ANALYSIS OF STOCHASTIC VOLATILITY SEQUENCES GENERATED BY PRODUCT AUTOREGRESSIVE MODELS

Thesis submitted to the
Cochin University of Science and Technology
for the Award of Degree of

DOCTOR OF PHILOSOPHY

under the Faculty of Science

by

SHIJI K.

Department of Statistics
Cochin University of Science and Technology
Cochin-682022

March 2014

CERTIFICATE

Certified that the thesis entitled **"Analysis of Stochastic Volatility Sequences Generated by Product Autoregressive Models "** is a bonafide record of work done by Smt. Shiji K. under my guidance in the Department of Statistics, Cochin University of Science and Technology and that no part of it has been included anywhere previously for the award of any degree or title.

Kochi- 22

March 2014

Dr. N. Balakrishna

Professor,

Department of Statistics,

Cochin University of

Science and Technology.

CERTIFICATE

Certified that all the relevant corrections and modifications suggested by the audience during pre-synopsis seminar and recommended by the Doctoral committee of the candidate has been incorporated in the thesis.

Kochi- 22

March 2014

Dr. N. Balakrishna

Professor,

Department of Statistics,

Cochin University of

Science and Technology.

DECLARATION

This thesis contains no material which has been accepted for the award of any other Degree or Diploma in any University and to the best of my knowledge and belief, it contains no material previously published by any other person, except where due references are made in the text of the thesis.

Kochi- 22 **SHIJI K.**

March 2014

ACKNOWLEDGEMENTS

I am deeply indebted to many who have generously helped me in completing this dissertation and I take this opportunity to express my sincere thanks to each and everyone of them.

First of all, I wish to express my deep sense of respect and gratitude to my supervising guide, Dr.N.Balakrishna, Professor and Formerly Head, Department of Statistics, Cochin University of Science and Technology (CUSAT) for supporting, guiding and working with me for the last four years. He has always been patient towards my shortcomings and kept encouraging me to perform in a better way.

I take this opportunity to record my sincere respect and heartiest gratitude to Dr.Asha Gopalakrishnan, Professor and Head, Department of Statistics, CUSAT and Dr.K.C.James, Professor and Formerly Head, Department of Statistics, CUSAT for their wholehearted support. I am obliged to Dr.K.R.Muraleedharan Nair, Professor and Formerly Dean, Faculty of Science, CUSAT, Dr.V.K.Ramachandran Nair, Professor, Dr. P.G.Sankaran, Professor and Dr.S.M.Sunoj, Associate Professor, Department of Statistics, CUSAT for their valuable suggestions and help to complete this endeavour. I remember with deep gratefulness all my former teachers.

I convey my sincere thanks to the non-teaching staff, Department of Statistics, CUSAT for the co-operation and help they had rendered.

I also would like to thank the Department of Science and Technology for the financial support provided during my research.

Discussions with my friends and other research scholars of the department often helped me during the work. I express my sincere thanks to all of them for their valuable suggestions and help.

I am deeply indebted to my beloved father, mother and sister for their encouragement, prayers and blessings given to me.

Above all, I bow before the grace of the Almighty.

SHIJI K.

To My Loving Parents

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

The classical methods of analysing time series by Box-Jenkins approach assume that the observed series fluctuates around changing levels with constant variance. That is, the time series is assumed to be of homoscedastic nature. However, the financial time series exhibits the presence of heteroscedasticity in the sense that, it possesses non-constant conditional variance given the past observations. So, the analysis of financial time series, requires the modelling of such variances, which may depend on some time dependent factors or its own past values. This lead to introduction of several classes of models to study the behaviour of financial time series. See Taylor (1986), Tsay (2005), Rachev et al. (2007). The class of models, used to describe the evolution of conditional variances is referred to as stochastic volatility models.

The stochastic models available to analyse the conditional variances, are based on either normal or log-normal distributions.

One of the objectives of the present study is to explore the possibility of employing some non-Gaussian distributions to model the volatility sequences and then study the behaviour of the resulting return series. This lead us to work on the related problem of statistical inference, which is the main contribution of the thesis.

1.2 Introduction

Time series is a sequence of observations on any variable of interest. Time series models are designed to capture various characteristics of time series data. These models have been widely used in many disciplines in the science, humanities, engineering etc. In particular, it has been found that time series models are very useful in analysing economic and financial data. The reports in the daily news papers, television and radio inform us for instance, of the latest stock market index values, currency exchange rates, gold prices etc. The reports often highlight substantial fluctuations in prices. It is often desirable to monitor price behaviour and try to understand the probable development of the prices in the future. The sequence of observations representing the prices or price indices are referred to as financial time series.

There are two main objectives of investigating financial time series. First, it is important to understand how prices behave over a period of time. The variance of the time series is particularly relevant to understand the presence of heteroscedasticity

in the system. Tomorrow's price is uncertain and it must therefore be described by a suitable probability distribution. This means that statistical methods are the natural way to investigate price behaviour. Usually one builds a model, which is a detailed description of how successive prices are evolving. The second objective is to use our knowledge of price behaviour to reduce risk or take better decisions. Time series models may for instance be used for forecasting, option pricing and risk management. This motivates more and more statisticians and econometricians to devote themselves to the development of new (or refined) time series models and methods.

Classical time series analysis, generally known as Box and Jenkins time series approach, deals with the modelling and analysis of finite variance linear time series models (see Box et al. (1994) and Brockwell and Davis (1987)). This approach of modelling time series heavily depends on the assumption that the series is a realization from a Gaussian sequence and the value at a time point t is a linear function of past observations. Box et al. (1994) proposed a four stage procedure for analysing a time series which includes model identification, parameter estimation, diagnostic checking and forecasting. The detailed discussion is given in Section 1.7.

In recent years a number of different models have been constructed for the generation of non-Gaussian processes in discrete time. The need for such models arises from the fact that many naturally occurring time series are clearly non-Gaussian. The usual techniques of transforming the data to use a Gaussian model also fail in certain situations (Lawrance (1991)). Hence, a number of non-Gaussian time series models have been introduced by different researchers during the last few years (see Gaver and Lewis (1980), Lawrance and Lewis (1985)). The study of non-Gaussian

time series is motivated mainly by two aspects. First is that, it gets stationary sequences having non-normal marginal random variables (rvs). Secondly, to study the point processes generated by the sequences of non-negative dependent rvs. One of the theoretical problems in non-Gaussian time series modelling is to identify the innovation distribution for a specified stationary marginal. In most of the cases, we cannot get a closed form expression for such distribution. For some other linear non-Gaussian time series models, one may refer Adke and Balakrishna (1992), Sim (1990) for gamma marginals, Balakrishna and Nampoothiri (2003) for Cauchy, Jayakumar and Pillai (1993) for Mittag-Leffler, etc.

The modelling of non-negative rv plays a major role in the study of financial time series, where one has to model the evolution of conditional variances known as Stochastic Volatility (see Tsay (2005)). The linear time series model for non-negative rvs lead to complicated form of the innovation distribution, which in turn makes the likelihood based inference intractable. As an alternative, McKenzie (1982) introduced a class of models with product structure which generates a Markov sequence of non-negative rvs. The contents of this thesis are on various aspects of modelling and analysis of non-Gaussian and non-negative time series in view of their applications in financial time series to model stochastic volatility.

1.3 Examples of Time series

Time series analysis deals with statistical methods for analysing and modelling an ordered sequence of observations. This modelling results in a stochastic process

model for the system which generated the data. The ordering of observations is most often, but not always, through time, particularly in terms of equally spaced time intervals. Time series occur in a variety of fields such as agriculture, business and economics, engineering, medical studies etc. In this section, we describe some examples of time series.

The first example is the daily exchange rate of Rupee to US dollar. The data consists of 273 observations from 1, January 2013 to 30, September 2013. The time series plot of the data is shown in Figure 1.1. It is obvious from the figure that the data exhibit a clear positive trend. This is a typical economic time series where time series analysis could be used to formulate a model for forecasting future values of the exchange rate.

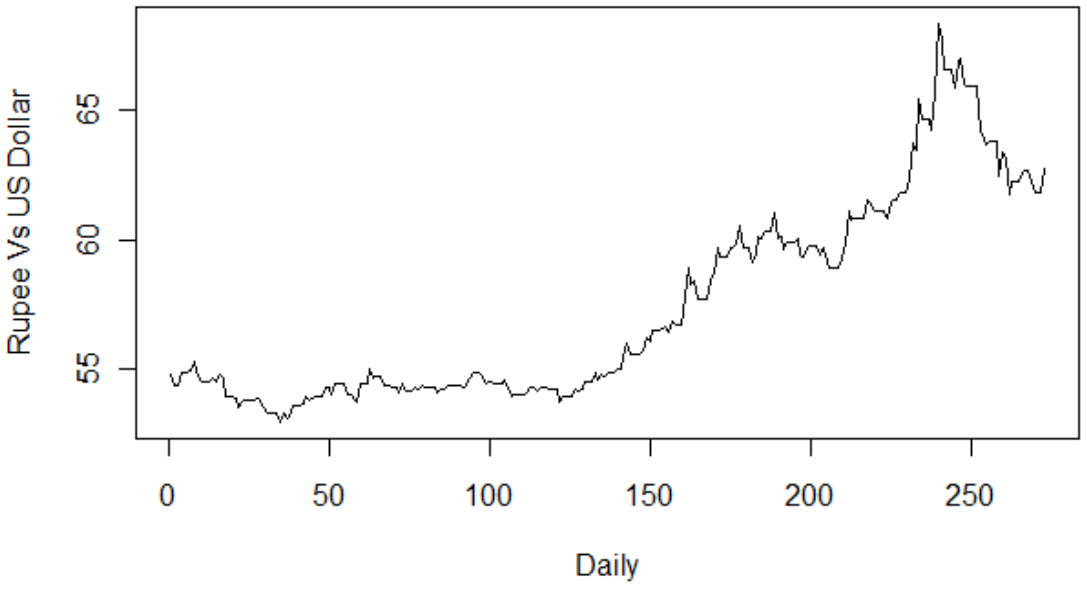

FIGURE 1.1: Time series plot of daily exchange rate of Rupee to US dollar for the period January 2013 to September 2013

Next, we consider the number of international passenger bookings per month on an airline in the United States. The data were obtained from the Federal Aviation Administration for the period 1949-1960 (Brown (1963)). The company used the data to predict future demand before ordering new aircraft and training aircrew. From the Figure 1.2, it is apparent that the number of passengers travelling on the airline is increasing with time, with some seasonal effects.

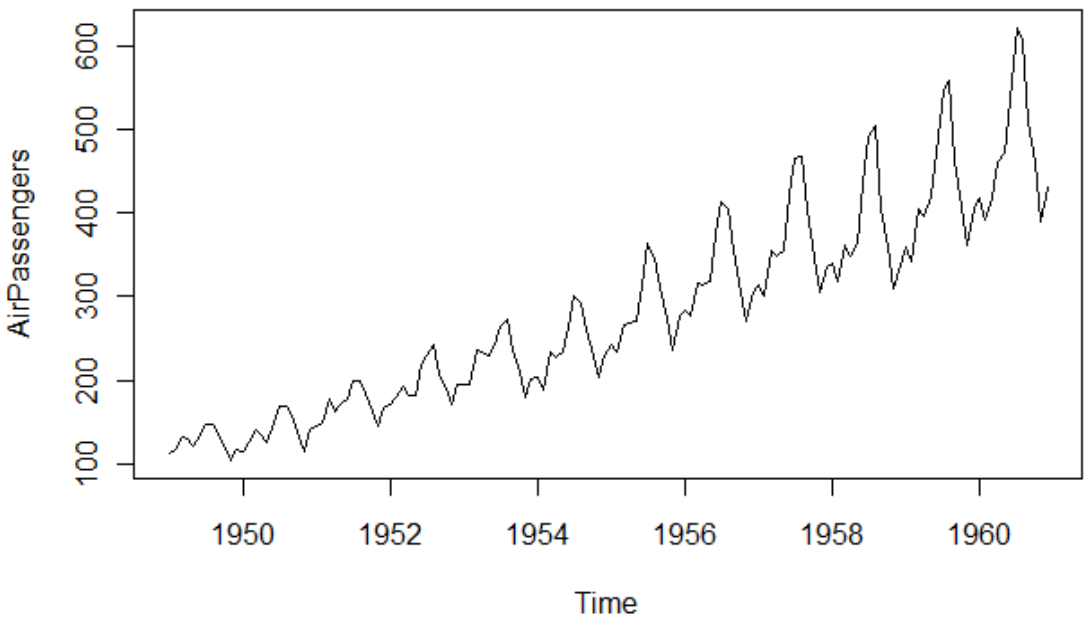

FIGURE 1.2: Time series plot of international air passenger bookings per month in the United States for the period 1949-1960

Other examples include (1) sales of a particular product in successive months, (2) the maximum temperature at a particular location on successive days, (3) electricity consumption in a particular area for successive one-hour periods, (4) daily closing stock prices, (5) weekly interest rates, and (6) monthly price indices, etc.

Time series analysis is done primarily for the purpose of making forecasts for future and also for the purpose of evaluating past performances. For example, an economist or a businessman is naturally interested in estimating the future figures of national income, population, prices and wages etc. In fact the success or the failure of an economist depends, to a large extent on the accuracy of his future forecasts. Forecasting for future is done by analysing the past behaviour of the variable under study. Thus, the future demand of a commodity or future profits of a concerned can be forecasted only by analysing the demand of the commodity or the profits of the concerned in the past years. Hence the analysis of time series assumes as great importance in the study of all economic problems.

In the upcoming sections, we list some of the basic concepts which facilitate the systematic development of the thesis.

1.4 Basic Concepts

1.4.1 Stochastic Processes

A stochastic processes is a family of time indexed random variables $X(\omega, t)$, where ω belongs to a sample space and t belongs to an index set. For a given ω, $X(\omega, t)$, as a function of t, is called a sample function or realization. The population that consists of all possible realizations is called the ensemble in stochastic processes and time series analysis. Thus, a time series is a realization or a sample function from a certain discrete time stochastic process. With proper understanding that a stochastic process, $X(\omega, t)$, is a set of time indexed random variables defined on

a sample space, we usually suppress the variable ω and simply write $X(\omega, t)$ as $X(t)$ or X_t. Thus, we may call $\{X_t\}$ as a stochastic process or a time series. The mean function and variance function of the process are defined as $\mu_t = E(X_t)$ and $\sigma_t^2 = V(X_t) = E(X_t - \mu_t)^2$.

1.4.2 Stationary Processes

A time series $\{X_t\}$ is said to be strictly stationary if the joint distribution of $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ is identical to that of $(X_{t_1+k}, X_{t_2+k}, ..., X_{t_n+k})$ for all t and k, where n is an arbitrary positive integer and $(t_1, t_2, ..., t_n)$ is a collection of n positive integers. In other words, strict stationarity requires that the joint distribution of $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ is invariant under time shift. This is a very strong condition that is hard to verify empirically. A weaker version of stationarity, which is often easy to verify is described below.

A time series $\{X_t\}$ is weakly stationary if X_t has constant mean, finite variance and the covariance between X_t and X_{t-k} depends only on k, where k is an arbitrary integer. From definitions, if $\{X_t\}$ is strictly stationary and its first two moments are finite, then it is also weakly stationary. The converse is not true in general.

1.4.3 Autocorrelation and Partial Autocorrelation Function

Let $\{X_t : t = 0, \pm1, \pm2, ...\}$ be a stochastic process (time series), the covariance between X_t and X_{t-k} is known as the autocovariance function at lag k and is defined

by

$$\gamma_X(k) = \text{Cov}\,(X_t,\, X_{t-k}) = E(X_t - E(X_t))(X_{t-k} - E(X_{t-k})).$$

Hence, the correlation coefficient between X_t and X_{t-k}, is called Autocorrelation function (ACF) at lag k, is given by

$$\rho_X(k) = \text{Corr}\,(X_t,\, X_{t-k}) = \frac{\text{Cov}(X_t,\, X_{t-k})}{\sqrt{V(X_t)}\sqrt{V(X_{t-k})}}, \tag{1.1}$$

where $V(.)$ is the variance function of the process.

For a strictly stationary process, since the distribution function is same for all t, the mean function $E(X_t) = E(X_{t-k}) = \mu$ is a constant, provided $E|X_t| < \infty$. Likewise, if $E(X_t^2) < \infty$, then $V(X_t) = V(X_{t-k}) = \sigma^2$ for all t and hence is also a constant.

The Partial Autocorrelation function (PACF) of a stationary process, $\{X_t\}$, denoted $\phi_{k\,,k}$ for $k = 1,\, 2, ...,$ is defined by

$$\phi_{1,\,1} = \text{Corr}(X_1,\, X_0) = \rho_1$$

and

$$\phi_{k,\,k} = \text{Corr}(X_k - \hat{X}_k,\, X_0 - \hat{X}_0), \;\; k \geq 2,$$

where $\hat{X}_k = l_1\, X_{k-1} + l_2\, X_{k-2} + \cdots + l_{k-1}\, X_1$ is the linear predictor. Both $(X_k, \hat{X}_k)$ and $(X_0, \hat{X}_0)$ are correlated with $\{X_1,\, X_2, ..., X_{k-1}\}$. By stationarity, the PACF, $\phi_{k,\,k}$, is the correlation between X_t and X_{t-k} obtained by fixing the effect of $X_{t-1}, ..., X_{t-(k-1)}$.

1.5 Linear Time Series Models

The classical set up of time series analysis asserts that the observed series is generated by a linear structure (Box-Jenkins method) and we call such time series as linear time series. The models introduced for such studies include Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average Models (ARIMA), etc.

1.5.1 Autoregressive Models

A stochastic model that can be extremely useful in the representation of certain practically occurring series is the autoregressive model. In this model, the current value of the process is expressed as a finite, linear aggregate of previous values of the process and a shock η_t. Let us denote the values of a process at equally spaced time $t,\ t-1,\ t-2$, ... by $X_t,\ X_{t-1},\ X_{t-2}, ...$, then X_t can be described by the expression:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + \eta_t. \tag{1.2}$$

Or equivalently,

$$\alpha(B)X_t = \eta_t \text{ with } \alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p,$$

where B is the back shift operator, defined by $BX_t = X_{t-1}$, $\{\eta_t\}$ is a sequence of uncorrelated random variables with mean zero and constant variance, termed as innovations and $\alpha(B)$ is referred to as the characteristic polynomial associated with

an AR(p) process. As X_t is a linear function of its own past p values, the process $\{X_t\}$ generated by (1.2) is referred to as an Autoregressive process of order (p) (AR(p)). This is rather like a multiple linear regression model, but X_t is regressed not on independent variables but on the past values of X_t; hence the prefix 'auto'. The resulting AR(p) process is weakly stationary if all the roots of the associated characteristic polynomial equation $\alpha(B) = 0$ lie outside the unit circle.

For a stationary AR(p) processes, the autocorrelation function, $\rho_X(k)$, can be found by solving a set of difference equations called the Yule-Walker equations given by

$$(1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p)\rho_X(k) = 0, \qquad k > 0.$$

The plot of ACF of a stationary AR(p) model would then show a mixture of damping sine and cosine patterns and exponential decays depending on the nature of its characteristic roots.

The autoregressive model of order 1 (AR(1)) is important as it has several useful features. It is defined by

$$X_t = \alpha X_{t-1} + \eta_t, \tag{1.3}$$

where $\{\eta_t\}$ is a white noise with mean 0 and variance σ^2. The sequence $\{X_t\}$ is weakly stationary AR(1) process, when $|\alpha| < 1$. Under stationarity, we have $E(X_t) = 0$, $V(X_t) = \sigma^2/(1 - \alpha^2)$ and the autocorrelation function is given by

$$\rho_X(k) = \alpha^k, \ k = 0, 1, 2, ...$$

This result says that the ACF of a weakly stationary AR(1) series decays exponentially in k. If we assume that the innovation sequence $\{\eta_t\}$ is independent and identically distributed (iid), then the AR(1) sequence is Markovian.

1.5.2 Moving Average Models

Another type of model of great practical importance in the representation of observed time-series is the finite moving average process. In this model, the observation at time t, X_t, is expressed as a linear function of the present and past shocks. A moving average model of order q (MA(q)) is defined by

$$X_t = \eta_t - \theta_1 \eta_{t-1} - \theta_2 \eta_{t-2} - ... - \theta_q \eta_{t-q}. \tag{1.4}$$

Or, $X_t = \Theta(B)\eta_t$, where $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$, is the characteristic polynomial associated with the MA(q) model, where θ_i's are constants, $\{\eta_t\}$ is a white noise sequence.

The definition implies that

$$E(X_t) = 0; V(X_t) = \sigma^2 \sum_{i=1}^{q} \theta_i^2$$

and the ACF is,

$$\rho_X(k) = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + ... + \theta_{q-k}\theta_q}{1+\theta_1^2+\theta_2^2+...++\theta_q^2}, & k = 1, 2, ..., q \\ 0, & k > q \end{cases}. \tag{1.5}$$

Hence, for a MA(q) model, its ACF vanishes after lag q.

In particular a MA(1) model for $\{X_t\}$ is defined by

$$X_t = \eta_t - \theta\,\eta_{t-1}.$$

So, X_t is a linear function of the present and immediately preceding shocks. The MA(q) process will always be stationary as it is a finite linear combination of shocks, but it is invertible if $|\theta| < 1$. The unconditional variance of a MA(1) process is given by $V(X_t) = (1 + \theta^2)\,\sigma^2$.

The ACF of the MA(1) process is

$$\rho_X(k) = \begin{cases} -\theta/(1+\theta^2), & k = 1 \\ 0, & k = 2, 3, ... \end{cases}.$$

1.5.3 Autoregressive Moving Average Models

A natural extension of the pure autoregressive and pure moving average processes is the mixed autoregressive moving average (ARMA) process. An ARMA model with p AR terms and q MA terms is called an ARMA(p, q) model. The advantage of ARMA process relative to AR and MA processes is that it gives rise to a more parsimonious model with relatively few unknown parameters.

A mixed process of considerable practical importance is the first order autoregressive first order moving average (ARMA(1, 1)) model,

$$X_t - \alpha X_{t-1} = \eta_t - \theta\,\eta_{t-1}. \tag{1.6}$$

The process is stationary if $|\alpha| < 1$ and invertible if $|\theta| < 1$. The mean, variance and the autocorrelation function of the ARMA(1, 1) model are respectively given by

$$E(X_t) = 0, \;\; Var(X_t) = \gamma_0 = E(X_t^2)$$

and

$$\rho_X(k) = \begin{cases} \frac{\alpha\theta^2 - \theta\,\alpha^2 + \alpha - \theta}{1 + \theta^2 - 2\theta\,\alpha}, & if\;\; k = 1 \\ \alpha\,\rho_{k-1}, & if\;\; k = 2, 3, ... \end{cases}. \tag{1.7}$$

Thus, the autocorrelation function decays exponentially from the starting value, $\rho(1)$, which depends on θ as well as on α.

A more general model that encompasses AR(p) and MA(q) model is the autoregressive moving average, or ARMA(p, q), model

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - ... - \alpha_p X_{t-p} = \eta_t - \theta_1\,\eta_{t-1} - \theta_2\,\eta_{t-2} - ... - \theta_q\,\eta_{t-q}. \tag{1.8}$$

The model is stationary if AR(p) component is stationary and invertible if MA(q) component is so. One may refer Box et al. (1994) for detailed analysis of linear time series models.

1.6 Product Autoregressive Model

The role of linear autoregressive models is well known in time series analysis when the variables take both positive and negative values. When the variables are non-negative, the additive form is not so natural and a multiplicative autoregressive form may be preferable. Let $\{Y_t, t \geq 0\}$ be a random sequence of non-negative rvs defined recursively by

$$Y_t = Y_{t-1}^{\alpha} V_t, \;\; 0 < \alpha < 1, \;\; t = 1, 2, ..., \tag{1.9}$$

where $\{V_t\}$ is a sequence of iid positive rvs. Further Y_0 is independent of V_1. The model (1.9) initially introduced by McKenzie (1982) is referred to as the Product Autoregressive model of order 1 (PAR(1)). McKenzie (1982) discusses the above model mainly for gamma random variables. One specifies a marginal distribution as the stationary distribution of the above sequence and investigates existence and form of distribution of V_t. Such processes are clearly Markovian, if V_t is a sequence of iid positive random variables.

The log-transform of (1.9) leads to

$$\log Y_t = \alpha \log Y_{t-1} + \log V_t, \;\; 0 < \alpha < 1, \tag{1.10}$$

which is an AR(1) model in $\log Y_t$. In terms of the Moment Generating Function (MGF), we may express (1.10) as

$$\phi_{\log V}(s) = \phi_{\log Y}(s) / \phi_{\log Y}(\alpha s), \tag{1.11}$$

where $\phi_U(s) = E(\exp(sU))$, is the MGF of U. Thus the model (1.9) defines a stationary sequence $\{Y_t\}$ if the right hand side of (1.11) is a proper MGF for every $\alpha \in (0,1)$. This happens if $\log Y_t$ is a self-decomposable rv. In fact the MGF of $\log Y_t$ may be expressed as the Mellin Transform (MT), $M_Y(s)$ of Y_t, defined by $M_Y(s) = E(Y_t^s),\ s \geq 0$. Thus, we can use the Mellin transform to identify the innovation distribution for PAR(1) models. The equation (1.11) now can be written in terms of MT as

$$M_V(s) = M_Y(s)/M_Y(\alpha s). \tag{1.12}$$

If V_t admits a density function $f_V(.)$, then the one step transition probability density function of $\{Y_t\}$ can be expressed as

$$f(y_{t+1}|y_t) = \frac{1}{y_t^\alpha} f_V(y_{t+1}/y_t^\alpha). \tag{1.13}$$

Conditional on the past observations, the mean and variance of Y_t in (1.9) depend just on Y_{t-1}, according to the formulae

$$E(Y_t|Y_{t-1}) = \mu_V\, Y_{t-1}^\alpha\,; \quad V(Y_t|Y_{t-1}) = \sigma_V^2\, Y_{t-1}^{2\alpha}, \tag{1.14}$$

where μ_V and σ_V^2 denote the mean and variance of V_t, respectively.

Instead of the usual linear expansion of the standard AR(1) model in terms of past innovations, for the PAR(1) model there is a multiplicative expansion in past innovations; for any chosen k, it takes the form

$$Y_t = \left(\prod_{i=0}^{k-1} V_{t-i}^{\alpha^i}\right) Y_{t-k}^{\alpha^k}. \tag{1.15}$$

Using this result, McKenzie (1982) gave the following ACF of PAR(1) sequence $\{Y_t\}$

$$\rho_Y(k) = Corr(Y_t, Y_{t-k}) = \frac{E(Y_t)\left\{E(Y_{t-k}^{\alpha^k+1}) - E(Y_{t-k}^{\alpha^k})E(Y_{t-k})\right\}}{E(Y_{t-k}^{\alpha^k})\,V(Y_t)}. \tag{1.16}$$

The ACF of the squared sequence, $\{Y_t^2\}$, is also important when we analyse the non-linear time series models. For the PAR(1) model, such ACF is given by

$$\rho_{Y^2}(k) = Corr(Y_t^2, Y_{t-k}^2) = \frac{E(Y_t^2)\left\{E(Y_{t-k}^{2\alpha^k+2}) - E(Y_{t-k}^{2\alpha^k})E(Y_{t-k}^2)\right\}}{E(Y_{t-k}^{2\alpha^k})\,V(Y_t^2)}. \tag{1.17}$$

The above ACFs depend only on the moments of stationary marginal distribution.

1.7 Box-Jenkins Modelling Techniques

This section highlights the Box-Jenkins methodology for model building and discusses its possible contribution to post-sample forecasting accuracy and therefore its need and value. A three step procedure is used to build a time series model. First, a tentative model is identified through analysis of historical data. Second, the unknown parameters of the model are estimated. Third, through residual analysis, diagnostic checks are performed to determine the adequacy of the model. We shall now briefly discuss each of these steps.

1.7.1 Model Identification

The primary tools for model identification are the plots of autocorrelation and the partial autocorrelation. The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behaviour of these plots when the order is known. Autocorrelation function of an autoregressive process of order p tail off and its partial autocorrelation function has a cut off after lag p. On the other hand, the autocorrelation function of moving average process cuts off after lag q, while its partial autocorrelation tails off after lag q. If both autocorrelation and partial autocorrelation tail off, a mixed process is suggested. Furthermore, the autocorrelation function for a mixed process, contains a p^{th} order AR component and q^{th} order moving average component, and is a mixture of exponential and damped sine waves after the first $q - p$ lags. The partial autocorrelation function for a mixed process is dominated by a mixture of exponential and damped sine waves after the first $q - p$ lags.

1.7.2 Parameter Estimation

Estimating the model parameters is an important aspect of time series analysis. There are several methods available in the literature for estimating the parameters, (see Box et al. (1994)). All of them produce very similar estimates, but may be more or less efficient for any given model. The main approaches to fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation. The Least Squares Estimator (LSE) of the parameter is obtained by minimizing

the sum of the squared residuals. For pure AR models, the LSE leads to the linear Ordinary Least Squares (OLS) estimator. If moving average components are present, the LSE becomes non-linear and has to be solved by numerical methods. The Maximum Likelihood Estimator (MLE) maximizes the (exact or approximate) log-likelihood function associated with the specified model. To do so, explicit distributional assumption for the innovations has to be made. Other methods for estimating model parameters are the Method of Moments (MM) and the Generalized Method of Moments (GMM), which are easy to compute but not very efficient.

1.7.3 Diagnosis Methods

After estimating the parameters one has to test the model adequacy by checking the validity of the assumptions imposed on the errors. This is the stage of diagnosis check. Model diagnostic checking involves techniques like over fitting, residual plots, and more importantly, checking that the residuals are approximately uncorrelated. This makes good modelling sense, since in the time series analysis a good model should be able to describe the dependence structure of the data adequately, and one important measurement of dependence is via the autocorrelation function. In other words, a good time series model should be able to produce residuals that are approximately uncorrelated, that is, residuals that are approximately white noise. Note that as in the classical regression case complete independence among the residuals is impossible because of the estimation process. However, the autocorrelations of the residuals should be close to being uncorrelated after taking into account the effect of estimation. As shown in the seminal paper by Box and Pierce

(1970), the asymptotic distribution of the residual autocorrelations plays a central role in checking out this feature. From the asymptotic distribution of the residual autocorrelations we can also derive tests for the individual residual autocorrelations and overall tests for an entire group of residual autocorrelations assuming that the model is adequate. These overall tests are often called portmanteau tests, reflecting perhaps that they are in the tradition of the classical Chi-square tests of Pearson. Nevertheless, portmanteau tests remain useful as an overall benchmark assuming the same kind of role as the classical Chi-square tests. Portmanteau tests and the residual autocorrelations are easy to compute and the rationale of using them is easy to understand. These considerations enhance their usefulness in applications.

Model diagnostic checking are often used together with model selection criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These two approaches actually complement each other. Model diagnostic checks can often suggest directions to improve the existing model while information criteria can be used in a more or less "automatic" way within the same family of models. Through the exposition on diagnostic checking methods, it is hoped that the practitioner should be able to grasp the relative merits of these models and how these different models can be estimated.

