint likelihood of realization ε_t through ε_T is:

$$L_t = \prod_{t=1}^{T} \left(\frac{1}{\sqrt{2\pi h_t}}\right) \exp\left(-\frac{\varepsilon_t^2}{2h_t}\right) \qquad 4.4$$

So, the log-likelihood function is:

$$lnL = -\frac{T}{2} \ln(2\pi) - 0.5 \sum_{t=1}^{T} \ln(h_t) - 0.5 \sum_{t=1}^{T} (\frac{\varepsilon_t^2}{h_t}) \qquad 4.5$$

Now substituting the value of h_t, that is GARCH (1, 1):

$$lnL = -\frac{T-1}{2} \ln(2\pi) - 0.5 \sum_{t=2}^{T} \ln(\alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}) - 0.5 \sum_{t=2}^{T} (\frac{\varepsilon_t^2}{\alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}}) \qquad 4.6$$

By the introduction ofε_{t-1}^2, the initial observation will be lost for ε_0 (outside of the sample). Now, it is possible to maximize lnL for all parameters.

It can also be written as Eq. 4.6 in matrix form by defining the vectors $\boldsymbol{y} = (y_1 \ldots y_t)'$ and$\boldsymbol{\alpha} = (\alpha_0 + \alpha_1)'$. Model parameters can be regrouped as $\Psi = (\boldsymbol{\alpha} + \beta)$ for simplification purposes. In turn, the diagonal matrix is $T \times T$:

$$\sum = \sum(\Psi) = diag((h_t(\Psi))_{t=1}^{T}) \qquad 4.7$$

Where;

$$h_t(\Psi) = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}(\Psi) \qquad 4.8$$

The ML function for $\Psi$is stated as follows:

$$lnL(\Psi \mid \text{y}) \propto (det\sum)^{-\frac{1}{2}}\text{exp}[-\frac{1}{2}y'\sum y] \qquad 4.9$$

It is very time-consuming to find out the solution for first-order conditions to the maximum. It is required for an analytical solution, but the numerical solution is straightforward.

4.2.2. Bayesian Approach

The methodology was explained by Ardia (2008). The likelihood function in Eq. 10 estimates the parameters where the probability is high, while in the case of the Bayesian approach in the first step, it needs priors. Thus, appropriate priors for parameters can be estimated from the following models:

$$p(\boldsymbol{\alpha}) \propto \mathcal{N}_2(\boldsymbol{\alpha}|\boldsymbol{\mu_\alpha}, \textstyle\sum_\alpha)\ \mathbb{I}_{(\boldsymbol{\alpha}>0)}$$

$$p(\boldsymbol{\beta}) \propto \mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\mu_\beta}, \textstyle\sum_{\boldsymbol{\beta}})\ \mathbb{I}_{(\boldsymbol{\beta}>0)} \qquad 4.10$$

Where $\boldsymbol{\mu}$ and $\sum$ are hyperparameters, $\mathbb{I}$ is an indicator function for the confirmation of covariance stationary condition, with a normal distribution, $\mathcal{N}$. Also, all parameter are independent of each other, hence implies $p(\Psi) = p(\boldsymbol{\alpha})p(\beta)$. By applying the Bayes' rule, it can be obtained posterior joint density as follows:

$$p(\Psi|y) \propto lnL(\Psi|y)p(\Psi) \qquad 4.11$$

Using this posterior density, it can be simulated the parameter and obtain the density of the estimated parameters.

4.3. Empirical Analysis

4.3.1. Basic statistics – PSX

Daily PSX is used as a real-world application to compare the DE (from the Frequentist approach) with RAM (from the Bayesian approach). The sample period is from March 20, 2002, to January 11, 2005, with 696 number of observations excluding weekdays and holidays. The data span covers three financial years and is suitable to apply the Maximum Likelihood Estimation method. Hence, the present study has a reliable comparison among different approaches for estimation of the GARCH (1, 1) model. The nominal returns are expressed in percent similar to Bollerslev and Ghysels (1996). Daily log-return of PSX is plotted in the upper plot of Fig. 4.1. The present study examines statistically, the existence of autocorrelation in the series of PSX and applied joint nullity of the autoregressive coefficient. The regression is estimated with the autoregressive coefficient up to 20^{th} lag. The p-value of the Wald statistics is 0.17, which does not support the presence of autocorrelation in the series of PSX. However, the upper plot of Fig. 4.1 depicts clusters of high and low volatility, i.e., volatility clustering. It is evident in the lower plot of the Fig. where the sample auto-correlogram of square observation is shown. This auto-correlogram supports the effect of volatility clustering in the log-return series. In this case, the autocorrelation is significant (bars are outside the bands), implying the GARCH effect in the series.

Moreover, the value of the ARCH statistic is 12.23 (P-value > 0.000), directly implying that the autoregressive part of PSX is conditional on heteroskedastic observations. The value of the Wald statistics is 207.274 (P-value >0.000), firmly rejecting the null hypothesis of no autocorrelation up to 20^{th} lag.

Furthermore, the P-value of the RUNS statistics is 0.22, which implies that there is no predictability in the series. ADF statistic is applied to test the stationarity of the log-return series of PSX. The null hypothesis of non-stationary is rejected, with calculated ADF -14.9087 at 1% significance level. From this essential testing, it is concluded that the log-return series of PSX is stationary with no evidence of autocorrelation. While the presence of ARCH and GARCH effect is suspected in the series.

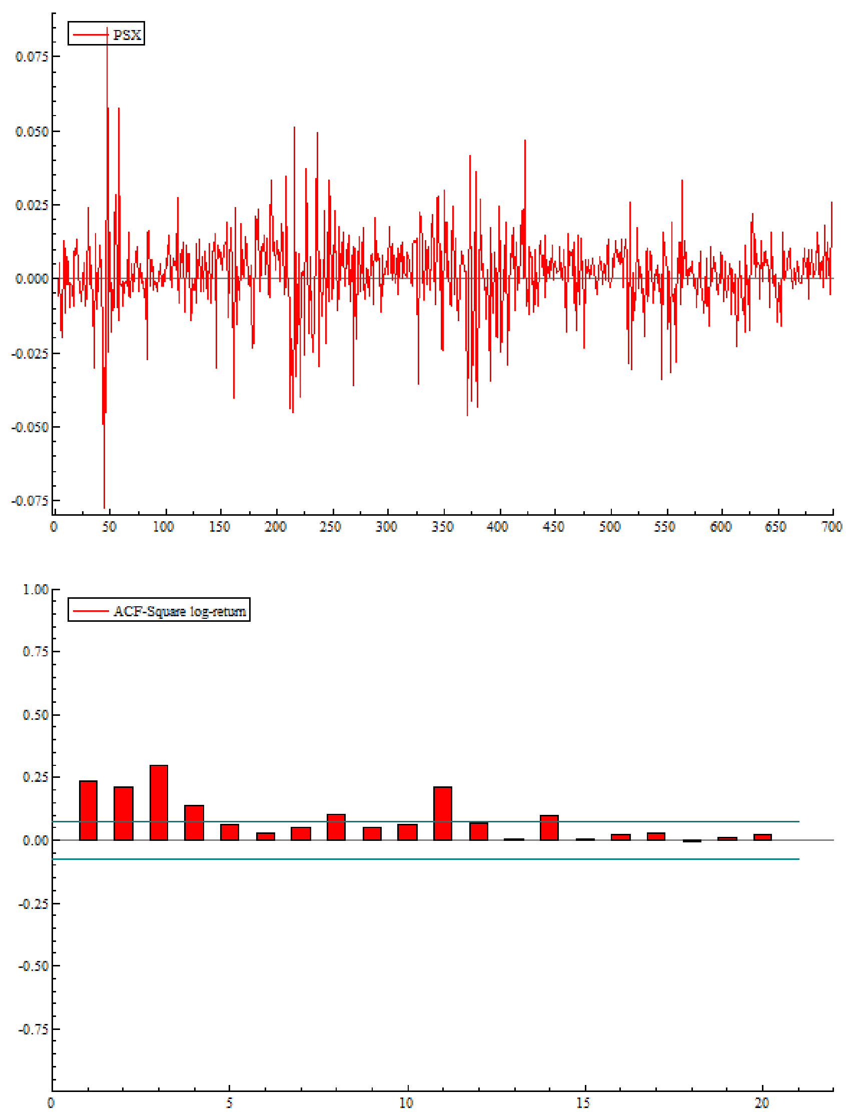


Figure 4.1: PSX daily log-returns (upper graph) and sample auto-correlogram of the squared log returns (lower graph)

4.3.2. Empirical results based on GARCH (1, 1)-Normal Distribution

The log return series of PSX is estimated with the parsimonious GARCH (1, 1), by assuming the normal distribution for the error term. As a prior density for the Bayesian estimation truncated, the normal distribution is chosen with zero mean vectors and diagonal covariance matrices. Finally, the joint prior is constructed by assuming that there is prior independence between parameters of the GARCH (1, 1) model. One chain is executed with the length of 5000 passes. The length of the chain is set long because of the adaptive behavior, i.e., it time some time to converge to a stationary point.

After estimation in the Bayesian approach, the first important step is to check the convergence of the chain, which is an indication of the reliability and accuracy of the estimated parameter. If the chain is not converging to a stationary point, then it explicitly implies that the estimated model is not a useful approximation data generating process. One way to achieve good approximation is to increase the length of the chain. An informal approach is to analyze the convergence of the chain by plotting the running mean of the chain. Fig. 4.2 represents the running means for the parameters of the GARCH (1, 1) model. For all parameter, the chain converges to a stationary point. Formal testing for convergence of chain has been introduced by Gelman and Rubin (1992). This is a formal approach based on the analysis of variance through which it is analyzed whether the path of chains should remain the same after convergence. Considering the $m$number of chains for each parameter in the model and is the real function of $\xi = \xi(\psi)$. Each chain has length J given by $\left\{\xi_i^{[j]}\right\}_{j=1}^{J}, i = 1, \dots, m$. The variance within the chain, W is:

$$W = \frac{1}{m(J-1)} \sum_{i=1}^{m} \sum_{j=1}^{J} \left(\xi_i^{[j]} - \bar{\xi}_i\right)^2 \qquad 4.12$$

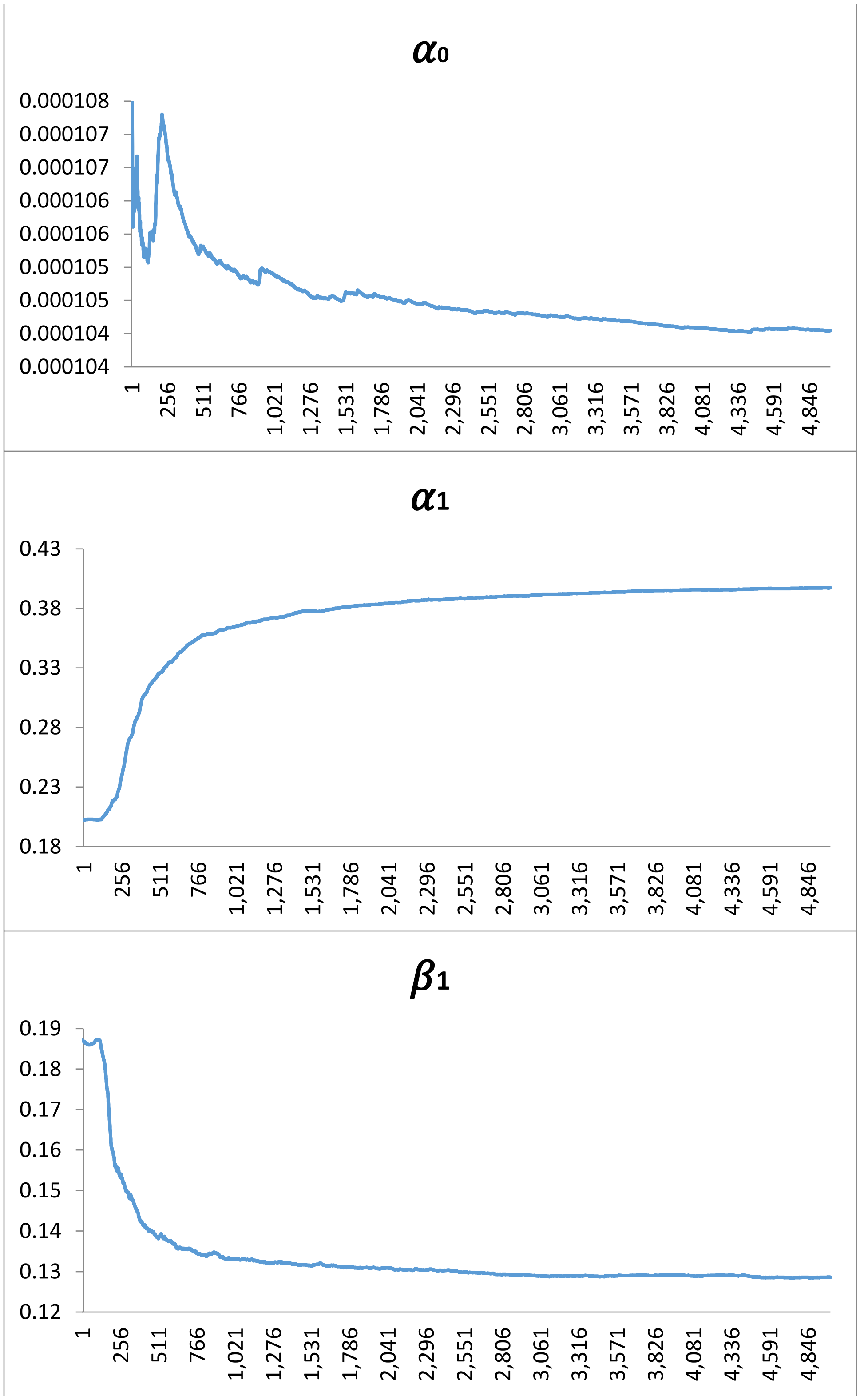


Figure 4.2: Running mean of the chain over 5,000 iterations. The acceptance rate is 0.98, and it seems that value converges to the stationary point after 2500 observation.

Where $\overline{\xi}_i$ is the average of observations of the ith chain, and $\bar{\xi}$ is the average of these averages. After convergence, all these mJ values for ξ_i are drawn from the posterior distribution, and δ_{ξ}^{2}, the variance ofξ, can be consistently estimated by the variance within and between chains. The variance between chains B is (if chains are more than one):

$$B = \frac{1}{m(J-1)} \sum_{i=1}^{m} \left(\bar{\xi}_i - \bar{\xi}\right)^2 \qquad 4.13$$

δ_{ξ}^{2} is the following weighted average:

$$\hat{\delta}_{\xi}^{2} = \left(\frac{J-1}{J}\right) W + \frac{1}{J} B \qquad 4.14$$

If the initial value is still influencing the trajectories, then the chain will not converge. In this case $\hat{\delta}_{\xi}^{2}$ will be overestimated due to the overdispersion. The overestimation of $\hat{\delta}_{\xi}^{2}$ can only be controlled by increasing the length of the chain and ultimately will converge. Before convergence, W tends to underestimate $\hat{\delta}_{\xi}^{2}$ because each chain will not have adequately traversed the entire state space. To handle this issue, Gelman and Rubin (1992) constructed a gauge for convergence. This indicator is an estimator of "*potential reduction factor*" given by:

$$\hat{R} = \frac{\hat{\delta}_{\xi}^{2}}{W} \qquad 4.15$$

If the iteration of chain converges, then the value of potential reduction factor approaches to one. Theoretically, if the value is below 1.2, it implies the convergence. Subsequently, asymptotic confidence intervals can be calculated because this indicator is subject to estimation error. For this purpose, the 97.5th percentile is used as a conservative point estimate.

In the context of the present study, it is executed one chain for each parameter. The element of B has not been studied in this case. It will be examined the convergence of chains by using the following functions:

$$\xi(\psi) = \alpha_0,\ \xi(\psi) = \alpha_1,\ \xi(\psi) = \beta \qquad 4.16$$

According to the procedure Gelman and Rubin (1992) to test the convergence of chains, it does not lead to the rejection of the convergence if the value of $\hat{R}$ is equal to 0.9998. The 97.5th percentile values of $\hat{R}$ belong to the interval [0.997, 1.002]. Therefore, it was possible to draw parameters from the joint posterior distribution.[12]

The estimated results of posterior statistics (RAM) and the point estimates (DE) are presented in Table 4.1. These estimated results by different approaches are approximately the same. However, there is a difference in the estimated value of the GARCH term. It is because the ML estimation method is point estimate, and DE is an algorithm that estimates the value of the parameter at the global maxima and has a built-in safeguard to avoid being trapped in local maxima. On the other hand, the Bayesian approach takes the average of multiple peaks, i.e., takes information from all maxima in the likelihood function. Therefore, there is a difference in the estimated value of the parameter of the GARCH (1, 1) model. Also, AIC and BIC values are approximately the same, implying the same level of goodness of fit for both models. Economically the sum of alpha and beta (0.538) is less than one, implying moderate persistence in PSX.