1.7.4 Forecasting

One of the objectives of analysing time series is to forecast its future behaviour. That is, based on the observation up to time t, we should be able to predict the value of the variable at a future time point. The method of Minimum Mean Square

Error (MMSE) forecasting is widely used when the time series follows a linear model. To derive the minimum mean square error forecasts, we first consider the stationary ARMA model,

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - ... - \alpha_p X_{t-p} = \eta_t - \theta_1 \eta_{t-1} - \theta_2 \eta_{t-2} - ... - \theta_q \eta_{t-q},$$

or, $\alpha(B)\, X_t = \Theta(B)\, \eta_t$.

We can rewrite it in a moving average representation,

$$X_t = \frac{\Theta(B)}{\alpha(B)}\, \eta_t = \psi(B)\, \eta_t = \sum_{j=0}^{\infty} \psi_j\, B^j\, \eta_t = \eta_t + \psi_1 \eta_{t-1} + \psi_2 \eta_{t-2} + \cdots \tag{1.18}$$

with $\psi_0 = 1$.

For $t = n + l$, we have

$$X_{n+l} = \sum_{j=0}^{\infty} \psi_j\, \eta_{n+l-j}. \tag{1.19}$$

Suppose, at time $t = n$, we have the observations X_n, X_{n-1}, $X_{n-2}, \ldots$ and wish to forecast l-step ahead value, X_{n+l}, as a linear combination of the observations X_n, $X_{n-1}, \ldots$. Since X_t for $t = n$, $n-1$, $n-2, \ldots$ can all be written in the form of (1.18), we can let the minimum mean square error forecast $\hat{X}_n(l)$ of X_{n+l} be

$$\hat{X}_n(l) = \psi_l^* \eta_n + \psi_{l+1}^* \eta_{n-1} + \psi_{l+2}^* \eta_{n-2} + \cdots,$$

where the ψ_j^* are to be determined. The mean square error of the forecast is

$$E(X_{n+l} - \hat{X}_n(l))^2 = \sigma^2 \sum_{j=0}^{l-1} \psi_j^2 + \sigma^2 \sum_{j=0}^{\infty} \left(\psi_{l+j} - \psi_{l+j}^*\right)^2,$$

which is seen to be minimized when $\psi^*_{l+j} = \psi_{l+j}$. Hence,

$$\hat{X}_n(l) = \psi_l \eta_n + \psi_{l+1}\eta_{n-1} + \psi_{l+2}\eta_{n-2} + \cdots .$$

But using (1.19) and the fact that

$$E(\eta_{n+j}|X_n,\, X_{n-1}, \ldots) = \begin{cases} 0, & j > 0, \\ \eta_{n+j}, & j \leq 0, \end{cases}$$

we have $E(X_{n+l}|X_n, X_{n-1}, \ldots) = \psi_l \eta_n + \psi_{l+1}\eta_{n-1} + \psi_{l+2}\eta_{n-2} + \cdots$.

Thus, the minimum mean square error forecast of X_{n+l} is given by its conditional expectation. That is,

$$\hat{X}_n(l) = E(X_{n+l}|X_n, X_{n-1}, \ldots).$$

$\hat{X}_n(l)$ is usually read as the l-step ahead of the forecast of X_{n+l} at the forecast origin n.

The forecast error is

$$e_n(l) = X_{n+l} - \hat{X}_n(l).$$

In the present study of financial time series, our goal is to forecast the volatility and we have to deal with non-linear models. Hence different approaches are adopted for different models and we will describe them as and when we need such methods.

1.8 Outline of the Thesis

The linear time series models available in the literature are not suitable to model the financial time series. So, new classes of models are introduced to deal with financial time series. Chapter 2 mainly discusses the characteristic of the financial time series. The models for financial time series may be broadly classified as *observation driven* and *parameter driven models*. In observation driven models, the conditional variance is assumed to be a function of the past observations, which introduces the heteroscedasticity in the model. The famous models such as Autoregressive Conditional Heteroscedastic (ARCH) model of Engle (1982) and Generalized ARCH model of Bollerslev (1986) are examples of these. While in the case of parameter driven models, the conditional variances are generated by some latent processes. The Stochastic Volatility (SV) model of Taylor (1986) is the example of parameter driven model. We summarize the properties of these models in Chapter 2. One of our objectives in this study is to identify some non-Gaussian time series models and study their suitability for modelling stochastic volatility.

We introduce a Gumbel Extreme Value Autoregressive (GEVAR) sequence $\{X_t\}$ in Chapter 3, with an idea to develop SV models induced by non-Gaussian volatility sequences. This Extreme value AR(1) model can be used to model the extreme events which includes the daily maximum/minimum of prices of assets, extreme floods and snowfalls, high wind speeds, extreme temperatures, large fluctuations in exchange rates, and market crashes. We have studied the second order properties and inference problems for this model. As the innovation distribution of the model does not admit a closed form expression, the problem of estimation becomes

complicated. We proposed the method of Conditional Least Squares (CLS), Quasi Maximum Likelihood (QML) and Maximum Likelihood (ML) for estimating the model parameters. A comparison study is made with respect to their efficiencies. Simulation studies are carried out to assess the performance of these methods. To illustrate the application of the proposed model, we have analysed two sets of data consisting of daily maximum of Bombay Stock Exchange (BSE) index and Standard and Poor 500 (S&P 500) index.

In Chapter 4, we study the details of the Product Autoregressive model of order one (PAR(1)) introduced by McKenzie (1982) to generate a non-negative Markov sequence. We developed the PAR(1) model for Weibull distribution and studied its statistical properties. As the innovation random variable does not admit closed form density, we use an approximation method to estimate the model parameters. Maximum Likelihood Estimators of the model parameters are obtained and their asymptotic properties are established.

We have considered the statistical analysis of Gumbel Extreme Value Stochastic Volatility (GEV-SV) model in Chapter 5. The volatility sequence are generated by GEVAR model, discussed in Chapter 3. The likelihood based inference of SV model is quite complicated because of the fact that the likelihood function involves the unobservable Markov dependent latent variables. Also, the innovation distribution of GEVAR model does not have a closed form and hence the other methods of estimation such as Bayesian estimation, Efficient importance sampling may not be appropriate. Thus, we employed the method of moments for parameter estimation.

Using the structure of the PAR(1) models we have constructed an absolutely continuous bivariate exponential distribution in Chapter 6. This bivariate distribution can be used for modelling two-dimensional renewal processes and queuing processes when arrival and service times are dependent. The basic properties of this model and problem of estimating its parameters are discussed. Some data sets are analysed to illustrate the applications of this model.

Chapter 2

Models for Financial Time Series

2.1 Introduction

Financial time series analysis is concerned with the theory and practice of asset valuation over time. One of the objectives of analysing financial time series is to model the volatility and forecast its future values. The volatility is measured in terms of the conditional variance of the random variables involved. Although volatility is not directly observable, it has some characteristics that are commonly seen in asset returns. First, there exist volatility clusters. Second, volatility evolves over time in a continuous manner - that is, volatility jumps are rare. Third, volatility does not diverge to infinity - that is, volatility varies within some fixed range. Statistically speaking, this means that volatility sequence is often stationary. Fourth, volatility seems to react differently to a big price increase or a big price drop, referred to

as the leverage effect. These properties play important role in the development of models for volatility.

The conditional variances in the case of financial time series are not constants. They may be functions of some known or unknown factors. This leads to the introduction of conditional heteroscedastic models for analysing financial time series. In financial markets, the data on price P_t of an asset at time t is available at different time points. However, in financial studies, the experts suggest that the series of returns be used for analysis instead of the actual price series, see Tsay (2005). For a given series of prices $\{P_t\}$, the corresponding series of returns is defined by

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - \ 1, \ \ t = 1, 2, \ldots .$$

The advantages of using the return series are, (1) for an investor, the return series is a scale free summary of the investment opportunity, (2) the return series are easier to handle than the price series because of their attractive statistical properties. Further consideration of the properties, suggested that, the log-return series defined by $r_t = \log\left(P_t/P_{t-1}\right)$ is more suitable for analysing the stochastic nature of the market behaviour. Hence, we focus our attention on the modelling and analysis of the log-return series in this thesis and $\{r_t = \log\left(P_t/P_{t-1}\right), t = 1, 2, ...\}$ is the financial time series of our interest.

Empirical studies on financial time series (See Mandelbrot (1963), Fama (1965) and Straumann (2005)) show that the series $\{r_t\}$ defined above is characterized by the properties such as

(i) Absence of autocorrelation in $\{r_t\}$.

(ii) Significant serial correlation in $\{r_t^2\}$.

(iii) The marginal distribution of $\{r_t\}$ is symmetric and heavy-tailed.

(iv) Conditional variance of r_t given the past is not constant.

The models described in the previous chapter are often very useful in modelling time series in general. However, they have the assumption of constant error variance. As a result, the conditional variance of the observation at any time given the past will remain a constant, a situation referred to as homoscedasticity. This is considered to be unrealistic in many areas of economics and finance as the conditional variances are non-constants. Therefore, Autoregressive Conditional Heteroscedastic (ARCH) model, Generalized ARCH (GARCH) model and Stochastic Volatility (SV) model which allow conditional variance to vary over time have been proposed, in particular to model financial market variables.

The chapter is split in to six sections. In the second section, we discuss the ARCH model and its properties. We surveyed the estimation procedure, model checking and volatility forecasting in the section. Generalized ARCH models are defined in Section 2.3. Section 2.4 introduces the mathematical representation of SV model. State-space approach for SV model are given in Section 2.5.

2.2 Autoregressive Conditional Heteroscedastic (ARCH) Model

The ARCH model introduced by Engle (1982) was a first attempt in econometrics to capture volatility clustering in time series data. In particular, Engle (1982) used conditional variance to characterize volatility and postulate a dynamic model for conditional variance. We will discuss the properties and some generalizations of the ARCH model in subsequent sections; for a comprehensive review of this class of models we refer to Bollerslev et al. (1992). ARCH models have been widely used in financial time series analysis and particularly in analysing the risk of holding an asset, evaluating the price of an option, forecasting time-varying confidence intervals and obtaining more efficient estimators under the existence of heteroscedasticity. Specifically, an ARCH(p) model assumes that

$$r_t = \sqrt{h_t}\varepsilon_t \ , \quad h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, r_{t-i}^2, \tag{2.1}$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed symmetric random variables with mean zero and variance 1, $\alpha_0 > 0$, and $\alpha_i \geq 0$ for $i > 0$. If $\{\varepsilon_t\}$ has standardized Gaussian distribution, r_t is conditionally normal with mean 0 and variance h_t. The Gaussian assumption of ε_t is not critical. We can relax it and allow for more heavy-tailed distributions, such as the Student's t -distribution, as is typically required in finance. Now we describe the properties of a first order ARCH model in detail.

2.2.1 ARCH(1) model and Properties

The structure of the ARCH model implies that the conditional variance h_t of r_t, evolves according to the most recent realizations of r_t^2 analogous to an AR(1) model. Large past squared shocks, $\{r_{t-i}^2\}_{i=1}^{p}$, imply a large conditional variance, h_t, for r_t. As a consequence, r_t tends to assume a large value which in turn implies that a large shock tends to be followed by another large shock. To understand the ARCH models, let us now take a closer look at the ARCH(1) model,

$$r_t = \sqrt{h_t}\,\varepsilon_t \ , \quad h_t = \alpha_0 + \alpha_1\, r_{t-1}^2, \tag{2.2}$$

where $\alpha_0 > 0$ and $\alpha_1 \geq 0$.

1. The unconditional mean of r_t is zero, since

$$E\left(r_t\right) = E\left(E\left(r_t|r_{t-1}\right)\right) = E\left(\sqrt{h_t}E\left(\varepsilon_t\right)\right) = 0.$$

2. The conditional variance of r_t is

$$E\left(r_t^2|r_{t-1}\right) = E\left(h_t\varepsilon_t^2|r_{t-1}\right) = h_t E\left(\varepsilon_t^2|r_{t-1}\right) = h_t = \alpha_0 + \alpha_1\, r_{t-1}^2.$$

3. The unconditional variance of r_t is

$$V\left(r_t\right) = E\left(r_t^2\right) = E\left(E\left(r_t^2|r_{t-1}\right)\right) = E\left(\alpha_0 + \alpha_1 r_{t-1}^2\right) = \alpha_0 + \alpha_1 E\left(r_{t-1}^2\right).$$

This implies that $V(r_t) = \alpha_0/(1-\alpha_1)$, $0 \leq \alpha_1 < 1$, because r_t is a stationary process with $E(r_t) = 0$, and $V(r_t) = V(r_{t-1}) = E(r_{t-1}^2)$.

4. Assuming that the fourth moment of r_t is finite, the kurtosis, K_r of r_t, is given by

$$K_r = \frac{E(r_t^4)}{E(r_t^2)^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3, \text{ provided } \alpha^2 < 1/3.$$

The ARCH model with a conditionally normally distributed r_t leads to heavy tails in the unconditional distribution. In other words, the excess kurtosis of r_t is positive and the tail distribution of r_t is heavier than that of the normal distribution.

5. The autocovariance of r_t is defined by

$$\begin{aligned} Cov(r_t, r_{t-k}) &= E(r_t\, r_{t-k}) - E(r_t)\, E(r_{t-k}) \\ &= E(r_t\, r_{t-k}) \\ &= E\left(\sqrt{h_t}\sqrt{h_{t-k}}\right) E(\varepsilon_t \varepsilon_{t-k}) = 0. \end{aligned}$$

Thus, the ACF of r_t is zero. The ACF of $\{r_t^2\}$ becomes $\rho_{r_t^2}(k) = \alpha_1^k$, and notice that $\rho_{r_t^2}(k) \geq 0$ for all k, a result which is common to all linear ARCH models.

Thus, the ARCH(1) process has a mean of zero, a constant unconditional variance, and a time-varying conditional variance. The $\{r_t\}$ is a stationary process when $0 \leq \alpha_1 < 1$ is satisfied, since the variance of r_t must be positive. These properties continue to hold for general ARCH models, but the formulae become more complicated for higher order ARCH models.

2.2.2 Estimation

The most commonly used estimation procedure for ARCH models has been the maximum likelihood approach. When the errors are normally distributed, the likelihood function of an ARCH(p) model is

$$f(r_1, r_2, ..., r_T|\alpha) = \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\, h_t}} \exp\left(-\frac{r_t^2}{2\, h_t}\right) f(r_1, r_2, ..., r_p|\alpha), \qquad (2.3)$$

where $\alpha = (\alpha_0, \alpha_1, ..., \alpha_p)'$ and $f(r_1, r_2, ..., r_p|\alpha)$ is the joint probability density function of $r_1, r_2, ..., r_p$. Since the exact form of $f(r_1, r_2, ..., r_p|\alpha)$ is complicated, it is commonly dropped from the prior likelihood function, especially when the sample size is sufficiently large. This results in using the conditional likelihood function

$$f(r_{p+1}, r_{p+2}, ..., r_T|\alpha,\ r_1, r_2, ..., r_p) = \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\, h_t}} \exp\left(-\frac{r_t^2}{2\, h_t}\right). \qquad (2.4)$$

Maximizing the conditional likelihood function is equivalent to maximizing its logarithm, which is easier to handle. The conditional log-likelihood function is

$$l(r_{p+1}, r_{p+2}, ..., r_T|\alpha,\ r_1, r_2, ..., r_p) = \sum_{t=p+1}^{T} \left(-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(h_t) - \frac{r_t^2}{2\, h_t}\right). \qquad (2.5)$$

A variety of alternative estimation methods can also be considered. Least squares and Quasi Maximum Likelihood (QML) estimations in ARCH models were considered in the seminal paper by Engle (1982). The Least Squares Estimator (LSE) for ARCH(p) models is simple to compute but requires existence of higher order moments. An important issue is the possible efficiency loss of the QMLE, resulting

from the use of an inappropriate Gaussian error distribution.

2.2.3 Model Checking

For a properly specified ARCH model, the standardized residuals

$$\tilde{\varepsilon}_t = \frac{r_t}{\sqrt{h_t}}, \ t = 1, 2, \ldots,$$

form a sequence of iid random variables. Therefore, one can check the adequacy of a fitted ARCH model by examining the series $\{\tilde{\varepsilon}_t\}$. In particular, the Ljung-Box statistics of $\tilde{\varepsilon}_t$ can be used to check the adequacy of the mean equation and that of $\tilde{\varepsilon}_t^2$ can be used to test the validity of the volatility equation. The skewness, kurtosis, and QQ-plot of $\{\tilde{\varepsilon}_t\}$ can be used to check the validity of the distribution assumption.

2.2.4 Forecasting

An important use of ARCH models is the evaluation of the accuracy of volatility forecasts. In standard time series methodology which uses conditionally homoscedastic ARMA processes, the variance of the forecast error does not depend on the current information set. If the series being forecasted displays ARCH effect, the current information set can indicate the accuracy by which the series can be forecasted. Engle and Kraft (1983) were the first to consider the effect of ARCH on forecasting. As the conditional variance is a linear function of the squares of the

past observations, one can use the Minimum Mean Square Error (MMSE) method for forecasting the volatility as in the case of classical AR models.

Using the MMSE method, the 1-step-ahead forecast of h_{n+1} at the forecast origin n, for the ARCH(p) model is,

$$h_n(1) = \alpha_0 + \alpha_1 r_n^2 + \cdots + \alpha_p r_{n+1-p}^2.$$

The 2-step-ahead forecast is

$$h_n(2) = \alpha_0 + \alpha_1 h_n(1) + \alpha_2 r_n^2 + \cdots + \alpha_p r_{n+2-p}^2,$$

and the l-step-ahead forecast for h_{n+l} is

$$h_n(l) = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_n(l-i),$$

where $h_n(l-i) = r_{n+l-i}^2$ if $l - i \leq 0$.

Despite the extensive literature on ARCH and related models, relatively little attention is being given to the issue of forecasting in models where time-dependent conditional heteroscedasticity is present. Bollerslev (1986), Diebold (1988), Granger et al. (1989) all discuss the construction of one-step-ahead prediction error intervals with time-varying variances. Engle and Kraft (1983) derive expressions for the multi-step prediction error variance in ARMA models with ARCH errors, but do not further discuss the characteristics of the prediction error distribution. The prediction error distribution is also analysed in Geweke (1989) within a Bayesian

framework using extensive simulation methods. In the next section, we will focus on the generalization of the observation driven models.

2.3 Generalized ARCH Models

The GARCH model is an extension of Engle's work by Bollerslev (1986) that allows the conditional variance to depend on the previous conditional variance and the squares of previous returns. The possibility that estimated parameters in ARCH model do not satisfy the stationarity condition, increases with lag. The GARCH(p,q) is defined by

$$r_t = \sqrt{h_t}\varepsilon_t \ , \quad h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, r_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{2.6}$$

where $\{\varepsilon_t\}$ is a sequence of iid symmetric random variables with mean 0 and variance 1, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$. The constraint on $\alpha_i + \beta_i$ implies that the unconditional variance of r_t is finite, whereas its conditional variance h_t evolves over time.

2.3.1 GARCH(1,1) model and properties

Let us now consider the simple GARCH(1,1) model, which is the most popular for modelling asset return volatility. We represent this model as

$$r_t = \sqrt{h_t}\varepsilon_t \, , \qquad h_t = \alpha_0 + \alpha_1 \, r_{t-1}^2 + \beta_1 h_{t-1}, \tag{2.7}$$

where $\varepsilon_t \sim N(0,1)$ and $0 \leq \alpha_1, \beta_1 \leq 1$, $\alpha_1 + \beta_1 < 1$.

1. The unconditional mean of r_t is zero, since

$$E(r_t) = E(E(r_t|r_{t-1})) = E\left(\sqrt{h_t}E(\varepsilon_t)\right) = 0.$$

2. The conditional variance of r_t is

$$E\left(r_t^2|r_{t-1}\right) = E\left(h_t\varepsilon_t^2|r_{t-1}\right) = h_t E\left(\varepsilon_t^2|r_{t-1}\right) = h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1}.$$

3. The unconditional variance of r_t is

$$\begin{aligned} V(r_t) &= E\left(r_t^2\right) = E\left(E\left(r_t^2|r_{t-1}\right)\right) \\ &= E\left(\alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1}\right) \\ &= \alpha_0 + \alpha_1 E\left(r_{t-1}^2\right) + \beta_1 E\left(h_{t-1}\right). \end{aligned}$$

   This implies that $V(r_t) = \alpha_0/1 - (\alpha_1 + \beta_1)$.

4. The kurtosis of r_t, K_r, is given by

$$K_r = \frac{3\left[1 - (\alpha_1 + \beta_1)^2\right]}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} > 3, \text{ if } 1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0.$$

   Consequently, similar to ARCH models, the tail distribution of GARCH(1,1) process is heavier than that of a normal distribution if $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$.

5. The ACF of $\{r_t\}$ is zero and the ACF of $\{r_t^2\}$ is given by

$$\rho_{r_t^2}(k) = (\alpha_1 + \beta_1)^{k-1} \frac{\alpha_1 (1 - \alpha_1\beta_1 - \beta_1^2)}{1 - 2\alpha_1\beta_1 - \beta_1^2}.$$

2.3.2 Forecasting

Forecasts of a GARCH model can be obtained using methods similar to those of an ARMA model. Consider the GARCH(1,1) model in (2.7) and assume that the forecast origin is n. For 1-step ahead forecast, we have

$$h_{n+1} = \alpha_0 + \alpha_1 r_n^2 + \beta_1 h_n,$$

where r_n and h_n are known at the time index n. Therefore, the 1-step ahead forecast is

$$h_n(1) = \alpha_0 + \alpha_1 r_n^2 + \beta_1 h_n.$$

For multi-step ahead forecasts, we use $r_t^2 = h_t\varepsilon_t^2$ and rewrite the volatility equation in (2.7) as

$$h_{t+1} = \alpha_0 + (\alpha_1 + \beta_1)h_t + \alpha_1 h_t(\varepsilon_t^2 - 1).$$

When $t = n + 1$, the equation becomes

$$h_{n+2} = \alpha_0 + (\alpha_1 + \beta_1)h_{n+1} + \alpha_1 h_{n+1}(\varepsilon_{n+1}^2 - 1).$$

Since $E(\varepsilon_{n+1}^2 - 1|r_{t-1}) = 0$, the 2-step ahead volatility forecast at the forecast origin n satisfies the equation

$$h_n(2) = \alpha_0 + (\alpha_1 + \beta_1)h_n(1).$$

In general, we have

$$h_n(l) = \alpha_0 + (\alpha_1 + \beta_1)h_n(l-1), \;\; l > 1.$$

The multi-step ahead volatility forecasts of a GARCH(1, 1) model converge to the unconditional variance of r_t as the forecast horizon increases to infinity provided that $V(r_t)$ exists, for details see Tsay (2005).

2.4 Stochastic Volatility Models

This is a class of parameter driven models, where the volatility is assumed to follow some latent stochastic process, such as an autoregressive model. An appealing feature of the SV model is its close relationship to financial economic theories. The univariate SV model proposed by Taylor (1986) is given by,

$$r_t = \varepsilon_t \exp\left(h_t/2\right), \quad h_t = \alpha + \beta h_{t-1} + \eta_t, \tag{2.8}$$

where ε_t and η_t are two independent Gaussian white noises, with variances 1 and σ_η^2, respectively. Adding the innovation η_t substantially increases the flexibility of the model in describing the evolution of h_t, but it also increase the difficulty in

parameter estimation. Due to the Gaussianity of η_t, this model is called a log-normal SV model. Its major properties are discussed in Taylor (1986), Taylor (1994).

As η_t is Gaussian, $\{h_t\}$ is a standard Gaussian autoregressive process. It will be (strictly and covariance) stationary if $|\beta| < 1$ with:

$$\mu_h = E(h_t) = \frac{\alpha}{1-\beta}; \quad \sigma_h^2 = V(h_t) = \frac{\sigma_\eta^2}{1-\beta^2}.$$

As $\{\varepsilon_t\}$ is always stationary, $\{r_t\}$ will be stationary if and only if $\{h_t\}$ is stationary, r_t being the product of two stationary process. Using the property of log-normal distribution, all the even moments of r_t exist if h_t is stationary and in particular the kurtosis is

$$K_r = \frac{E(r_t^4)}{E(r_t^2)^2} = 3 \ \exp\left(\sigma_h^2\right) \ \geq 3,$$

which shows that the SV model has fatter tails than the corresponding normal distribution. The dynamic properties of r_t are easy to find. First, as $\{\varepsilon_t\}$ is iid, $\{r_t\}$ is a martingale difference and is a white noise if $|\beta| < 1$. As h_t is a Gaussian AR(1),

$$\begin{aligned} Cov\left(r_t^2, r_{t-k}^2\right) &= E\left(r_t^2 r_{t-k}^2\right) - E\left(r_t^2\right) E\left(r_{t-k}^2\right) \\ &= E\left(\exp\left(h_t + h_{t-k}\right)\right) - \left(E\left(\exp\left(h_t\right)\right)\right)^2 \\ &= \exp\left(2\mu_h + \sigma_h^2\right)\left(\exp\left(\sigma_h^2\beta^k\right) - 1\right) \end{aligned}$$

and so

$$\rho_{r_t^2}(k) = \frac{Cov\left(r_t^2, r_{t-k}^2\right)}{V\left(r_t^2\right)} = \frac{\exp\left(\sigma_h^2\beta^k\right) - 1}{3\exp\left(\sigma_h^2\right) - 1} \simeq \frac{\exp\left(\sigma_h^2\right) - 1}{3\exp\left(\sigma_h^2\right) - 1}\beta^k.$$

Note that, if $\beta < 0$, $\rho_{r_t^2}(k)$ can be negative, unlike ARCH models. This resembles the autocorrelation function of an ARMA(1,1) process. Thus, the SV model behaves in a manner similar to the GARCH(1,1) model. Finally, note that there is no need for non-negativity constraints or for bounded kurtosis constraints on the coefficients. This is a great advantage with respect to GARCH models. A review of the properties of SV model may be found in Taylor (1994) and Tsay (2005).

Despite theoretical advantages, the SV models have not been popular as the ARCH models in practical applications. The main reason is that the likelihood function for the SV models is not easy to evaluate unlike for the ARCH models. The likelihood function of the parameters $\theta = (\alpha, \beta, \sigma_\eta^2)$ based on $(r_1, r_2, ..., r_T)$ may be written as

$$L(\theta) = \int_h P_\theta(r|h)P_\theta(h)dh$$

$$= \int_{h_T} \int_{h_{T-1}} ... \int_{h_1} \prod_{t=1}^{T} P_\theta(r_t|h_t)P_\theta(h_t|h_{t-1})dh_1...dh_T.$$

The multiple integral $L(\theta)$ cannot be factored in to a product of T one-dimensional integrals because of the dependence of h_t on the past. Exact evaluation of the likelihood is possible through Kalman filter only for the case in which both $P_\theta(r_t|h_t)$ and $P_\theta(h_t|h_{t-1})$ are Gaussian. But Gaussian distributions poorly represent volatilities, because they are defined over the complete real line where as the volatility distribution is thought to be highly skewed on the right. In the SV model, the positivity constraint on h_t makes distributions such as log-normal suitable for $P_\theta(h_t|h_{t-1})$. Due to the difficulties in obtaining explicit forms of MLE for a SV model, several numerical methods are proposed by several authors. Markov Chain Monte Carlo

(MCMC) procedures (see Jacquier et al. (1994) and Kim et al. (1998)) is a commonly used method for numerical estimation. A variety of other estimation procedures has also been proposed, including, the Generalized Method of Moments (GMM) used by Melino and Turnbull (1990), the Quasi Maximum Likelihood (QML) approach followed by Harvey et al. (1994) and Ruiz (1994) and the Efficient Method of Moments (EMM) applied by Gallant et al. (1997). For a survey of these estimation procedures, one can refer Ghysels et al. (1996) and Broto and Ruiz (2004).

2.5 State-Space Approach and Kalman Filter

The state-space model provides a flexible approach to time series analysis. A wide range of all linear and many non-linear time series models can be handled, including regression models with changing coefficients, autoregressive integrated moving average (ARIMA) models and unobserved component models. A state-space model consists of a state equation and an observation equation. While the state equation formulates the dynamics of the state variables, the observation equation relates the observed variables to the unobserved state vector. The state variable and the parameters have to be estimated from the data. Maximum likelihood estimates of the parameters can be obtained by applying the Kalman filter. It is a recursive algorithm that computes estimates for the unobserved components at time t, based on the available information at the same time. Many dynamic time series models in economics and finance can be represented in state-space form.

Once a model is put into state-space form, the Kalman filter can be employed to update knowledge of the state variable recursively when a new data point becomes available. Here, we briefly discuss the derivation of Kalman filter and smoothing algorithms using local trend model. The Kalman filter method assumes that the distributions of underlying rvs are normal.

Consider the univariate time series w_t satisfying

$$w_t = \mu_t + e_t; \;\; \mu_t = \mu_{t-1} + a_{t-1}, \tag{2.9}$$

where $\{e_t\}$ and $\{a_t\}$ are two independent Gaussian sequences with mean 0 and variances σ_e^2, and σ_a^2 respectively.

Let $F_t = \{w_1, w_2, \ldots, w_T\}$ be the information available at time t. The conditional mean and variance of μ_t given F_j is respectively given by $\mu_{t|j} = E(\mu_t|F_j)$ and $\Omega_{t|j} = V(\mu_t|F_j)$. Similarly, $w_{t|j}$ denotes the conditional mean of w_t given F_j. Furthermore, let $\upsilon_t = w_t - w_{t|t-1}$ and $f_t = V(\upsilon_t|F_{t-1})$ be the 1-step ahead forecast error and the variance of w_t given F_{t-1}, respectively. Note that the forecast error υ_t is independent of F_{t-1} so that the conditional variance is same as the unconditional variance; that is, $V(\upsilon_t|F_{t-1}) = V(\upsilon_t)$. From (2.9),

$$w_{t|t-1} = E(w_t|F_{t-1}) = E\left(\mu_t + e_t|F_{t-1}\right) = E\left(\mu_t|F_{t-1}\right) = \mu_{t|t-1}.$$

Consequently,

$$\upsilon_t = w_t - w_{t|t-1} = w_t - \mu_{t|t-1}$$

and

$$f_t = V(\upsilon_t|F_{t-1}) = V(w_t - \mu_{t|t-1}|F_{t-1}) = V(\mu_t + e_t - \mu_{t|t-1}|F_{t-1})$$
$$= V(\mu_t - \mu_{t|t-1}|F_{t-1}) + V(e_t|F_{t-1}) = \Omega_{t|t-1} + \sigma_e^2.$$

To derive the Kalman filter, it suffices to consider the joint conditional distribution of $(\mu_t, \upsilon_t)'$ given F_{t-1}. The conditional distribution of υ_t given F_{t-1} is normal with mean zero and variance f_t and that of μ_t given F_{t-1} is normal with mean $\mu_{t|t-1}$ and variance $\Omega_{t|t-1}$. Furthermore, the joint distribution of $(\mu_t, \upsilon_t)'$ given F_{t-1} is also normal. From the definition, we have $Cov(\mu_t, \upsilon_t|F_{t-1}) = \Omega_{t|t-1}$. Now, we state a theorem in multivariate normal distribution, whose proof may be found in Anderson (1954).