ML estimation method assumes asymptotic normality for the estimated parameters; this assumption is valid under regularity conditions, which do not hold for the

[12] This procedure of Gelman and Rubin (1992), is computed in MS Excel.

GARCH model with boundary constraints. Also, if the confidence interval is constructed for these parameters, they will be symmetric by definition. However, if the actual distribution of the estimated parameter is skewed in either direction; ML estimation method either over or under-estimates the actual parameters. To analyze this issue, Kernel density plots are constructed by using the chains of RAM, as shown in Fig. 4.3. These plots explain the reason why the assumption of asymptotically normal distribution does not hold for the estimated parameters. The Bayesian approach does not suffer from this deficiency since the posterior distribution is directly estimated and can be symmetric or otherwise. Kernel density estimate of the posterior show that there is no normality for α_0 and β. However, for α_1 there is normality. Statistically, the values of skewness for these parameters are 1.99, 0.37 and 2.69, respectively. These skewness statistics are statistically significant at 1% significance level. Therefore, the ML estimation method has a propensity to undervalue the right boundary of the 95% confidence interval of these parameters. Hence, these results indicate that the assumption of normality does not hold for the estimated parameters through ML estimation method, even in case of a large set of observations, i.e. 696.

Table 4.1: Empirical results of GARCH (1, 1) with a normal distribution, for the series of PSX

	ML	**Bayesian**	
	Differential Evolution (Point Estimate)	Robust Adaptive Metropolis (Posterior Mean)	Corrected Robust Adaptive Metropolis (Posterior Mode)
$\boldsymbol{\alpha_0}$	0.0001***	0.0001	0.0001
$\boldsymbol{\alpha_1}$	0.3970**	0.3974	0.3950
$\boldsymbol{\beta}$	0.1413**	0.1286	0.1242
AIC	-4095.53	-4093.52	
BIC	-4081.90	-4079.88	

$***$, $**$, and $*$ indicate the significances at 1%, 5%, and 10% levels, respectively.

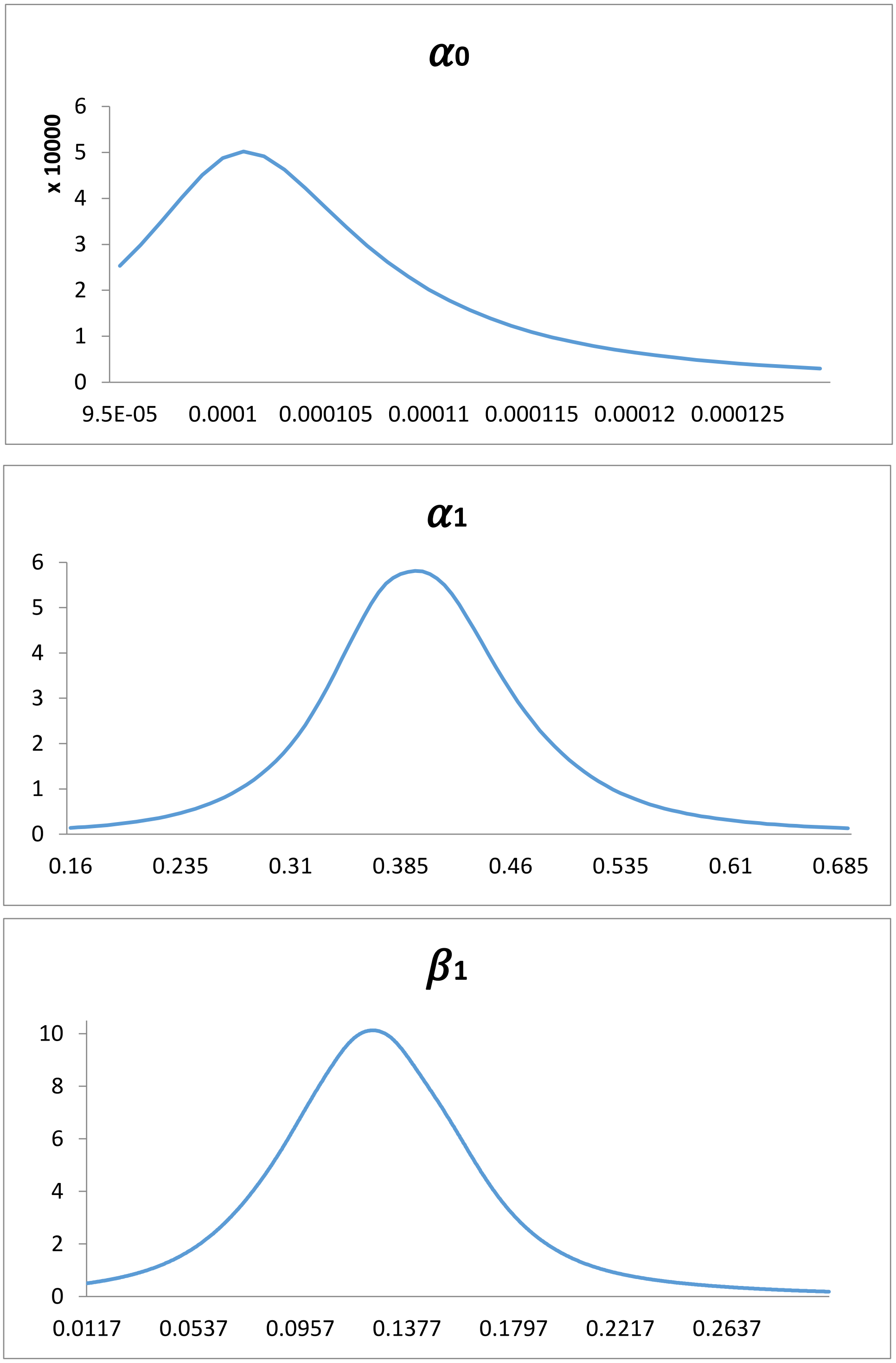


Figure 4.3: Kernel density function is constructed for the parameter of GARCH (1, 1) model, by using the chains of RAM

The posterior mean statistics are presented in Table 4.1 by using the chains of RAM. When the distribution is not symmetric, then the mean statistics could be misleading. One of the reasons can be illustrated in Fig. 4.2. It shows the chains are converging to a stationary point, after many iterations. Mean statistics of the posterior as a point estimate is valid if the distribution of the chain is normal. As the chain is adaptive, i.e., the current draw of the chain takes information from the previous draws, due to which it takes time to converge at a certain level. It is also possible that if an inconsistent value of prior draw is chosen, even then it might take some steps to converge. This type of convergence in the chain from one step to the other could make the distribution skewed or multimodal, as shown in Fig. 4.3. Therefore, the mean statistic of the posterior distribution could lead to misleading inferences, and hence, implications.

To avoid these issues by constructing the kernel densities of chains, as shown in Fig. 4.3, and by choosing the point with maximum frequency or mode as posterior statistics. These mode statistics of the posterior density are presented in the last column of Table 4.1. Moreover, it is named as Corrected RAM.[13] Apparently, in this case, there is no significant difference between the mean and mode statistic of posterior, but in later chapters, it will become apparent. For example, in the case of student-t distribution, there is a significant difference in estimated parameters and Corrected RAM.

[13] Histogram could be used, to pick the point with maximum frequency. However, there are number of issue with histogram: bin size, couple of bins with same frequency etc. For detail, to analyze issues in histogram, and how they are resolve in kernel density see, https://sites.google.com/site/introstats4muslims/textbook/density

4.4. Estimation of GARCH (1, 1) – Student-t distribution

> *"This development (i.e., the* student-t distribution) *permits a distinction between conditional heteroscedasticity and conditional leptokurtic distribution, either of which could account for the observed unconditional kurtosis in the data."*
>
> *-Tim Bollerslev*

The theory of finance is based on the individual behavior of assessment of risk. To measure risk, variance and covariance are calculated. However, depending upon the distribution of the return series, the variance is not valid statistics. The distributional properties are essential to get accurate results and have many significant implications for several financial models. Mandelbert (1963) discussed the idea of leptokurtic and heavy-tailed empirical distribution for stock prices. The distribution of stock prices is unimodal but more peaked than the normal distribution, and there is a possibility of outliers in the return series of stock prices, which make the tails of the distribution extraordinary long relative to the normal distribution. Also, the return series are uncorrelated, but they are not independent, i.e., volatility clustering. Later on, Fama (1965) discussed this idea in detail. Statistically, the variance of the return series is not constant because of the volatility clustering, whereas many standard statistical techniques are based on the assumption of constant variance. Engle (1982) introduced the ARCH model with a normal distribution. This simple model of ARCH has been generalized by Bollerslev (1986) who replaced normal distribution with student-t distribution by Bollerslev (1987).

4.5. Methodology

4.5.1. Frequentist GARCH (1, 1) – Student-t distribution

$$y_t = \varepsilon_t h_t^{1/2}$$

$$\varepsilon_t \sim i.i.d\ \mathcal{N}(0,1)$$

$$h_t^{1/2} \sim t\ (0\ ,1\ ,v)$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} \qquad 4.17$$

Where v is the degree of freedom for student-t distribution. In the case of student-t distribution, Geweke (1993) formulated a model via data argument. Under this model, the likelihood function can be defined as a vector $y = (y_1, \ldots, y_T)'$, $w = (w_1, \ldots, w_T)'$ and $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)'$. This study is regrouped the model parameters into the vector $\psi = (\boldsymbol{\alpha}, \beta, v)$. Then, upon defining the $T \times T$ diagonal matrix

$$\Sigma = \Sigma(\psi, w) = diag\left(\left\{w_t \frac{v-2}{v} h_t(\boldsymbol{\alpha}, \beta)\right\}_{t=1}^{T}\right), \qquad 4.18$$

Where,

$$h_t(\boldsymbol{\alpha}, \beta) = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1}(\boldsymbol{\alpha}, \beta) \qquad 4.19$$

It can be expressed the likelihood function of (ψ, w) as:

$$lnL(\psi, w|\,y) \propto (det\Sigma)^{-\frac{1}{2}} \exp\left[-\frac{1}{2} y' \Sigma y\right] \qquad 4.20$$

4.5.2. Bayesian GARCH (1, 1) – Student-t distribution

The Bayesian approach considers parameters as a random variable which is characterized by a prior density denoted by $p(\psi, w)$. Hyperparameters are used to quantify the prior and are initially assumed to be known and constant. Then, by using the Bayes rule, the likelihood is coupled with prior density to get the posterior density $p(\psi, w\,|y)$ as follows:

$$p(\psi, w\,|y) = \frac{\mathcal{L}(\psi,w\,|y)p(\psi,w)}{\int \mathcal{L}(\psi,w\,|y)p(\psi,w)d\psi dw} \qquad 4.21$$

This posterior is a numerical and probabilistic description of the knowledge about the model parameters after observing the data. This study used normal priors on the GARCH parameters $\boldsymbol{\alpha}$ and β

$$p(\boldsymbol{\alpha}) \propto \emptyset\, N_2(\boldsymbol{\alpha}|u_{\boldsymbol{\alpha}}\,, \Sigma_{\boldsymbol{\alpha}})\; 1\{\boldsymbol{\alpha} \in R_+^2\}$$

$$p(\beta) \propto \emptyset\, N_1\left(\beta|u_\beta\,, \Sigma_\beta\right)\; 1\{\beta \epsilon\, R_+^2\} \qquad 4.22$$

Where $u.$ and $\Sigma_.$ are the hyperparameters. $1\{.\}$ is the indicator function and $\emptyset\, N_d$ is the d-dimensional normal density.

The prior distribution of vector w conditional on v is found by noting the components. w_t are independent and identically distributed from the inverted gamma density, which yields:

$$p(w|v) = \left(\frac{v}{2}\right)^{\frac{Tv}{2}} \left[\gamma\frac{v}{2}\right]^{-T} \left(\prod_{t=1}^{T} w_t\right)^{-\frac{v}{2}-1} exp\left[-\frac{1}{2}\sum_{i=1}^{T}\frac{v}{w_t}\right] \qquad 4.23$$

4.6. Empirical Analysis

4.6.1. Empirical Results Based on GARCH (1, 1) – Student-t Distribution

It is estimated the log-return series of PSX with the parsimonious GARCH (1, 1). In this section, it has been assumed that error follows the student-t distribution. The joint prior is constructed by assuming that the parameters of the model are independent of each other. One chain is executed with a length of 10000 passes. In this section, student-t distribution is assumed instead of the normal distribution, while all other criteria remain the same for comparison.

After the estimation of the posterior statistics, the first important step is to test the convergence of chains. An informal approach to analyze the convergence of chains is to construct the running mean plot, as presented in Fig. 4.4. Plots of running means are showing slow convergence for all parameters. According to the formal testing procedure defined by Gelman and Rubin (1992) to test the convergence of chain, it does not lead to the rejection of the convergence, with the value of $\hat{R}$ equal to 0.9999. The 97.5th percentile values of $\hat{R}$ belong to the interval [0.95, 1.05]. Therefore, it was possible to draw parameters from the joint posterior distribution.

Table 4.2 presents the empirical results of PSX by assuming a student-t distribution for the GARCH (1, 1) model. The point estimates of the ML estimation method and mean of the posterior statistics are approximately similar, except α_1. First, the ML estimation method assumes the symmetric distribution around the point estimate. It is not true in the case of the Bayesian approach. Fig. 4.5 presents the kernel densities function for each parameter. These kernel densities are constructed by using the chains. Each of these plots implies skewed distribution for each parameter. Statistically, the values of skewness are 2.05, 0.40, 0.51, and 1.17, respectively. All the skewness statistics are significant at 1% significance level. Hence, the ML estimation method has a propensity to undervalue the right boundary of the 95% confidence interval of these parameters. Therefore, these results indicate that the assumption of asymptotic normality does not hold, even in case of a large set of observations.

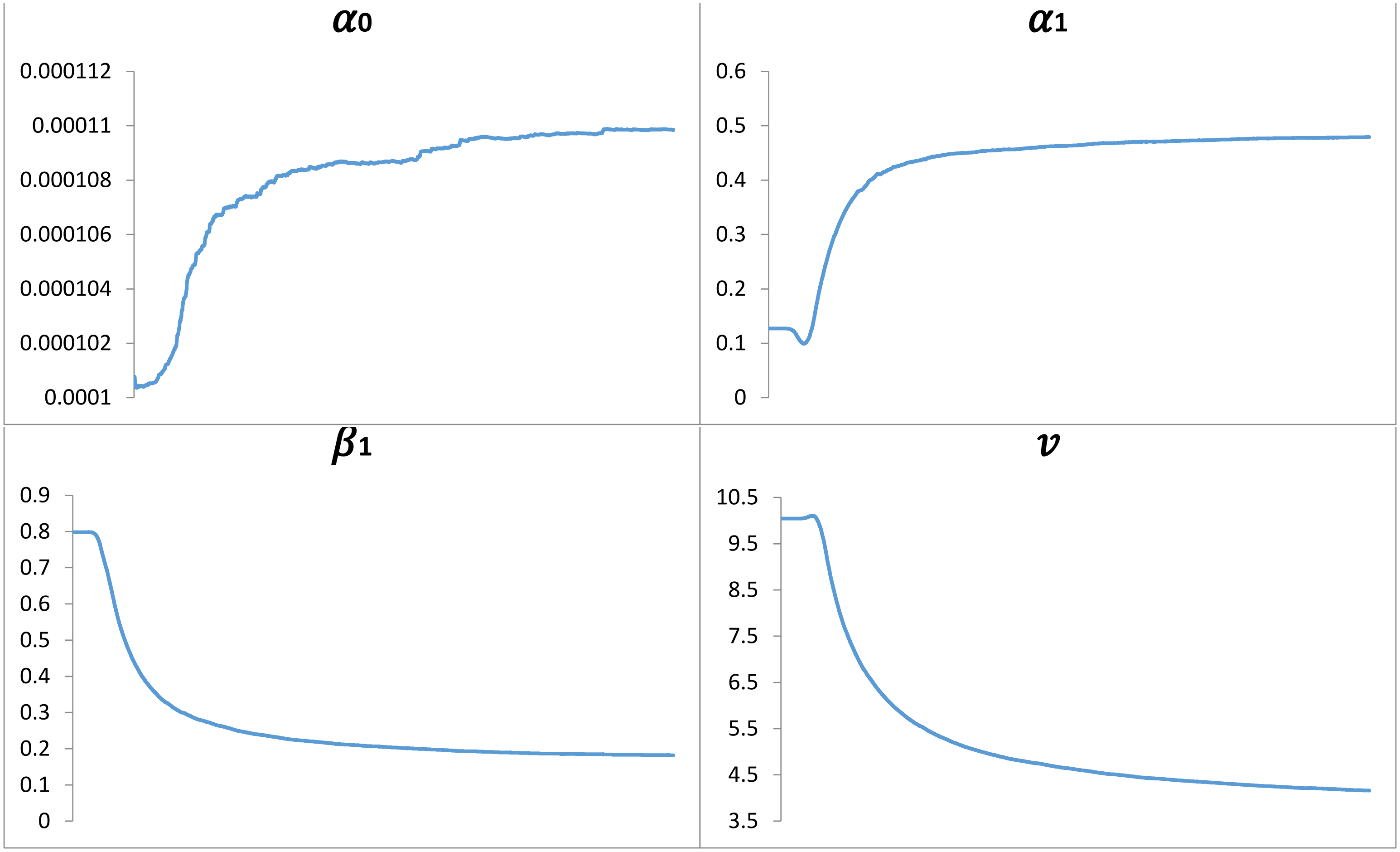


Figure 4.4: The running mean of chains for 10000 iterations, showing convergence to the stationary point

Table 4.2: Empirical results of GARCH (1, 1) with student-t distribution, for the series of PSX

	ML	**Bayesian**	
	Differential Evolution (Point Estimates)	Robust Adaptive Metropolis (Posterior Mean)	Corrected Robust Adaptive Metropolis (Posterior Mode)
$\boldsymbol{\alpha_0}$	0.0001***	0.0001	0.0001
$\boldsymbol{\alpha_1}$	0.4640**	0.4795	0.4650
$\boldsymbol{\beta}$	0.1889**	0.1818	0.1334
$\boldsymbol{v}$	3.6787**	4.1591	3.4700
AIC	-4166.05	-4159.52	
BIC	-4147.87	-4141.34	

$***$, $**$, and $*$ indicate the significances at 1%, 5%, and 10% levels, respectively.