Theorem 2.1. *Let the components of X be divided into two groups composing the sub-vectors $X^{(1)}$ and $X^{(2)}$. Suppose that the mean of X, μ is similarly divided into $\mu^{(1)}$ and $\mu^{(2)}$, and the covariance matrix Ω of X is divided into Ω_{11}, the covariance matrix of $X^{(1)}$, Ω_{12}, the covariance matrix of $X^{(1)}$ and $X^{(2)}$, and Ω_{22}, the covariance matrix of $X^{(2)}$. Then if the distribution of X is normal, the conditional distribution of $X^{(1)}$ given $X^{(2)} = x^{(2)}$ is normal with mean $\mu^{(1)} + \Omega_{12}\Omega_{22}^{-1}(x^{(2)} - \mu^{(2)})$ and covariance matrix $\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$.*

Using this theorem for the above state-space model, it follows that the conditional distribution of μ_t given F_t is normal with mean and variance

$$\mu_{t|t} = \mu_{t|t-1} + \Omega_{t|t-1}\upsilon_t/f_t = \mu_{t|t-1} + K_t\upsilon_t,$$

and

$$\Omega_{t|t} = \Omega_{t|t-1} - \Omega_{t|t-1}^2/f_t = \Omega_{t|t-1}(1 - K_t),$$

where $K_t = \Omega_{t|t-1}/f_t$ is commonly referred to as the Kalman gain.

Next, one can make use of the knowledge of μ_t given F_t to predict μ_{t+1}. We have

$$\mu_{t+1|t} = E(\mu_t + a_t|F_t) = E(\mu_t|F_t) = \mu_{t|t},$$

$$\Omega_{t+1|t} = V(\mu_{t+1}|F_t) = V(\mu_t|F_t) + V(a_t) = \Omega_{t|t} + \sigma_a^2.$$

Once the new data w_{t+1} is observed, one can repeat the above procedure to update the knowledge on μ_{t+1}. This is the famous Kalman filter algorithm proposed by Kalman (1960). We use this procedure for estimating unobservable volatility in Chapter 5.

Chapter 3

Gumbel Extreme Value Autoregressive Model

3.1 Introduction

The problem of modelling extreme events arises in many areas where such events can have serious consequences on the event generating system. Some examples of such events include extreme floods and snowfalls, high wind speeds, extreme temperatures, large fluctuations in exchange rates, and market crashes. To develop appropriate probabilistic models and assess the risks caused by these events, business analysts and engineers frequently use the extreme value distributions. Extreme value distributions are generally considered to comprise the following three families, (see Johnson et al. (1994)),

Type 1:

$$F(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), \quad x \in \mathbf{R} \tag{3.1a}$$

Type 2:

$$F(x) = \begin{cases} 0, & x < \mu, \\ \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-k}\right), & x \geq \mu \end{cases} \tag{3.1b}$$

Type 3:

$$F(x) = \begin{cases} \exp\left(-\left(\frac{\mu-x}{\sigma}\right)^{k}\right), & x \leq \mu \\ 1, & x > \mu \end{cases} \tag{3.1c}$$

where μ, $\sigma(>0)$ and $k(>0)$ are parameters. The Type 1 distribution is referred to as the Gumbel-type distribution; the Type 2 distribution is called as the Fréchet-type distribution; and the Type 3 distribution as the Weibull-type distribution. The Type 2 and Type 3 distributions are closely related to the Type 1 distribution by the simple transformations

$$W = \log(X - \mu), \quad W = -\log(\mu - X),$$

respectively. Type 1 is sometimes called the log-Weibull distribution. Extreme value distributions were obtained as limiting distributions of greatest (or least) values in random samples of increasing size.

Among these distributions, Gumbel distribution played an important role in analysing

the extreme value data. The Type 1 (Gumbel) distribution is obtained as the limiting distribution of $M_n = \max(X_1, X_2, ..., X_n)$ after properly centralizing and normalizing, where $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) rvs. In this chapter, we develop a first order autoregressive (AR(1)) process with Type 1 distribution as the stationary marginal distribution.

Corresponding distribution for the minimum can be obtained using the relation,

$$m_n = \min(X_1, X_2, ..., X_n) = -\max(-X_1, -X_2, ..., -X_n).$$

Thus, the cumulative limiting distribution function corresponding to the minimum (or least) values is given by

$$P(X \leq x) = 1 - \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right).$$

As far as the properties of the distribution are concerned, there is no much difference between Gumbel minimum and Gumbel maximum distribution. We employ Gumbel minimum distribution for modelling volatility in Chapter 5.

In the classical method of analysis, it is customary to assume that the variables generating events are statistically independent. However, such assumptions hold rarely in practice and the more practical study requires the method to deal with sequences of dependent random variables. Statistical analysis of time series is an appropriate area where one can deal with such problems. In the classical set up, the time series analysis assumes that the observed series is a realization of a Gaussian

time series generated by a linear model. But in situations where extreme events are expected as stated above, the normal time series is not a suitable choice.

Hughes et al. (2007) studied the time series modelling of monthly extreme temperatures in the Antarctic Peninsula using generalized extreme value innovations. However, there was no discussion on the explicit form of the stationary marginal distribution of the time series. Toulemonde et al. (2010) applied autoregressive models with stationary marginal Gumbel Extreme Value (GEV) distribution to model the maximum of certain environmental variables. The authors use the method of moments to estimate the model parameters. The detailed analysis of class of extreme value distributions and their applications in modelling interest rates and other financial related characteristics may be found in Bali (2003) and the references there in. Nakajima et al. (2012) proposed a state-space approach to model the time dependence in an extreme value process, assuming that the state variables follow ARMA models with Gumbel extreme value distributed innovations. This distribution also plays an important role in the study of limiting distribution of the extremes of stationary time series. A recent survey on statistical analysis of extremes of time series may be found in Chavez-Demoulin and Davison (2012).

In this chapter, we provide a systematic development of the AR(1) model which generates a stationary sequence of GEV distributed random variables and study its probabilistic and statistical properties. In Section 3.2, we discuss the properties of Gumbel extreme value AR(1) model. Section 3.3 discusses the problem of parameter estimation. A simulation study is carried out to assess the performance of the estimators and the findings are described in Section 3.4. An application of

the proposed model is illustrated using Bombay Stock Exchange (BSE) index and Standard and Poor 500 (S&P 500) index in Section 3.5.

3.2 Model and Properties

Let $\{\eta_t\}$ be a sequence of independent and identically distributed rvs and define a stationary AR(1) sequence $\{X_t\}$ by

$$X_t = \alpha X_{t-1} + \eta_t, \;\; 0 \leq \alpha < 1, \quad t = 1, 2, ..., \tag{3.2}$$

such that X_0 is independent of η_1. We assume the marginal distribution of $\{X_t\}$ to be an extreme value Type 1 with cdf given in (3.1a). A major problem in the analysis of non-Gaussian time series is to find the distribution of the innovation rv η_t for a specified marginal distribution of $\{X_t\}$.

Theorem 3.1. *The stationary marginal distribution of* $\{X_t\}$ *in model* (3.2) *is GEV distribution with parameters* μ *and* σ *iff the distribution of* η_t *is specified by*

$$\eta_t \stackrel{L}{=} (1-\alpha)\mu + \sigma Z, \text{ and } Z \stackrel{L}{=} -\log\left(U^{-\alpha}\right),$$

where $\stackrel{L}{=}$ *denotes the equality in distribution and* U *denotes a positive stable rv.*

Proof: It is well known that, for a specified marginal distribution, F of $\{X_t\}$ in model (3.2), there exist a proper distribution for the innovation η_t if and only if F is self-decomposable.

In terms of the characteristic function, $\phi_X(s) = E\left(\exp(isX)\right)$, the random variable

X is said to be self-decomposable if the ratio $\phi_X(s)/\phi_X(\alpha s)$ is a well defined characteristic function for every α, $0 \leq \alpha < 1$.

Now, the characteristic function of the rv X with cdf (3.1a) is given by

$$\phi_X(s) = e^{is\mu}\,\Gamma(1 - is\sigma), \tag{3.3}$$

where $\Gamma(.)$ is the gamma integral and $i = \sqrt{-1}$ is the complex number.

The definition of the model (3.2) implies that the rvs X_{t-1} and η_t are statistically independent for every t and hence in terms of characteristic functions, (3.2) leads to the equation

$$\begin{aligned} \phi_X(s) = \phi_X(\alpha s)\phi_\eta(s) \Rightarrow \phi_\eta(s) &= \frac{\phi_X(s)}{\phi_X(\alpha s)} \\ \Rightarrow \phi_\eta(s) &= e^{is(1-\alpha)\mu}\frac{\Gamma(1 - is\sigma)}{\Gamma(1 - i\alpha s\sigma)}, \end{aligned} \tag{3.4}$$

using (3.3). The cdf of η_t can be obtained by inverting the characteristic function on the right hand side of (3.4). Using the results of Brockwell and Brown (1978) it is readily verified that

$$\eta_t \stackrel{L}{=} (1 - \alpha)\mu + \sigma Z, \quad \text{with} \quad Z \stackrel{L}{=} -\log\left(U^{-\alpha}\right), \tag{3.5}$$

where $\stackrel{L}{=}$ denotes the equality in distribution and U denotes a positive stable rv with Laplace Transform (L.T.), $\varphi(s) = \exp(-s^\alpha)$.

Conversely, if η_t is independent and identically distributed with distribution specified by (3.5) and X_0 follows an arbitrary distribution, then we have to prove that $\{X_t\}$ is asymptotically a stationary process with GEV marginal distribution.

The model (3.2) can be written in terms of η_t's as,

$$X_t = \alpha^t X_0 + \sum_{k=0}^{t-1} \alpha^k \eta_{t-k}.$$

Then the characteristic function of X_t is given by

$$\begin{aligned}
\phi_X(s) &= \phi_{X_0}\left(\alpha^t s\right) \prod_{k=0}^{t-1} \phi_\eta(\alpha^k s) \\
&= \phi_{X_0}\left(\alpha^t s\right) \prod_{k=0}^{t-1} e^{i\alpha^k s(1-\alpha)\mu} \frac{\Gamma(1 - i\alpha^k s\sigma)}{\Gamma(1 - i\alpha^{k+1} s\sigma)} \\
&= \phi_{X_0}\left(\alpha^t s\right) e^{\left(i\, s(1-\alpha)\mu \sum_{k=0}^{t-1} \alpha^k\right)} \frac{\Gamma(1 - i\alpha^0 s\sigma)}{\Gamma(1 - i\alpha^t s\sigma)} \\
&\to e^{i\, s\mu}\, \Gamma(1 - is\sigma), \qquad \text{as } t \to \infty \text{ and } \alpha^t \to 0, \quad \text{where } 0 \le \alpha < 1.
\end{aligned}$$

Hence the theorem.

The probability density function (pdf) of Z in (3.5) can be expressed as an infinite series,

$$f_Z(z;\alpha) = \frac{1}{\pi} \sum_{k=1}^{\infty} (-1)^{k-1} \frac{\Gamma(k\alpha)}{\Gamma(k)} (e^{-z})^k \sin(k\pi\alpha), \tag{3.6}$$

which is also referred to as log-Mittag-Leffler density in the literature (cf, Blumenfeld and Mandelbrot (1997)). The mean and variance of η_t are given by

$$E(\eta_t) = (1-\alpha)\,(\mu + \sigma\,\gamma) = \mu^*, \quad \text{(say)} \quad \text{and} \quad V(\eta_t) = (1-\alpha^2)\,\frac{\pi^2\sigma^2}{6} = \sigma^{2*}, \quad \text{(say)}, \tag{3.7}$$

where $\gamma \approx 0.5772$ is the Euler's constant. We refer the stationary AR(1) sequence with GEV marginal distribution as GEVAR(1) (Gumbel extreme value autoregressive of order one) sequence. The second order properties of the GEVAR(1) process

are summarized below.

$$E(X_t) = \mu + \gamma\sigma, \;\; V(X_t) = \frac{\pi^2}{6}\sigma^2 \;\text{ and the ACF, }\; \rho_{\mathrm{k}} = \alpha^{\mathrm{k}}, \;\; \mathrm{k} = 1, 2, \ldots. \tag{3.8}$$

The regression of X_t on X_{t-1} is given by

$$\begin{aligned} E(X_t|X_{t-1}) &= \alpha X_{t-1} + E(\eta_t|X_{t-1}) \\ &= \alpha X_{t-1} + (1-\alpha)(\mu + \sigma\gamma) \\ &= g(\theta; X_{t-1}) \;\text{ with }\; \theta = (\alpha, \mu, \sigma)'. \end{aligned} \tag{3.9}$$

Now, the conditional cdf of X_t given $X_{t-1} = x_{t-1}$ is expressed as,

$$\begin{aligned} P(X_t \leq x_t|X_{t-1} = x_{t-1}) &= P(\alpha X_{t-1} + \eta_t \leq x_t|X_{t-1} = x_{t-1}) \\ &= P(\eta_t \leq x_t - \alpha\, x_{t-1}). \end{aligned}$$

Then the one-step transition density function of $\{X_t\}$ defined by (3.2) can be obtained as,

$$f(x_t|x_{t-1}) = f_{\eta_t}(x_t - \alpha x_{t-1}) = \frac{1}{\sigma} f_Z\left(\frac{x_t - \alpha x_{t-1} - (1-\alpha)\mu}{\sigma}; \alpha\right). \tag{3.10}$$

The form of the transition density (3.10) shows that the likelihood function of the model parameters for the Markov sequence $\{X_t\}$ will not have a closed form expression. Hence, we have to employ some numerical methods for obtaining the Maximum Likelihood Estimate (MLE) of the parameters. In the next section, we propose the method of Conditional Least Squares introduced by Klimko and Nelson

(1978) to estimate θ and then consider its further improvement in the subsequent sections.

3.3 Estimation of Model Parameters

The estimation of model parameters is one of the important problems involved in modelling of non-Gaussian time series. If the stationary marginal distribution of an AR(p) model is not specified, then the estimation methods such as Conditional Least Squares (CLS) and Quasi Maximum Likelihood (QML) may be preferred. For a specified marginal distribution, one can go for maximum likelihood method of estimation, if there is a closed form density for the innovation rv. The innovation rvs do not have closed form densities for most of the non-Gaussian AR models. In view of this, we include the detailed analysis of CLSE and QMLE for the proposed AR(1) model. Even though, the innovation density does not have a closed form, we propose the ML method of estimation using the properties of positive stable distribution.

3.3.1 Parameter estimation by the method of Conditional Least Squares

Klimko and Nelson (1978) developed an estimation procedure for dependent observations based on the minimization of a sum of squared deviations about conditional expectations. This approach referred to as "conditional least squares", provides a unified treatment of estimation problems for widely used classes of stochastic

models. The method is implicit in the observation of Mann and Wald (1943), Durbin (1960) and others that the assumption of normally distributed error terms in autoregressive models renders maximum likelihood estimation equivalent to the minimization of a sum of squares. On a light note, here we briefly explain about the method of conditional least squares developed by Klimko and Nelson (1978).

Let $\{X_t,\ t = 1,\ 2, ...\}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, P_\theta)$, whose distribution depends on an unknown parameter vector $\theta = (\theta_1,\ \theta_2,\ ...,\ \theta_p)'$. Let $\{\mathcal{F}_{t-1}\}_{t=1}^{\infty}$ denote a sequence of sub-sigma fields with $\mathcal{F}_{t-1}$ generated by an arbitrary subset of $\{X_1,\ X_2,\ ...,\ X_{t-1}\}$, $t > 1$. Then the CLS estimator of the parameters are obtained by minimizing the conditional sum of squares,

$$Q_T(\theta) = \sum_{t=1}^{n} \left[X_t - g(\theta;\ \mathcal{F}_{t-1})\right]^2,$$

with respect to the parameter vector $\theta = (\theta_1,\ \theta_2,\ ...,\ \theta_p)'$, where $g(\theta;\ \mathcal{F}_{t-1}) = E(X_t|\ \mathcal{F}_{t-1})$. The estimates are obtained by solving the least squares equations

$$\frac{\partial\, Q_T(\theta)}{\partial\, \theta_i} = 0,\ i = 1,\ 2,\ ...,\ p.$$

Under a set of regularity conditions, Klimko and Nelson (1978) proved that the CLS estimators are strongly consistent and asymptotically jointly normally distributed. The rate of convergence of the estimators is found to be $((\log \log T)/T)^{1/2}$. The assumptions made concern the application of strong laws, central limit theorems and laws of iterated logarithm to sums of dependent random variables. A wide variety of conditions under which these hold may be found in Stout (1974), McLeish (1974), and Heyde and Scott (1973). These conditions are generally a trade-off among

moment assumptions, stationarity, the martingale property, mixing, ergodicity and the Markov property; with no one set of assumptions being most universal.

Let $\{X_t\}_{t=0}^{\infty}$ be a stationary ergodic sequence of integrable rvs. Then the function $g = g(\theta;\, \mathcal{F}_{t-1})$ satisfies the following regularity conditions:

(i) $\partial g/\partial\theta_i$, $\partial^2 g/\partial\theta_i\,\partial\theta_j$ and $\partial^3 g/\partial\theta_i\,\partial\theta_j\,\partial\theta_k$ exists and are continuous for all θ, $i \leq p$, $j \leq p$, $k \leq p$;

(ii) For $i \leq p$, $j \leq p$, $E\,|(X_t - g)\,\partial g/\partial\theta_i| < \infty$, $E\,|(X_t - g)\,\partial^2 g/\partial\theta_i\,\partial\theta_j| < \infty$ and $E\,|\partial g/\partial\theta_i\;\partial g/\partial\theta_j| < \infty$ where g and its partial derivatives are evaluated at true value of parameters θ and X_{t-1};

(iii) For i, j, $k = 1, 2, ..., p$, there exist functions

$H^{(0)} \equiv H^{(0)}(X_{t-1}, X_{t-2}, ..., X_0)$, $H_i^{(1)} \equiv H_i^{(1)}(X_{t-1}, X_{t-2}, ..., X_0)$,
$H_{ij}^{(2)} \equiv H_{ij}^{(2)}(X_{t-1}, X_{t-2}, ..., X_0)$, $H_{ijk}^{(3)} \equiv H_{ijk}^{(3)}(X_{t-1}, X_{t-2}, ..., X_0)$,
such that
$|g| \leq H^{(0)}$, $|\partial g/\partial\theta_i| \leq H_i^{(1)}$, $|\partial^2 g/\partial\theta_i\,\partial\theta_j| \leq H_{ij}^{(2)}$, $|\partial^3 g/\partial\theta_i\,\partial\theta_j\,\partial\theta_k| \leq H_{ijk}^{(3)}$,
for all θ and $E\left|X_t\,H_{ijk}^{(3)}\right| < \infty$, $E\{H^{(0)}\,H_{ijk}^{(3)}\} < \infty$, $E\{H_i^{(1)}\,H_{jk}^{(2)}\} < \infty$.

Then we have the following lemma.

Lemma 3.2. *Let $\{X_t\}$ be a stationary and ergodic Markov sequence with finite third order moments. Under the regularity conditions listed above, the CLS estimator $\hat{\theta}$ of θ is consistent and asymptotically normal (CAN). That is, as $T \to \infty$*

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{L} N_p(0,\, V^{-1}W\,V^{-1})$$

where V and W are $p \times p$ matrices, whose $(i,\, j)^{th}$ elements are respectively given by

$$V_{ij} = E\left(\frac{\partial\, g(\theta;\, \mathcal{F}_{t-1})}{\partial\theta_i}.\frac{\partial\, g(\theta;\, \mathcal{F}_{t-1})}{\partial\theta_j}\right),\ i,\, j = 1,\, 2, ..., p$$

and

$$W_{ij} = E\left(u_t^2(\theta)\frac{\partial\, g(\theta;\, \mathcal{F}_{t-1})}{\partial\theta_i}\,\frac{\partial\, g(\theta;\, \mathcal{F}_{t-1})}{\partial\theta_j}\right), \quad i,\, j = 1,\, 2, ...,\, p,$$

where $u_t = X_t - g(\theta;\, \mathcal{F}_{t-1})$.

We employ this method to estimate θ of our GEVAR(1) model. Let $(x_1, x_2, ..., x_T)$ be a realization from the stationary GEVAR(1) sequence discussed in the earlier section. The CLS estimate of the parameter vector $\theta = (\alpha, \mu, \sigma)'$ is obtained by minimizing the conditional error sum of squares

$$Q(\theta; x_1, ..., x_T) = \sum_{t=1}^{T} \left(x_t - g(\theta; x_{t-1})\right)^2, \tag{3.11}$$

where $g(.;.)$ is the conditional expectation given in (3.9). We assume that the scale parameter σ is known to avoid the problem of identification and fix $\sigma = 1$. The CLS estimates of $\theta = (\theta_1, \theta_2)' = (\alpha, \mu)'$ are obtained by solving the conditional least squares equations,

$$\begin{aligned}\frac{\partial Q(\theta; x_1, ..., x_T)}{\partial\alpha} = 0 \Rightarrow \frac{\partial}{\partial\alpha}\left(\sum_{t=1}^{T}\left(x_t - \alpha x_{t-1} - (1-\alpha)(\mu+\gamma)\right)^2\right) = 0 \\ \Rightarrow 2\sum_{t=1}^{T}\left(x_t - \alpha x_{t-1} - (1-\alpha)(\mu+\gamma)\right)\left(-x_{t-1} + (\mu+\gamma)\right) = 0\end{aligned}$$

$$\Rightarrow -\sum_{t=1}^{T}(x_t-\alpha x_{t-1})\,x_{t-1}+\sum_{t=1}^{T}(x_t-\alpha x_{t-1})\,(\mu+\gamma)$$
$$+(1-\alpha)(\mu+\gamma)\sum_{t=1}^{T}x_{t-1}-T(1-\alpha)(\mu+\gamma)^2=0.$$

$$\frac{\partial Q(\theta;x_1,...,x_T)}{\partial\mu}=0\Rightarrow\frac{\partial}{\partial\mu}\left(\sum_{t=1}^{T}(x_t-\alpha x_{t-1}-(1-\alpha)(\mu+\gamma))^2\right)=0$$
$$\Rightarrow 2\sum_{t=1}^{T}(x_t-\alpha x_{t-1}-(1-\alpha)(\mu+\gamma))\,(-(1-\alpha))=0$$
$$\Rightarrow\sum_{t=1}^{T}(x_t-\alpha x_{t-1})\,-T(1-\alpha)(\mu+\gamma)=0.$$

The solutions are given by

$$\hat{\alpha}=\frac{\sum_{t=1}^{T}x_tx_{t-1}-(1/T)\sum_{t=1}^{T}x_t\sum_{t=1}^{T}x_{t-1}}{\sum_{t=1}^{T}x_{t-1}^2-(1/T)\left(\sum_{t=1}^{T}x_{t-1}\right)^2}\ \text{ and }\ \hat{\mu}=\frac{\sum_{t=1}^{T}(x_t-\hat{\alpha}x_{t-1})}{T(1-\hat{\alpha})}-\gamma, \tag{3.12}$$

where $\gamma\approx 0.5772$.

Since the GEVAR(1) process is stationary and ergodic with all moments finite, it is readily verified that all the regularity conditions of Klimko and Nelson (1978) are satisfied. We briefly indicate the verification here.

To begin with, consider

$$\frac{\partial g}{\partial\alpha}=X_{t-1}-(\mu+\gamma)\,;\ \ \frac{\partial g}{\partial\mu}=1-\alpha\,;\ ,$$

and conclude that $\partial g/\partial\theta_i, i=1,2$ exist and are continuous for all θ.

Similarly, we can show that $\partial^2 g/\partial\theta_i\,\partial\theta_j, i, j = 1, 2$ also exist and are continuous for all θ. Hence, condition (i) is satisfied.

In order to prove conditions (ii) and (iii), we need to find the appropriate H functions. If we define

$$H^{(0)} = \alpha\, X_{t-1} + \mu + \gamma,$$

it follows that, since $0 \leq \alpha < 1$, $|g| < H^{(0)}$ for $\sigma = 1$. Define

$$H_i^{(1)} = \begin{cases} X_{t-1}, & i = 1 \\ 1, & i = 2 \end{cases}.$$

Then we see that, for all θ, $|\partial\, g/\partial\theta_i| < H_i^{(1)}, i = 1, 2$. The $H_i^{(1)}$ were obtained by omitting the negative terms in $\partial\, g/\partial\theta_i, i = 1, 2$. Similarly, by omitting the negative terms from the expressions for $\partial^2 g/\partial\theta_i\,\partial\theta_j, i, j = 1, 2$, we can obtain the function $H_{ij}^{(2)}, i,\ j = 1, 2$ (Billard and Mohamed (1991)). In our case, it is zero.

Since the H functions are linear in X_t's, we have all the H functions are square integrable. Hence, from the Cauchy-Schwartz inequality, the integrability requirements of conditions (ii) and (iii) are satisfied.

Hence, the CLS estimators obtained above are CAN for the corresponding parameters. Now applying the Lemma 3.2, we have proved that,

$$\sqrt{T}\,((\hat{\alpha}, \hat{\mu})' - (\alpha, \mu)') \xrightarrow{L} N_2(0, D),$$

where $N_2(0, D)$ stands for a bivariate normal distribution with mean vector 0 and dispersion matrix $D = V^{-1}WV^{-1}$ with

$$V = \begin{pmatrix} E\left(\frac{\partial g(\theta;X_{t-1})}{\partial \alpha}\right)^2 & E\left(\frac{\partial g(\theta;X_{t-1})}{\partial \alpha} \ \frac{\partial g(\theta;X_{t-1})}{\partial \mu}\right) \\ E\left(\frac{\partial g(\theta;X_{t-1})}{\partial \alpha} \ \frac{\partial g(\theta;X_{t-1})}{\partial \mu}\right) & E\left(\frac{\partial g(\theta;X_{t-1})}{\partial \mu}\right)^2 \end{pmatrix}$$

$$W = \begin{pmatrix} E\left(u_t \ \frac{\partial g(\theta;X_{t-1})}{\partial \alpha}\right)^2 & E\left(u_t^2 \ \frac{\partial g(\theta;X_{t-1})}{\partial \alpha} \ \frac{\partial g(\theta;X_{t-1})}{\partial \mu}\right) \\ E\left(u_t^2 \ \frac{\partial g(\theta;X_{t-1})}{\partial \alpha} \ \frac{\partial g(\theta;X_{t-1})}{\partial \mu}\right) & E\left(u_t \ \frac{\partial g(\theta;X_{t-1})}{\partial \mu}\right)^2 \end{pmatrix},$$

where $u_t = X_t - g(\theta;\, X_{t-1})$.

For a stationary GEVAR(1) sequence defined by (3.2), the elements of the matrix V are

$$\begin{aligned} V_{11} &= E\left(\frac{\partial}{\partial \alpha}\left(\alpha X_{t-1} + (1-\alpha)(\mu+\gamma)\right)\right)^2 = E\left(X_{t-1} - (\mu+\gamma)\right)^2 = \frac{\pi^2}{6}. \\ V_{12} &= E\left(\frac{\partial}{\partial \alpha}\left(\alpha X_{t-1} + (1-\alpha)(\mu+\gamma)\right)\frac{\partial}{\partial \mu}\left(\alpha X_{t-1} + (1-\alpha)(\mu+\gamma)\right)\right) \\ &= E\left(\left(X_{t-1} - (\mu+\gamma)\right)(1-\alpha)\right) \\ &= (1-\alpha)E\left(X_{t-1} - (\mu+\gamma)\right) = 0 = V_{21}. \\ V_{22} &= E\left(\frac{\partial}{\partial \mu}\left(\alpha X_{t-1} + (1-\alpha)(\mu+\gamma)\right)\right)^2 = (1-\alpha)^2. \end{aligned}$$

and that of W are

$$\begin{aligned} W_{11} &= E\left(u_t \ \frac{\partial}{\partial \alpha}\left(\alpha X_{t-1} + (1-\alpha)(\mu+\gamma)\right)\right)^2 \\ &= E\left(\left(X_t - \alpha X_{t-1} - (1-\alpha)(\mu+\gamma)\right)\left(X_{t-1} - (\mu+\gamma)\right)\right)^2 \\ &= (1-\alpha^2)\left(\frac{\pi^2}{6}\right)^2. \end{aligned}$$

$$
\begin{aligned}
W_{12} &= E\left(u_t^2 \frac{\partial}{\partial \alpha}\left(\alpha X_{t-1} + (1-\alpha)(\mu+\gamma)\right) \frac{\partial}{\partial \mu}\left(\alpha X_{t-1} + (1-\alpha)(\mu+\gamma)\right)\right) \\
&= E\left(\left(X_t - \alpha X_{t-1} - (1-\alpha)(\mu+\gamma)\right)^2 \left(X_{t-1} - (\mu+\gamma)\right)(1-\alpha)\right) = 0 = W_{21}. \\
W_{22} &= E\left(u_t \frac{\partial}{\partial \mu}\left(\alpha X_{t-1} + (1-\alpha)(\mu+\gamma)\right)\right)^2 \\
&= E\left(\left(X_t - \alpha X_{t-1} - (1-\alpha)(\mu+\gamma)\right)(1-\alpha)\right)^2 \\
&= (1-\alpha)^2 E\left(X_t - \alpha X_{t-1} - (1-\alpha)(\mu+\gamma)\right)^2 \\
&= (1-\alpha)^2(1-\alpha^2)\frac{\pi^2}{6}.
\end{aligned}
$$

Therefore, V and W can be written as

$$
V = \begin{pmatrix} \frac{\pi^2}{6} & 0 \\ 0 & (1-\alpha)^2 \end{pmatrix};
$$

$$
\begin{aligned}
W &= \begin{pmatrix} (1-\alpha^2)\left(\frac{\pi^2}{6}\right)^2 & 0 \\ 0 & (1-\alpha)^2(1-\alpha^2)\left(\frac{\pi^2}{6}\right) \end{pmatrix} \\
&= (1-\alpha^2)\left(\frac{\pi^2}{6}\right)\begin{pmatrix} \frac{\pi^2}{6} & 0 \\ 0 & (1-\alpha)^2 \end{pmatrix} = (1-\alpha^2)\left(\frac{\pi^2}{6}\right) V.
\end{aligned}
$$

Then the above dispersion matrix becomes

$$
D = \begin{pmatrix} 1-\alpha^2 & 0 \\ 0 & \frac{\pi^2(1+\alpha)}{6(1-\alpha)} \end{pmatrix}. \tag{3.13}
$$

That is, the CLSE of α and μ are asymptotically independent. Even though the

CLS estimates can be easily computed, they are biased as can be seen from the simulation results listed in Table 3.1 of the Section 3.4. Further, one of the model parameters is not identifiable. These drawbacks suggest that we need to look for better estimates.

3.3.2 Method of Quasi Maximum Likelihood Estimation

The log-likelihood function of $\theta = (\alpha, \mu, \sigma)'$ based on a realization $(x_1, x_2, ..., x_T)$ from the stationary GEVAR(1) model (3.2) is given by

$$\begin{aligned} L_T(\theta; x_1, x_2, ..., x_T) &= \log f_X(x_1; \theta) + \sum_{t=2}^{T} \log f(x_t; \theta | x_{t-1}) \\ &= \log f_X(x_1; \theta) + \sum_{t=2}^{T} \log \left(\frac{1}{\sigma} f_Z \left(\frac{x_t - \alpha x_{t-1} - (1-\alpha)\mu}{\sigma}; \alpha \right) \right), \end{aligned} \tag{3.14}$$

where $f_X(.)$ is the pdf corresponding to the cdf (3.1a) and $f_Z(.)$ is the density function given by (3.6). The structure of the innovation distribution described in Section 3.2 indicates that it is difficult to express the likelihood function in a closed form. The method of quasi maximum likelihood (QML) is proposed to estimate the model parameters when the exact form of the error distribution is unknown or when it does not have a tractable form. For a general discussion on the QML method and asymptotic properties of the resulting estimators one may refer White (1982). A recent paper by Aue et al. (2006) obtained the QML estimates of the parameters of random coefficient autoregressive models. We follow their method to obtain QMLE of the parameters of our GEVAR(1) model described above. Accordingly we write

down the likelihood function for the model (3.2) by assuming that the innovations are normally distributed with mean and variance given in (3.7). Thus the pdf of η_t becomes,

$$f(\eta_t) = \frac{\sqrt{3}}{\sqrt{\pi^3}} \frac{1}{\sigma\sqrt{1-\alpha^2}} \exp\left\{ -\frac{3}{\pi^2\sigma^2(1-\alpha^2)} (\eta_t - (1-\alpha)(\mu + \sigma\,\gamma))^2 \right\}. \tag{3.15}$$

Since $\{\eta_t\}$ is an iid normal sequence, the stationary marginal distribution of $\{X_t\}$ is also normal with mean $\mu^*/(1-\alpha)$ and variance $\sigma^{2*}/(1-\alpha^2)$. The quasi maximum likelihood estimator of $\theta = (\alpha,\, \mu,\, \sigma)'$ now can be obtained by maximizing the log-likelihood function (3.14). The QMLE of α denoted by $\hat{\alpha}$ is obtained as a solution of the equation

$$\begin{aligned}
&\left(\frac{T-1}{T}\right)\left(\frac{T(\alpha-1)^2 - 2\alpha^2}{2\alpha + T - T\alpha}\right)\left((1+\alpha)x_1 + \sum_{t=2}^{T}(x_t - \alpha x_{t-1})\right)^2 \\
&\quad + \alpha(1+\alpha)\left(x_1^2 + \frac{1}{1-\alpha^2}\sum_{t=2}^{T}(x_t - \alpha x_{t-1})^2\right) \\
&\quad = \frac{1}{1-\alpha}\left(\alpha\sum_{t=2}^{T}x_t^2 + \alpha\sum_{t=2}^{T}x_{t-1}^2 - (1+\alpha^2)\sum_{t=2}^{T}x_t x_{t-1}\right) \\
&\quad - \frac{\alpha-1}{2\alpha + T - T\alpha}\left(\sum_{t=2}^{T}x_t + \sum_{t=2}^{T}x_{t-1}\right)\left((1+\alpha)x_1 + \sum_{t=2}^{T}(x_t - \alpha x_{t-1})\right),
\end{aligned} \tag{3.16}$$

which needs to be solved numerically.

Then the QMLE of σ^2 and μ are respectively given by

$$\hat{\sigma}^2 = \frac{6}{T\,\pi^2}\left[x_1^2 + \frac{1}{1-\hat{\alpha}^2}\sum_{t=2}^{T}(x_t - \hat{\alpha}x_{t-1})^2 - \frac{1}{(1+\hat{\alpha})(2\hat{\alpha}+T-T\hat{\alpha})}\left((1+\hat{\alpha})x_1 + \sum_{t=2}^{T}(x_t - \hat{\alpha}x_{t-1})\right)^2\right], \quad (3.17)$$

$$\hat{\mu} = \frac{1}{2\hat{\alpha}+T-T\hat{\alpha}}\left((1+\hat{\alpha})x_1 + \sum_{t=2}^{T}(x_t - \hat{\alpha}x_{t-1})\right) - \hat{\sigma}\,\gamma. \quad (3.18)$$

Next, we discuss the asymptotic properties of the QMLE. Billingsley (1961) and Basawa and Rao (1980) proved under certain regularity conditions, listed below, that the MLE is consistent and asymptotically normal. Let $\{X_t\}$ be a stationary Markov sequence with one-step transition density function $f(x_t;\theta|x_{t-1})$ and the initial density $f_X(x_1;\theta)$ and Θ be the parameter space.

(i) $\log f(x_t;\theta|x_{t-1})$ is thrice differentiable with respect to θ for all θ in a neighbourhood I of θ_0.

(ii) $E|\partial^2 \log f(x_t;\theta_0|x_{t-1})/\partial\theta_i\,\partial\theta_j| < \infty$;
$E|\partial^2 \log f(x_t;\theta_0|x_{t-1})/\partial\theta_i\,\partial\theta_j + (\partial \log f(x_t;\theta_0|x_{t-1})/\partial\theta_i)^2\,| < \infty$.

(iii) There exist sequences $\{K(T)\}$ and $\{M(T)\}$ of positive constants with $K(T) \to \infty$ and $M(T) \to \infty$, as $T \to \infty$ such that

(1) $M(T)\,\{K(T)\}^{-1}\sum_{t=1}^{T}\partial \log f(x_t;\theta_0|x_{t-1})/\partial\theta_i \xrightarrow{L} N(0,\ B(\theta_0))$ for some non-random function $B(\theta_0) > 0$,

(2) $\{K(T)\}^{-1}|\sum_{t=1}^{T}\partial^2 \log f(x_t;\theta_0|x_{t-1})/\partial\theta_i\,\partial\theta_j| \xrightarrow{P} A(\theta_0)$ for some non-random function $A(\theta_0) > 0$, and

(3) for all $\varepsilon > 0$ and for all $\nu > 0$, there exists $\delta = \delta(\varepsilon, \nu)$ and $N = N(\varepsilon, \nu)$ such that for all $T > N$,

$$P\left[\{K(T)\}^{-1}\left|\sum_{t=1}^{T}\left(\frac{\partial^2 \log f(x_t;\theta^*|x_{t-1})}{\partial\theta_i\,\partial\theta_j} - \frac{\partial^2 \log f(x_t;\theta_0|x_{t-1})}{\partial\theta_i\,\partial\theta_j}\right)\right| > \nu\right] < \varepsilon,$$

whenever, $|\theta^* - \theta_0| < \delta$, where $\theta^* = \theta_0 + r(\theta - \theta_0)$ with $r = r(T, \theta_0)$ satisfying $|r| < 1$.

Hence, there exists a root $\hat{\theta}$ of the likelihood equation with P_{θ_0}-probability approaching one which is consistent for θ_0 as $n \to \infty$

Under the conditions (i)-(iii), any consistent solution of the maximum likelihood equation is asymptotic normal (CAN). That is,

$$M(T)\left(\hat{\theta} - \theta_0\right) \xrightarrow{L} N(0, C(\theta_0)),$$

where $C(\theta_0) = ((C_{i\,j}(\theta_0))) = A(\theta_0)^{-1}\,B(\theta_0)\,A(\theta_0)^{-1}$ and

$A_{i\,j}(\theta_0) = E\left(\partial^2 \log f(x_t;\theta_0|x_{t-1})/\partial\theta_i\,\partial\theta_j\right)$,

$B_{i\,j}(\theta_0) = E\left((\partial \log f(x_t;\theta_0|x_{t-1})/\partial\theta_i)\,(\partial \log f(x_t;\theta_0|x_{t-1})/\partial\theta_j)\right)$.

White (1982) proved under the above mentioned regularity conditions that the QMLE $\hat{\theta}$ is consistent and asymptotically normal for the corresponding parameter. As the likelihood function in (3.14) is differentiable with respect to parameters and all the moments of $\{X_t\}$ are finite, the required regularity conditions defined above hold good. If we take $\theta = (\alpha, \mu, \sigma)' = (\theta_1, \theta_2, \theta_3)'$, then

$$\sqrt{T}\left(\hat{\theta} - \theta\right) \xrightarrow{L} N(0, C(\theta)), \tag{3.19}$$

where

$$C(\theta) = ((C_{ij}(\theta))) = A(\theta)^{-1}\, B(\theta)\, A(\theta)^{-1} \tag{3.20}$$

with $(i,j)^{th}$ elements of $A(\theta)$ and $B(\theta)$ are respectively given by

$$A_{ij}(\theta) = E\left(\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\log\, f\right) \text{ and } B_{ij}(\theta) = E\left(\frac{\partial}{\partial\theta_i}\log f\,\frac{\partial}{\partial\theta_j}\log f\right).$$

Let us now compute the elements of the variance-covariance matrix. Based on the discussion above, we have

$$\log f = C - \log\sigma - \frac{1}{2}\log(1-\alpha^2) - \frac{3}{\pi^2\sigma^2(1-\alpha^2)}\left[x_t - \alpha x_{t-1} - (1-\alpha)(\mu+\sigma\,\gamma)\right]^2,$$

where $C = \log\frac{\sqrt{3}}{\sqrt{\pi^3}}$.

The elements of $A(\theta)$ are obtained as,

$$A_{11} = E\left(\frac{\partial^2\log f}{\partial\alpha^2}\right) = -\frac{1+\alpha^2}{(1-\alpha^2)^2};$$

$$A_{12} = E\left(\frac{\partial^2\log f}{\partial\alpha\,\partial\mu}\right) = 0 = A_{21};$$

$$A_{13} = E\left(\frac{\partial^2\log f}{\partial\alpha\,\partial\sigma}\right) = \frac{2\alpha}{\sigma(1-\alpha^2)} = A_{31};$$

$$A_{22} = E\left(\frac{\partial^2\log f}{\partial\mu^2}\right) = -\frac{6(1-\alpha)}{\pi^2\sigma^2(1+\alpha)};$$

$$A_{23} = E\left(\frac{\partial^2\log f}{\partial\mu\,\partial\sigma}\right) = -\frac{6\,\gamma(1-\alpha)}{\pi^2\sigma^2(1+\alpha)} = A_{32};$$

$$A_{33} = E\left(\frac{\partial^2\log f}{\partial\sigma^2}\right) = -\frac{2}{\sigma^2} - \frac{6\,\gamma^2(1-\alpha)}{\pi^2\sigma^2(1+\alpha)};$$

and that of $B(\theta)$ are

$$B_{11} = E\left(\frac{\partial \log f}{\partial \alpha}\,\frac{\partial \log f}{\partial \alpha}\right) = \frac{(5+7\alpha^2)(1+\alpha^2)}{5\,(1-\alpha^2)^3};$$

$$B_{12} = E\left(\frac{\partial \log f}{\partial \alpha}\,\frac{\partial \log f}{\partial \mu}\right) = \frac{36\,\alpha\,(1-\alpha^3)\,\psi_2(1)}{\pi^4\sigma\,(1+\alpha)\,(1-\alpha^2)^2} = B_{21};$$

$$B_{13} = E\left(\frac{\partial \log f}{\partial \alpha}\,\frac{\partial \log f}{\partial \sigma}\right) = \frac{36\,\alpha\,\gamma\,(1-\alpha^3)\,\psi_2(1)}{\pi^4\sigma\,(1+\alpha)\,(1-\alpha^2)^2} + \frac{\alpha}{\sigma\,(1-\alpha^2)} - \frac{36\,\alpha\,(9-\alpha^2)}{60\,\sigma\,(1-\alpha^2)^2};$$

$$B_{22} = E\left(\frac{\partial \log f}{\partial \mu}\,\frac{\partial \log f}{\partial \mu}\right) = \frac{6\,(1-\alpha)}{\pi^2\sigma^2\,(1+\alpha)};$$

$$B_{23} = E\left(\frac{\partial \log f}{\partial \mu}\,\frac{\partial \log f}{\partial \sigma}\right) = \frac{6\,\gamma\,(1-\alpha)}{\pi^2\sigma^2\,(1+\alpha)} - \frac{36\;(1-\alpha^3)\psi_2(1)}{\pi^4\,\sigma^2\,(1+\alpha)\,(1-\alpha^2)} = B_{32};$$

$$B_{31} = B_{13};$$

$$B_{33} = E\left(\frac{\partial \log f}{\partial \sigma}\,\frac{\partial \log f}{\partial \sigma}\right) = -\frac{1}{\sigma^2} + \frac{6\,\gamma^2(1-\alpha)}{\pi^2\sigma^2\,(1+\alpha)} - \frac{36\,\gamma\,(1-\alpha^3)\,\psi_2(1)}{\pi^4\sigma^2\,(1+\alpha)\,(1-\alpha^2)} + \frac{3\,(9-\alpha^2)}{5\,\sigma^2\,(1-\alpha^2)};$$

Then the expression of $C_{ij}(\theta) = C_{ij}$, for $i, j = 1, 2, 3$ are obtained as,

$$C_{11} = 1 - \alpha^2 + \frac{36\,\alpha^2\gamma\,(1+\alpha+\alpha^2)\,\psi_2(1)}{\pi^4(1+\alpha)^2};$$

$$C_{12} = -\alpha\,\gamma\,\sigma - \frac{18\,\alpha\,\gamma^2\sigma\,(1+\alpha^2)\,(1+\alpha+\alpha^2)\,\psi_2(1)}{\pi^4\,(1-\alpha^2)\,(1+\alpha)^2} = C_{21};$$

$$C_{13} = \alpha\,\sigma + \frac{18\,\alpha\,\gamma\,\sigma\,(1+\alpha^2)\,(1+\alpha+\alpha^2)\,\psi_2(1)}{\pi^4\,(1-\alpha^2)\,(1+\alpha)^2} = C_{31};$$

$$C_{22} = \frac{11\,\gamma^2\,\sigma^2(1+\alpha^2)}{10\,(1-\alpha^2)} + \frac{6\,\gamma\,\sigma^2\,(1+\alpha+\alpha^2)\,\psi_2(1)}{\pi^2\,(1-\alpha^2)} + \frac{\pi^2\,\sigma^2(1+\alpha)^2}{6\,(1-\alpha^2)} + \frac{9\,\gamma^3\,\sigma^2\,(1+\alpha^2)^2(1+\alpha+\alpha^2)\,\psi_2(1)}{\pi^4\,(1-\alpha^2)^2\,(1+\alpha)^2};$$

$$C_{23} = -\frac{11\,\gamma\,\sigma^2(1+\alpha^2)}{10\,(1-\alpha^2)} - \frac{3\,\sigma^2\,(1+\alpha+\alpha^2)\,\psi_2(1)}{\pi^2\,(1-\alpha^2)} - \frac{9\,\gamma^2\,\sigma^2\,(1+\alpha^2)^2(1+\alpha+\alpha^2)\,\psi_2(1)}{\pi^4\,(1-\alpha^2)^2\,(1+\alpha)^2};$$

$$= C_{32};$$

$$C_{33} = \frac{11\,\sigma^2(1+\alpha^2)}{10\,(1-\alpha^2)} + \frac{9\,\gamma\,\sigma^2\,(1+\alpha^2)^2(1+\alpha+\alpha^2)\,\psi_2(1)}{\pi^4\,(1-\alpha^2)^2\,(1+\alpha)^2},$$

where $\psi_2(1) = -2.4041$, and $\gamma = 0.5772$ is the Euler's constant.

The simulation study reported in Table 3.2 of Section 3.4 indicates that the estimates are biased and a large sample is required to get a better estimate, which is expected under the QML set up. The estimation methods described in this and the previous section, provide easily computable estimators. But they are not as efficient as the MLE. Even though, we do not have closed form expressions for the MLE, due to the infinite series form of the innovation distribution, we study their properties in the next section and compare their performance with the earlier estimates by simulation.

3.3.3 Method of Maximum Likelihood Estimation

As noted earlier, the complex structure of the transition density function of the GEVAR(1) sequence, the likelihood based inference is difficult to manage. This lead us to study the properties of CLSE and QMLE as some alternative estimators

for the parameters. Since the transition density function exists, unlike in the case of many other non-Gaussian AR models, it is worth trying for parameter estimation by maximum likelihood method. The explicit form of the log-likelihood function of $\theta = (\alpha, \mu, \sigma)'$ based on $(x_1, x_2, ..., x_T)$ can be obtained from (3.6) and (3.14) and is given by

$$L_T(\theta;\, x_1, x_2, ..., x_T) = -T\log(\sigma) - \frac{(x_1-\mu)}{\sigma} - \exp\left(-\frac{(x_1-\mu)}{\sigma}\right) + (T-1)\log\left(\frac{1}{\pi}\right) + \sum_{t=2}^{T}\log(R(x_{t-1}, x_t)),$$

where $R(x_{t-1}, x_t) = \sum_{k=1}^{\infty}(-1)^{k-1}\frac{\Gamma(k\alpha)}{\Gamma(k)}\, e^{-\frac{k}{\sigma}(x_t-\alpha x_{t-1}-(1-\alpha)\mu)}\sin(k\pi\alpha)$.
The form of the likelihood function indicates that we have to maximize it by some numerical methods. This requires us to approximate the infinite series by truncating it at a finite number of terms. Let us approximate $R(x_{t-1}, x_t)$ by

$$R^N(x_{t-1}, x_t) = \sum_{k=1}^{N}(-1)^{k-1}\frac{\Gamma(k\alpha)}{\Gamma(k)}\, e^{-\frac{k}{\sigma}(x_t-\alpha x_{t-1}-(1-\alpha)\mu)}\sin(k\pi\alpha),$$

and choose N for which $|R^N(x_{t-1}, x_t) - R^{N+1}(x_{t-1}, x_t)| < \delta$. In our computation, we have chosen $\delta = 10^{-10}$. The value of this sum depends on N as well as on a particular realization, $\{x_t, t > 0\}$.

Now the MLE $\tilde{\theta} = (\tilde{\alpha}, \tilde{\mu}, \tilde{\sigma})'$ of $\theta = (\alpha, \mu, \sigma)'$ can be obtained by solving the following likelihood equations,

$$\sum_{t=2}^{T}\frac{R_\alpha(x_{t-1}, x_t)}{R(x_{t-1}, x_t)} = 0 \tag{3.21}$$

$$1-\exp(-(x_1-\mu)/\sigma)+(1-\alpha)\sum_{t=2}^{T}\frac{R_\mu(x_{t-1},x_t)}{R(x_{t-1},x_t)}=0 \tag{3.22}$$

$$x_1-\mu-n\sigma-(x_1-\mu)\exp\left(-\frac{(x_1-\mu)}{\sigma}\right)+\sum_{t=2}^{T}\frac{(x_t-\alpha x_{t-1}-(1-\alpha)\mu)R_\sigma(x_{t-1},x_t)}{R(x_{t-1},x_t)}=0 \tag{3.23}$$

where

$$R_\alpha(x_{t-1},x_t)=\sum_{k=1}^{\infty}(-1)^{k-1}\frac{\Gamma(k\alpha)}{\Gamma(k)}\ k\,e^{-\frac{k}{\sigma}(x_t-\alpha\,x_{t-1}-(1-\alpha)\mu)}$$

$$\{[((x_{t-1}-\mu)/\sigma)+\Psi(k\alpha)]\ \sin(k\pi\alpha)+\pi\cos(k\pi\alpha)\}$$

$$R_\mu(x_{t-1},x_t)=\sum_{k=1}^{\infty}(-1)^{k-1}\frac{\Gamma(k\alpha)}{\Gamma(k)}k\,e^{-\frac{k}{\sigma}(x_t-\alpha\,x_{t-1}-(1-\alpha)\mu)}\sin(k\pi\alpha)=R_\sigma(x_{t-1},x_t)$$

and $\Psi(.)$ is a digamma function defined by $\Psi(k)=\frac{d}{dk}\ln\Gamma(k)=\frac{\Gamma'(k)}{\Gamma(k)}$.

We applied the standard theory on the properties of MLE for a stationary Markov sequences, discussed in Section 3.3.2, to prove that $\tilde{\theta}$ is consistent and asymptotic normal with mean vector 0 and covariance matrix $C(\theta)$, defined by (3.20). However, the elements of the matrices in (3.20) are to be evaluated for the GEVAR(1) model, which do not have closed form expression for its innovation density. The elements of the matrices can be evaluated using the following relations. For $i,j=1,2,3$,

$$A_{ij}(\theta)=E\left(\frac{\partial^2}{\partial\theta_i\theta_j}\log\ f_Z(x_1,x_2)\right);\ \ B_{ij}(\theta)=E\left(\frac{\partial\log f_Z(x_1,x_2)}{\partial\theta_i}\,\frac{\partial\log f_Z(x_1,x_2)}{\partial\theta_j}\right). \tag{3.24}$$

These expectations need to be evaluated numerically as the underlying density is expressed as infinite series. While evaluating the expectations, we truncate the series as discussed above. In the expressions (3.24), we replace (x_{t-1},x_t) by (x_1,x_2)

due to the stationarity of the GEVAR(1) sequence. The elements of the matrices presented below will be used for computation in the next section.

$$A_{11} = E\left(\frac{\partial^2}{\partial\alpha^2}\log\, f\right); A_{12} = A_{21} = E\left(\frac{\partial^2}{\partial\alpha\,\partial\mu}\log f\right);$$
$$A_{13} = A_{31} = E\left(\frac{\partial^2}{\partial\alpha\,\partial\sigma}\log f\right); A_{22} = E\left(\frac{\partial^2}{\partial\mu^2}\log f\right);$$
$$A_{23} = A_{32} = E\left(\frac{\partial^2}{\partial\mu\,\partial\sigma}\log f\right); A_{33} = E\left(\frac{\partial^2}{\partial\sigma^2}\log f\right);$$
$$B_{11} = E\left(\frac{\partial}{\partial\alpha}\log f\right)^2; B_{12} = B_{21} = E\left(\frac{\partial}{\partial\alpha}\log f\frac{\partial}{\partial\mu}\log f\right);$$
$$B_{13} = B_{31} = E\left(\frac{\partial}{\partial\alpha}\log f\frac{\partial}{\partial\sigma}\log f\right); B_{22} = E\left(\frac{\partial}{\partial\mu}\log f\right)^2;$$
$$B_{23} = B_{32} = E\left(\frac{\partial}{\partial\mu}\log f\frac{\partial}{\partial\sigma}\log f\right); B_{33} = E\left(\frac{\partial}{\partial\sigma}\log f\right)^2.$$

In the next section, we simulate observations from the GEVAR(1) model and obtain the estimates under different methods described above.

3.4 Simulation Study

We carried out simulation studies to illustrate the numerical methods for computing the estimates proposed in the earlier sections and also to examine their performances. To simulate a realization from the model (3.2) we used the link between the innovation rv η_t and the positive stable rv U described in (3.5). Accordingly for specified values of the parameters we simulated iid sequence of positive stable

rvs $\{U_t\}$ using the relation:

$$U = E^{-(1-\alpha)/\alpha}(\sin\xi)^{-1/\alpha}\sin(\alpha\xi)\sin((1-\alpha)\xi)^{(1-\alpha)/\alpha},$$

proposed by McKenzie (1982), where ξ is a uniform rv over $(0,\pi)$ and E is a unit exponential rv independent of ξ. Then generate η_t using the equation (3.5) and finally obtain the sequence of GEVAR(1) sequence from the equation (3.2). Based on this realization, we obtain the estimates using the procedures described in Section 3.3. The detailed algorithm (in MATLAB code) for the computations is given in Appendix A. The estimates along with their bias and root mean square errors (RMSE) are listed in the Tables 3.1 – 3.3 under different method of estimation.

In Tables 3.1 – 3.4, the entries in the column with title "Asym Std dev" are asymptotic standard deviations computed at the theoretical values of the parameters. They are obtained as the square root of the diagonal elements of the corresponding asymptotic dispersion matrices. These values could be compared with the estimated RMSE of the respective estimates given in the parenthesis. For a better estimation method these values become closer.

The Table 3.1 summarizes the computation of the conditional least squares estimates (CLSE) of α and μ when the scale parameter is fixed as $\sigma = 1$. The CLSE of α and μ are obtained by using the equations given in (3.12) for specified sample sizes and are listed in the Table 3.1. For each specified value of the parameter, we repeated the experiment 50 times for computing the estimates and then averaged them over the repetitions. The final values are entered in Table 3.1 along with the bias and the RMSE in the parenthesis.

TABLE 3.1: The average estimates, the bias, the corresponding RMSE and the asymptotic standard deviation for the CLS estimates

	True Values		Conditional Least Squares Estimates			
n	α	μ	$\hat{\alpha}$ (Bias, RMSE)	Asym Std dev	$\hat{\mu}$ (Bias, RMSE)	Asym Std dev
100	0.80	2.00	0.7722 (-0.0278, 0.0532)	0.0600	1.8616 (-0.1384, 0.3817)	0.3848
	0.70	3.00	0.7054 (0.0054, 0.0597)	0.0714	2.9130 (-0.0870, 0.2949)	0.3053
	0.60	0.00	0.5735 (-0.0265, 0.0805)	0.0800	0.0145 (0.0145,0.2912)	0.2565
	0.50	3.00	0.5265 (0.0265, 0.0780)	0.0866	2.9504 (-0.0496 , 0.2449)	0.2221
	0.30	2.00	0.3409 (0.0409, 0.0861)	0.0954	1.9579 (-0.0421, 0.1516)	0.1748
	0.20	1.00	0.1891 (-0.0109, 0.0873)	0.0980	0.9982 (-0.0018, 0.1508)	0.1571
200	0.80	2.00	0.7874 (-0.0126, 0.0422)	0.0424	1.9724 (-0.0276, 0.1908)	0.2721
	0.70	3.00	0.7015 (0.0015, 0.0444)	0.0504	2.9560 (-0.0440, 0.2125)	0.2159
	0.60	0.00	0.5753 (-0.0247, 0.0663)	0.0566	0.0082 (0.0082, 0.1853)	0.1814
	0.50	3.00	0.5169 (0.0169, 0.0450)	0.0612	2.9691 (-0.0309, 0.1720)	0.1571
	0.30	2.00	0.3110 (0.0110, 0.0693)	0.0675	1.9819 (-0.0181, 0.1024)	0.1236
	0.20	1.00	0.1969 (-0.0031, 0.0736)	0.0693	0.9932 (-0.0068, 0.1074)	0.1110
500	0.80	2.00	0.7964 (-0.0036, 0.0277)	0.0268	2.0046 (0.0046, 0.1904)	0.1721
	0.70	3.00	0.6916 (-0.0084, 0.0313)	0.0319	2.9819 (-0.0181, 0.1278)	0.1365
	0.60	0.00	0.5912 (-0.0088, 0.0350)	0.0358	0.0133 (0.0133, 0.1147)	0.1147
	0.50	3.00	0.5062 (0.0062, 0.0347)	0.0387	2.9710 (-0.0290, 0.0902)	0.0993
	0.30	2.00	0.3040 (0.0040, 0.0390)	0.0427	1.9932 (-0.0068, 0.0844)	0.0782
	0.20	1.00	0.1898 (-0.0102, 0.0435)	0.0438	0.9975 (-0.0025, 0.0668)	0.0702

From the table, it is observed that the estimates of (α, μ) perform well as sample size increases. For instance, if $\mu = 2$ and $\alpha = 0.8$, the average $\hat{\mu}$ is 1.8616 and the average $\hat{\alpha}$ is 0.7722, for $n = 100$. When the sample size is increased to 500, the estimates of μ and α are respectively 2.0046 and 0.7964. Further, the bias, the RMSE and the asymptotic standard deviation of the estimates decrease when sample size increases. The other parameter combinations of μ and α behave in a similar way. It is observed from the Table 3.1 that as α decreases, the asymptotic standard deviation of the $\hat{\alpha}$ increases and that of $\hat{\mu}$ decreases.

In Table 3.2, we present the QMLE of θ based on the simulated samples. The estimates are obtained by solving the QML equations (3.16), (3.17), and (3.18). As before, we repeated the computation 50 times, and the final estimates are the averages over these repetitions. Corresponding root mean square errors and the bias are given in the parenthesis. The asymptotic standard deviations of the estimators are also reported in Table 3.4. We observed that the estimates are close to the true values when the sample size is large. Also, the bias, RMSE and the asymptotic standard deviation of the estimates decrease when sample size increases.

Finally, the Table 3.3 summarizes the calculations of the MLE of θ. For the specified sample sizes, we computed the MLE by solving the likelihood equations (3.21), (3.22) and (3.23). The averages of the estimates obtained by repeating the experiment 50 times along with their bias, and RMSE are presented in Table 3.3. Also, the asymptotic standard deviations are entered in Table 3.4.

TABLE 3.2: The average estimates, the bias and the corresponding RMSE for the QMLE.

	True Values			Quasi Maximum Likelihood Estimates		
n	α	μ	σ	$\hat{\alpha}$ (Bias, RMSE)	$\hat{\mu}$ (Bias, RMSE)	$\hat{\sigma}$ (Bias, RMSE)
	0.80	2.00	3.00	0.7775	2.2574	2.8408
				(-0.0225,0.0586)	(0.2574, 0.9226)	(-0.1592, 0.6585)
100	0.70	3.00	2.00	0.6514	3.1659	1.9139
				(-0.0486, 0.0737)	(0.1659, 0.5874)	(-0.0861, 0.2954)
	0.60	0.00	1.00	0.5528	0.0056	0.9496
				(-0.0472, 0.0847)	(0.0056, 0.1946)	(-0.0504, 0.1331)
	0.50	3.00	2.00	0.4746	2.9705	1.9300
				(-0.0254, 0.0856)	(-0.0295, 0.3664)	(-0.0700, 0.2474)
	0.30	2.00	0.50	0.2830	1.9974	0.4943
				(-0.0170, 0.0912)	(-0.0026, 0.0709)	(-0.0057, 0.0597)
	0.20	1.00	2.00	0.1770	0.9601	1.9181
				(-0.0230, 0.0959)	(-0.0399, 0.2612)	(-0.0819, 0.1854)
	0.80	2.00	3.00	0.7935	2.1389	3.0199
				(-0.0065, 0.0379)	(0.1389, 0.6842)	(0.0199, 0.4409)
	0.70	3.00	2.00	0.6738	3.1506	1.9902
200				(-0.0262, 0.0492)	(0.1506, 0.3589)	(-0.0098, 0.2248)
	0.60	0.00	1.00	0.5795	0.0303	0.9761
				(-0.0205, 0.0520)	(0.0303, 0.1622)	(-0.0239, 0.1020)
	0.50	3.00	2.00	0.4942	2.9683	1.9909
				(-0.0058, 0.0529)	(-0.0317, 0.2468)	(-0.0091, 0.1615)
	0.30	2.00	0.50	0.2966	2.0177	0.4991
				(-0.0034, 0.0634)	(0.0177, 0.0509)	(-0.0009, 0.0369)
	0.20	1.00	2.00	0.1805	1.0200	1.9890
				(-0.0195, 0.0752)	(0.0200, 0.1994)	(-0.0110, 0.1445)
	0.80	2.00	3.00	0.7928	2.1093	2.9716
				(-0.0072, 0.0247)	(0.1093, 0.4459)	(-0.0284, 0.3187)
	0.70	3.00	2.00	0.6929	3.0085	1.9936
500				(-0.0071, 0.0314)	(0.0085, 0.2617)	(-0.0064, 0.1716)
	0.60	0.00	1.00	0.5908	0.0187	0.9954
				(-0.0092, 0.0390)	(0.0187, 0.0981)	(-0.0046, 0.0519)
	0.50	3.00	2.00	0.4959	3.0047	1.9916
				(-0.0041, 0.0374)	(0.0047, 0.1684)	(-0.0084, 0.1276)
	0.30	2.00	0.50	0.2941	2.0022	0.4990
				(-0.0059, 0.0371)	(0.0022, 0.0337)	(-0.0010, 0.0219)
	0.20	1.00	2.00	0.2050	0.9973	1.9978
				(0.0050, 0.0481)	(-0.0027, 0.1236)	(-0.0022, 0.1072)

TABLE 3.3: The average estimates, the bias and the corresponding RMSE for the MLE.