The chains, along with the distribution of these chains, are presented in Fig. 4.5. The trace plots show that plot chains are converging to a new level after about 1000 iterations. Because of this convergence, the distribution of these parameters shows multi-modality. Therefore, in this case, the mean statistics of the posterior statistics are not useful and hence can lead to misleading inferences. To avoid this issue, the kernel densities are constructed, as shown in Fig. 4.6. The bins were chosen on the bases of modal frequency. These values are presented in Table 4.2 under the heading of Corrected RAM. These statistics are significantly different from other listed models.

In the comparison of empirical results of a normal distribution with student-t distribution, a significant difference is found the empirical results across these distributions. First, on the basics of model significance, i.e., AIC and SBC, a model with student-t distribution is better. Second, in the case of student-t distribution, the convergences of chains need a longer length.

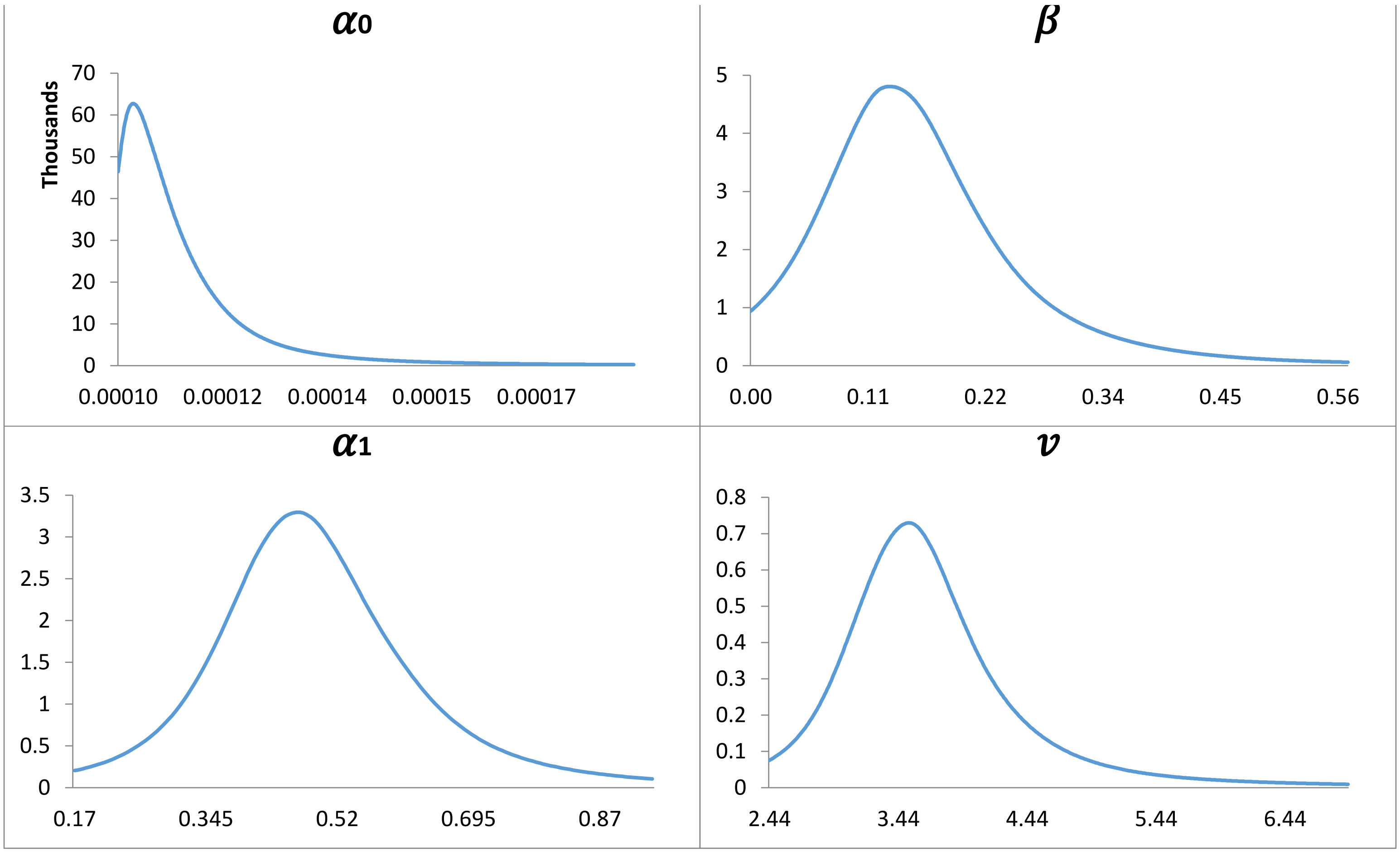


Figure 4.5: Kernel density function, which is constructed by using the chains of the RAM. On the x-axis points, there are different points. Each point contains relative frequency on y-axis. These kernel density function help to know about the skewness of parameter, and at the point with the highest frequency.

4.7. Chapter Summary

PSX is used as a real-world application to analyze the significance of the difference in inference and prediction, in the presence of multimodality. The estimates of the DE are significantly different from BFGS. It is because of multimodality in the likelihood function of PSX, and BFGS often trap into the local maxima, due to the single starting value. While DE uses multiple starting values with various chains, that automatically avoid local maxima and converge to global maxima.

Furthermore, the fundamental objective of the GARCH type model is to forecast. The accuracy of the forecast depends upon the standard error of the estimated parameters. The empirical results of PSX imply that the standard error is quite precise in the case of DE. Therefore the forecasting based on these estimated parameters will be more accurate, as compared to BFGS.

Based on the frequentist approach, estimated parameters are point estimates, and the distribution of these estimated parameters is hypothetically assumed to be normally distributed. While the posterior distribution in a Bayesian approach implies that the distribution of the estimated parameter is skewed. If the distribution of the estimated parameter is skewed, then the frequentist approach either over/underestimate the estimated parameters. Therefore, the estimated parameter could be biased in either direction, hence inference and predictions.

CHAPTER 5

Asymmetric GARCH Type Models and Forecasting

Simple GARCH model assumes the symmetric effect of past information on the current volatility, i.e., good and bad news in the financial market effect current volatility with the same magnitude. However, in the case of an assets price, it has been observed that the bad news appears to have a more intensive effect of volatility than good news (Francq et al., 2019). For many stocks, there is a strong negative correlation between the current return and future volatility. In the financial markets, this phenomenon is known as leverage or asymmetric effect. The idea of the leverage effect is captured in Fig.. 5.1. The dashed line represents the effect of information on the volatility is symmetric, i.e., the same magnitude on both sides around the volatility. Conversely, the solid line represents the effect of information on volatility is asymmetric; i.e., the magnitude of volatility on the negative side is higher than the positive side.

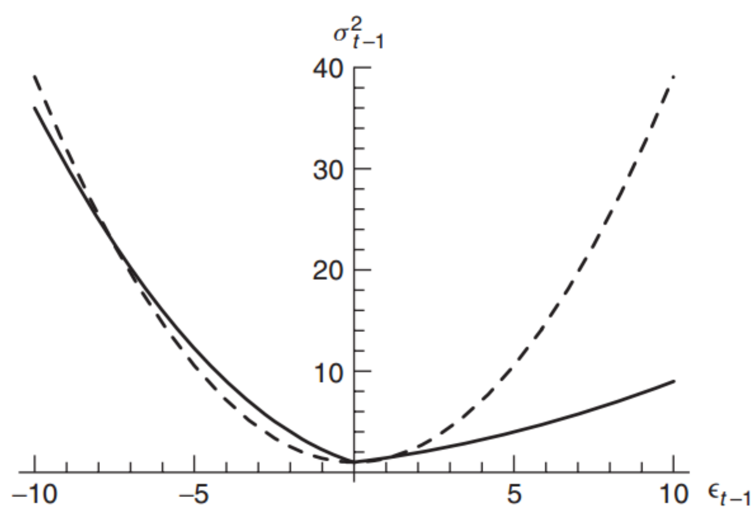


Figure 5.1: News impact curves for symmetric (dash line) and asymmetric (solid line) GARCH type models. The x-axis represents information, and the y-axis represents volatility.

Mathematically, the symmetric GARCH model assumes that there is no covariance between the current value of volatility and past information, i.e.

$$cov(\sigma_t, \varepsilon_{t-i}) = 0 \qquad 5.1.1$$

Where $i > 0$, because of the current value of σ_t volatility is a function of ε_{t-i}. By dividing the random shocks, ε_t into positive and negative shocks,

$$\varepsilon_t^+ = \max(\varepsilon_t, 0), \ \varepsilon_t^- = \min(\varepsilon_t, 0) \qquad 5.1.2$$

It can easily be understood that symmetric assumption only holds, when

$$cov(\varepsilon_t^+, \varepsilon_{t-i}) = cov(\varepsilon_t^-, \varepsilon_{t-i}) = 0 \qquad 5.1.3$$

This assumption of symmetric autocovariance can easily be tested empirically. In the case of financial time series, this assumption of autocovariance is violated most of the time.

$$cov(\varepsilon_t^+, \varepsilon_{t-i}) < cov(\varepsilon_t^-, \varepsilon_{t-i}) \qquad 5.1.4$$

It can also be written as:

$$cov(\sigma_t, \varepsilon_{t-i}^+) < cov(\sigma_t, \varepsilon_{t-i}^-) \qquad 5.1.5$$

This last equation can be interpreted as there is a high impact on the decrease in past prices on the current volatility as compared to the increase in recent prices. In the field of finance, this phenomenon is known as the leverage effect. Volatility tends to increase dramatically following the bad news and increase moderately following the good news.

The purpose of this chapter is to test the leverage effect for the same window of PSX and then extend the empirical model in the presence of the asymmetric model. The last section concluded and confirmed the choice of algorithms in the presence of multimodality in the likelihood function. Therefore in this chapter, the selected algorithm will be used for the estimation of asymmetric GARCH type models.

This chapter is divided into two sections; the first section of the chapter is based on the selection of asymmetric GARCH model, after testing the presence of leverage effect for the window of PSX statistically. The second section of the chapter extends the selected asymmetric GARCH model with the skewed student t distribution. Then finally, this section is based on the fundamental aim of the GARCH type modelling, i.e., forecasting. Forecasting analysis is based on Value at Risk, Expected Short Fall, and Predictive density.

5.1. ASYMMETRIC MODELS

In literature, there are three asymmetric GARCH type models. These are as follow;

5.1.1. Exponential GARCH

Let ε_t is an i.i.d sequence, such that $E(\varepsilon_t) = 0$ and $var\ (\varepsilon_t) = 1$, then ε_t is said to be exponential GARCH (EGARCH), if it satisfies an equation of the form:

$$\log h_t^2 = \omega + \sum_{i=1}^{q} \alpha_i\ g(\varepsilon_{t-i}) + \sum_{j=1}^{p} \beta_j\ log\ h_{t-j}^2 \qquad 5.1.6$$

where,

$$g(\varepsilon_{t-i}) = \theta\varepsilon_{t-i} + \varsigma(|\varepsilon_{t-i}| - E|\varepsilon_{t-i}|) \qquad 5.1.7$$

where ω, α_i,β_i, $\theta\ and\ \varsigma$ are real numbers.

Following are the features of the EGARCH model:

1. The standard GARCH model imposes a necessary condition of positivity on the estimated parameter, which is not a requisition under these specifications.
2. Innovations of the large modulus should increase volatility. This entail constraint on the coefficients: for instance, if $log\sigma_t^2 = \omega + \theta h_{t-1} + \varsigma(|h_{t-1}| - E|h_{t-1}|)$, σ_t^2 increases with $|h_{t-1}|$, the sign of h_{t-1} being fixed, if and only if $-\varsigma < \theta < \varsigma$. In the general case it suffices to impose:

$$-\varsigma < \theta < \varsigma \, , \alpha_i > 0 \, , \beta_j > 0 \qquad 5.1.8$$

3. The asymmetry property is taken into account through the coefficient θ. For instance, let $\theta < 0$ and $log\sigma_t^2 = \omega + \theta$: if $h_{t-1} < 0$, the variable $log\sigma_t^2$ will be larger than its mean ω, and it will be smaller if $\varepsilon_{t-1} > 0$. Thus, obtain the common asymmetry property of financial time series.

5.1.2. Threshold GARCH models

A natural way to introduce asymmetry is to specify the conditional variance as a function of the positive and negative parts of the past innovations.

$$\varepsilon_t^+ = \max(\varepsilon_t, 0)\,,\ \varepsilon_t^- = \min(\varepsilon_t, 0) \qquad 5.1.9$$

and $\varepsilon_t = \varepsilon_t^+ + \varepsilon_t^-$. The threshold GARCH (TGARCH) class of model introduces a threshold effect into the volatility.

Let h_t is an i.i.d series of random variables, such that $E(h_t) = 0$ and $var\ (h_t) = 1$, then ε_t is said to be TGARCH (p, q) if it satisfies the equation of the form:

$$\sigma_t = \omega + \sum_{i=1}^{q}(\alpha_{i,+}\varepsilon_t^+ - \alpha_{i,-}\varepsilon_t^-) + \sum_{j=1}^{p}\beta_j\sigma_{t-1} \qquad 5.1.10$$

Where ω,$\alpha_{i,+}$, $\alpha_{i,-}$ and β_i are real numbers. Moreover, the following are the constraints on these parameters:

$$\omega > 0,\ \alpha_{i,+} \geq 0\ ,\ \alpha_{i,-} \geq 0,\ \beta_i \geq 0 \qquad 5.1.11$$

The variable σ_t is strictly positive and represents the conditional standard deviation of ε_t. In general, the conditional standard deviation of the $\varepsilon_t$is $|\sigma_t|$: imposing the positivity of σ_t is not required.

5.1.3. Glosten, Jagannathan, and Runkle – GARCH

Glosten, Jagannathan, and Runkle (1993) showed how to allow good and bad news to have different effects on volatility. In a sense, $\varepsilon_{t-1} = 0$ is a threshold such that shocks greater than the threshold have different effects than shocks below the threshold. Consider the threshold GARCH process as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_{i,}\varepsilon_{t-1}^2 + \gamma_i\, d_{t-1}\varepsilon_{t-1}^2 + \sum_{j=1}^{p} \beta_j\sigma_{t-1} \qquad 5.1.12$$

Where d_{t-1} is a dummy variable that is equal to one if $\varepsilon_{t-1} < 0$ and is equal to zero if $\varepsilon_{t-1} \geq 0$. Then intuition behind the TARCH model is that positive values of ε_{t-1} are associated with a zero value of d_{t-1}. Hence, if $\varepsilon_{t-1} \geq 0$, the effect of an ε_{t-1} shock on h_t is $(\alpha_1 + \gamma_1)\varepsilon_{t-1}^2$. If $\gamma_1 > 0$, adverse shocks will have a more significant effect on volatility than positive shocks. The dummy variable d_t moreover, the product $d_{t-1}\varepsilon_{t-1}^2$ can easily be created. If the coefficient γ_1 is statistically different from zero; it can be concluded that the data contains a threshold effect.

The primary difference between E-GARCH and T-GARCH is simple to understand. In the case of the GJR model, if $d_{t-1} = 0$, then model reduces to $h_t = \omega +$

$\alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$. This implies that if past information has a positive impact then the value of h_t depends upon $\omega, \alpha_1, \beta_1$. However, if $d_{t-1} = 1$, then $h_t = \omega + (\alpha_1 + \gamma_1)\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$. In this case, past information has a negative impact and the value of h_t depends upon an additional parameter γ_1. Also, the value of γ_1 is also be positive which ultimately creates a jump in the variable h_t. In an expression of EGARCH is a *log-linear form,* which increases or decreases exponentially as the dashed line shown in Fig. 5.1. Therefore, before applying the proper methodology, it is necessary to check either the leverage effect is of threshold or exponential nature.