	True Values			Maximum Likelihood Estimates		
n	α	μ	σ	$\hat{\alpha}$ (Bias, RMSE)	$\hat{\mu}$ (Bias, RMSE)	$\hat{\sigma}$ (Bias, RMSE)
100	0.80	2.00	3.00	0.7708 (-0.0292, 0.0372)	1.9136 (-0.0864, 0.7578)	2.6582 (-0.3418, 0.4722)
	0.70	3.00	2.00	0.6865 (-0.0135, 0.0472)	2.9815 (-0.0185, 0.4071)	1.9237 (-0.0763, 0.2695)
	0.60	0.00	1.00	0.5947 (-0.0053, 0.0572)	-0.0098 (-0.0098, 0.1724)	0.9844 (-0.0156, 0.1269)
	0.50	3.00	2.00	0.5015 (0.0015, 0.0564)	3.0853 (0.0853, 0.3390)	1.9860 (-0.0140, 0.2043)
	0.30	2.00	0.50	0.3092 (0.0092, 0.0722)	2.0164 (0.0164, 0.0707)	0.4911 (-0.0089, 0.0524)
	0.20	1.00	2.00	0.1969 (-0.0031, 0.0775)	1.0188 (0.0188, 0.2037)	1.9894 (-0.0106, 0.1759)
200	0.80	2.00	3.00	0.7866 (-0.0134, 0.0241)	2.1328 (0.1328, 0.7144)	2.8619 (-0.1381, 0.3347)
	0.70	3.00	2.00	0.6946 (-0.0054, 0.0286)	3.0662 (0.0662, 0.3745)	1.9707 (-0.0293, 0.2074)
	0.60	0.00	1.00	0.5980 (-0.0020, 0.0342)	0.0313 (0.0313, 0.1588)	0.9999 (-0.0001, 0.0840)
	0.50	3.00	2.00	0.4997 (-0.0003, 0.0398)	3.0607 (0.0607, 0.2748)	2.0071 (0.0071, 0.1508)
	0.30	2.00	0.50	0.3074 (0.0074, 0.0546)	2.0143 (0.0143, 0.0534)	0.5057 (0.0057, 0.0324)
	0.20	1.00	2.00	0.2051 (0.0051, 0.0559)	1.0563 (0.0563, 0.1905)	2.0207 (0.0207, 0.1228)
500	0.80	2.00	3.00	0.7930 (-0.0070, 0.0137)	2.1133 (0.1133, 0.4267)	2.9476 (-0.0524, 0.2100)
	0.70	3.00	2.00	0.6986 (-0.0014, 0.0246)	3.0573 (0.0573, 0.2702)	2.0169 (0.0169, 0.1455)
	0.60	0.00	1.00	0.5963 (-0.0037, 0.0214)	0.0284 (0.0284, 0.1025)	0.9997 (-0.0003, 0.0536)
	0.50	3.00	2.00	0.4957 (-0.0043, 0.0247)	3.0503 (0.0503, 0.1808)	1.9984 (-0.0016, 0.0972)
	0.30	2.00	0.50	0.2957 (-0.0043, 0.0306)	2.0095 (0.0095, 0.0364)	0.4993 (-0.0007, 0.0211)
	0.20	1.00	2.00	0.1964 (-0.0036, 0.0343)	1.0312 (0.0312, 0.1325)	1.9963 (-0.0037, 0.0805)

TABLE 3.4: The asymptotic standard deviations under the QML and ML methods

	True Values			QML Method			ML Method		
n	α	μ	σ	Asymp std dev (α)	Asymp std dev (μ)	Asymp std dev (σ)	Asymp std dev (α)	Asymp std dev (μ)	Asymp std dev (σ)
100	0.80	2.00	3.00	0.0336	0.9529	0.5203	0.0292	0.7942	0.5084
	0.70	3.00	2.00	0.0565	0.5096	0.3088	0.0388	0.4963	0.2663
	0.60	0.00	1.00	0.0706	0.2151	0.1376	0.0456	0.2112	0.1144
	0.50	3.00	2.00	0.0806	0.3734	0.2495	0.0540	0.3696	0.2112
	0.30	2.00	0.50	0.0934	0.0734	0.0540	0.0668	0.0776	0.0409
	0.20	1.00	2.00	0.0971	0.2632	0.2061	0.0729	0.2653	0.1786
200	0.80	2.00	3.00	0.0238	0.6738	0.3679	0.0178	0.6342	0.2978
	0.70	3.00	2.00	0.0399	0.3603	0.2184	0.0281	0.3443	0.1841
	0.60	0.00	1.00	0.0499	0.1521	0.0973	0.0250	0.1554	0.0623
	0.50	3.00	2.00	0.0570	0.2640	0.1764	0.0361	0.2653	0.1388
	0.30	2.00	0.50	0.0660	0.0519	0.0382	0.0499	0.0511	0.0310
	0.20	1.00	2.00	0.0686	0.1861	0.1457	0.0518	0.1850	0.1140
500	0.80	2.00	3.00	0.0150	0.4261	0.2327	0.0109	0.4026	0.1901
	0.70	3.00	2.00	0.0253	0.2279	0.1381	0.0192	0.2272	0.1403
	0.60	0.00	1.00	0.0316	0.0962	0.0616	0.0161	0.0970	0.0416
	0.50	3.00	2.00	0.0361	0.1670	0.1116	0.0292	0.1632	0.1109
	0.30	2.00	0.50	0.0418	0.0328	0.0241	0.0347	0.0321	0.0217
	0.20	1.00	2.00	0.0434	0.1177	0.0922	0.0399	0.1151	0.0867

The MLE of all parameters perform well even for a small sample of size $n = 100$. For example, at $\mu = 3$, $\alpha = 0.7, \sigma = 2$, $\hat{\mu} = 2.9815$, $\hat{\alpha} = 0.6865$ and $\hat{\sigma} = 1.9237$. The RMSE is close to the asymptotic standard deviation. Also, as α decreases, the RMSE and the asymptotic standard deviations of $\hat{\alpha}$ increase.

Overall, it is observed that the estimates of α, μ and σ behave reasonably well for large sample sizes under CLS and QML methods in terms of their bias and RMSE. However, from the Table 3.3, we can see that there is significant reduction in bias

and RMSE of MLE even for reasonable sample sizes. Hence, the use of likelihood estimates is recommended compared to other methods for GEVAR(1) model.

3.5 Application

To illustrate the application of the proposed time series model, an AR(1) model with Gumbel marginal distribution has been fitted to two time series of daily maximum of Bombay Stock Exchange (BSE) index and Standard and Poor 500 (S&P 500) index. The data consists of 375 observations of BSE index values from 4, January 2010 to 30, June 2011 and 252 of S&P 500 index from 2, January 2009 to 31, December 2009. The data are obtained from the website of Yahoo Finance. The time series plots and ACF of the data are displayed in Figure 3.1 and 3.2 respectively.

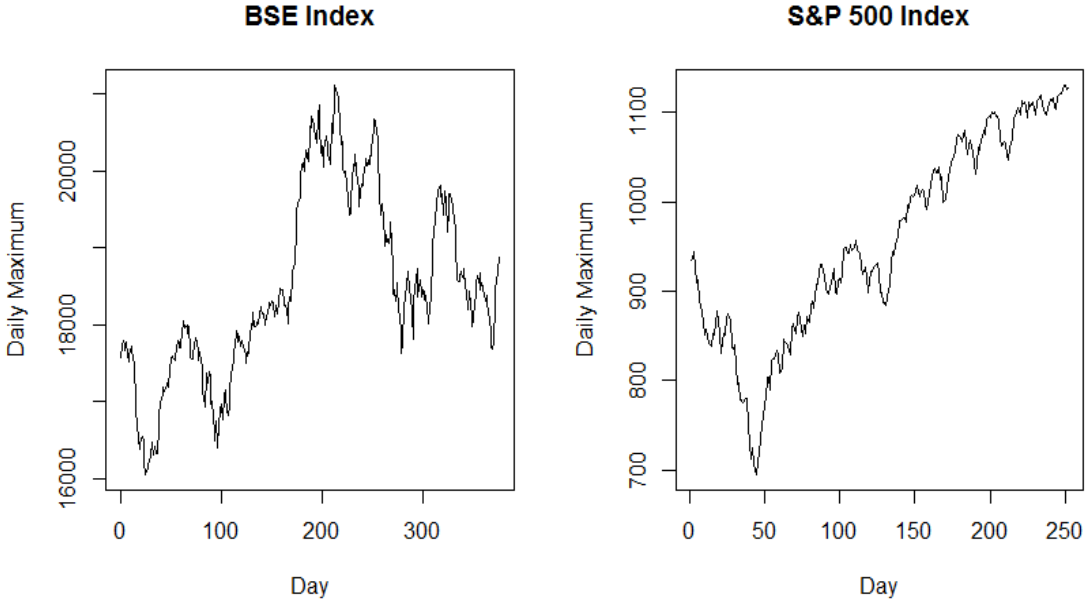

FIGURE 3.1: Time series plot of BSE and S&P 500 indices

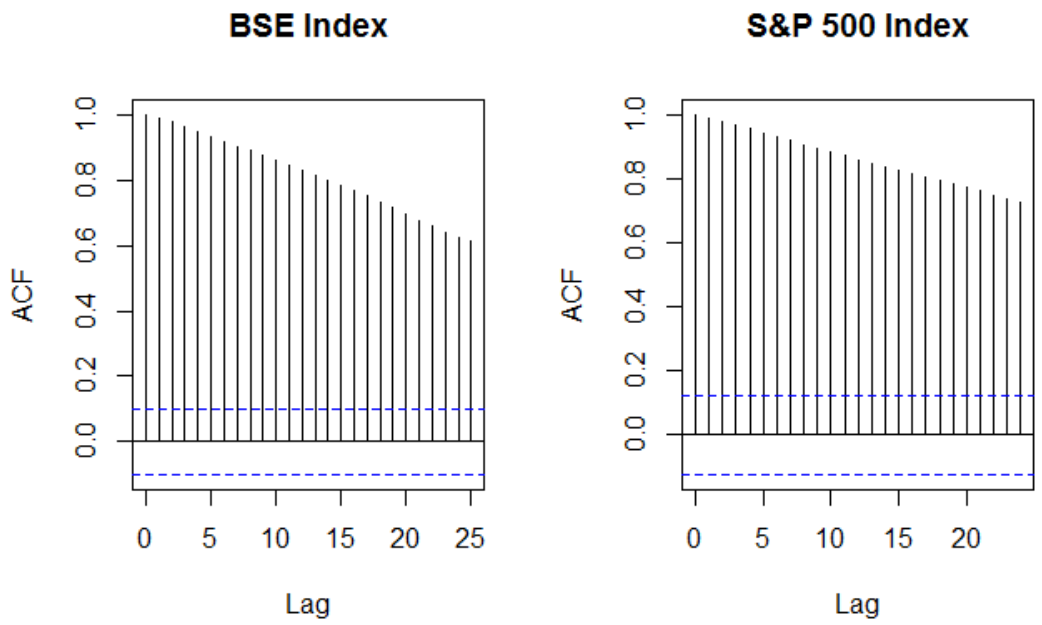

FIGURE 3.2: ACF of BSE and S&P 500 indices

From the time series and ACF plots, it is observed that both the series show non-stationary behaviour. This is further confirmed using *Augmented Dickey-Fuller* (ADF) test for the BSE data. The calculated value of the ADF test statistic for the BSE data is given in Table 3.5. ADF tests whether a unit root is present in an autoregressive model. In this method, one can check for unit roots using either the normalized bias test, given by $T(\hat{\alpha} - 1)$ or the usual t-statistic. If the calculated t value exceeds the 5% critical value, then we accept the null hypothesis of unit roots. However, for the S&P data, the ADF rejects the non-stationarity hypothesis due to unit root, but the ACF plots exhibits non-stationarity. Hence, we adopted exponential smoothing method for removing non-stationarity and confirmed the stationarity of the residual obtained by exponential smoothing using ADF test.

Exponential smoothing is a technique that can be applied to time series data, either

TABLE 3.5: ADF statistic along with associated p values (in brackets) for the BSE data

BSE Index	ADF Statistic
Original series	-1.5354(0.7732)
Log-transformed series	-1.5586(0.7634)
exponentially smoothed series	-14.5202(0.0050)

to produce smoothed data, or to make forecasts. When the sequence of observations begins at time $t = 0$, the simplest form of exponential smoothing is given by the formulae

$$S_0 = X_0; \;\; S_t = cX_t + (1 - c)\, S_{t-1}, \;\; t > 0,$$

where c is the smoothing factor, and $0 < c < 1$. The value of c is chosen by minimizing the residual sum of squares. In the present case, it is found that $c = 0.9$. Then we fit a GEVAR(1) model to the residuals which are calculated as $X_t - S_t$, the actual series minus the exponential smooth. We call the resulting residual series as the exponentially smoothed series and are plotted in Figure 3.3.

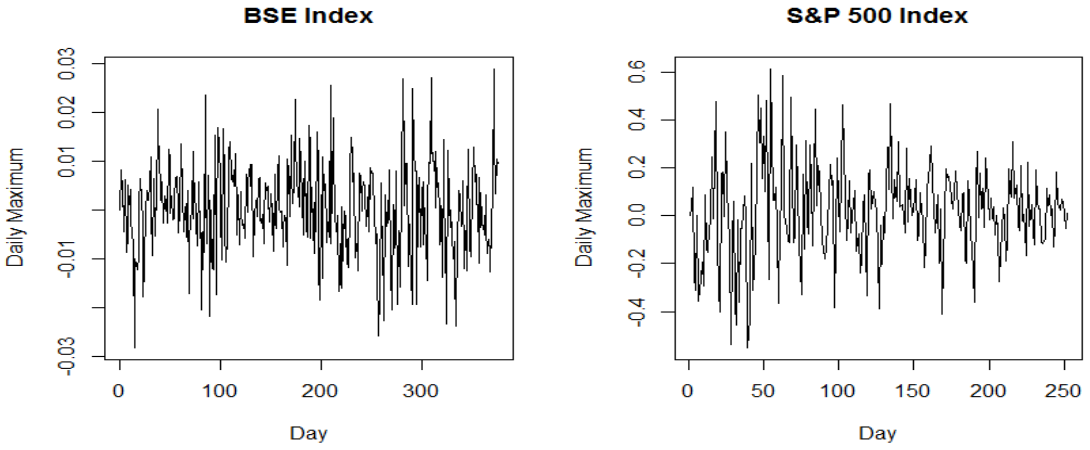

FIGURE 3.3: Time series plot of exponentially smoothed series

Next, we have to identify an appropriate model for the smoothed series. The graphs of autocorrelation function and partial autocorrelation function in Figure 3.4 suggest that the series may have an AR(1) structure.

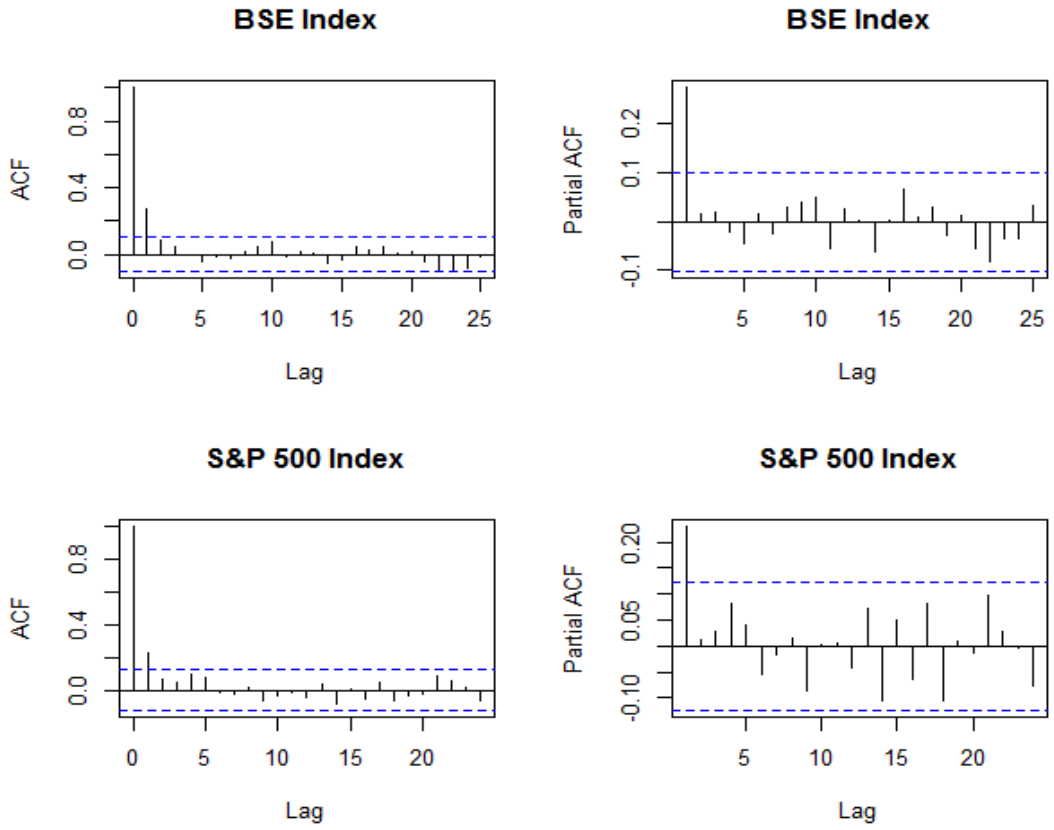

FIGURE 3.4: ACF and PACF of smoothed series

We estimate the parameters α, μ and σ by using the methods: conditional least squares, quasi maximum likelihood and the maximum likelihood and the results are summarized in Table 3.6.

Then fit the Gumbel distribution to the data to check the marginal stationarity. The histogram of the data and the true density are superimposed in the Figures 3.5 and 3.6 respectively for BSE and S&P indices.

TABLE 3.6: The estimated values of the parameters for two data sets under three methods

Estimation Methods	BSE Index	S&P 500 Index
CLSE	$\hat{\alpha}$ =0.2740	$\hat{\alpha}$ = 0.2292
($\sigma = 1$)	$\hat{\mu}$ =-0.5770	$\hat{\mu}$ = -0.5649
	$\hat{\alpha}$ =0.2733	$\hat{\alpha}$ = 0.2283
QMLE	$\hat{\mu}$ =-0.0039	$\hat{\mu}$ = -0.0756
	$\hat{\sigma}$ = 0.0072	$\hat{\sigma}$ = 0.1526
	$\hat{\alpha}$ = 0.2698	$\hat{\alpha}$ =0.1947
MLE	$\hat{\mu}$ =-0.0038	$\hat{\mu}$ =-0.0606
	$\hat{\sigma}$ = 0.0075	$\hat{\sigma}$ = 0.1363

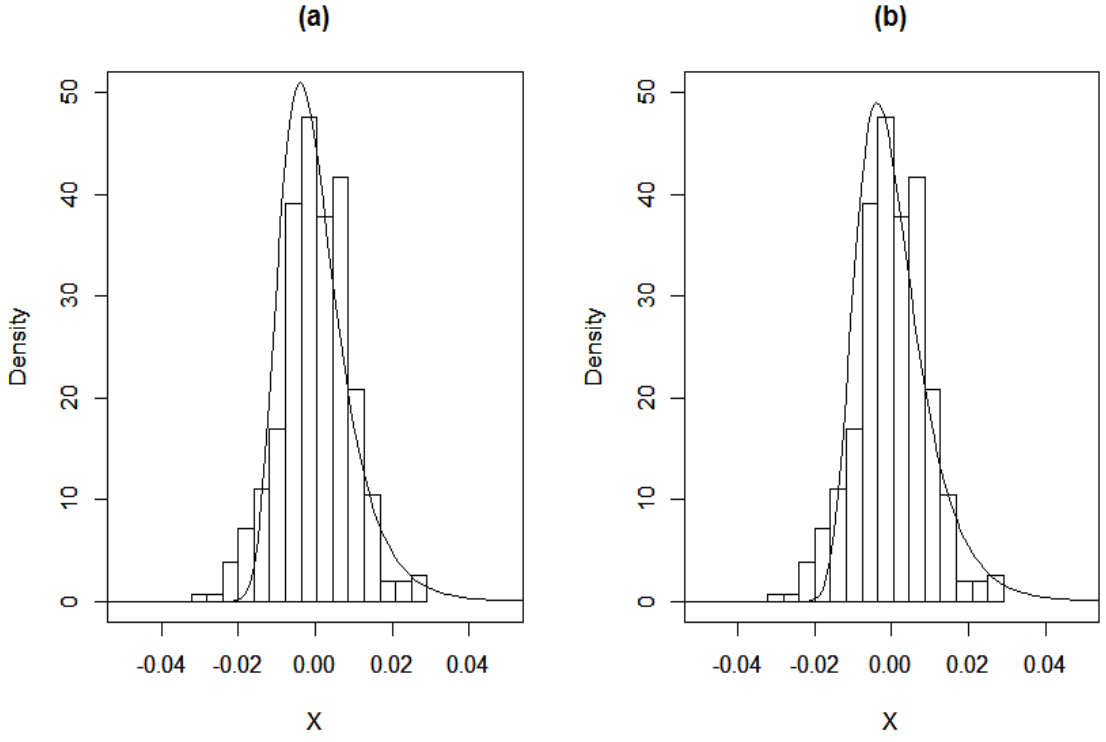

FIGURE 3.5: Histogram of smoothed BSE data with super imposed extreme value distribution based on the parameters estimated by (a): QMLE, (b) MLE

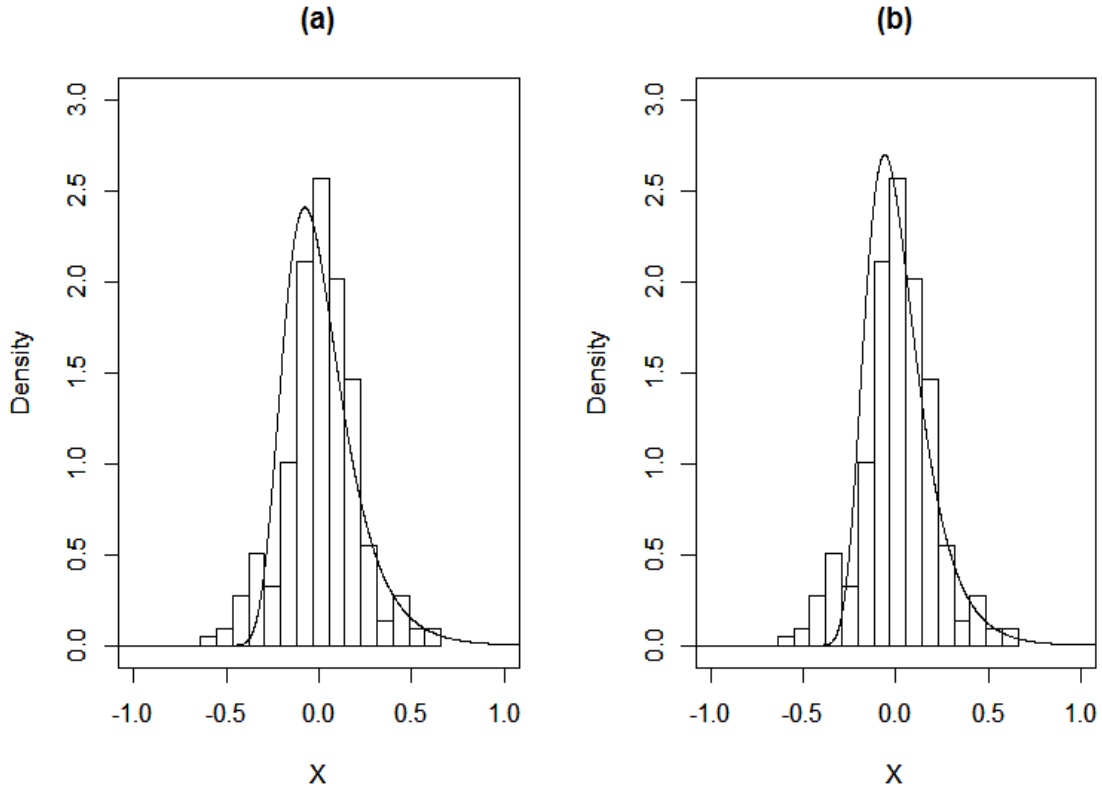

FIGURE 3.6: Histogram of smoothed S&P 500 data with super imposed extreme value distribution based on the parameters estimated by (a): QMLE, (b) MLE

The panel (b) in the Figure 3.5 and Figure 3.6 show that there is a close agreement between the observed and fitted density function whose parameters are estimated by the maximum likelihood method.

For a visual check, we plot in Figure 3.7, the ACF of the residuals from the model. It is observed that there is no significant correlation among the residuals.

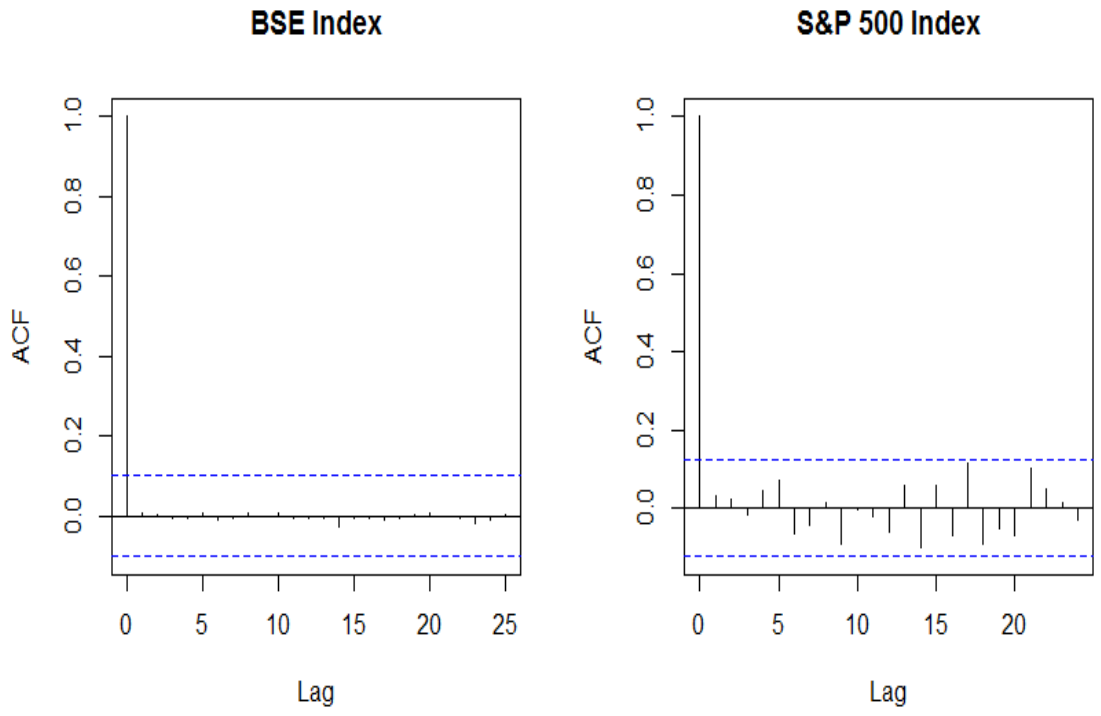

FIGURE 3.7: ACF of the residual series

Next, we have to test the model adequacy by checking the validity of the assumptions imposed on the errors. Under normal circumstances, we propose Portmanteau tests and Ljung-Box test for diagnosing model validity. Since we are working with non-Gaussian set up, the above mentioned tests may not be suitable for our model. Therefore we apply a non-parametric test based on permutation entropy proposed by Matilla-García and Ruiz Marín (2008) for testing the independence of errors.

Recently, serial independence has been increasingly studied by using entropy measures; see for example, Matilla-García and Ruiz Marín (2008). These measures avoid restrictive parametric assumptions on the probability distribution generating the data, and they can capture the dependence present in a time series. In this

method, given a time series, we study the dependence present in the series by translating the problem into symbolic dynamic and then, we use the entropy measure associated to these symbols to test the dependence present in the time series. More concretely, we study all $m!$ permutations (symbols) π of length m in the symmetric group S_m^* which are considered here as possible order types of m different numbers. Afterwards, we give the distribution followed by the mentioned symbols and define the entropy measure associated to them. This entropy measure is called permutation entropy. Moreover, under the null hypothesis of independence we prove that an affine transformation of the permutation entropy is asymptotically χ^2 distributed.

In this method, we first construct a symmetric group corresponding to the embedding dimension $m = 4$. For a given set of T observations, the embedding dimension will be the largest m that satisfies $5m! \leq T$ with $m = 2, 3, ...$ Then find the relative frequency p_π of a symbol $\pi \in S_m^*$ using

$$p(\pi) := p_\pi = \frac{\#\{t \in T | t \text{ is of } \pi - \text{type}\}}{T - m + 1}.$$

Now, under this setting we can define the permutation entropy of a time series for an embedding dimension $m \geq 2$. This entropy is defined as the Shannon's entropy of the $m!$ distinct symbols as follows:

$$h(m) = -\sum_{\pi \in S_m^*} p_\pi \ln(p_\pi).$$

Permutation entropy, $h(m)$, is the information contained in comparing m consecutive values of the time series. It is clear that $0 \leq h(m) \leq \ln(m!)$ where the lower

bound is attained for an increasing or decreasing sequence of values and the upper bound for a completely random system (iid sequence) where all $m!$ possible permutations appear with the same probability.

Then, we can define the affine transformation of the permutation entropy, under the null hypothesis of independence, as

$$G(m) = 2(T - m + 1)(Ln(m!) - h(m)),$$

which is asymptotically $\chi^2_{m!-1}$ distributed.