5.1.4. Diagnostics for Leverage Effects

1. A specific diagnostic test to determine whether there are any leverage effects in residuals.

   a. Estimate ARCH or GARCH model, calculate standardized residual as:

$$s_t = \frac{\hat{\varepsilon}_t}{\hat{h}_t^{\frac{1}{2}}} \qquad 5.1.13$$

   b. Estimate the following regression:

$$s_t^2 = \alpha_0 + \alpha_1 s_{t-1} + \alpha_2 s_{t-2} + \alpha_3 s_{t-3} + \cdots \qquad 5.1.14$$

   c. If there is no leverage effect the standardized residual, the squared residual should be uncorrelated. By applying F statistics with the null hypothesis $\alpha_1 = \alpha_2 = \alpha_3 = \cdots$, if the null is significantly rejected, it implies that there is evidence for leverage effect.

2. Another test was developed by Engle and Ng (1993).

   a. Let d_{t-1} is a dummy variable,

$$d_{t-1} = 1, if \quad \hat{\varepsilon}_{t-1} < 0 \qquad 5.1.15$$

$$d_{t-1} = 0, \; if \quad \hat{\varepsilon}_{t-1} > 0 \qquad 5.1.16$$

b. The test is to determine whether the estimated squared residuals can be predicted using the dummy sequence.

$$s_t^2 = \alpha_0 + \alpha_1 d_{t-1} + \varepsilon_{1t} \qquad 5.1.17$$

c. If α_1 is statistically significant, then the current period shock helps predict the conditional volatility.

d. General regression can be estimated as:

$$s_t^2 = \alpha_0 + \alpha_1 d_{t-1} + \alpha_2 d_{t-1} s_{t-1} + \alpha_3 (1 - d_{t-1}) s_{t-1} + \varepsilon_{1t} \qquad 5.1.18$$

e. The presence of $d_{t-1}s_{t-1}$ moreover, $(1 - d_{t-1})s_{t-1}$ is designed to determine whether the effects of positive and negative shocks also depend on their size. F statistic is used to test the significance level for the null hypothesis $\alpha_1 = \alpha_2 = \alpha_3 = 0$. After the confirmation of the leverage effect in series, apply asymmetric GARCH type models.

5.1.5. Confirmation of Leverage Effect

Before testing the empirical asymmetric model, the first step is to test the existence of leverage effect. For this purpose, this study applied the following tests. First, a simple test is applied based on Eq. 5.1.3. Empirically it is found that the $-8.7E - 06 \; < 1.35E - 05$, which can be interpreted as a high impact on the of the decrease in past prices on the current volatility as compared to the increase in recent prices. In other words, this test confirms the presence of leverage effect for the window of PSX.

Two formal tests are also applied for the identification of leverage effect. The first formal method is based on Eq. 5.1.14. Empirically, two lags of residual are introduced into this model. Empirical results are presented in Table 5.1. First lags of the standardized residual are statistically significant and different from zero, while the second lag is statistically insignificant. The null hypothesis $\alpha_1 = \alpha_2 = 0$ is used to test the significance of both lags jointly. The calculated F value is statistically significant at the 5% significance level. It confirms the existence of the leverage in the return series of PSX.

Finally, this study also applied the procedure of Engle and Ng (1993). Empirical results of Eq. 5.1.18 are presented in Table 5.2. The estimated parameters are statistically significant at the 5% significance level, except α_3. The null hypothesis $\alpha_1 = \alpha_2 = \alpha_3 = 0$ is used to test the joint significance. The null hypothesis of joint significance is also rejected in this case. It implies the existence of a leverage effect for the window of PSX. Three types of test are used to check the presence of leverage effect, and all these statistics confirm the presence of leverage effect in the series.

Table 5.1: Leverage effect is tested by regressing the squared standardized residual on the lag of the residual. A probability value of F statistic is less than 5%, which indicate the presence leverage effect in PSX series

	Coefficient	SE	t-value	P-value
$\boldsymbol{\alpha_0}$	1.04695***	0.08552	12.2	0.0000
$\boldsymbol{\alpha_1}$	-0.23112***	0.08534	-2.71	0.0069
$\boldsymbol{\alpha_2}$	-0.10532	0.08537	-1.23	0.2177
F(2,690)	4.6198**			

$***$, $**$, and $*$ indicate the significances at 1%, 5%, and 10% levels, respectively.

Table 5.2: by using Engle and Ng (1993) procedure, to test the presence of leverage effect in the series. A probability value of F statistic is less than 5%, which indicate the presence leverage effect in PSX series

	Coefficient	SE	t-value	P-value
α_0	0.9020***	0.1090	8.28	0.0000
α_1	0.2950*	0.1719	1.72	0.0866
α_2	-22.9386**	9.0460	-2.54	0.0114
α_3	-10.0526	8.3380	-1.21	0.2284
F(3,689)	3.5517**			

$***$, $**$, and $*$ indicate the significances at 1%, 5%, and 10% levels, respectively.

5.1.6. Empirical Results for Asymmetric Models

In chapter 4, it was concluded that student t distribution is more appropriate for the estimation of the GARCH (1, 1) model as compared to the normal distribution. Therefore, in this chapter, the student t distribution is assumed of the prior density. Also, prior independence between the model's parameters is assumed and executed one chain with the length of 10,000 passes. Lastly, the model is estimated through the ML estimation method for comparison between the Bayesian and the frequentist approach.

Estimated results for asymmetric models are presented in Table 5.3. At first, the empirical results are compared with each other, on the bases of AIC and BIC. On the comparison, it is found that the absolute value of AIC and BIC are higher for the EGARCH model, as compare to GJR and TGARCH, in case of DE estimation method. However, based on AIC and BIC apparently, it seems that the TGARCH in case of Bayesian is also a good approximation. However, this point is not consistent. Because, in the case of the Bayesian approach, the Acceptance rate and the procedure of Gelman and Rubin (1992) implies that the EGARCH is a good approximation for the window of PSX. Hence, the

Bayesian and the frequentist approach, both implies that the EGARCH model is a preferable model, as compared to other asymmetric models.

Now at the second stage, the empirical results of this chapter are compared with the previous chapter on the bases of AIC and BIC in case of DE, and in cases of Bayesian approach, it is compared on the bases of acceptance rate and confidence interval of Gelman and Rubin (1992) procedure. The value of AIC and BIC in the previous chapter is -4166.05 and -4147.87 with is quite low as compared to the 4199 and 4177 in this chapter, respectively. Similarly, in the case of the Bayesian approach, it implies that the EGARCH model is a good approximation for the window of PSX, as compared to simple GARCH. According to the formal procedure of Gelman and Rubin (1992), $\hat{R}$ is 0.99 for EGARCH models that do not lead to the rejection of the chain. The 97.5th percentile values of $\hat{R}$ belong to different confidence intervals. The confidence intervals are precise for EGARCH, while in other cases, the upper value is closer to 1.2; consequently, it indicates weak convergence. The Acceptance Rate is higher than other models, and the intervals are more consistent in comparison to other models as well. Running means plots are presented in appendix B. Running means are showing slow convergence, especially for the GJR parameters. Also, kernel densities are also skewed for the GJR.

Table 5.3: Results for the asymmetric model based on three different econometric models

	EGARCH			GJR			TGARCH		
	Differential Evolution	Robust Adaptive Metropolis	Corrected Robust Adaptive Metropolis	Differential Evolution	Robust Adaptive Metropolis	Corrected Robust Adaptive Metropolis	Differential Evolution	Robust Adaptive Metropolis	Corrected Robust Adaptive Metropolis
$\boldsymbol{\alpha_0}$	-1.094***	-1.355	-1.240	0.000***	0.000	0.000	0.002***	0.002	0.002
$\boldsymbol{\alpha_1}$	0.382**	0.411	0.400	**0.000**	0.164	0.157	**0.000**	0.131	0.124
$\boldsymbol{\alpha_2}$	-0.145**	-0.160	-0.155	0.954	0.654	0.741	0.354*	0.394	0.375
$\boldsymbol{\beta}$	0.873*	0.843	0.855	0.339	0.201	0.171	0.675*	0.641	0.658
$\boldsymbol{v}$	5.196**	6.534	5.100	3.313**	5.494	3.470	5.111**	6.571	5.278
AIC	**-4199**	-4197		-4174	-4163		-4191	-4199	
BIC	**-4177**	-4175		-4151	-4140		-4168	-4176	
Acceptance Rate		**0.992**			0.971			0.989	
Gelman and Rubin (1992)		[0.89, 1.11]			[0.86, 1.14]			[0.87, 1.13]	

∗∗∗, ∗∗, and ∗ indicate the significances at 1%, 5%, and 10% levels, respectively.

5.2. SKEWED Student-t DISTRIBUTION EGARCH

It is a well-known fact that returns from financial variables follow the non-normal distribution. The empirical distribution of such returns is leptokurtic and has more massive tails than the normal distribution. This implies that considerable changes in return occur with a higher frequency than under normality. Moreover, it is often skewed having one heavy and one semi-heavy or more Gaussian-like tail. This phenomenon of a skewed distribution is explained by prospect theory (Barberis et al., 2016). One set of distributions for modelling skewed and heavy-tailed data is the skew Student's t-distribution (Aas & Haff, 2006; Adcock et al., 2015). Hansen (1994) was the first to propose a skewed extension to the Student's t-distribution for modelling financial returns. The choice of distribution for the empirical models is essential because it helps to fulfil the criteria of parsimonious and reliable inference and prediction as well.

In the first section of chapter 5, the asymmetric model is selected for the window of PSX, i.e., EGARCH. In this section, the selected model, i.e., EGARCH will be extended with skewed student t distribution, and then forecasting will be analyzed based on the final selected model. While all other models will be not considered in this section, because if they are not estimating well the data generating process of PSX, ultimately the forecast will be biased on the bases of biased estimates. Furthermore, the empirical results of the previous chapter show that the convergence of the chains is quite slow. In the current section, extend the EGARCH model with the skewed student-t distribution. For forecasting purpose, three different measure will be used, i.e., Value at Risk, Expected Shortfall and Predictive density.

5.2.1. Methodology

Consider the simple EGARCH model,

$$y_t = \varepsilon_t h_t^{1/2}$$

$$\varepsilon_t \sim i.i.d\ \mathcal{N}(0,1)$$

$$h_t \sim f_r(t) = \frac{2}{r+\frac{1}{r}}\left\{f\left(\frac{t}{r}\right)1_{[0,\infty)}(t) + f(rt)1_{(-\infty,0)}(t)\right\} \qquad 5.2.1$$

Here f_r is symmetric student-t distribution if $r = 1$, and the same model is approached, which has already been explained at the start of the chapter. f_r is skewed student-t distribution towards left if $r < 1$, otherwise it is right skewed (see Klar et al., 2012). Let ε_t is an iid sequence, such that $E(\varepsilon_t) = 0$ and $var\ (\varepsilon_t) = 1$, then ε_t is said to be exponential GARCH (EGARCH), if it satisfies the equation of the form:

$$\log h_t^2 = \omega + \sum_{i=1}^{q} \alpha_i\ g(\varepsilon_{t-i}) + \sum_{j=1}^{p} \beta_j\ log\ h_{t-j}^2 \qquad 5.2.2$$

where,

$$g(\varepsilon_{t-i}) = \theta\varepsilon_{t-i} + \varsigma(|\varepsilon_{t-i}| - E|\varepsilon_{t-i}|) \qquad 5.2.3$$

where, $\omega, \alpha_i, \beta_i,\ \theta\ and\ \varsigma$ are real numbers.

5.2.2. Empirical Results Based on Skewed Student-t Distribution

The empirical result of EGARCH with skewed student t distribution is presented in Table 5.2.1. In comparison with the empirical results of the previous section, it is found that these results are more consistent. For the model comparison, AIC and BIC are used. The absolute value of AIC and BIC are higher in the present section as compare to the previous section. In the previous section, the value of AIC and BIC are 4199 and 4177, are lower as compare the value 4226.35 and 4199.08 of the current section.

Furthermore, all the parameters are significant. It implies a relatively good approximation for the window of PSX. Moreover, there is somehow a difference in the value of estimated parameters, across Bayesian and frequentist approach. It is because of the DE estimation method estimate value of the parameter at the global maxima while the Bayesian approach takes an average of all the peak in the likelihood function.

Running means are presented in Appendix C. Running means showed fast convergence to the steady state point after about 2000 points. This convergence of the parameter is relatively faster as compared to the previous models because of the use of the most appropriate estimated model along with the choice of distribution. According to the formal procedure of Gelman and Rubin (1992) for the convergence of chains, it does not lead to the rejection of the chain when the value of $\hat{R}$ is 0.9999. The confidence interval with 97.5th percentile values of $\hat{R}$ belong to the interval [0.98, 1.01]. These confidence intervals are much precise as compared to all other previous models. Also, the chain for the student-t distribution is converging to the stable point, and the issue is solved as compared to the previous model.

Table 5.4: Empirical Results of EGARCH Model by Assuming Skewed Student-t Distribution

	Differential Evolution (ML)	Robust Adaptive Metropolis (Posterior Mean)	Corrected Robust Adaptive Metropolis (Posterior Mode)
$\boldsymbol{\omega}$	-1.085***	-1.363	-1.270
$\boldsymbol{\alpha_1}$	0.411**	0.451	0.435
$\boldsymbol{\alpha_2}$	-0.135**	-0.155	-0.148
$\boldsymbol{\beta_1}$	0.869**	0.836	0.850
$\boldsymbol{\theta}$	4.064**	4.212	4.100
$\boldsymbol{\varsigma}$	0.764**	0.772	0.768
AIC:	-4226.35	-4225.54	
BIC:	-4199.08	-4198.27	

$***$, $**$, and $*$ indicate the significances at 1%, 5%, and 10% levels, respectively.

5.2.3. Value at Risk and Decision Theory

"Density forecasting is fast becoming an important tool for decision makers in situations where loss functions are asymmetric, and forecast errors follow non-Gaussian distributions."

Allan Timmermann

Value at risk (VaR) is an essential tool in risk management to allocate capital for a particular asset. VaR is easy to implement and understand. It is a measure of expected loss for an asset. It gives information, (i) For a given period and (ii) for a given confidence level, p. The expected loss which increases with the probability $p^e = (1 - p)$. From a statistical point of view, it is the percentile of the profit and loss distribution over a fixed horizon. In the real world application, if the distribution of profit and loss is unknown, VaR can only be estimated from sample data. In case of the normal distribution, VaR is interpreted as one day ahead forecast; VaR is a given percentile of the standard normal distribution scaled by the conditional standard deviation.[14]

VaR density varies from model to model.[15] Therefore, the EGARCH model with the skewed student-t distribution, which is verified at the start of this section. VaR is presented in Fig. 5.2.1, for frequentist (upper plot) and Bayesian approach (low plot). In both approaches, the VaR density seems alike because the estimated parameters of EGARCH models are approximately the same. Furthermore, one day ahead forecast value of VaR is presented in Table 5.2.2. Following two points is infer from these results; (i) one day ahead forecast value is higher for DE than RAM (in absolute term), which implies there is a higher level of risk in DE. (ii) Also, if the level of risk is increased from 95% to 99%, the value of DE increases (in absolute term) as compared to the RAM.

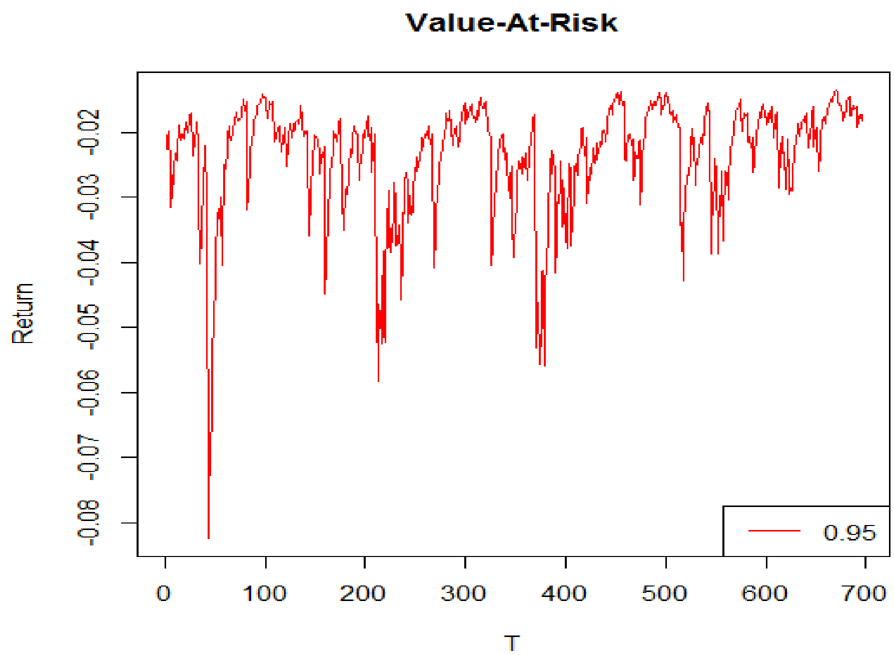
5.2.4. The Expected Short Fall Risk Measure

The VaR is a standard tool to measure the risk in the field of financial risk management. However, VaR has been criticized in the literature for majorly two reasons, in particular:

a. The VaR does not provide information about the potential size of the loss that exceeds its level, and as a result, it is flawed;
b. The VaR is not a comprehensible measure of risk. In particular, it lacks the property of sub-additivity.

---

[14] For detail, see Ardia (2008).
[15] For detail, see Guidolin and Timmermann (2006).



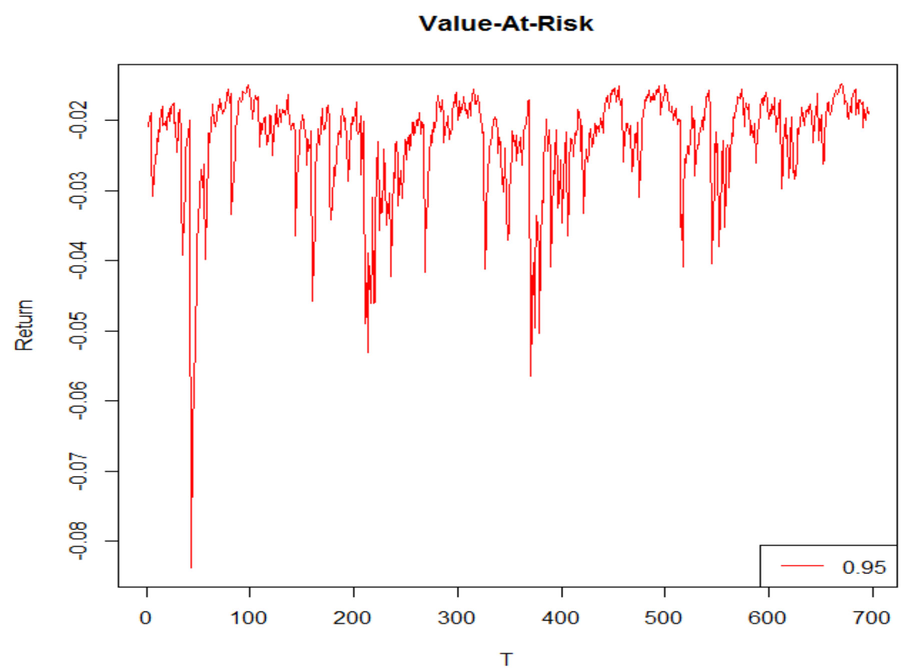


Figure 5.2: Value at risk (VaR) estimated by using DE (upper part), and VaR estimated by using RAM (lower part).

The concept of Expected Shortfall (ES) has been introduced To overcome the issues mentioned above. The ES risk measure is the expectation of the profit and loss below the VaR level. Estimated ES is presented in Table 5.5. In comparison with the results of VaR, the estimated parameters for ES are lower than VaR, as expected. Furthermore, the estimated parameters for ES in the Bayesian approach are lower than the frequentist approach, which implies that the Bayesian approach has a higher chance to minimize the risk of profit and loss.

Table 5.5: Estimated value at risk (VaR) and expected shortfall (ES)

	Value at Risk		**Expected Short Fall**	
	0.95	0.99	0.95	0.99
Differential Evolution	-0.0228	-0.0426	-0.0359	-0.0606
Robust Adaptive Metropolis	-0.0203	-0.0377	-0.0434	-0.0772

5.2.5. Predictive Density

One of the essential tools to evaluate consistency is predictive distribution. The density of predictive distribution is similar to the prior density except that the prior is replaced by the posterior. If the time series experiment is repeated, it will yield the predictive density for the outcome of the repeated experiment.[16] The purpose of this predictive testing is not about rejection and non-rejection of models; instead, it just gives information about the new model (Box, 1980).

[16] For detail, see J. Geweke and Whiteman. (2006).

The predictive density of DE and RAM, respectively, is presented in Fig. 5.2.5. Both predictive densities are left-skewed. These left-skewed distributions give information about the new model for PSX that it is an uncertain choice. According to prospect theory, if the probability of the occurrence at the left tail is low, it would lead the investor to the worst situation. In short, in terms of investment choice, PSX is a risky asset.

On the other hand, this predictive density would not give information about the rejection or non-rejection of the model, because both predictive densities are the same. There is a significant difference in the left tail. Left tail of RAM is more massive than the left tail of DE. It implies that the predictive density in case of RAM gives more weight to the unlikely events. Offering more probability to the unlikely events will ultimately minimize the risk of the loss. These results are consistent with VaR and ES.

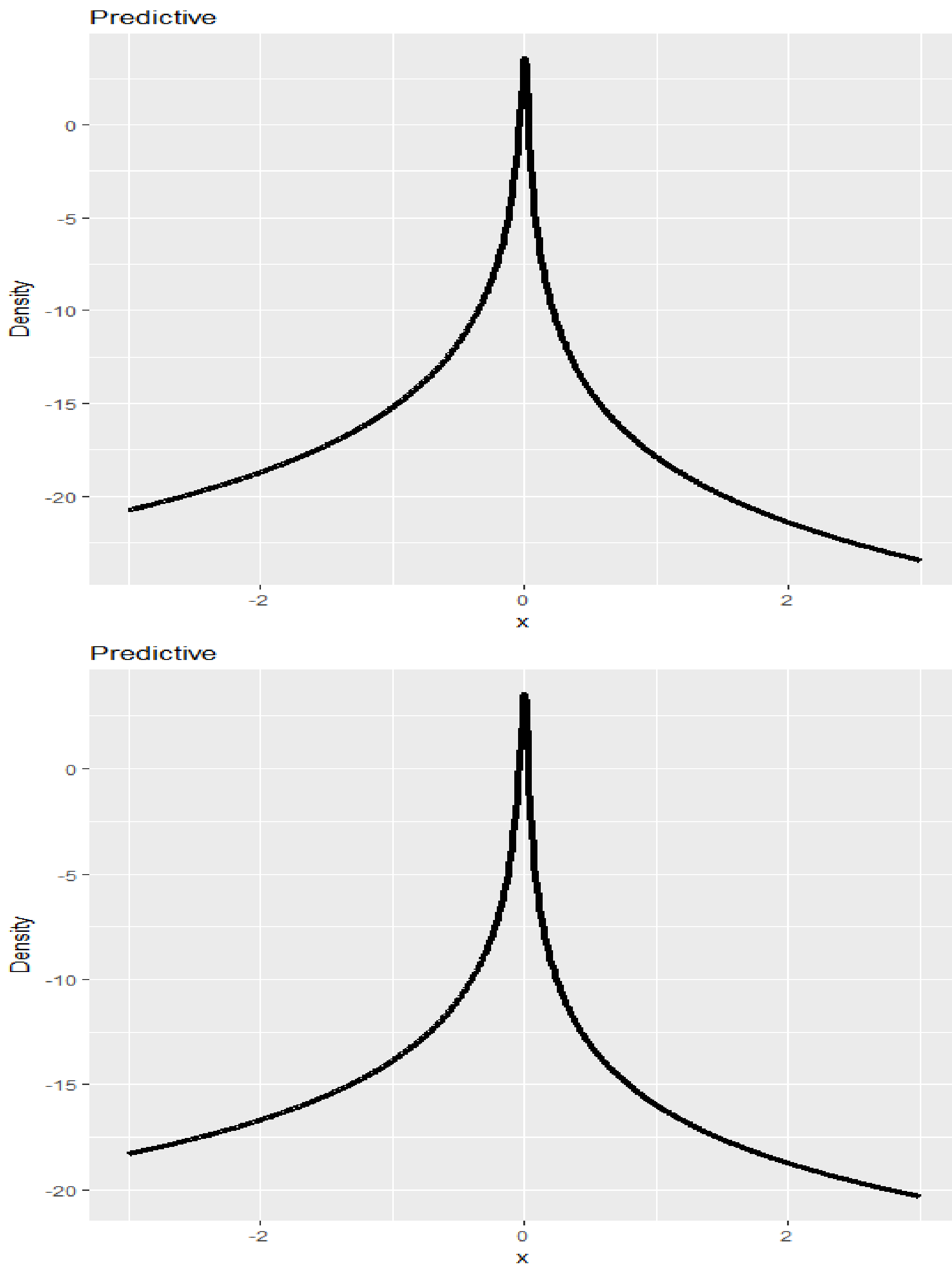


Figure 5.3: predictive densities for DE (upper graph), and robust adaptive metropolis (lower graph). Both predictive densities are left-skewed. In case of the RAM, the left tail is more massive than DE.

5.3. Chapter Summary

The primary purpose of this chapter is to study the issue of forecasting. Therefore, in the first section of this chapter, this study has estimated three asymmetric models for series of PSX, and then these estimated models are compared with each other to select the best model from three. The essential purpose of this comparison to select that appropriate model, which help in forecasting because a biased selected model will give an explicitly biased forecast. Moreover, then in the second chapter, this empirical model is extended with the skewed student t distribution. This finally selected empirical model is then used for forecasting purpose.

It is the essential requirement of the asymmetric GARCH type models, to test leverage effect statistically. Empirically, it is found that the series of PSX follow the E-GARCH model, i.e., which implies that the stock market is not perfect and leads to exponential change into the mean level of PSX. Furthermore, in the case of any negative shock in the stock market, will smoothly converge to the new level rather any sudden shift.

The final section of this chapter is about forecasting, which is the primary purpose of GARCH type modeling. For this purpose, Value at Risk, Expected Shortfall, and Predictive density is measured. Empirically it is found that the decision made on the bases of the Bayesian approach is more precise as compared to the frequentist approach. Hence, the Bayesian approach is more valuable to avoid risk, and also the preferable choice of estimation for the risk-averse investor.

CHAPTER 6

CONCLUSION

The fundamental objective of the study is to compare the Frequentist and Bayesian approaches in the presence of multimodality in the likelihood function of GARCH models. For this purpose, the present study has selected two estimation method from each approach. To evaluate the estimation power of each algorithm around the true parameter, this study is employed used following steps; (i) true data generating process, (ii) true data generating process with different distributions, (iii) surface analysis comparison. However, before comparing approaches, it is essential to compare algorithms within each approach. Simulation is used as a standard to evaluate the performance of each algorithm around the true data generating process. The experimental result of simulation implies that DE is preferable estimation method as compared to BFGS, across different data generating process and along the entire surface of the contour plot. The main reasons for this difference between BFGS and DE estimation methods are just because of the multimodality in the likelihood function, and choice of single starting value and multiple starting values in each estimation method, respectively.

Contrary, in the Bayesian approach, RAM is the preferable choice of estimation for GARCH type model, as compared to the MH estimation methods. This preference is consistent not only across different data generating process but also consistent along the entire surface of the contour plot. Because MH estimation method uses the method independent MCMC, where the current draw of the posterior distribution does not take any information from the previous draw, therefore **it** might not converge to a single point. While RAM uses the technique of Adaptive MCMC, where it receives data from the earlier draws of the posterior distribution, and ultimately converge to a stationary point. Therefore, RAM has a lower level of bias as compared to the MH estimation method.

PSX is used as a real-world application to analyze the significance of the difference in inference and prediction, in the presence of multimodality in the likelihood function of the GARCH model. The estimates of the DE are significantly different from BFGS. It is because of multimodality in the likelihood function of PSX, and BFGS often trap into the local maxima, due to the single starting value. While DE uses multiple starting values with various chains, that automatically avoid local maxima and converge to global maxima.

Furthermore, the fundamental objective of the GARCH type model is to forecast. The accuracy of the forecast depends upon the standard error of the estimated parameters. The empirical results of PSX imply that the standard error is quite precise in the case of DE as compare to BFGS. Therefore the forecasting based on the estimated parameters of DE will be more accurate, as compared to BFGS.

Based on the frequentist approach, estimated parameters are point estimates, and the distribution of these estimated parameters is hypothetically assumed to be normally distributed while the posterior distribution in a Bayesian approach implies that the distribution of the estimated parameter is skewed. If the distribution of the estimated parameter is skewed, then the frequentist approach either over or underestimate the estimated parameters. Therefore, the estimated parameter could be biased in either direction, hence inference and predictions.

Empirically, often the Bayesian approach uses "means" of the posterior distribution as a measure of the point estimate. Mean is an appropriate measure if the distribution is normal. However, if the distribution is non-standard, then the mean is not a good approximation. By constructing kernel density, and then selecting the model frequency for the posterior distribution can resolve this issue. It is analyzed that there is a significant difference in the value of the estimated parameters from RAM and selecting the model frequency point. Hence, it is concluded that Corrected-RAM is preferable in comparison to RAM algorithms.

As already mentioned that the main objective of GARCH type modeling is forecasting. Therefore, at first step asymmetric model is selected for the window of PSX. Empirically, it is found that the series of PSX follow the E-GARCH model, i.e., which implies that the stock market is not perfect and leads to exponential change into the mean level of PSX. Furthermore, in the case of any negative shock in the stock market, will smoothly converge to the new level rather any sudden shift.

Finally, the forecasting is done, by applied using; Value at Risk, Expected Shortfall, and Predictive density is measured. Empirically it is found that the decision made on the bases of the Bayesian approach is more precise as compared to the frequentist approach. Hence, the Bayesian approach is more valuable to avoid risk, and also the preferable choice of estimation for the risk-averse investor.

Hence, the Bayesian approach has the following advantage over the frequentist approach; (i) completed the distribution of the estimated parameter is obtained at the low cost of simulation. (ii) Through this distribution of the estimated parameter, the properties of the estimated parameters can easily be discussed. (iii) Therefore, in the case of the frequentist approach, it either over or underestimated the actual value of parameters. (iv) Forecasting based on a Bayesian approach is preferable than the frequentist approach.

RECOMMENDATIONS

This study sets a new base for GARCH type modeling, in the presence of multimodality in the likelihood function. Furthermore, it is proved that literature has this issue as well, so, the complete literature based on financial econometrics can be revised in light of the present study. Moreover, it can be used to analyze the level of significant difference in empirical results of literature and the results based on the methodology of this study.

APPENDIX A

A.1. Normal Distribution:

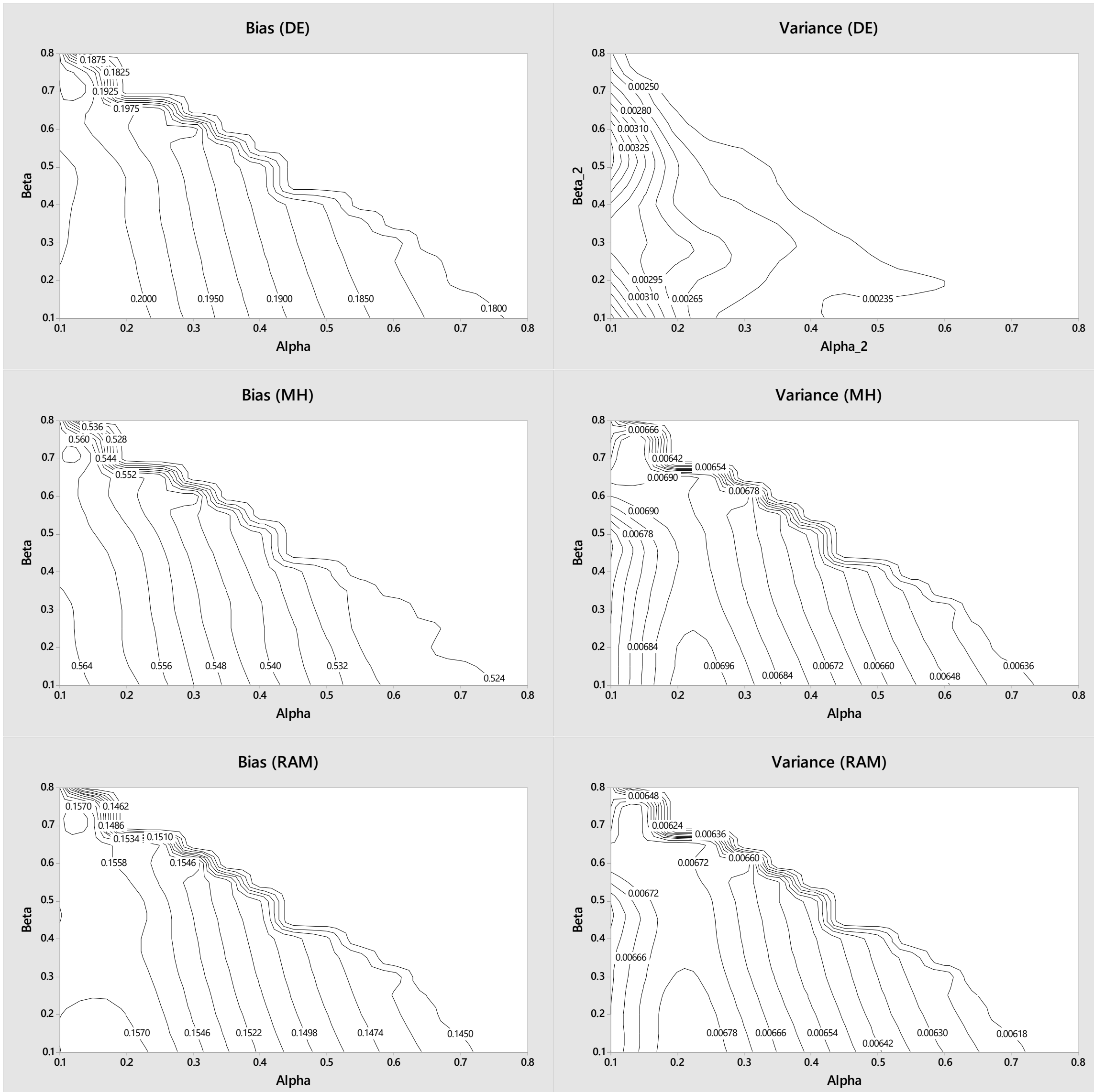


Figure A.1. Contour plots represent the level of bias (left plots) and variance (right plots) for DE, MH, and RAM, for the parameters of GARCH (1, 1) model, respectively. The x-axis represents the change in the value of the Alpha parameter, from 0.1 to 0.9, while the y-axis represents the change in the value of Beta.