Concerning the BSE index, the computed value of $G(m)$ is equal to 11.2364. The computed value of $G(m)$ in the S&P data is equal to 17.0191. Both values are less than the 5% Chi-square critical value at degrees of freedom 23. Hence, we conclude that there is no significant serial dependence among the residuals. That is, the GEVAR(1) is adequate for the above data sets.

Note:

In the above analysis, we have considered the BSE index values from 4, January 2010 to 30, June 2011 for the model construction purpose. The time series analysis revealed that the model is adequate for the data. With an objective to validate the model, we again fitted the data from 1, July 2011 to 31, December 2013 and found that the model works well.

The results of this chapter are reported in the paper Balakrishna and Shiji (2013).

Chapter 4

Weibull Product Autoregressive Models

4.1 Introduction

In the context of analysing financial time series, it is important to study the behaviour of market volatility, which is measured in terms of the conditional variance of return given the past. One way of analysing the stochastic volatility is by modelling the squares or absolute values of the sequence of returns. These are non-negative variables and exhibit a tendency to follow heavy-tailed and skewed distributions such as log-normal, Pareto, Weibull, etc. It is well known that the return series are serially dependent and hence the classical method of data analysis based on random samples is of no use. So, it is necessary to introduce stochastic models to generate sequences of non-negative dependent random variables.

Linear autoregressive models have played a significant role in modelling the dependence structure in the study of Gaussian and non-Gaussian time series. When the time series of interest is a sequence of non-negative random variables as mentioned above, the product form of the models are preferable compared to their linear counterparts. In this chapter, we study the properties of a model with product structure which generates a sequence of Markov dependent Weibull random variables. The role of Weibull distribution in modelling non-negative variables such as failure time, sea wave height, wind velocity, etc., is well known in the literature. The Weibull distribution is also used to model extreme events such as floods and snowfall, extreme temperatures, large fluctuations in exchange rates, market crashes, etc.

Even though, the statistical analysis of Weibull distribution is well studied based on random samples, not much research is available on the problem of modelling when the observations are serially dependent. Sim (1986) introduced a model with minification structure, to define a sequence of Markov dependent Weibull random variables. But there are difficulties in handling the inference problems, (cf. Balakrishna and Jacob (2003)).

Rest of the chapter is organized as follows. In Section 4.2, we discuss some of the useful properties of the sequence generated by the model. We propose an approximate method to analyse the innovation random variable in Section 4.3. The method of maximum likelihood estimation is discussed in Section 4.4. The simulation results to assess the performance of the estimates are presented in Section 4.5. Finally, in Section 4.6, we apply the model to a real life data.

4.2 Model and Properties

The general form of the first order product autoregressive (PAR(1)) model is given by

$$Y_t = Y_{t-1}^{\alpha} V_t, \;\; t = 1, 2, ..., \;\; 0 \leq \alpha < 1\,, \tag{4.1}$$

where $\{V_t\}$ is a sequence of independent and identically distributed non-negative innovation random variables. Assume that $Y_0 > 0$, is independent of V_1 and hence it follows that Y_t depends on V_j for $j \leq t$ and is independent of V_j for $j > t$. Under these conditions, $\{Y_t\}$ generates a stationary Markov sequence. For an explicit analysis of the model it is important to know the stationary marginal distribution of $\{Y_t\}$. This in turn requires to identify the distribution of $\{V_t\}$ for a specified marginal distribution of $\{Y_t\}$, a problem common in the study of non-Gaussian time series models. If the innovation random variable admits a density function then the one-step transition density function for the Markov sequence $\{Y_t\}$ can be expressed as

$$f(y_t|y_{t-1}) = \frac{1}{y_{t-1}^{\alpha}} f_V\left(\frac{y_t}{y_{t-1}^{\alpha}}\right)\,, \tag{4.2}$$

where, $f_V(.)$ is the marginal density function of $\{V_t\}$. McKenzie (1982) introduced this model to define a stationary sequence of gamma random variables through the properties of linear gamma AR(1) model of Gaver and Lewis (1980). However, the form of the innovation distribution was not obtained explicitly. McKenzie's interest was to establish a characterizing property of the gamma sequence, namely $\{Y_t\}$ and $\{\log Y_t\}$ have the same autocorrelation structure. Later, Abraham and Balakrishna (2012) obtained the explicit form of the innovation distribution for the gamma PAR(1) model and studied its properties.

Balakrishna and Lawrence (2012) discussed the PAR(1) model with gamma marginal distribution by approximating the innovation densities. They also identify some more members of the family of rvs which admit explicit solutions for V_t in the model (4.1). In the present chapter, we concentrate on the properties and applications of the Weibull PAR(1) model. Balakrishna and Lawrence (2012) proved that the random variable Y_t defined by (4.1) follows a Weibull distribution denoted by $Weibull(\theta, \lambda)$ with probability density function,

$$f_Y(y) = \lambda\theta y^{\theta-1} \exp\left(-\lambda y^{\theta}\right), \; y \geq 0, \; \lambda > 0, \; \theta > 0, \tag{4.3}$$

if the distribution of the corresponding innovation V_t is given by that of $\left(\lambda^{-(1-\alpha)}U^{-\alpha}\right)^{1/\theta}$, where U has a positive stable distribution, with Laplace Transform (LT), $\varphi_U(s) = \exp\left(-s^{\alpha}\right)$. The second order properties of our Weibull PAR(1) sequence are stated below for ready reference. The mean, variance and the ACF are respectively given by

$$E(Y_t) = \lambda^{-1/\theta}\Gamma(\theta^{-1}+1); \quad V(Y_t) = \lambda^{-2/\theta}\left\{\Gamma\left(2\theta^{-1}+1\right) - \Gamma\left(\theta^{-1}+1\right)^2\right\}; \tag{4.4}$$

$$\rho_Y(j) = \frac{\Gamma(\theta^{-1}+1)\left\{\Gamma[\theta^{-1}(\alpha^j+1)+1] - \Gamma(\theta^{-1}\alpha^j+1)\Gamma(\theta^{-1}+1)\right\}}{\left(\Gamma\left(\theta^{-1}\alpha^j+1\right)\left\{\Gamma(2\theta^{-1}+1) - \Gamma(\theta^{-1}+1)^2\right\}\right)}, \quad j = 1, 2, \ldots. \tag{4.5}$$

Note that the innovation random variable, V_t, for the Weibull model does not have a closed form for its density function. Hence, the likelihood based inference becomes difficult. However moments of all orders exist and are given by

$$E(V_t^k) = \lambda^{-k(1-\alpha)/\theta}\Gamma(k\theta^{-1}+1)/\Gamma(\alpha k\theta^{-1}+1), \tag{4.6}$$

so that one can obtain moment estimators of the parameters. But, the moment estimators are not efficient in general. We have computed the moment estimates of the parameters based on simulated samples and found that they are seriously biased. In our further analysis, we will consider these estimators as our initial estimates. Next, we propose a method for analysing the above model based on an approximate innovation.

4.3 Approximation to Innovation Variable

Aiming at the likelihood based inference for the model parameters, we propose an approximation approach to the problem. According to this approach we assume that the innovation V_t also follows a Weibull distribution with its scale and shape parameters are chosen by matching the first two moments of the random variables on both sides of the equation (4.1). We justify the choice of Weibull distribution for the innovation through a result of Brockwell and Brown (1978). That is, if U is a positive stable random variable with LT, $\varphi_U(s) = \exp(-s^\alpha)$, then $U^{-\alpha}$ converges to an exponential random variable with unit mean as $\alpha \to 0$. Hence, $V_t = \left(\lambda^{-(1-\alpha)}U^{-\alpha}\right)^{1/\theta}$ tends to a $Weibull(\theta, \lambda)$ random variable as $\alpha \to 0$. This result motivated us to assume $Weibull(\beta, \delta)$ distribution for the innovation. We chose the parameters δ and β by equating the first two moments of the stationary random variable Y_t and $Y_{t-1}^{\alpha}V_t$ as the distribution is completely determined by these two parameters. The required moments can be obtained using (4.4) and (4.6) along with the independence of Y_{t-1} and V_t . The proposed approximation works very well and which is confirmed by numerical comparison. By the above method

the expressions for the innovation parameters β and δ are obtained by solving the following equations for given values of α, θ, λ,

$$\frac{B(\beta^{-1}, \beta^{-1})}{2\beta} = \frac{B(\theta^{-1}, \theta^{-1})}{\alpha\, B(\alpha\theta^{-1}, \alpha\theta^{-1})} \quad \text{and} \quad \delta = \left(\frac{\alpha\lambda^{(1-\alpha)/\theta}}{\beta}\right)^{\beta} \left(\frac{\Gamma(\alpha\theta^{-1})\Gamma(\beta^{-1})}{\Gamma(\theta^{-1})}\right)^{\beta}, \tag{4.7}$$

where $B(.,.)$ denotes a beta function, given by $B(x, y) = \Gamma(x)\ \Gamma(y)/\Gamma(x + y)$. The probability density function of $Y_{t-1}^{\alpha} V_t$ can be numerically evaluated using the equation,

$$f_{Y_{t-1}^{\alpha} V_t}(x) = \frac{\lambda\theta\delta\beta}{\alpha} y^{\frac{\theta}{\alpha}-1} \int_0^{\infty} v^{\beta-\frac{\theta}{\alpha}-1} \exp(-\delta v^{\beta} - \lambda(y/v)^{\theta/\alpha})\ dv. \tag{4.8}$$

This density is taken as an approximation to that of the $Weibull(\theta, \lambda)$ density function.

The plots in Figure 4.1 compare the exact $Weibull(\theta, \lambda)$ probability density function with the approximate one for selected parameter values. We have noted the similar behaviour for several other combinations of the parameters.

The Figure 4.2 is the histogram of simulated sample of size 5000 from the exact Weibull PAR(1) model superimposed on the Weibull density using the approximate innovation. The plots in panel (a) and (b) of Figure 4.2 can be paired with panel (a) and (d) of Figure 4.1 respectively. These figures further confirm that the proposed approximation is acceptable for practical purposes.

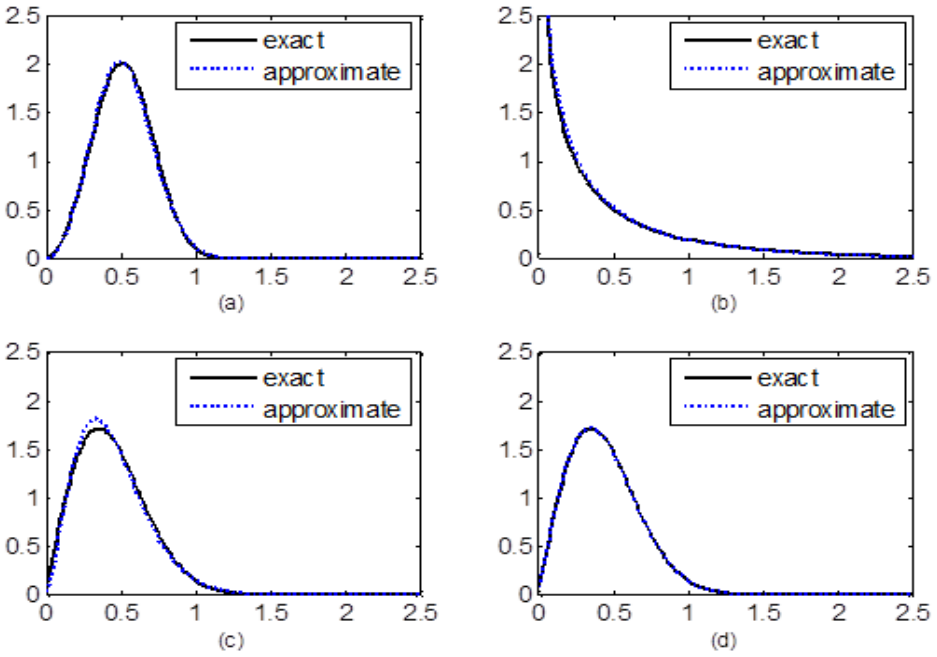

FIGURE 4.1: Comparison of exact and approximate Weibull density when (a) α=0.95, θ=3, λ=5. (b) α=0.8, θ=0.7, λ=2. (c) α=0.5, θ=2, λ=4. and (d) α=0.1, θ=2, λ=4.

4.4 Maximum Likelihood Estimation

If we have an explicit form for the innovation density function then the likelihood based inference is possible for the PAR(1) model (4.1). We have already observed that for a Weibull PAR(1) model the approximate innovation density works very well. Let $(y_1, y_2, ..., y_T)$ be an observed realization from the model (4.1) with $Weibull(\theta, \lambda)$ marginal density. Since the innovation random variable admits a density function we have an explicit expression for the one-step transition density function, $f(y_t|y_{t-1})$ of the Markov sequence defined by (4.1). Then the likelihood

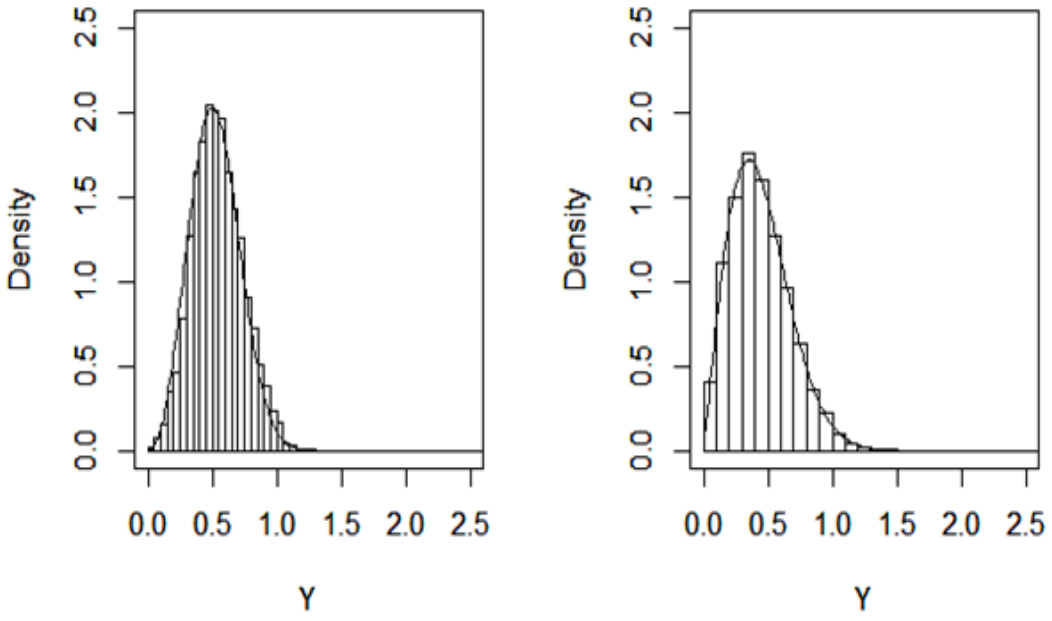

FIGURE 4.2: Theoretical Weibull marginal density function based on approximate innovation and its simulated version from the exact Weibull PAR(1) model for (a) α=0.95, θ=3, λ=5 and (b) α=0.1, θ=2, λ=4.

function of the parameter vector $\mu^* = (\alpha, \beta, \delta)' = (\mu_1^*, \mu_2^*, \mu_3^*)'$, (say) can be expressed as,

$$L(\mu^*|y_1,\, y_2,\, ...,\, y_T) = \; f_Y(y_1) \prod_{t=2}^{T} f(y_t|y_{t-1}) = \; f_Y(y_1) \prod_{t=2}^{T} f_V\left(\frac{y_t}{y_{t-1}^{\alpha}}\right) \frac{1}{y_{t-1}^{\alpha}}, \quad (4.9)$$

where $f_Y(.)$ and $f_V(.)$ are the marginal density functions of $\{Y_t\}$ and $\{V_t\}$ respectively. Once the MLE of μ^* is found from the sample, the MLE of the original parameter $\mu = (\alpha, \theta, \lambda)' = (\mu_1, \mu_2, \mu_3)'$, (say) can be obtained using the relation (4.7). Further analysis of MLE is carried out by skipping the first term on the right

side of (4.9) as it will not affect its asymptotic properties. Hence, the likelihood function we considered is given by

$$L(\mu^*|y_1,\, y_2,\, ...,\, y_T) = \prod_{t=2}^{T} \delta\, \beta\, \left(\frac{y_t}{y_{t-1}^{\alpha}}\right)^{\beta-1} \exp\left(-\delta\, \left(\frac{y_t}{y_{t-1}^{\alpha}}\right)^{\beta}\right)\, \frac{1}{y_{t-1}^{\alpha}}.$$

The log-likelihood function can be written as,

$$\log\, L(\mu^*|y_1,\, y_2, ..., y_T) = (T-1)\log\delta + (T-1)\log\beta + (\beta-1)\sum_{t=2}^{T}\log\left(\frac{y_t}{y_{t-1}^{\alpha}}\right) - \delta\sum_{t=2}^{T}\left(\frac{y_t}{y_{t-1}^{\alpha}}\right)^{\beta} - \alpha\sum_{t=2}^{T}\log y_{t-1}.$$

Now, the MLE of μ^* can be obtained by solving the following system of likelihood equations:

$$(T-1)\left(\sum_{t=2}^{T}\left(\frac{y_t}{y_{t-1}^{\alpha}}\right)^{\beta}\right)^{-1}\sum_{t=2}^{T}\left(\frac{y_t}{y_{t-1}^{\alpha}}\right)^{\beta}\log\, y_{t-1} - \sum_{t=2}^{T}\log\, y_{t-1} = 0. \tag{4.10}$$

$$\frac{(T-1)}{\beta} + \sum_{t=2}^{T}\log\left(\frac{y_t}{y_{t-1}^{\alpha}}\right) - (T-1)\left(\sum_{t=2}^{T}\left(\frac{y_t}{y_{t-1}^{\alpha}}\right)^{\beta}\right)^{-1}\sum_{t=2}^{T}\left(\frac{y_t}{y_{t-1}^{\alpha}}\right)^{\beta}\log\left(\frac{y_t}{y_{t-1}^{\alpha}}\right) = 0. \tag{4.11}$$

$$\delta = (T-1)\left(\sum_{t=2}^{T}\left(\frac{y_t}{y_{t-1}^{\alpha}}\right)^{\beta}\right)^{-1}. \tag{4.12}$$

These equations have to be solved by numerical methods and are illustrated using simulated samples in Section 4.5.

Next, we discuss the asymptotic properties of the MLE. Billingsley (1961), Basawa and Rao (1980) developed the standard theory on the properties of MLE for a stationary Markov sequences. It can be applied to prove that $\hat{\mu}^*$ is consistent

and asymptotic normal with mean vector 0 and covariance matrix $C(\mu^*)$, which is defined by (3.20) in Chapter 3.

That is,

$$\sqrt{n}\,(\hat{\mu}^* - \mu^*) \xrightarrow{L} N(0, C(\mu^*)),$$

where

$$C(\mu^*) = ((C_{i\,j}(\mu^*))) = A(\mu^*)^{-1}\,B(\mu^*)\,A(\mu^*)^{-1}$$

with $(i,j)^{th}$ elements of $A(\mu^*)$ and $B(\mu^*)$ are respectively given by

$$A_{i\,j}(\mu^*) = E\left(\frac{\partial^2}{\partial\mu_i^*\,\partial\mu_j^*}\log\,f\right) \quad \text{and} \quad B_{i\,j}(\mu^*) = E\left(\frac{\partial}{\partial\mu_i^*}\log f\,\frac{\partial}{\partial\mu_j^*}\log f\right),$$

where f is the conditional density function given in (4.2).

Now, we compute the elements of the variance-covariance matrix. We have

$$\log\,f = \log\delta + \log\,\beta + (\beta-1)\log\left(\frac{y_t}{y_{t-1}^{\alpha}}\right) - \delta\left(\frac{y_t}{y_{t-1}^{\alpha}}\right)^{\beta} - \alpha\log y_{t-1}.$$

Then, the elements of $A(\mu^*)$ are obtained as

$$
\begin{aligned}
A_{11} &= E\left(\frac{\partial^2 \log f}{\partial \alpha^2}\right) = -\beta^2\left(\frac{\pi^2}{6\,\theta^2} + \left(\frac{\gamma + \log\lambda}{\theta}\right)^2\right);\\
A_{12} &= E\left(\frac{\partial^2 \log f}{\partial \alpha\,\partial \beta}\right) = -\left(\frac{\gamma + \log\lambda}{\theta}\right)(1 - \gamma - \log\delta) = A_{21};\\
A_{13} &= E\left(\frac{\partial^2 \log f}{\partial \alpha\,\partial \delta}\right) = -\frac{\beta}{\delta}\left(\frac{\gamma + \log\lambda}{\theta}\right) = A_{31};\\
A_{22} &= E\left(\frac{\partial^2 \log f}{\partial \beta^2}\right) = -\frac{1}{\beta^2}\left(\frac{\pi^2}{6} + (1 - \gamma - \log\delta)^2\right);\\
A_{23} &= E\left(\frac{\partial^2 \log f}{\partial \beta\,\partial \delta}\right) = -\frac{1}{\beta\,\delta}(1 - \gamma - \log\delta) = A_{32};\\
A_{33} &= E\left(\frac{\partial^2 \log f}{\partial \delta^2}\right) = -\frac{1}{\delta^2};
\end{aligned}
$$

and that of $B(\mu^*)$ are

$$
\begin{aligned}
B_{11} &= E\left(\frac{\partial \log f}{\partial \alpha}\,\frac{\partial \log f}{\partial \alpha}\right) = \beta^2\left(\frac{\pi^2}{6\,\theta^2} + \left(\frac{\gamma + \log\lambda}{\theta}\right)^2\right);\\
B_{12} &= E\left(\frac{\partial \log f}{\partial \alpha}\,\frac{\partial \log f}{\partial \beta}\right) = \left(\frac{\gamma + \log\lambda}{\theta}\right)(1 - \gamma - \log\delta) = B_{21};\\
B_{13} &= E\left(\frac{\partial \log f}{\partial \alpha}\,\frac{\partial \log f}{\partial \delta}\right) = \frac{\beta}{\delta}\left(\frac{\gamma + \log\lambda}{\theta}\right) = B_{31};\\
B_{22} &= E\left(\frac{\partial \log f}{\partial \beta}\,\frac{\partial \log f}{\partial \beta}\right) = \frac{1}{\beta^2} + \frac{\pi^2}{2\beta^2} - \frac{1}{\beta^2}(\gamma + \log\delta)(2 - \gamma - \log\delta);\\
B_{23} &= E\left(\frac{\partial \log f}{\partial \beta}\,\frac{\partial \log f}{\partial \delta}\right) = \frac{1}{\beta\,\delta}(1 - \gamma - \log\delta) = B_{32};\\
B_{33} &= E\left(\frac{\partial \log f}{\partial \delta}\,\frac{\partial \log f}{\partial \delta}\right) = \frac{1}{\delta^2}.
\end{aligned}
$$

Then, the expression of $C_{i\,j}(\mu^*) = C_{ij}$, for $i, j = 1, 2, 3$ are obtained as

$$
\begin{aligned}
C_{11} &= \frac{6\,\theta^2}{\pi^2\,\beta^2}; \quad C_{12} = 0 = C_{21};\\
C_{13} &= -\frac{6\,\theta\,\delta}{\pi^2\,\beta}(\gamma + \log\lambda) = C_{31};\\
C_{22} &= \frac{18\,\beta^2}{\pi^4}\left(\pi^2 - 2(\log\delta)^2 + 2(\gamma-2)\log\lambda + 2(2-\gamma+\log\lambda)\log\delta\right);\\
C_{23} &= \frac{18\,\beta\,\delta}{\pi^4}(\gamma + \log\delta - 1)\left(\pi^2 - 2(\log\delta)^2 + 2(\gamma-2)\log\lambda + 2(2-\gamma+\log\lambda)\log\delta\right);\\
&= C_{32};\\
C_{33} &= \frac{\delta^2}{\pi^4}\{\pi^2[\pi^2 + 6(3 - 6\gamma + 4\gamma^2)] - 36(\log\delta)^4 - 36(\log\delta)^3[3\gamma - 4 - \log\lambda]\}\\
&\quad + \frac{\delta^2}{\pi^4}\{12[3\gamma^3 - 12\gamma^2 + \gamma(15+\pi^2) - 6]\log\lambda + 6\pi^2(\log\lambda)^2\}\\
&\quad + \frac{\delta^2}{\pi^4}\{36(\gamma-1)[\pi^2 - 2 + 3\gamma - \gamma^2 + (3\gamma-5)\log\lambda]\log\delta\}\\
&\quad + \frac{\delta^2}{\pi^4}\{18[\pi^2 - 10 + 16\gamma - 6\gamma^2 + (6\gamma-8)\log\lambda](\log\delta)^2\},
\end{aligned}
$$

where $\gamma = 0.5772$ is the Euler's constant.

The Fisher information matrix, $C = ((C_{ij}\,(\mu)))$, in terms of the original parameter μ can be obtained using (4.7). If σ_i^2 is the i^{th} diagonal element of C, then the asymptotic variance of $\hat{\mu}_i$ is σ_i^2/n. Then, for a specified value of significance level τ, the $100(1-\tau)\%$ confidence interval for μ_i is given by $\hat{\mu}_i \pm Z_{\tau/2}(\hat{\sigma}_i/\sqrt{n})$, where $\hat{\sigma}_i$ is the i^{th} diagonal element of C, with unknown parameters replaced by the corresponding MLE and $Z_{\tau/2}$ is an appropriate percentile point of standard normal distribution.

4.5 Simulation

In this section, we carry out a simulation study to assess the finite sample performance of the MLE of the parameters. We generate a sample of size T from the Weibull PAR(1) model (4.1) using the innovation random variable described in Section 4.3. This requires the simulation of observations from a positive stable random variable U_t, which is detailed in Section 3.4 and thus obtain $V_t = \left(\lambda^{-(1-\alpha)} U_t^{-\alpha}\right)^{1/\theta}$, $t = 1, 2, ...,$ for specified values of the parameters. Then form the sequence $\{Y_t\}$ using (4.1). We used this simulated sample of Weibull series to obtain the MLE of the parameters by solving the likelihood equations (4.10) - (4.12). Then we repeated the computation of MLE for 100 samples and took their arithmetic mean as the final estimate. The computational algorithm are given in Appendix B.

The computations are carried out for several combinations of parameters and summarized in Table 4.1 and 4.2. The values in the parenthesis are the standard error of the respective MLE calculated over the repetitions. As can be seen from the table, for series length 200 estimation is reasonably satisfactory and becomes more accurate with increasing length.

TABLE 4.1: The MLE and corresponding standard errors (in parenthesis) based on simulated observations of sample sizes n=200, 500.

True values			MLE Based on n=200			MLE Based on n=500		
α	θ	λ	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\lambda}$
0.90	0.70	1.00	0.8974 (0.0127)	0.6733 (0.1235)	1.2545 (0.4152)	0.8963 (0.0071)	0.6755 (0.0649)	1.2394 (0.2723)
0.80	2.00	1.50	0.7992 (0.0246)	2.1719 (0.3100)	1.5573 (0.3937)	0.7991 (0.0124)	2.1951 (0.2133)	1.5270 (0.2389)
0.70	2.00	2.00	0.6866 (0.0309)	2.1075 (0.3671)	2.0762 (0.3690)	0.7004 (0.0186)	2.0823 (0.1179)	2.0279 (0.2049)
0.70	0.80	2.00	0.6938 (0.0314)	0.7150 (0.0974)	2.2270 (0.3647)	0.6970 (0.0179)	0.7187 (0.0463)	2.1590 (0.1812)
0.60	2.00	3.00	0.5937 (0.0337)	2.0176 (0.2956)	3.1407 (0.4240)	0.5976 (0.0220)	2.0390 (0.1231)	3.0900 (0.2813)
0.50	2.00	0.80	0.4864 (0.0428)	2.0004 (0.1622)	0.8305 (0.1228)	0.4929 (0.0264)	2.0276 (0.0985)	0.8043 (0.0733)
0.40	3.00	4.00	0.3977 (0.0453)	3.1056 (0.2207)	4.1465 (0.4568)	0.4012 (0.0275)	3.0501 (0.1209)	4.1304 (0.2920)
0.20	3.00	5.00	0.2042 (0.0528)	3.0173 (0.1770)	5.0319 (0.5128)	0.1998 (0.0323)	3.0369 (0.1111)	5.0436 (0.3199)

TABLE 4.2: The MLE and corresponding standard errors (in parenthesis) based on simulated observations of sample sizes n=1000, 2000.

True values			MLE Based on n=1000			MLE Based on n=2000		
α	θ	λ	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\lambda}$
0.90	0.70	1.00	0.8992 (0.0060)	0.6600 (0.0524)	1.2150 (0.1809)	0.8992 (0.0041)	0.6569 (0.0319)	1.2199 (0.1285)
0.80	2.00	1.50	0.7993 (0.0099)	2.1945 (0.1193)	1.5224 (0.1564)	0.7987 (0.0068)	2.1942 (0.0691)	1.5187 (0.1131)
0.70	2.00	2.00	0.6988 (0.0119)	2.0764 (0.0981)	2.0727 (0.1340)	0.6993 (0.0093)	2.0737 (0.0683)	2.0683 (0.1151)
0.70	0.80	2.00	0.6988 (0.0119)	0.7102 (0.0381)	2.1810 (0.1250)	0.7009 (0.0089)	0.7076 (0.0235)	2.1517 (0.0977)
0.60	2.00	3.00	0.5983 (0.0161)	2.0471 (0.0753)	3.0521 (0.1743)	0.5994 (0.0115)	2.0244 (0.0587)	3.0831 (0.1425)
0.50	2.00	0.80	0.4989 (0.0190)	2.0126 (0.0739)	0.8067 (0.0530)	0.4991 (0.0128)	2.0103 (0.0488)	0.8002 (0.0347)
0.40	3.00	4.00	0.3959 (0.0202)	3.0573 (0.0856)	4.1438 (0.1922)	0.3990 (0.0136)	3.0594 (0.0542)	4.1021 (0.1393)
0.20	3.00	5.00	0.1988 (0.0222)	3.0091 (0.0758)	5.0239 (0.2228)	0.1984 (0.0169)	3.0139 (0.0543)	5.0506 (0.1594)

4.6 Data Analysis

We now apply the Weibull PAR(1) model described in the earlier sections to a real data set. The data consists of 734 observations of the *daily maximum* of BSE (Bombay Stock Exchange) index values from 2, January 2007 to 31, December 2009. The observations are not available on Saturday, Sunday and other National holidays. The data is downloaded from the website *Yahoo Finance*. Figure 4.3 provides the time series plot of the data and indicates that the time series is not stationary. The

time series plot of the first order difference of the log-transformed data is given in Figure 4.4.