A.2. Student-t Distribution:

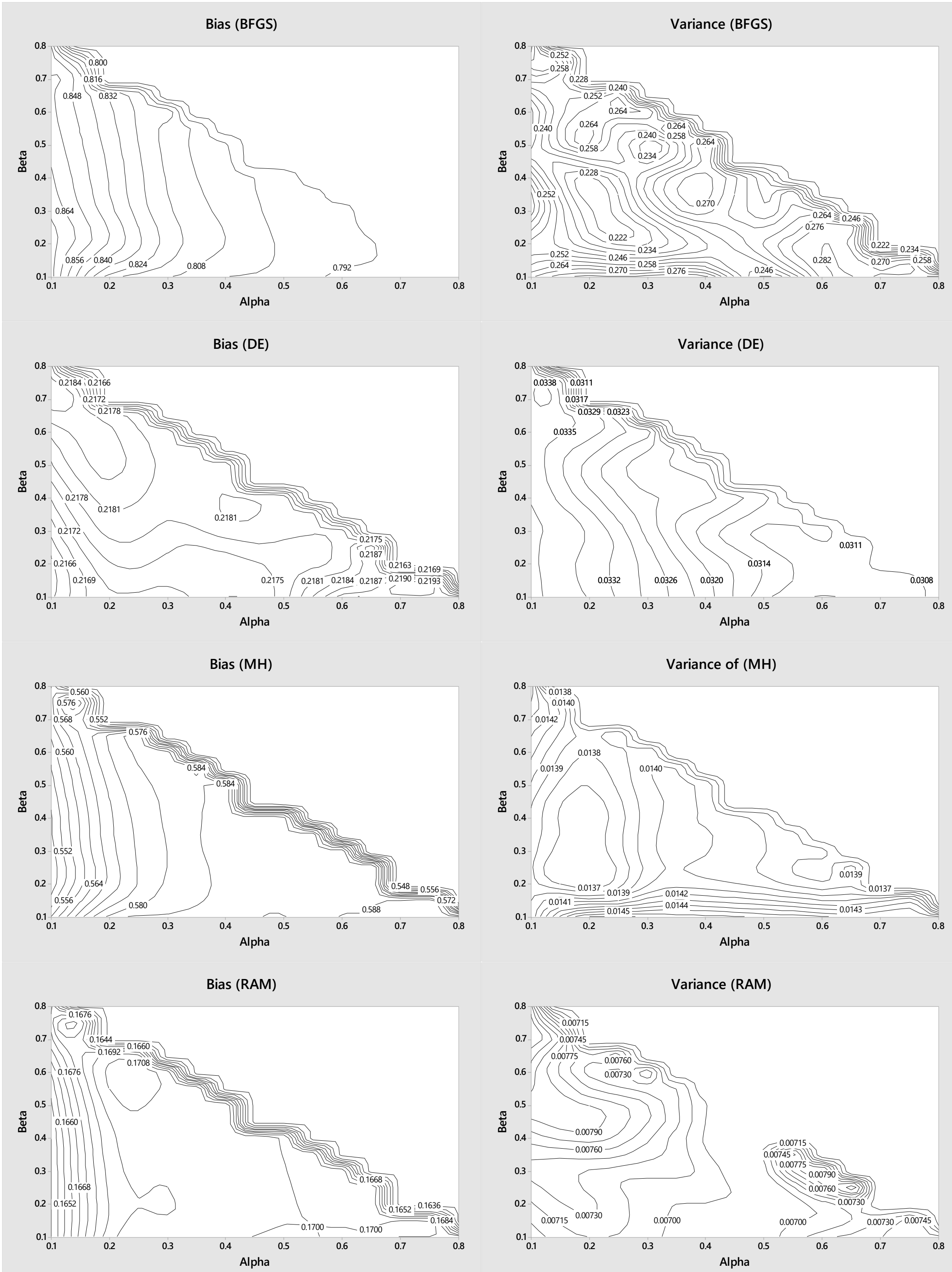
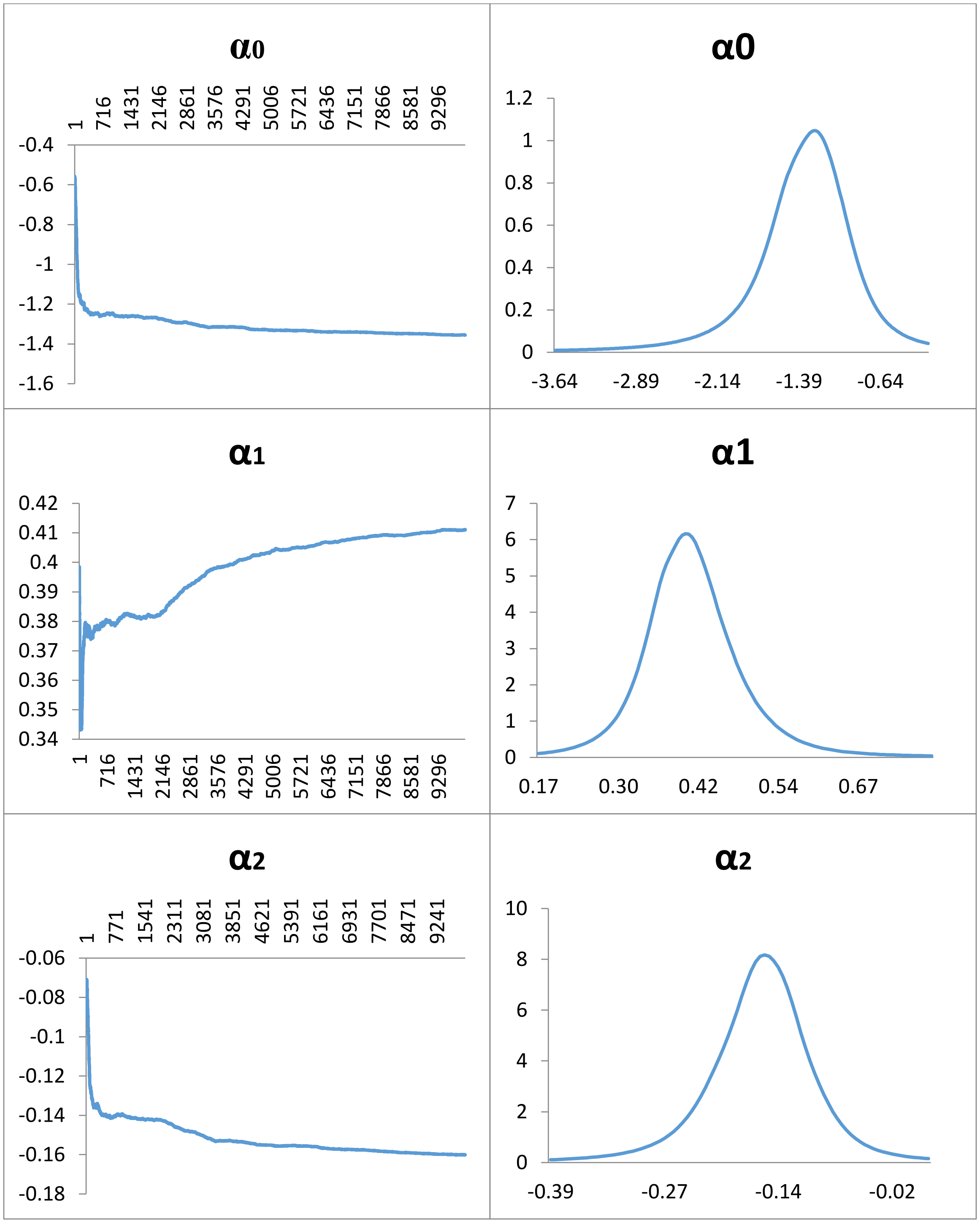


Figure A.2. Contour plots represent the level of bias (left plots) and variance (right plots) for BFGS, DE, MH, and RAM, for the parameters of GARCH (1, 1) model, respectively. The x-axis represents the change in the value of the Alpha parameter, from 0.1 to 0.9, while, the y-axis represents the change in the value of Beta.

Empirically it has been found that the likelihood function of the GARCH model is multimodel. Due to this numerical algorithm trap into local maxima like BFGS (for details, see chapter 3). Whereas DE has built-in safeguards to avoid local maxima and converge to global maxima. Because of this, the level of bias and variance from the true parameter is lower for DE as compared to BFGS. Moreover, this difference becomes more prominent when the distribution of data changes for normal to student-t distribution. Similarly, in the case of the Bayesian approach, RAM was found to be more prominent over MH. Hence, in the case of the Frequentist approach, DE is relatively better to estimate the GARCH model, and in the Bayesian approach, RAM is relatively better than MH.

APPENDIX B

1. E-GARCH



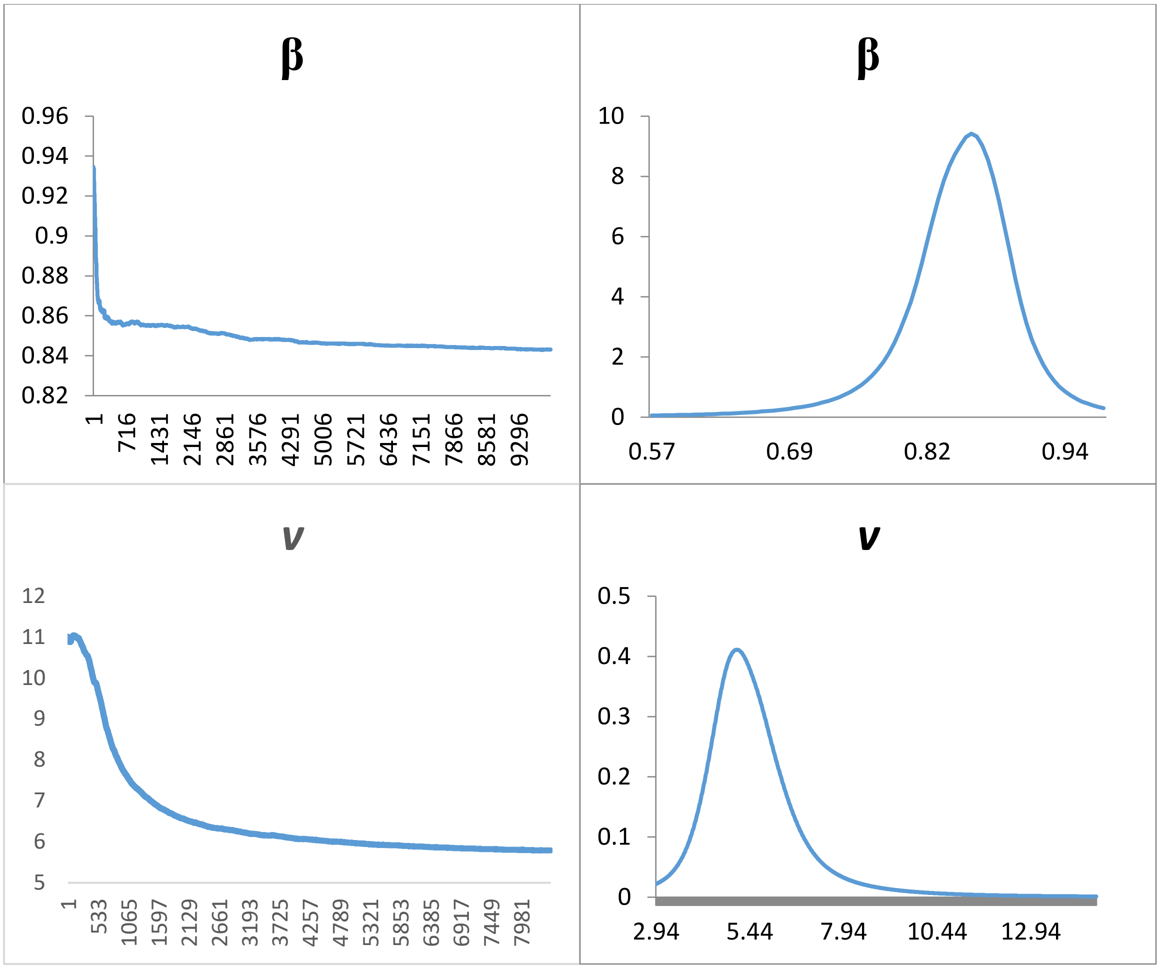
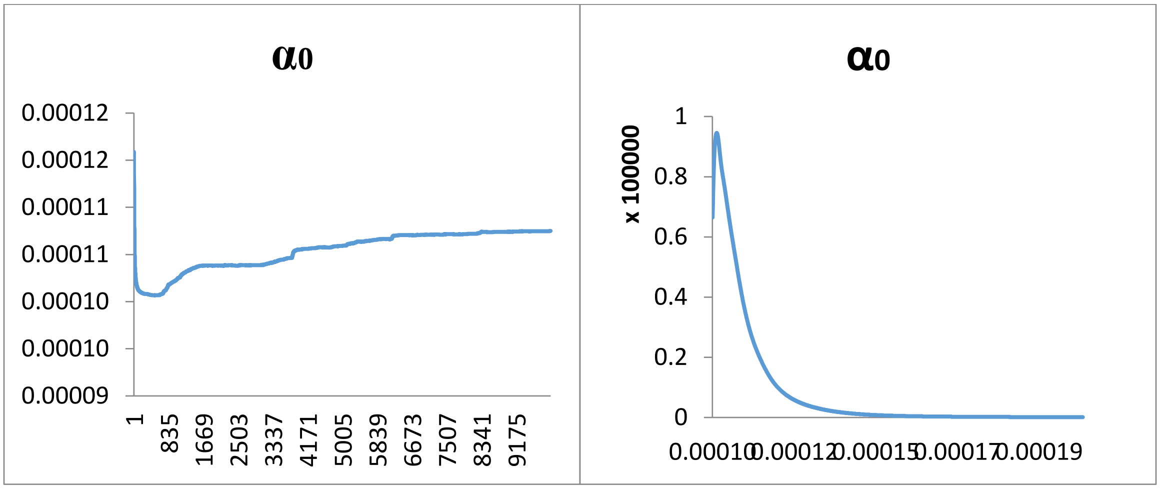
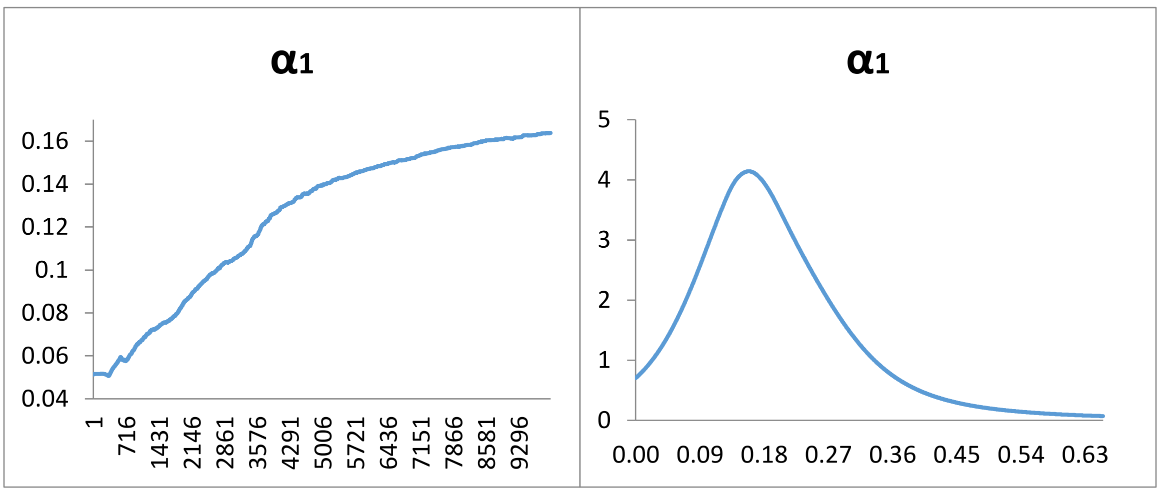
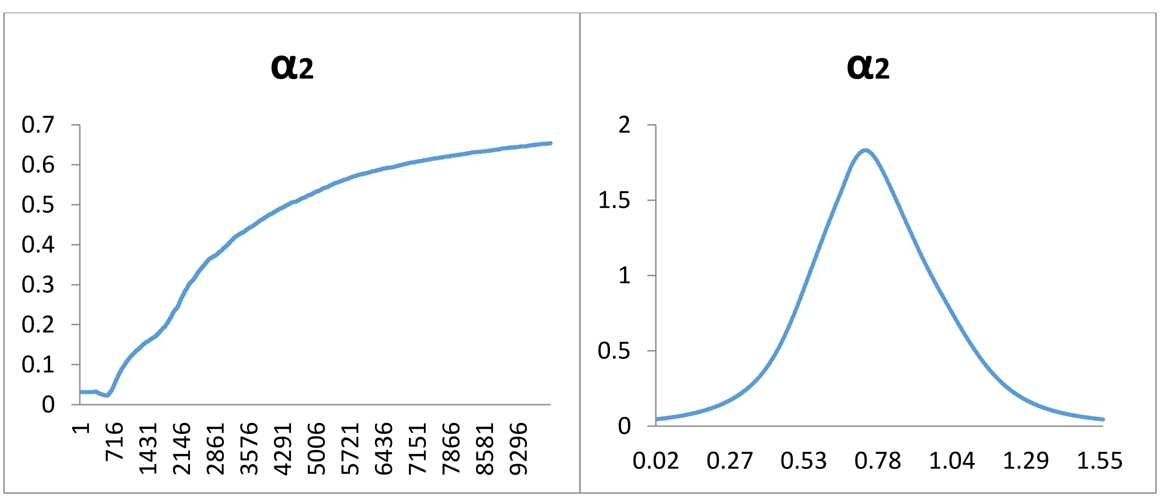


Figure B.1. Running means along with the distribution of each parameter is constructed by using chains of RAM. Running means for each chain are converging to a stationary point. Also, the distribution for each parameter is approximately asymmetric.

2. GJR


α0
0.00012
0.00012
0.00011
0.00011
0.00010
0.00010
0.00009
1
835
1669
2503
3337
4171
5005
5839
6673
7507
8341
9175
α0
x 100000
1
0.8
0.6
0.4
0.2
0
0.00010
0.00012
0.00015
0.00017
0.00019


α1
0.16
0.14
0.12
0.1
0.08
0.06
0.04
1
716
1431
2146
2861
3576
4291
5006
5721
6436
7151
7866
8581
9296
α1
5
4
3
2
1
0
0.00
0.09
0.18
0.27
0.36
0.45
0.54
0.63


α2
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
716
1431
2146
2861
3576
4291
5006
5721
6436
7151
7866
8581
9296
α2
2
1.5
1
0.5
0
0.02
0.27
0.53
0.78
1.04
1.29
1.55

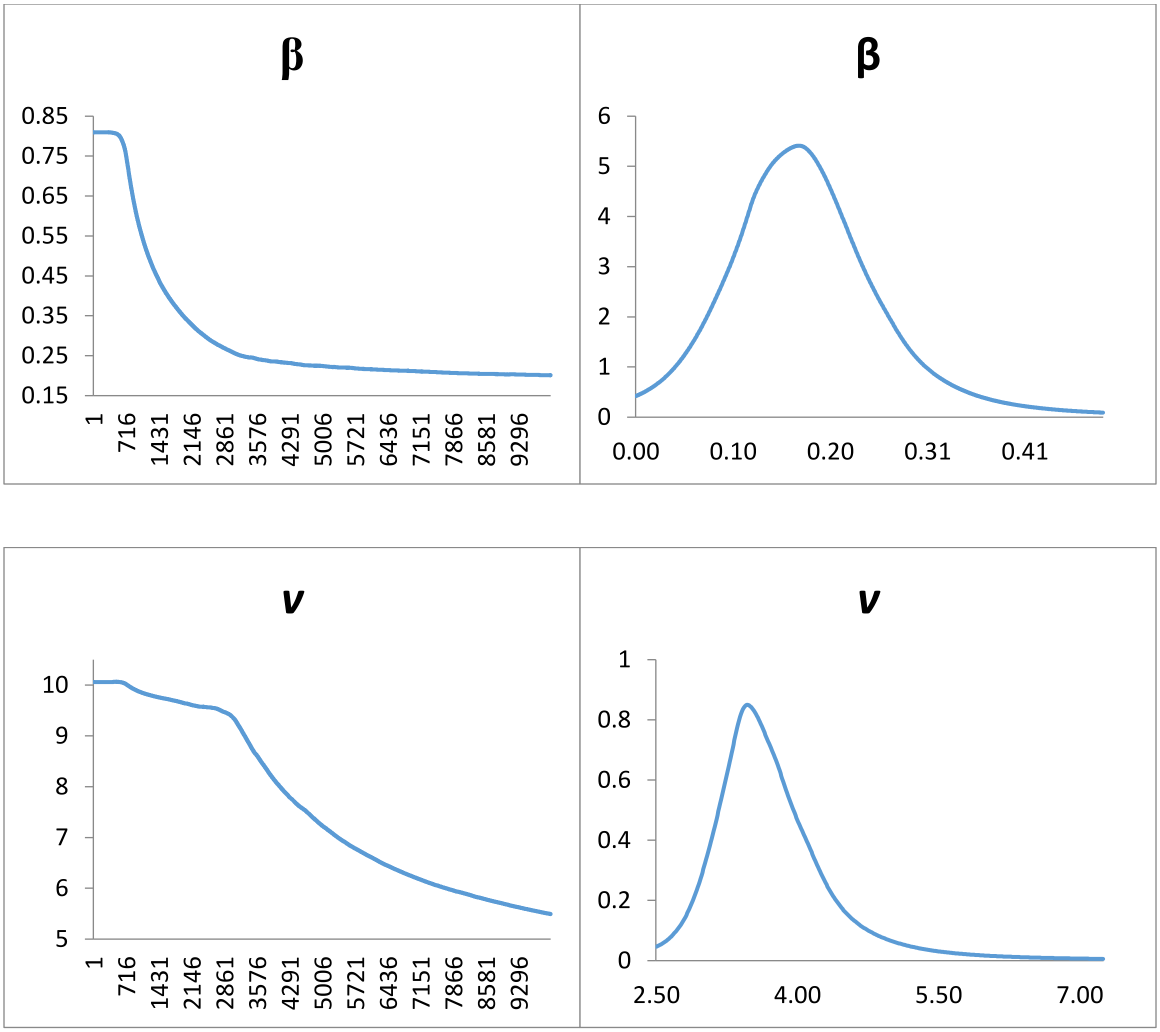
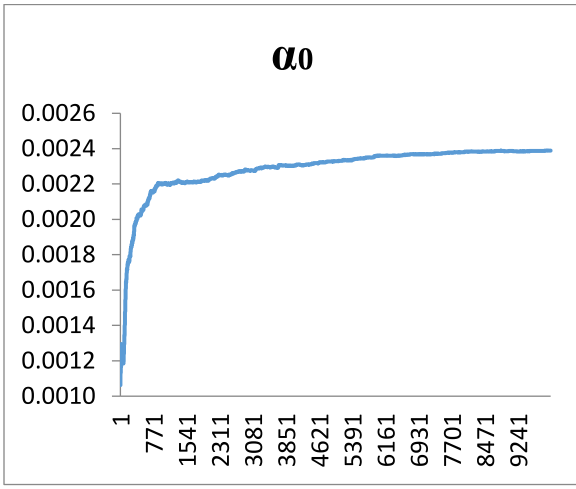
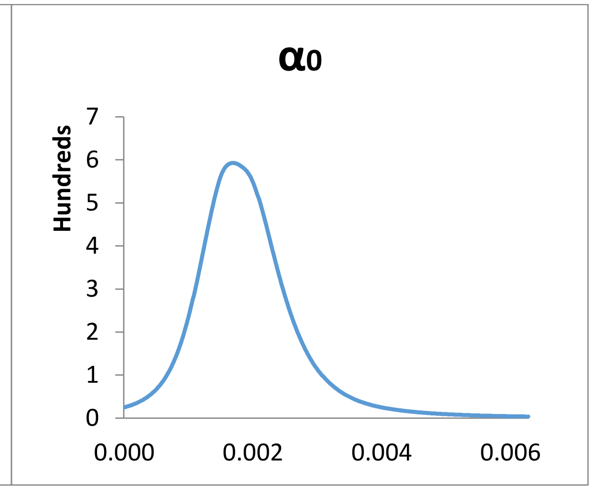
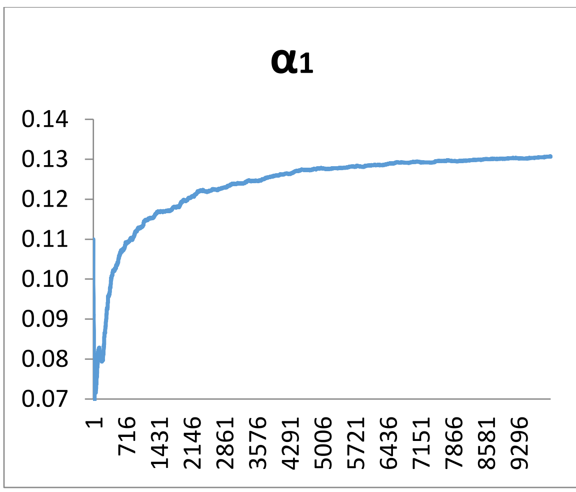


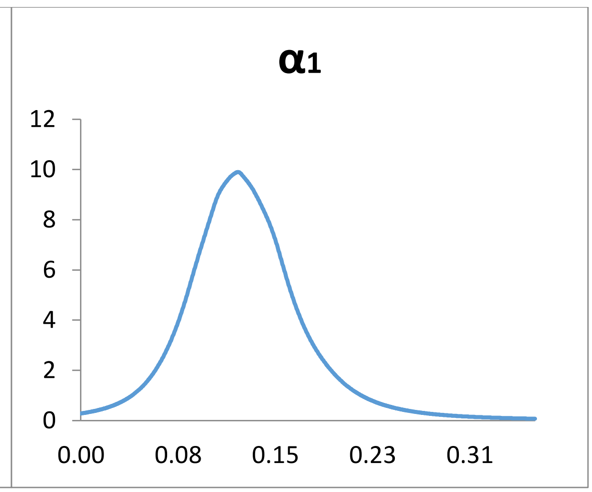
Figure B.2. Running means along with the distribution of each parameter is constructed by using chains of RAM. Running means for each chain is approximately not converging to a stationary point. Also, the distribution for each parameter is approximately asymmetric.

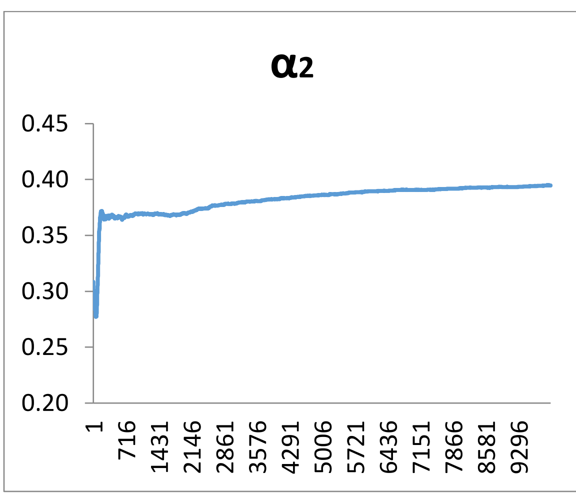
3. TGARCH

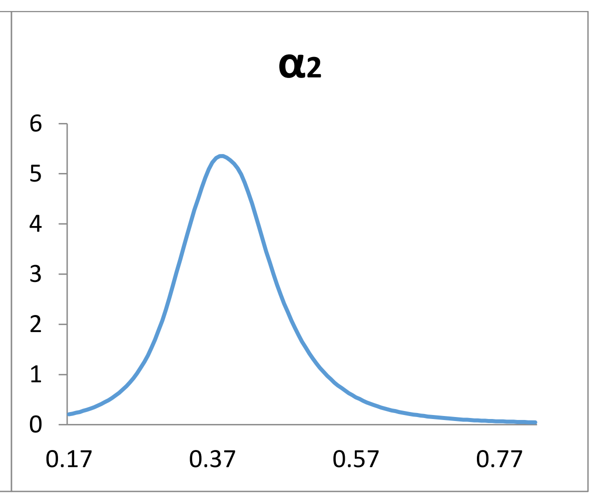

α0
0.0026
0.0024
0.0022
0.0020
0.0018
0.0016
0.0014
0.0012
0.0010
1
771
1541
2311
3081
3851
4621
5391
6161
6931
7701
8471
9241


α0
Hundreds
7
6
5
4
3
2
1
0
0.000
0.002
0.004
0.006


α1
0.14
0.13
0.12
0.11
0.10
0.09
0.08
0.07
1
716
1431
2146
2861
3576
4291
5006
5721
6436
7151
7866
8581
9296


α1
12
10
8
6
4
2
0
0.00
0.08
0.15
0.23
0.31


α2
0.45
0.40
0.35
0.30
0.25
0.20
1
716
1431
2146
2861
3576
4291
5006
5721
6436
7151
7866
8581
9296