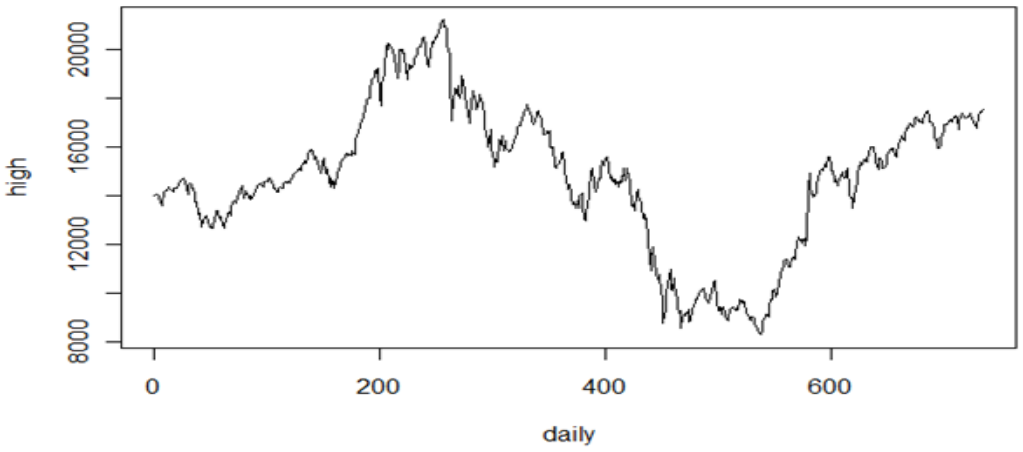

FIGURE 4.3: Time series plot of the daily maximum of BSE index values

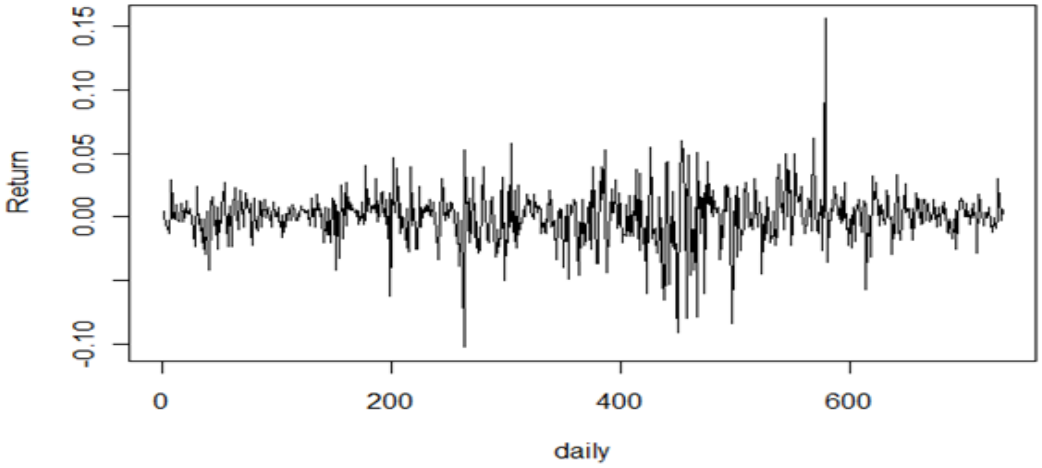

FIGURE 4.4: Time series plot of the first order difference of the log-transformed BSE data

We form the return series $\{R_t\}$ by considering the log-transform of the successive indices and then obtain its exponential smoothing, say $\{S_t\}$, using the formula $S_t = cR_t + (1-c)S_{t-1}$, for $t = 1, 2, ...$, $0 < c < 1$ and S_0 is taken as the arithmetic mean of the return series. The value of c is chosen by minimizing the residual sum of squares and in the present case, it is found that $c = 0.9$. Then obtain, $y_t = |R_t - S_t|$, $t = 1, 2, ...$, and analyse this series by Weibull PAR(1) model. We use the absolute values of the smoothed data to retain the non-negativity. Figure 4.5 is the plot of the absolute values of the exponentially smoothed series.

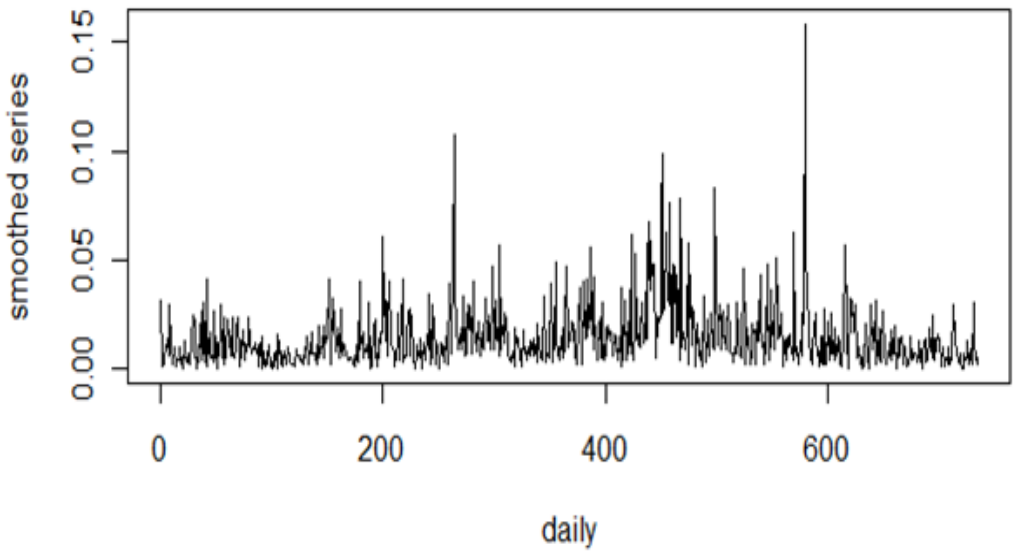

FIGURE 4.5: Time series plot of absolute value of the smoothed BSE data

The MLE of the parameters are obtained as $\hat{\alpha} = 0.1490$, $\hat{\theta} = 1.0676$, $\hat{\lambda} = 0.0146$. The first panel of Figure 4.6 shows the histogram of the series, $\{y_t\}$ super-imposed by the Weibull density function evaluated with these estimated parameters. The

plot in the second panel of Figure 4.6 is the histogram of the residuals obtained by fitting a Weibull PAR(1) model superimposed by the corresponding Weibull density function.

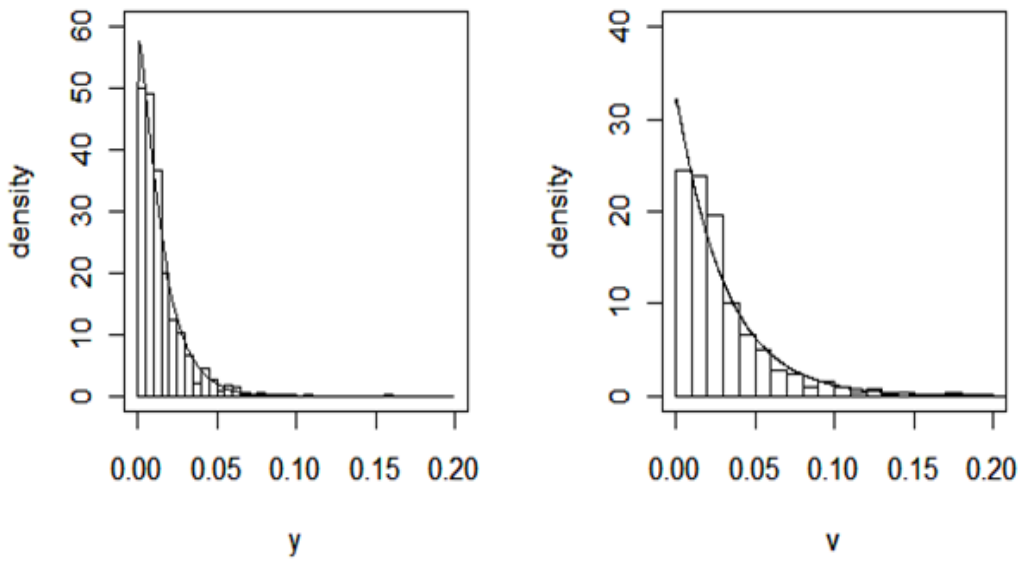

FIGURE 4.6: Histogram of smoothed BSE data with superimposed weibull density with parameters α= 0.1490; θ=1.0676; λ=0.0146 and histogram of residuals with superimposed Weibull (1.0077, 0.0299) density

These plots show that the Weibull PAR(1) model fits well for the given data. Yet, we have to check formally the goodness-of-fit by some statistical procedure. Since the classical Chi-square test of goodness-of-fit developed under the normal assumption, we use the Chi-square test of goodness-of-fit for dependent random variables proposed by Bhat (1961). In the next section, we briefly discuss the method proposed by Bhat (1961).

4.6.1 Bhat's Chi-square goodness-of-fit test

The fundamental assumption underlying the classical Chi-square test of goodness-of-fit is the independent and unordered nature of the sample observations. Patankar (1954) developed the goodness-of-fit of frequency distribution obtained from stochastic processes. In Patankar's test, the data are divided into r mutually exclusive groups and the corresponding group frequencies will be denoted by n_1, n_2, ..., n_r. If the group frequencies expected from the parent distribution are m_1, m_2, ..., m_r, then Pearson's measure of deviation is defined by,

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - m_i)^2}{m_i}. \tag{4.13}$$

Then, the test is provided by the fact that as sample size tends to infinity, the sampling distribution of χ^2 tends to that of χ^2 distribution with $r-1$ degrees of freedom. The importance of this test is that the χ^2 criterion can be shown to be equivalent to the likelihood-ratio test in large samples.

Patankar (1954) derived the approximate asymptotic distribution of (4.13) based on the assumption that for a large sequence of observations the marginal frequencies are asymptotically multivariate normal. He specified the asymptotic distribution of (4.13) completely by its first two moments and derived the mean and variance of (4.13) which is given by,

$$E(\chi^2) = \sum_{i=1}^{r} \frac{\sigma_i^2}{m_i} = A, \text{ say, and } \mathrm{V}(\chi^2) = 2 \sum_{\mathrm{i,j=1}}^{\mathrm{r}} \frac{\sigma_{\mathrm{ij}}^2}{\mathrm{m_i\, m_j}} = 2\,\mathrm{B}, \text{ say.}$$

Then, a better approximation to the standard χ^2 is obtained by taking $A\,\chi^2/B$ to have an asymptotic χ^2 distribution with A^2/B degrees of freedom. This is called the modified χ^2 test of goodness-of-fit.

To obtain the expression of σ_i^2 and $\sigma_{i\,j}$, Patankar (1954) employed a familiar device of putting,

$$X_{(t)\,i} = \begin{cases} 1, & \text{if at any time } t, \text{ the system is in the } i^{th} \text{ group} \\ 0, & \text{otherwise} \end{cases}$$

where $i = 1,\, 2, ...,\, r\,;\, t = 1,\, 2, ...,\, n$.

Then, $n_i = \sum_{t=1}^{n} X_{(t)\,i}$ and after some algebra, we get the expression,

$$\sigma_i^2 = m_i - m_i^2 + 2\sum_{s=1}^{n-1}(n-s)\,P_{ii}^{(s)}, \quad \text{and} \quad \sigma_{i\,j} = -m_i\,m_j + \sum_{s=1}^{n-1}(n-s)P_{ij}^{(s)} + \sum_{s=1}^{n-1}(n-s)\,P_{ji}^{(s)}, \tag{4.14}$$

where $P_{i\,j}^{(s)}$ is the probability of obtaining the observations belonging to the i^{th} and j^{th} groups at times t and $t+s$, respectively, for all t (assuming stationarity).

Bhat (1961) modified the Patankar's method by using any arbitrary distribution for marginal frequencies. Let n_u be the frequency of the t-tuple, $u = (u_1, u_2, ..., u_t)$ in a sequence of length $n + t - 1$ from an m^{th} order stationary Markov chain; and let m_u be its expected value in a new sequence of the same length. To test whether the chain has a specified transition probability matrix, one may construct the statistic

$$\chi_t^2 = \sum_u \frac{(n_u - m_u)^2}{m_u} \tag{4.15}$$

and test the goodness-of-fit for n_u. In (4.15) the summation extends over those values of u for which m_u does not vanish.

He obtained the expression for $E(\chi_t^2)$ and $V(\chi_t^2)$ as in the similar way Patankar (1954) derived. The expressions are

$$E(\chi_t^2) = k_t - n + 2\sum_{s=1}^{n-1} \frac{n-s}{n} \operatorname{tr}(P_t^s),$$

$$V(\chi_t^2) = 2\left\{k_t - n^2 + 4\sum_{s} \frac{n-s}{n} \operatorname{tr}(P_m^s) + 4\sum_{s,t} \frac{n-s}{n}\,\frac{n-t}{n} \operatorname{tr}(P_m^{s+t})\right\},$$

where $\operatorname{tr}(A)$ is the trace of matrix A, and k_t is the number of t-tuples for which m_u does not vanish. These expressions can be calculated if we know $\operatorname{tr}(P_1^r)$. Thus $\operatorname{tr}(P_1^r)$ is expressed as

$$\operatorname{tr}(P_1^r) \cong \int_{-\infty}^{\infty} f_r(x - \rho^r x)\, dx$$

and because f_{r} is a density function, $\operatorname{tr}(P_1^r)$ equals $(1-\rho^r)^{-1}$. Because all the expected frequencies may be assumed non-zero, we have

$$E(\chi_t^2) = k_t - n + 2\sum_{r=1}^{n-r} \frac{n-r}{n}\ (1-\rho^r)^{-1}$$

$$V(\chi_t^2) = 2\left\{k_t - n^2 + \sum_{r} \frac{n-r}{n}\,\frac{1}{1-\rho^r} + 4\sum_{r,t} \frac{n-r}{n}\,\frac{n-t}{n}\,\frac{1}{1-\rho^{r+t}}\right\}.$$

These are the same as those derived by Patankar (1954) for the special case when the X's are normal variates and the class intervals equal. Thus, his results are true even when the X's follow a general class of continuous distributions.

Using this procedure, we check whether the series $\{y_t\}$ follows a Weibull distribution

or not. The value of Chi-square statistics obtained is 9.3353. We can therefore accept the hypothesis that $\{y_t\}$ follows a Weibull distribution and the model is suitable for the data with p-value greater than 0.001. Further, it is observed that the ACF of the resulting residuals is negligible.

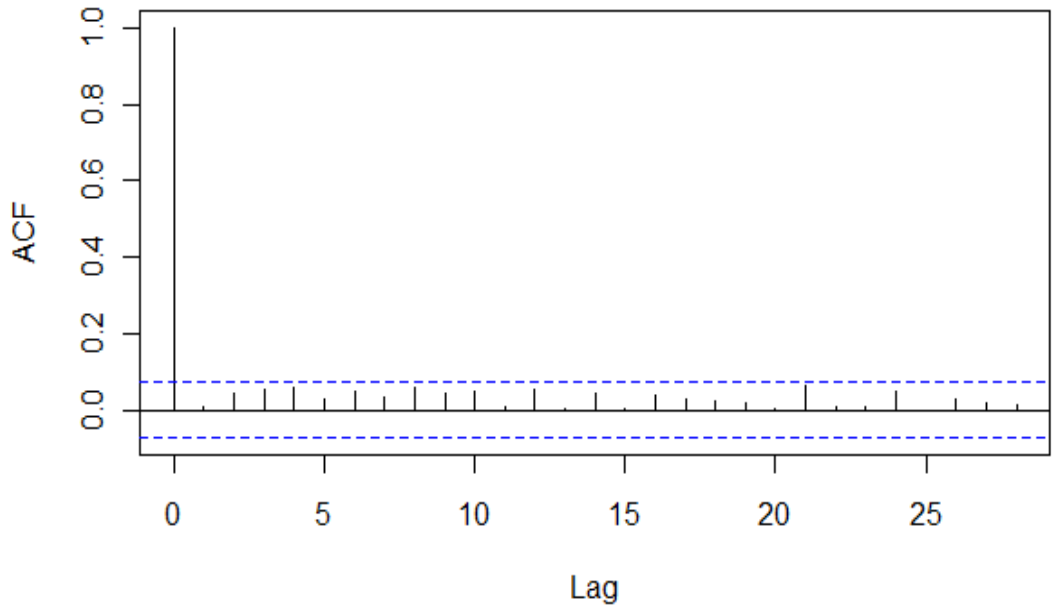

FIGURE 4.7: ACF of the residuals

The summary of this chapter is published in Balakrishna and Shiji (2010). In the next chapter, we discuss the properties of the stochastic volatility models induced by the Weibull PAR(1) model or exponentiated Extreme Value AR(1) model.

Chapter 5

Stochastic Volatility process generated by Gumbel Extreme Value Autoregressive model

5.1 Introduction

The Stochastic Volatility (SV) model introduced by Taylor (1986) is used to account for the well-documented autoregressive behaviour in the volatility of financial time series. The literature in this area mainly deals with the models with normal-lognormal distributions. An exponentiated Gaussian autoregressive sequence provides a Markov dependent sequence of log-normal random variables to describe the conditional variances, see Tsay (2005). Taylor (1994) suggested several alternative models to describe the evolution of conditional variances while modelling stochastic

volatilities. As quoted by Shephard (1996), volatility models provide an excellent testing ground for the development of new non-linear and non-Gaussian time series techniques. A number of autoregressive models are introduced for non-negative rvs in the context of non-Gaussian time series. In principle, one can very well use these autoregressive models to describe the evolution of time-dependent volatilities. As an alternative to normal-lognormal SV models, Abraham et al. (2006) proposed a SV model in which the volatility sequence is generated by a gamma AR(1) sequence of Gaver and Lewis (1980).

In this chapter, we study the properties of Gumbel Extreme Value Stochastic Volatility (GEV-SV) model where the volatilities are generated by GEVAR(1) process discussed in Chapter 3. One difficulty we faced in the analysis of GEV Markov sequence is that the transition distribution does not admit closed form density and hence the likelihood based inference is not an easy task. We used numerical techniques to optimize the likelihood function. We considered the GEV distribution corresponding to maximum rvs as marginal for the GEVAR(1) process in Chapter 3. When we employ this distribution for modelling volatility, we are getting the kurtosis of the return series in the form,

$$K_r = 3\,\frac{\Gamma(1-2\sigma)}{[\Gamma(1-\sigma)]^2},\ \sigma \leq 0.5.$$

Since there is restriction on the parameter σ, we consider Gumbel distribution corresponding to minimum rv as an alternative.

The GEV-SV model and its second order properties are described in Section 5.2 of this chapter. We discussed the estimation procedure by the method of moments in

Section 5.3. A simulation study is carried out in Section 5.4. In Section 5.5, we present the results on data analysis using our model.

5.2 Model and Properties

Let $\{r_t\}$ be a sequence of returns on certain financial asset and the volatilities are generated by a Markov sequence $\{\exp(h_t)\}$ of non-negative rvs. Define the SV model

$$r_t = \exp\left(h_t/2\right)\varepsilon_t; \tag{5.1}$$

$$h_t = \alpha h_{t-1} + \eta_t, \;\; t = 1, 2, ..., \;\; 0 \leq \alpha < 1, \tag{5.2}$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed standard normal random variables. We assume that the sequence $\{\varepsilon_t\}$ is independent of h_t and η_t for every t. Here we assume that for every t, the volatility, h_t, is a GEV rv with probability density function

$$f_{h_t}(x;\, \mu,\, \sigma) = \frac{1}{\sigma}\,\mathrm{Exp}\left(\frac{x-\mu}{\sigma}\right)\,\mathrm{Exp}\left(-\mathrm{Exp}\left(\frac{x-\mu}{\sigma}\right)\right), \tag{5.3}$$

$-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ and denoted by an $GEV\,(\mu,\ \sigma)$ distribution. Note that if $\{h_t\}$ is an Extreme value autoregressive process, then $\exp(h_t/2)$ has a Weibull distribution for every t. In order to have this marginal distribution for $\{h_t\}$ defined by (5.2) we need to have the distribution of η_t expressed by

$$\eta_t \stackrel{L}{=} (1-\alpha)\mu - \sigma\, Z, \text{ and } Z \stackrel{L}{=} -\log\left(U^{-\alpha}\right), \tag{5.4}$$

where $\underline{L}$ denotes the equality in distribution, U denotes a positive stable rv with Laplace transform, $\varphi(s) = e^{-s^{\alpha}}$, $0 \leq \alpha < 1$, whose density function does not admit a closed form expression. The mean and variance of $\{\eta_t\}$ are respectively given by

$$E(\eta_t) = (1-\alpha)(\mu - \sigma\gamma) = \mu^*, \text{ (say) and } \mathrm{V}(\eta_t) = (1-\alpha^2)\frac{\pi^2\sigma^2}{6} = \sigma^{2*}, \text{ (say)}, \tag{5.5}$$

where $\gamma \approx 0.5772$ is the Euler's constant. Since the sequence $\{\varepsilon_t\}$ follows standard normal distribution, the odd moments of r_t are zero and its even moments are given by

$$E\left(r_t^{2r}\right) = (2r-1)(2r-3)...3.1.e^{r\mu}\,\Gamma(r\sigma+1)\ , \quad r = 1, 2, ...$$

Then $V(r_t) = e^{\mu}\ \Gamma(\sigma+1)$ and the kurtosis of r_t becomes

$$K_r = 3\ \frac{\Gamma(2\sigma+1)}{[\Gamma(\sigma+1)]^2}. \tag{5.6}$$

By choosing different values for σ, one can get a distribution with larger kurtosis as shown in Figure 5.1.

The structure of the model (5.1) implies that the ACF of $\{r_t\}$ is zero and that of $\{r_t^2\}$ is significant. Thus, the variance and covariance function of the squared return series are obtained as

$$\begin{aligned} V\left(r_t^2\right) &= E\left(r_t^4\right) - \left(E\left(r_t^2\right)\right)^2 \\ &= e^{2\mu}\left(3\,\Gamma(2\sigma+1)\ - [\Gamma(\sigma+1)]^2\right). \\ \gamma_{r_t^2}(k) &= e^{2\mu}\left\{\frac{\Gamma(\sigma+1)\ \Gamma\left[\left(\alpha^k+1\right)\sigma+1\right]}{\Gamma(\alpha^k\sigma+1)}\ - [\Gamma(\sigma+1)]^2\right\}. \end{aligned}$$

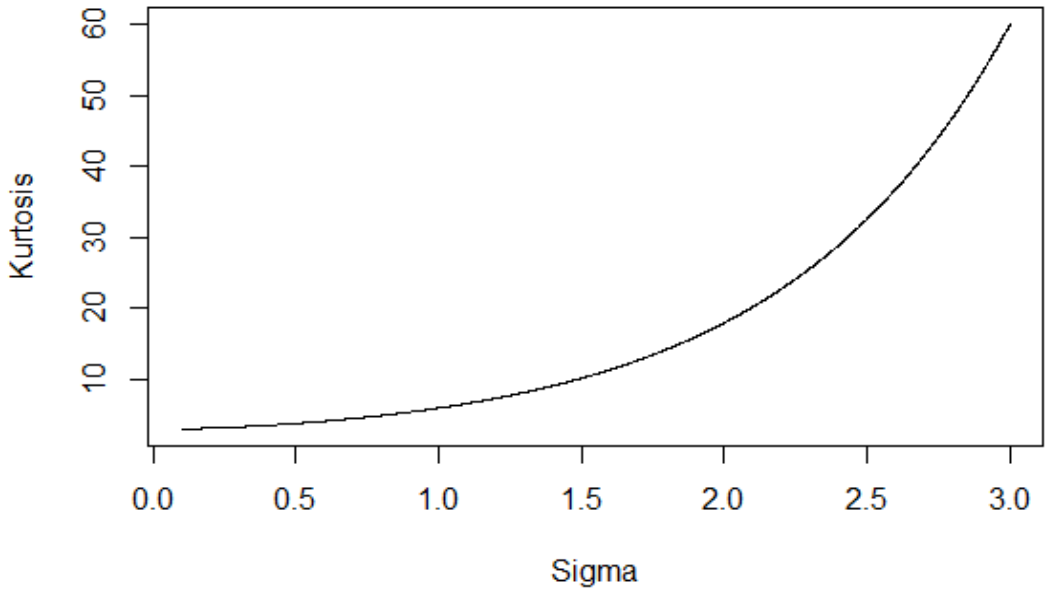

FIGURE 5.1: The plot of kurtosis, K_r, of r_t.

Hence, the lag k autocorrelation of the squared sequence $\{r_t^2\}$ is

$$\begin{aligned}\rho_{r_t^2}(k) &= Corr\left(r_t^2, r_{t-k}^2\right) \\ &= \frac{\Gamma(\sigma+1)\,\Gamma\left[\left(\alpha^k+1\right)\sigma+1\right] - \left[\Gamma(\sigma+1)\right]^2\,\Gamma\left(\alpha^k\sigma+1\right)}{\Gamma\left(\alpha^k\sigma+1\right)\left\{3\,\Gamma(2\sigma+1) - \left[\Gamma(\sigma+1)\right]^2\right\}}.\end{aligned} \tag{5.7}$$

The ACF is an exponentially decreasing function of the lags for different values of the parameters, as can be seen in Figure 5.2.

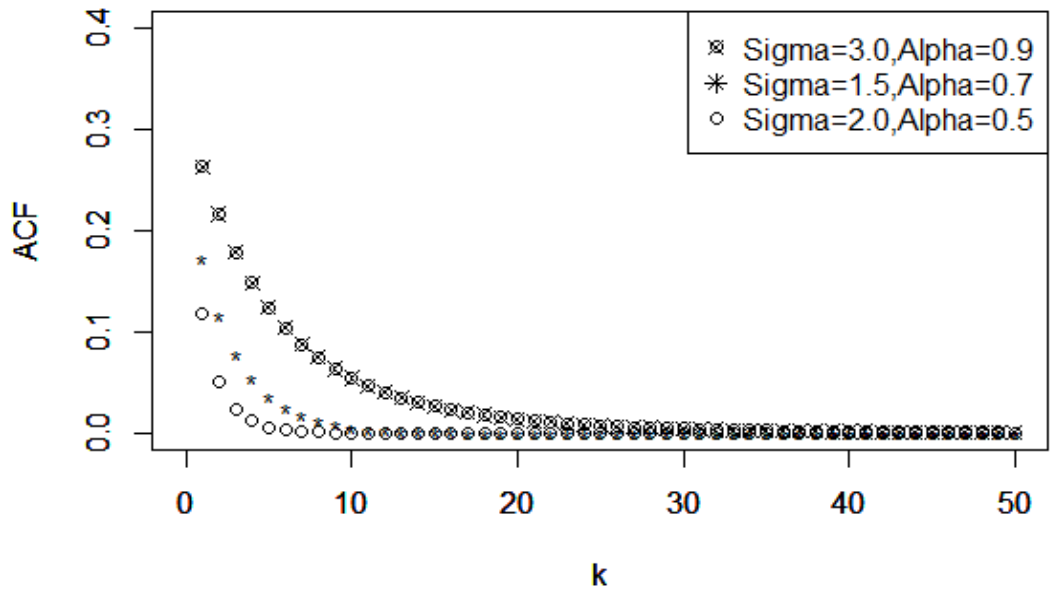

FIGURE 5.2: The ACF of squared return for different combinations of the parameters

5.3 Parameter Estimation

One of the difficulties with the statistical inference for SV models is that the likelihood function involves the unobservable Markov dependent latent variables. These variables have to be integrated out using multiple integrals and this complicates the parameter estimation by the method of maximum likelihood. Moreover, in the present case, the probability density function of η_t in expression (5.4) does not have a closed form and hence the other methods of estimation such as Bayesian estimation, Efficient importance sampling etc., may not be appropriate. In view of this, we employ the Generalized Method of Moments (GMM) introduced by Hansen (1982)

for estimation.

Let $(r_1, r_2, ..., r_T)$ be a realization of length T from the GEV-SV model (5.1) and $\theta = (\mu, \sigma, \alpha)'$ be the parameter vector to be estimated. We use the moments $E(r_t^2) = e^{\mu}\Gamma(\sigma+1)$, $E(r_t^4) = 3e^{2\mu}\Gamma(2\sigma+1)$ and $E(r_t^2 r_{t-1}^2) = e^{2\mu}\frac{\Gamma(\sigma+1)\Gamma[(\alpha+1)\sigma+1])}{\Gamma(\alpha\sigma+1)}$ to estimate the parameters.

If we define

$$f(r_t, r_{t-1}, \theta) = \begin{pmatrix} r_t^2 - e^{\mu}\Gamma(\sigma+1) \\ r_t^4 - 3e^{2\mu}\Gamma(2\sigma+1) \\ r_t^2 r_{t-1}^2 - e^{2\mu}\frac{\Gamma(\sigma+1)\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\sigma+1)} \end{pmatrix}. \tag{5.8}$$

Then the moment estimator $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\alpha})'$ of θ may be obtained by solving

$$\frac{1}{T}\sum_{t=1}^{T} f(r_t, r_{t-1}, \theta) = 0.$$

The resulting moment equations for μ, σ and α are expressed as

$$\hat{\mu} = \text{Log}\left(\frac{\bar{Y}_2}{\Gamma(\hat{\sigma}+1)}\right); \quad \frac{\bar{Y}_2^2}{\bar{Y}_4} = \frac{\Gamma(\hat{\sigma}+1)^2}{3\Gamma(2\hat{\sigma}+1)} \text{ and } \frac{\bar{Y}_{22}}{e^{2\mu}\Gamma(\sigma+1)} = \frac{\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\sigma+1)}, \tag{5.9}$$

where $\bar{Y}_2 = (1/T)\sum_{t=1}^{T} r_t^2$, $\bar{Y}_{22} = (1/T)\sum_{t=1}^{T} r_t^2 r_{t-1}^2$ and $\bar{Y}_4 = (1/T)\sum_{t=1}^{T} r_t^4$.

Hansen (1982) proved that under the following stated assumptions, the generalized moment estimators are consistent and asymptotically normal (CAN).

(i) $\{r_t : -\infty < t < \infty\}$ is stationary and ergodic sequence.

(ii) The parameter space Θ is an open subset of R^q that contains the true parameter θ_0.

(iii) $f(.,\theta)$ and $\partial f(.,\theta)/\partial\theta$ are Borel-measurable for each $\theta \in \Theta$ and $\partial f(r,.)/\partial\theta$ is continuous on Θ for each $r \in R^q$.

(iv) $\partial f_1/\partial\theta$ is first moment continuous at θ_0, $D = E[\partial f(r_t, \theta_0)/\partial\theta]$ exists, is finite, and has full rank.

(v) Let $\omega_t = f(r_t, \theta_0)$, $-\infty < t < \infty$ and $\vartheta_j = E(\omega_0|\omega_{-j}, \omega_{-j-1}, ...) - E(\omega_0|\omega_{-j-1}, \omega_{-j-2}, ...)$, $j \geq 0$. The assumptions are that $E(\omega_0\, \omega_0')$ exists and is finite, $E(\omega_0\,|\omega_{-j}, \omega_{-j-1}, ...)$ converges in mean square to zero and $\sum_{j=0}^{\infty} E(\vartheta_j'\vartheta_j)^{1/2}$ is finite.

Now, we have the following theorem, proved by Hansen (1982).