α2
6
5
4
3
2
1
0
0.17
0.37
0.57
0.77

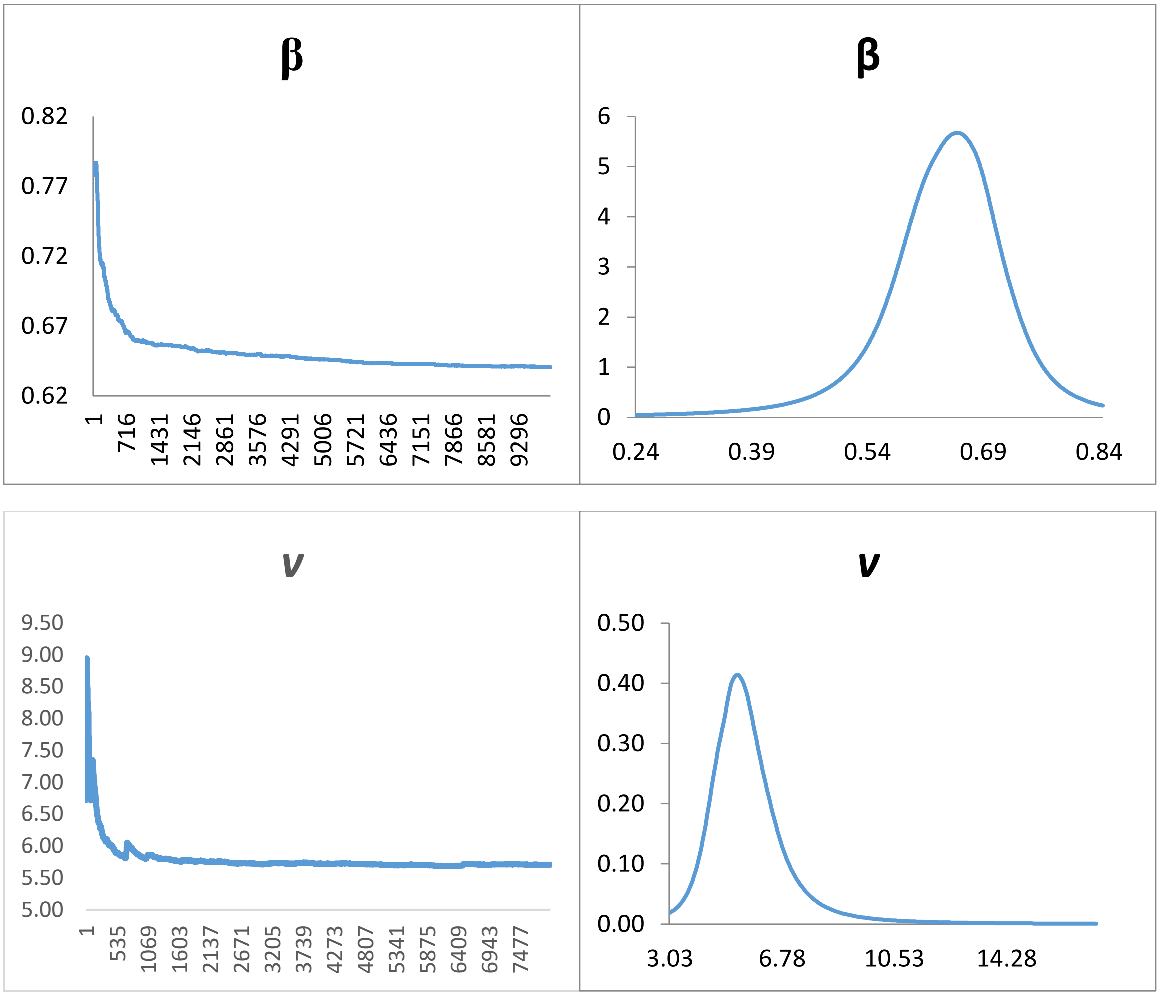
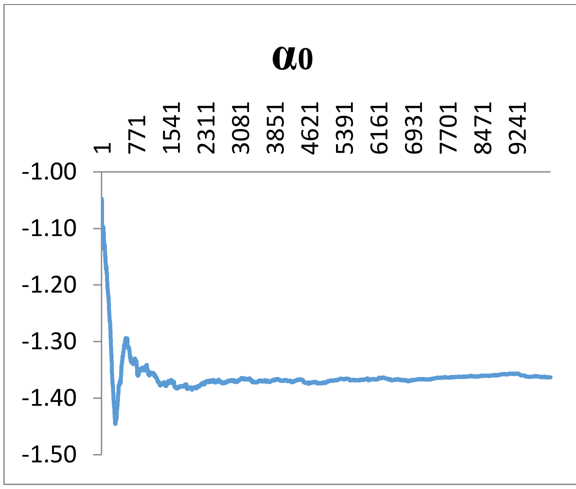
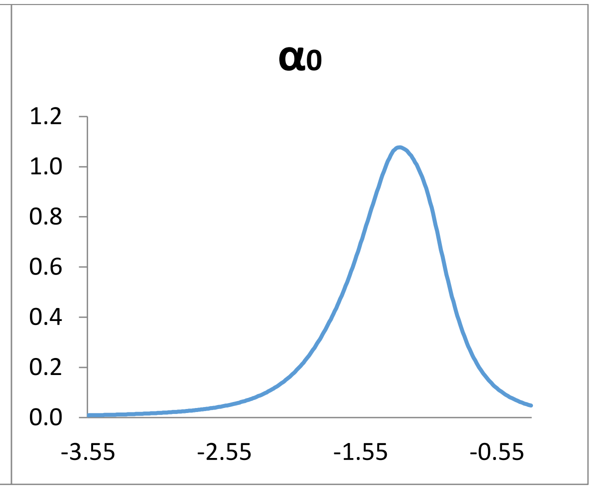
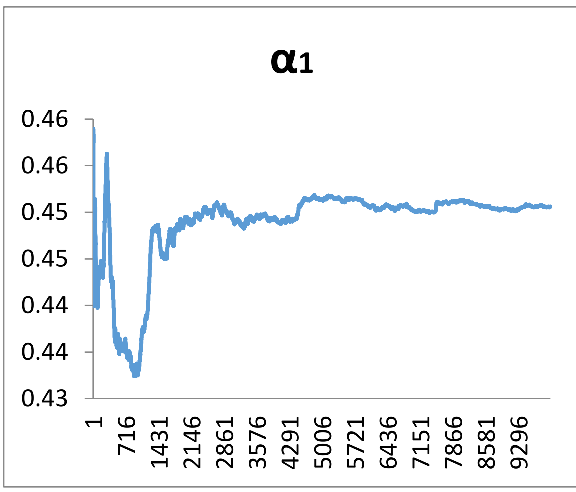


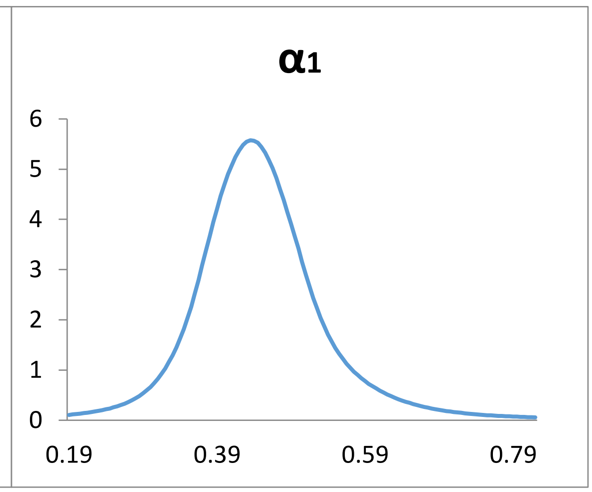
Figure B.3. Running means along with the distribution of each parameter is constructed by using chains of RAM. Running means for each chain is approximately converging to a stationary point. Also, the distribution for each parameter is approximately asymmetric.

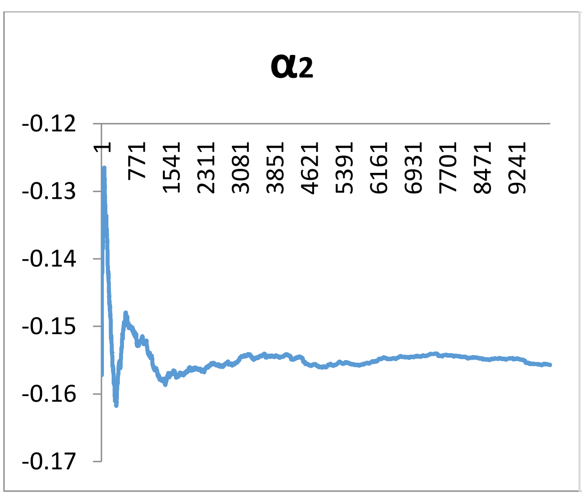
APPENDIX C

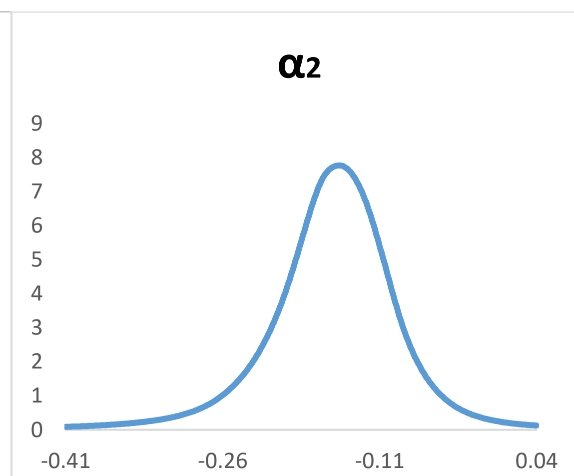

α0
1
771
1541
2311
3081
3851
4621
5391
6161
6931
7701
8471
9241
-1.00
-1.10
-1.20
-1.30
-1.40
-1.50


α0
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-3.55
-2.55
-1.55
-0.55


α1
0.46
0.46
0.45
0.45
0.44
0.44
0.43
1
716
1431
2146
2861
3576
4291
5006
5721
6436
7151
7866
8581
9296


α1
6
5
4
3
2
1
0
0.19
0.39
0.59
0.79


α2
-0.12
-0.13
-0.14
-0.15
-0.16
-0.17
1
771
1541
2311
3081
3851
4621
5391
6161
6931
7701
8471
9241


α2
9
8
7
6
5
4
3
2
1
0
-0.41
-0.26
-0.11
0.04

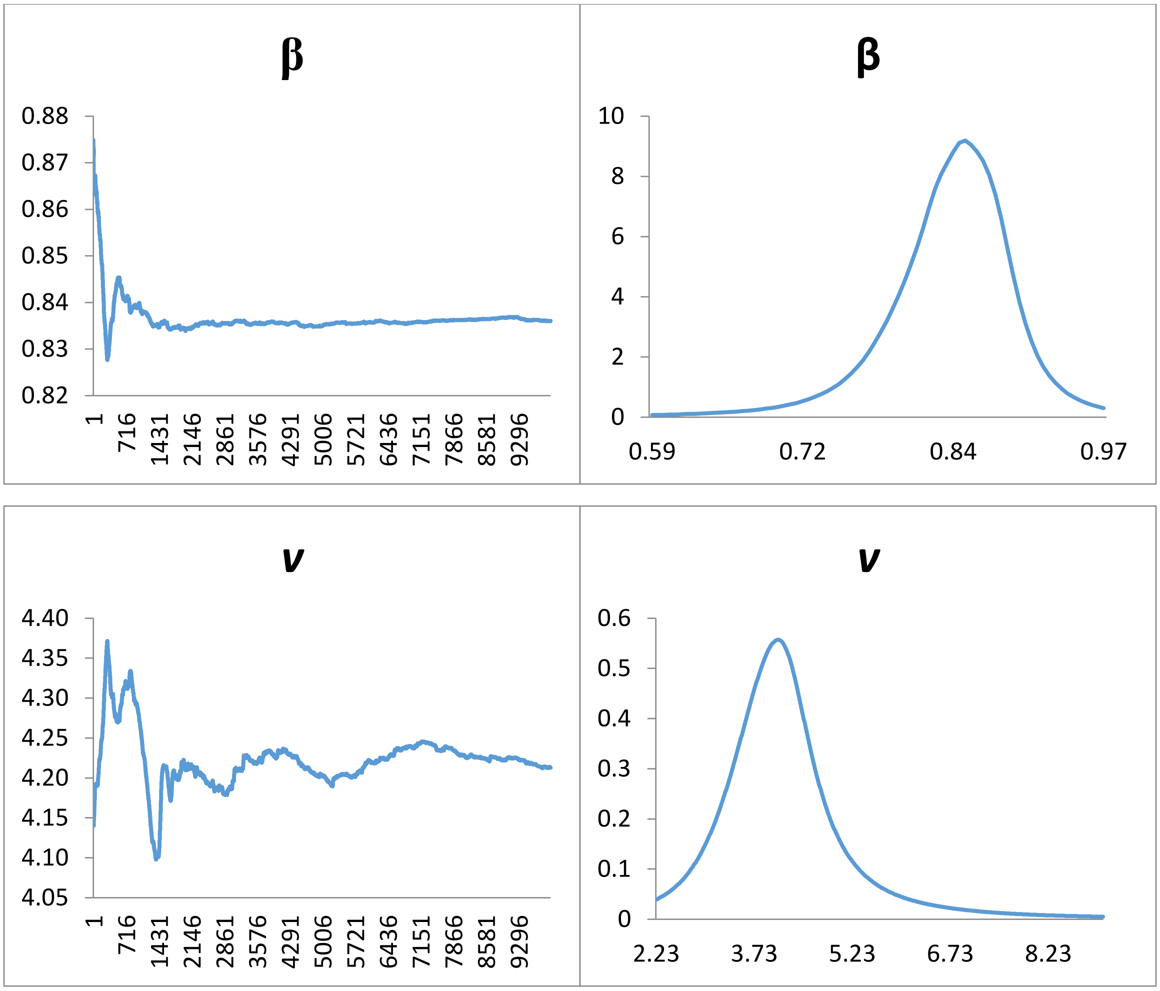


Figure C.3. Running means along with the distribution of each parameter is constructed by using chains of RAM. These chains are obtained by estimating the EGARCH model, by assuming skewed student-t distribution. Running means for each chain is approximately converging to a stationary point. Also, the distribution for each parameter is approximately asymmetric.

REFERENCES

Aas, K., & Haff, I. H. (2006). The generalized hyperbolic skew student's t-distribution. *Journal of financial econometrics*, *4*(2), 275-309.

Adcock, C., Eling, M., & Loperfido, N. (2015). Skewed distributions in finance and actuarial science: a review. *The European Journal of Finance*, *21*(13-14), 1253-1281.

Akima, H. (1969). A method of smooth curve fitting (No. 101-ITS-73). *Institute for Telecommunication Sciences Boulder CO.*

Ardia, D. (2008). Financial risk management with Bayesian estimation of GARCH models. *Heidelberg: Springer.*

Ardia, D., & Boudt, K. (2015). Implied expected returns and the choice of a mean-variance efficient portfolio proxy. *Journal of Portfolio Management*, *41*(4), 68.

Ardia, D., & Hoogerheide, L. F. (2009). Bayesian estimation of the GARCH (1, 1) model with student-t innovations. *The R Journal*, *2*(2), 41-47.

Ardia, D., & Hoogerheide, L. F. (2010). Efficient Bayesian estimation and combination of GARCH-type models. *Working Paper (TI 2010-046/4)*.

Ardia, D., Bluteau, K., Boudt, K., Catania, L., & Trottier, D. A. (2016). Markov-switching GARCH models in R: The MSGARCH package. *Journal of Statistical Software*, *Forthcoming*.

Ausín, M. C., & Galeano, P. (2007). Bayesian estimation of the Gaussian mixture GARCH model. *Computational Statistics & Data Analysis*, *51*(5), 2636-2652.

Barberis, N., Mukherjee, A., & Wang, B. (2016). Prospect theory and stock returns: An empirical test. *The Review of Financial Studies*, *29*(11), 3068-3107.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*(3), 307-327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, *69*(3), 542-547.

Bollerslev, T., & Ghysels, E. (1996). Periodic autoregressive conditional heteroscedasticity. *Journal of Business & Economic Statistics*, *14*(2), 139-151.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. *Handbook of Econometrics*, *4*, 2959-3038.

Box, G. E. (1980). Sampling and Bayes' inference in scientific modeling and robustness. *Journal of the Royal Statistical Society*. *Series A (General)*, 383-430.

Cairo, A. (2016). The truthful art: Data, charts, and maps for communication. *New Riders.*

Doornik, J. A., & Ooms, M. (2000). Multimodality and the GARCH Likelihood. In *World Congress of the Econometric Society,* Seattle.

Doornik, J. A., & Ooms, M. (2008). Multimodality in GARCH regression models. *International Journal of Forecasting*, *24*(3), 432-448.

Enders, W. (2008). Applied econometric time series. *John Wiley & Sons*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, *50*(4), 987-1007.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The journal of finance*, *48*(5), 1749-1778.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, *38*(1), 34-105.

Fiorentini, G., Calzolari, G., & Panattoni, L. (1996). Analytic derivatives and the computation of GARCH estimates. *Journal of applied econometrics*, *11*(4), 399-417.

Francq, C., & Zakoian, J. M. (2019). GARCH models: structure, statistical inference and financial applications. *Wiley.*

Gallant, A. R., & Tauchen, G. (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica: Journal of the Econometric Society*, 1091-1120.

Gelman, A., & Rubin, D. B. (1992). A single series from the Gibbs sampler provides a false sense of security. *Bayesian statistics*, *4,* 625-631.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457-472.

Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of applied econometrics*, *8*(S1), S19-S40.

Geweke, J., & Whiteman, C. (2006). Bayesian forecasting. *Handbook of economic forecasting*, *1*, 3-80.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, *48*(5), 1779-1801.

Granger, C. W. J., & Andersen, A. (1978). Non-linear time series modeling. *Proceedings of the First Applied Time Series Symposium*. Tulsa, Oklahoma.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, *35*(3), 705-730.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97-109.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: Springer.

Jerrell, M. E., & Campione, W. A. (2001). Global optimization of econometric functions. *Journal of Global Optimization*, *20*(3-4), 273-295.

Kennedy, P. (2003). A guide to econometrics. *Massachusetts, USA: MIT Press*.

Klar, B., Lindner, F., & Meintanis, S. G. (2012). Specification tests for the error distribution in GARCH models. *Computational Statistics & Data Analysis*, *56*(11), 3587-3598.

Mandelbrot, B. (1963). New methods in statistical economics. *Journal of political economy*, *71*(5), 421-440.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, *21*(6), 1087-1092.

Mullen, K., Ardia, D., Gil, D. L., Windover, D., & Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, *40*(6), 1-26.

Pagan, A. R., & Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of econometrics*, *45*(1-2), 267-290.

Price, K., Storn, R. M., & Lampinen, J. A. (2006). Differential evolution: a practical approach to global optimization. *Springer Science & Business Media*.

Ritter, C., & Tanner, M. A. (1992). Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association*, *87*(419), 861-868.

Särkkä, S. (2013). Bayesian filtering and smoothing. *Cambridge, UK: Cambridge University Press*.

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, *11*(4), 341-359.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, *22*(4), 1701-1728.

Tsay, R. S. (2010). Analysis of Financial Time Series. *New York: John Wiley & Sons*.

Vihola, M. (2012). Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, *22*(5), 997-1008.

Virbickaite, A., Ausín, M. C., &Galeano, P. (2015). Bayesian inference methods for univariate and multivariate GARCH models: A survey. *Journal of Economic Surveys*, *29*(1), 76-96.

Viterbi, A. (1967). Performance of an M-ary orthogonal communication system using stationary stochastic signals. *IEEE Transactions on Information Theory*, *13*(3), 414-422.

Zaman, A. (1996). Statistical foundations for econometric techniques. *Academic Press*.

Zaman, A. (2002). Maximum likelihood estimates for the Hildreth–Houck random coefficients model. *The Econometrics Journal*, *5*(1), 237-262.