Theorem 5.1. *Suppose that the sequence* $\{r_t : -\infty < t < \infty\}$ *satisfies the assumptions (i) - (v). Then* $\left\{\sqrt{T}\left(\hat{\theta} - \theta\right), T \geq 1\right\}$ *converges in distribution to a normal random vector with mean 0 and dispersion matrix* $\left[D\, S^{-1} D'\right]^{-1}$, *where D is as given in (iv) and* $S = \sum_{k=-\infty}^{\infty} \Gamma_{(k)}$, $\Gamma_{(k)} = E\left(\omega_t\, \omega_{t-k}'\right)$.

Since the sequence $\{h_t\}$ defined by (5.2) is stationary and ergodic, it follows that the sequence $\{r_t\}$ given in (5.1) also possesses these properties. Further, all the moments of r_t and h_t are finite. We can shown that

$$\frac{\partial f}{\partial \mu} = \begin{pmatrix} -e^{\mu}\,\Gamma(\sigma+1) \\ -6\,e^{2\mu}\,\Gamma(2\sigma+1) \\ -2e^{2\mu}\,\frac{\Gamma(\sigma+1)\Gamma[(1+\alpha)\sigma+1]}{\Gamma(\alpha\sigma+1)} \end{pmatrix};$$

$$\frac{\partial f}{\partial \sigma} = \begin{pmatrix} -e^{\mu}\,\Psi(\sigma+1)\Gamma(\sigma+1) \\ -6\,e^{2\mu}\,\Psi(2\sigma+1)\Gamma(2\sigma+1) \\ -e^{2\mu}\,\frac{\Gamma(\sigma+1)\Gamma[(1+\alpha)\sigma+1]}{\Gamma(\alpha\sigma+1)}\,(\Psi(\sigma+1)-\alpha\,\Psi(\alpha\,\sigma+1)+(1+\alpha)\Psi((1+\alpha)\sigma+1)) \end{pmatrix};$$

and

$$\frac{\partial f}{\partial \alpha} = \begin{pmatrix} 0 \\ 0 \\ -e^{2\mu}\,\frac{\sigma\,\Gamma(\sigma+1)\Gamma[(1+\alpha)\sigma+1]}{\Gamma(\alpha\sigma+1)}\,(\Psi((1+\alpha)\sigma+1)-\ \Psi(\alpha\,\sigma+1)) \end{pmatrix},$$

where $\Psi(.)$ is a digamma function.

Thus, we see that $\partial f/\partial\theta$ exists and continuous for all θ. Similarly, we can show that $E\,(\partial f/\partial\theta)$ and $E\,(\omega_0\,\omega_0')$ exists and finite. Hence, the regularity conditions hold good for our SV model. Next, we derive the elements of the dispersion matrix which are required to compute the asymptotic standard errors of the estimators.

Let

$$\Gamma_{(k)} = \begin{pmatrix} \gamma_{11}^{(k)} & \gamma_{12}^{(k)} & \gamma_{13}^{(k)} \\ \gamma_{21}^{(k)} & \gamma_{22}^{(k)} & \gamma_{23}^{(k)} \\ \gamma_{31}^{(k)} & \gamma_{32}^{(k)} & \gamma_{33}^{(k)} \end{pmatrix}, \quad k = 0, \pm1, \pm2, ...,$$

and $\Gamma_{(k)} = \Gamma_{(-k)},\ k = 1, 2, \ldots$. Then the 3×3 matrix S is given by $S = \Gamma_{(0)} + 2\sum_{k=1}^{\infty}\Gamma_{(k)}$.

When $k = 0$, the elements of $\Gamma_{(0)} = E\left(\omega_t\,\omega_t'\right)$ are obtained as

$$\gamma_{11}^{(0)} = e^{2\mu}\left\{3\ \Gamma(2\sigma+1) - [\Gamma(\sigma+1)]^2\right\};$$

$$\gamma_{12}^{(0)} = e^{3\mu}\left\{15\,\Gamma(3\sigma+1) - 3\,\Gamma(\sigma+1)\,\Gamma(2\sigma+1)\right\};$$

$$\gamma_{13}^{(0)} = e^{3\mu}\left\{3\,\frac{\Gamma[(2\alpha+1)\sigma+1]\;\Gamma(2\sigma+1)}{\Gamma(2\alpha\sigma+1)} - \frac{[\Gamma(\sigma+1)]^2\,\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\sigma+1)}\right\};$$

$$\gamma_{21}^{(0)} = \gamma_{12}^{(0)};$$

$$\gamma_{22}^{(0)} = e^{4\mu}\left\{105\;\Gamma(4\sigma+1) - 9\,[\Gamma(2\sigma+1)]^2\right\};$$

$$\gamma_{23}^{(0)} = e^{4\mu}\left\{15\,\frac{\Gamma[(3\alpha+1)\sigma+1]\;\Gamma(3\sigma+1)}{\Gamma(3\alpha\sigma+1)} - 3\,\frac{\Gamma(\sigma+1)\;\Gamma(2\sigma+1)\;\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\sigma+1)}\right\};$$

$$\gamma_{31}^{(0)} = \gamma_{13}^{(0)};\;\; \gamma_{32}^{(0)} = \gamma_{23}^{(0)};$$

$$\gamma_{33}^{(0)} = e^{4\mu}\left\{9\,\frac{\Gamma[(2\alpha+2)\sigma+1]\;\Gamma(2\sigma+1)}{\Gamma(2\alpha\sigma+1)} - \frac{[\Gamma(\sigma+1)]^2\,\{\Gamma[(\alpha+1)\sigma+1]\}^2}{[\Gamma(\alpha\sigma+1)]^2}\right\}.$$

Similarly, the following are the elements of $\Gamma_{(k)}$ for $k = 1, 2, ...,$

$$\gamma_{11}^{(k)} = e^{2\mu}\left\{\frac{\Gamma(\sigma+1)\,\Gamma[(\alpha^k+1)\sigma+1]}{\Gamma(\alpha^k\sigma+1)} - [\Gamma(\sigma+1)]^2\right\};$$

$$\gamma_{12}^{(k)} = 3\,e^{3\mu}\left\{\frac{\Gamma(\sigma+1)\,\Gamma[(\alpha^k+2)\sigma+1]}{\Gamma(\alpha^k\sigma+1)} - \Gamma(\sigma+1)\,\Gamma(2\sigma+1)\right\};$$

$$\gamma_{13}^{(k)} = e^{3\mu}\left\{\frac{\Gamma[(\alpha^{k+1}+\alpha+1)\sigma+1]\,\Gamma[(\alpha^k+1)\sigma+1]\,\Gamma(\sigma+1)}{\Gamma(\alpha^k\sigma+1)\,\Gamma[(\alpha^{k+1}+\alpha)\sigma+1]} - \frac{[\Gamma(\sigma+1)]^2\,\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\sigma+1)}\right\};$$

$$\gamma_{21}^{(k)} = 3\,e^{3\mu}\left\{\frac{\Gamma(2\sigma+1)\,\Gamma[(2\alpha^k+1)\sigma+1]}{\Gamma(2\alpha^k\sigma+1)} - \Gamma(\sigma+1)\,\Gamma(2\sigma+1)\right\};$$

$$\gamma_{22}^{(k)} = 9e^{4\mu}\left\{\frac{\Gamma(2\sigma+1)\,\Gamma[(2\alpha^k+2)\sigma+1]}{\Gamma(2\alpha^k\sigma+1)} - [\Gamma(2\sigma+1)]^2\right\};$$

$$\gamma_{23}^{(k)} = 3\,e^{4\mu}\left\{\frac{\Gamma(2\sigma+1)\,\Gamma[(2\alpha^{k}+1)\sigma+1]\,\Gamma[(2\alpha^{k+1}+\alpha+1)\sigma+1]}{\Gamma[(2\alpha^{k+1}+\alpha)\sigma+1]\,\Gamma(2\alpha^{k}\sigma+1)} - \frac{\Gamma(2\sigma+1)\,\Gamma(\sigma+1)\,\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\sigma+1)}\right\};$$

$$\gamma_{31}^{(k)} = e^{3\mu}\left\{\frac{\Gamma(\sigma+1)\,\Gamma[(\alpha+1)\sigma+1]\,\Gamma[(\alpha^{k}+\alpha^{k-1}+1)\sigma+1]}{\Gamma[(\alpha^{k}+\alpha^{k-1})\sigma+1]\,\Gamma(\alpha\sigma+1)} - \frac{[\Gamma(\sigma+1)]^{2}\,\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\sigma+1)}\right\};$$

$$\gamma_{32}^{(k)} = 3\,e^{4\mu}\left\{\frac{\Gamma(\sigma+1)\,\Gamma[(\alpha+1)\sigma+1]\,\Gamma[(\alpha^{k}+\alpha^{k-1}+2)\sigma+1]}{\Gamma[(\alpha^{k}+\alpha^{k-1})\sigma+1]\;\Gamma(\alpha\sigma+1)} - \frac{\Gamma(2\sigma+1)\;\Gamma(\sigma+1)\,\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\sigma+1)}\right\};$$

$$\gamma_{33}^{(k)} = e^{4\mu}\left\{\frac{\Gamma(\sigma+1)\Gamma[(\alpha+1)\sigma+1]\Gamma[(\alpha^{k}+\alpha^{k-1}+1)\sigma+1]\Gamma[(\alpha^{k+1}+\alpha^{k}+\alpha+1)\sigma+1]}{\Gamma[(\alpha^{k+1}+\alpha^{k}+\alpha)\sigma+1)\Gamma[(\alpha^{k}+\alpha^{k-1})\sigma+1]\Gamma(\alpha\sigma+1)} - \frac{[\Gamma(\sigma+1)]^{2}\,\{\Gamma[(\alpha+1)\sigma+1]\}^{2}}{[\Gamma(\alpha\sigma+1)]^{2}}\right\}.$$

The 3×3 matrix D is evaluated using the form $D = E\left(d\,f\left(r_t,\,\theta\right)/d\theta\right)$ and its elements are:

$$D_{11} = -\,e^{\mu}\,\Gamma(\sigma{+}1);$$

$$D_{12} = -\,6\,e^{2\mu}\,\Gamma(2\sigma{+}1);$$

$$D_{13} = -\,2\,e^{2\mu}\frac{\Gamma(\sigma+1)\;\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\sigma+1)};$$

$$D_{21} = -\,e^{\mu}\,\Gamma(\sigma{+}1)\;\Psi(1{+}\sigma);$$

$$D_{22} = -\,6\,e^{2\mu}\,\Gamma(2\sigma{+}1)\,\Psi(1{+}2\sigma);$$

$$D_{23} = -\,e^{2\mu}\frac{\Gamma(\sigma+1)\,\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\sigma+1)}\,\{\Psi(1+\sigma)+(1+\alpha)\,\Psi[1+(1+\alpha)\sigma] - \alpha\,\Psi(1+\alpha\,\sigma)\}\,;$$

$$D_{31} = D_{32} = 0;$$

$$D_{33} = -\,e^{2\mu}\frac{\Gamma(\sigma+1)\;\Gamma[(\alpha+1)\sigma+1]}{\Gamma(\alpha\,\sigma+1)}\,[(1+\alpha)\Psi\,(1+(\alpha+1)\sigma) - \alpha\,\Psi\,(1+\alpha\,\sigma)]\,;$$

Hence the asymptotic dispersion matrix becomes $\frac{1}{T}\Sigma$, where

$$\Sigma = \left[D\,S^{-1}D'\right]^{-1}. \tag{5.10}$$

The diagonal elements of this matrix are used to compute the asymptotic standard errors of the estimators in the next section.

5.4 Simulation Study

We carry out a simulation study to understand the performance of the estimators with sample sizes 1000 and 2000. First, we generate a sample of size T from the GEV Markov sequence, $\{h_t\}$, specified in (5.2) using the innovation random variable described in (5.4). The method of generating GEV Markov sequence is described in Section 3.4. Based on this realization, we simulate the sequence $\{r_t\}$ using (5.1). Then we obtain the estimates by solving (5.9). The algorithm are given in Appendix C.

TABLE 5.1: The average estimates and the corresponding root mean square errors of the moment estimates based on simulated observations of sample size n=1000. The estimates of asymptotic standard deviations are also given.

True Values			$\hat{\mu}$			$\hat{\sigma}$			$\hat{\alpha}$		
μ	σ	α	Mean	Std Dev	Asymp. Std.Dev	Mean	Std Dev	Asymp. Std.Dev	Mean	Std Dev	Asymp. Std.Dev
0.5	1.5	0.9	0.5334	0.2627	0.3229	1.4061	0.2471	0.3875	0.9158	0.2764	0.2593
		0.8	0.5342	0.1943	0.2709	1.4085	0.2220	0.3602	0.8207	0.2677	0.2642
		0.7	0.5287	0.1789	0.2513	1.4175	0.2255	0.3509	0.7173	0.2534	0.2702
		0.5	0.5137	0.1719	0.2347	1.4424	0.2340	0.3438	0.4935	0.2259	0.2872
		0.2	0.5237	0.1563	0.2253	1.4321	0.2246	0.3408	0.1995	0.1768	0.3408
0.7	1	0.9	0.7016	0.1699	0.1704	0.9611	0.1710	0.2038	0.9480	0.3436	0.1820
		0.8	0.7022	0.1227	0.1307	0.9628	0.1598	0.1912	0.8339	0.3174	0.1871
		0.7	0.7027	0.1083	0.1145	0.9629	0.1488	0.1869	0.7301	0.2956	0.1934
		0.5	0.6950	0.0974	0.0996	0.9739	0.1613	0.1835	0.5157	0.2606	0.2122
		0.2	0.6984	0.0834	0.0901	0.9757	0.1545	0.1819	0.2063	0.1962	0.2778
1.5	2	0.9	1.5969	0.3552	0.6871	1.8065	0.3163	0.7483	0.9021	0.2579	0.4191
		0.8	1.6098	0.2998	0.6196	1.8343	0.3134	0.6940	0.8162	0.2565	0.4256
		0.7	1.5939	0.2789	0.5960	1.8458	0.3199	0.6757	0.7171	0.2766	0.4335
		0.5	1.5872	0.2511	0.5776	1.8633	0.3003	0.6623	0.5013	0.2311	0.4545
		0.2	1.6020	0.2506	0.5690	1.8640	0.3083	0.6574	0.1937	0.1829	0.5117
2	1	0.9	1.9917	0.1615	0.1809	0.9688	0.1679	0.2258	0.9336	0.3432	0.4216
		0.8	1.9999	0.1283	0.1372	0.9645	0.1702	0.2025	0.8268	0.3152	0.3993
		0.7	1.9979	0.1102	0.1191	0.9598	0.1601	0.1941	0.7307	0.3129	0.3954
		0.5	2.0030	0.0915	0.1021	0.9694	0.1583	0.1871	0.5312	0.2649	0.4035
		0.2	2.0017	0.0857	0.0910	0.9698	0.1531	0.1830	0.2078	0.2017	0.4561

TABLE 5.2: The average estimates and the corresponding root mean square errors of the moment estimates based on simulated observations of sample size n=2000. The estimates of asymptotic standard deviations are also given.

True Values			$\hat{\mu}$			$\hat{\sigma}$			$\hat{\alpha}$		
μ	σ	α	Mean	Std Dev	Asymp. Std.Dev	Mean	Std Dev	Asymp. Std.Dev	Mean	Std Dev	Asymp. Std.Dev
0.5	1.5	0.9	0.5191	0.1953	0.2283	1.4366	0.1858	0.2740	0.9238	0.2151	0.1834
		0.8	0.5126	0.1504	0.1916	1.4600	0.1876	0.2547	0.8137	0.2104	0.1868
		0.7	0.5083	0.1453	0.1778	1.4687	0.1942	0.2481	0.7069	0.2050	0.1912
		0.5	0.5128	0.1269	0.1660	1.4681	0.1743	0.2431	0.5024	0.1736	0.2031
		0.2	0.5126	0.1215	0.1593	1.4676	0.1803	0.2411	0.2028	0.1303	0.2410
0.7	1	0.9	0.7023	0.1189	0.1205	0.9715	0.1242	0.1441	0.9245	0.2432	0.1287
		0.8	0.6998	0.0904	0.0924	0.9853	0.1224	0.1352	0.8157	0.2323	0.1323
		0.7	0.6978	0.0821	0.0809	0.9884	0.1219	0.1321	0.7205	0.2149	0.1368
		0.5	0.6987	0.0656	0.0704	0.9835	0.118	0.1297	0.5181	0.1876	0.1500
		0.2	0.6988	0.0623	0.0637	0.9799	0.1127	0.1286	0.1982	0.1428	0.1964
1.5	2	0.9	1.5809	0.2847	0.4859	1.8644	0.2584	0.5291	0.9140	0.2101	0.2963
		0.8	1.5856	0.2396	0.4381	1.8823	0.2563	0.4907	0.8123	0.2057	0.3009
		0.7	1.5647	0.2356	0.4214	0.9142	0.2586	0.4778	0.7153	0.1990	0.3065
		0.5	1.5569	0.2373	0.4084	1.9137	0.2694	0.4683	0.5057	0.1838	0.3213
		0.2	1.5544	0.2229	0.4023	1.9198	0.2676	0.4649	0.1918	0.1364	0.3618
2	1	0.9	2.0062	0.1203	0.1279	0.9747	0.1215	0.1597	0.9223	0.2456	0.2981
		0.8	1.9999	0.0884	0.0970	0.9795	0.1167	0.1431	0.8284	0.2240	0.2824
		0.7	1.9984	0.0809	0.0842	0.9844	0.1180	0.1372	0.7235	0.2116	0.2796
		0.5	2.0024	0.0653	0.0722	0.9770	0.1115	0.1323	0.5074	0.1889	0.2853
		0.2	2.0009	0.0588	0.0643	0.9788	0.1086	0.1294	0.2002	0.1447	0.3225

For each specified value of the parameter, we repeat the experiment 1000 times for computing the estimates and then averaged them over the repetitions. The average estimates and corresponding Root Mean Square Errors (RMSE) based on simulated observations are reported in Tables 5.1 and 5.2.

We also calculate the asymptotic standard deviation based on the theoretical values of the parameters. They are obtained as the square root of the diagonal elements of the corresponding asymptotic dispersion matrices given in (5.10). These values are compared with the estimated RMSE of the respective estimates. For a better estimation method these values become closer. From the above tables, we observe that the estimates $\hat{\mu}$, $\hat{\alpha}$, and $\hat{\sigma}$ are slightly biased. When the sample size is large, the estimates perform reasonably well and there is a significant reduction in asymptotic standard deviations and root mean square errors. Hence, we claim that the method of moment estimation yields good estimates for the parameters involved.

5.5 Data Analysis

To illustrate the application of the proposed model, we analysed the daily stock price index returns using the model defined above. We use three sets of data for the analysis purpose: (1) the closing index data of Bombay Stock Exchange (BSE) for the period 2000-2010, (2) the opening index data of Standard and Poor 500 (S&P 500) for the period 2000-2009 and (3) low index data of S&P 500 for the period 2000-2009. The time series plots of these data are given in Figure 5.3. The left

panels show the plots of actual data series and the log-return series are on the right panels.

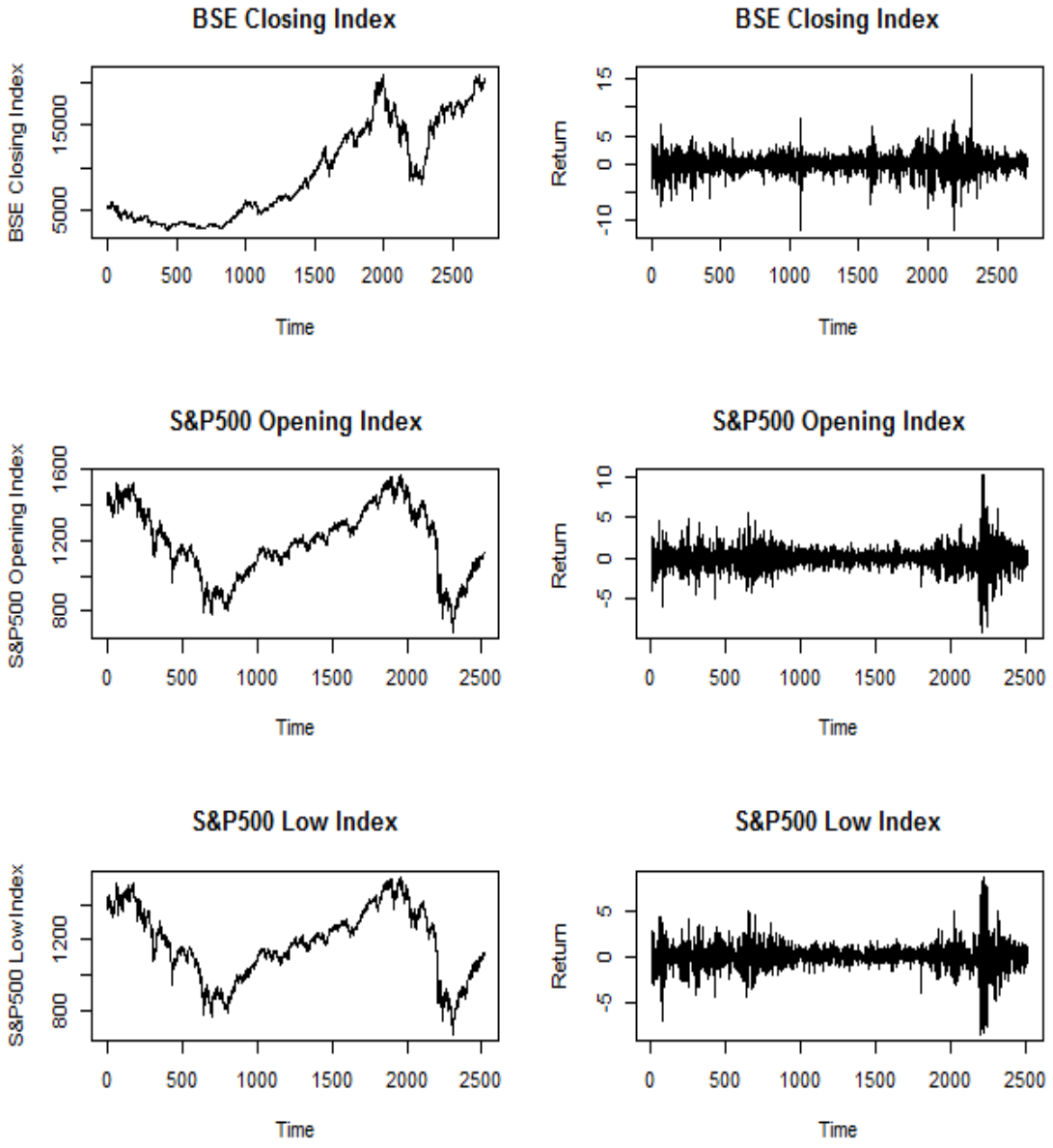

FIGURE 5.3: Time series plot of the stock prices and the return

Denoting the daily price index by p_t, the returns are transformed into continuously compounded rates centered around their sample mean:

$$r_t = 100\left[\ln\left(\frac{p_t}{p_{t-1}}\right) - \left(\frac{1}{T}\right)\sum_{t=1}^{T}\ln\left(\frac{p_t}{p_{t-1}}\right)\right].$$

The summary statistics of the return series are reported in Table 5.3, where $Q(20)$ and $Q^{(2)}(20)$ are the Ljung-Box statistic for return and squared return series with lag 20, respectively. The corresponding χ^2 table value at 5% significance level is 10.117. Hence, the test suggests that the return series is serially uncorrelated whereas the squared return series has significant serial correlation. The kurtosis of the returns for all the series is greater than three which implies that the distribution of the returns is leptokurtic in nature.

TABLE 5.3: Summary statistics of the return series

Statistics	BSE Closing Index	S&P500 Opening Index	S&P500 Low Index
Sample size	2725	2515	2515
Std. Dev	1.7331	1.3572	1.2654
Kurtosis	9.0457	10.1669	11.9490
Minimum	-11.8583	-9.1044	-8.5578
Maximum	15.9408	10.1499	8.7272
$Q(20)$	5.9814	0.0293	0.0826
$Q^{(2)}(20)$	13.9558	87.1349	47.0954

From the ACF of the returns plotted in Figure 5.4, it is observed that serial correlations in the return series are insignificant where as the ACF of the squared returns in the bottom panel declines slowly with increasing lags.

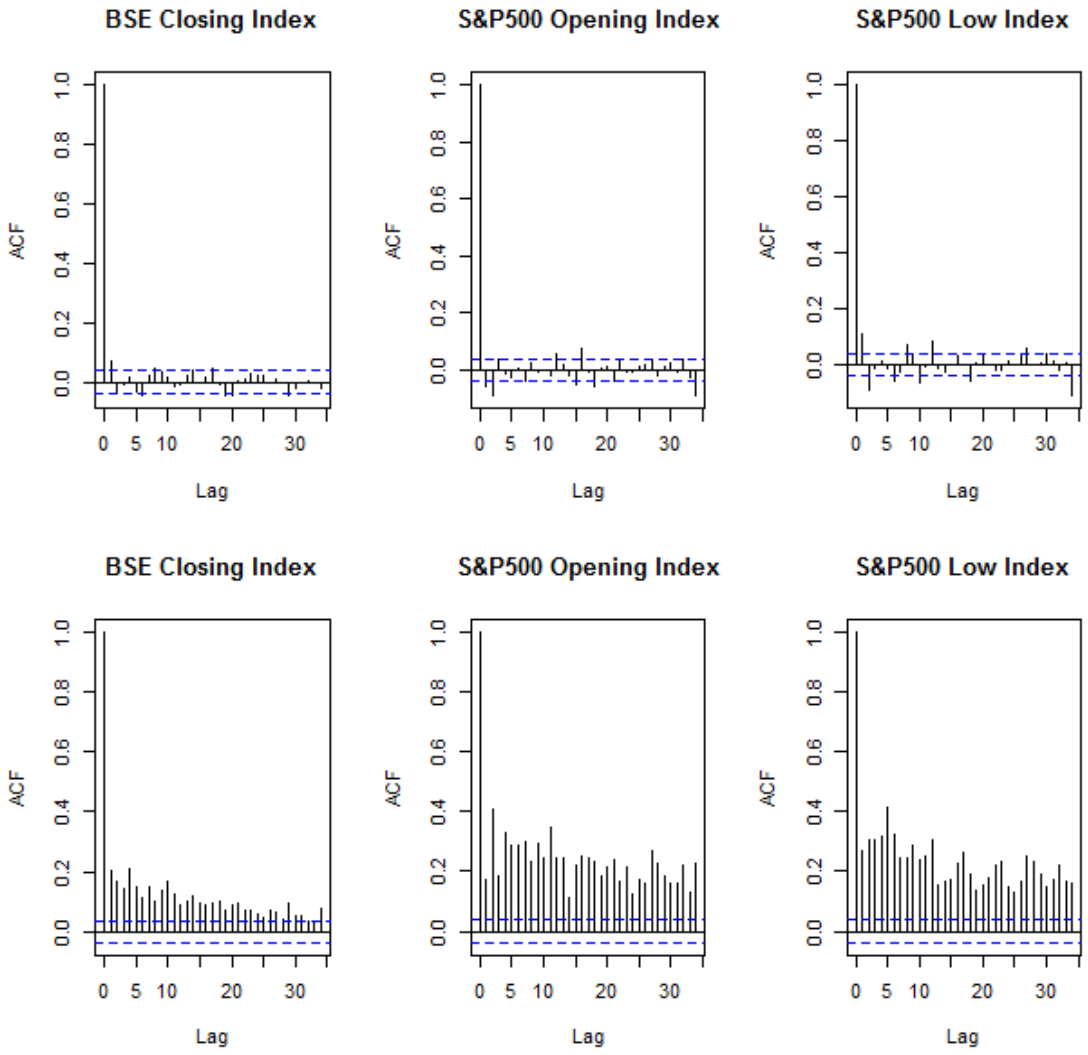

FIGURE 5.4: ACF of the returns(top panels) and the squared returns(bottom panels)

In Table 5.4, we present the parameter estimates for each of the return series. The values of $\hat{\alpha}$ in the table suggest that there is a significant persistence of volatility in the data series.

TABLE 5.4: Parameter estimates using Method of Moments

Parameters	BSE Closing Index	S&P500 Opening Index	S&P500 Low Index
$\hat{\mu}$	0.8868	0.3265	0.0759
$\hat{\sigma}$	1.3927	1.4986	1.6459
$\hat{\alpha}$	0.8311	0.7006	0.9967

Once the estimates of parameters are obtained, the next stage is the model diagnostic checking. That is, we need to check whether the assumptions on the model (5.1) are satisfied with respect to the data we have analysed. Note that the model (5.1) is in terms of the volatilities h_t, which are unobservable. This aspect makes the diagnosis problem difficult. One of the methods suggested in such cases is to employ Kalman filtering by rewriting the model (5.1) in the state-space form. The state-space approach and Kalman filter algorithm for the local trend model are discussed in Chapter 2. We can very well use this procedure for estimating the unobservable volatility, h_t. For more details on Kalman filter method and associated theory, one can refer Jacquier et al. (1994) and Tsay (2005). Since the Kalman filter method is developed under the normality assumptions, we approximate the distribution of η_t specified in (5.2) by a normal distribution and then adopt Kalman filter method for estimating the volatilities. Using these estimated volatilities, we can compute the residuals using the equation (5.1).

The state-space representation of the SV model given in (5.1) can be written as

$$\log r_t^2 = -1.27 + h_t + \nu_t\,, \qquad E(\nu_t) = 0,\ V(\nu_t) = \frac{\pi^2}{2}; \tag{5.11}$$

and

$$h_t = \alpha\, h_{t-1} + \eta_t,$$

where η_t is assumed to be normally distributed with mean $(1-\alpha)(\mu - \sigma\,\gamma)$ and variance $(1-\alpha^2)\,(\pi^2\sigma^2)/6$ which are given in (5.5). If the distribution of ν_t is approximated by a normal distribution then the preceding system (5.11) becomes a standard dynamic linear model, to which the Kalman filter can be applied. Let $\bar{h}_{t\,|\,t-1}$ be the prediction of h_t based on the information available at time $t-1$ and $\Omega_{t\,|\,t-1}$ be the variance of the predictor. Here we are making a presumption that the update that uses the information at time t as $\bar{h}_{t\,|\,t}$ and the variance of the update as $\Omega_{t\,|\,t}$. The equations that recursively compute the predictions and updating are given by

$$\bar{h}_{t\,|\,t-1} = \alpha\,\bar{h}_{t-1\,|\,t-1} + (1-\alpha)\,(\mu - \sigma\,\gamma)$$

$$\Omega_{t\,|\,t-1} = \alpha^2\,\Omega_{t-1\,|\,t-1} + (1-\alpha^2)\,\frac{\pi^2\sigma^2}{6}$$

and

$$\bar{h}_{t\,|\,t} = \bar{h}_{t\,|\,t-1} + \frac{\Omega_{t\,|\,t-1}}{f_t}\left[\log r_t^2 + 1.27 - \bar{h}_{t\,|\,t-1}\right]$$

$$\Omega_{t\,|\,t} = \Omega_{t\,|\,t-1}\left(1 - \frac{\Omega_{t\,|\,t-1}}{f_t}\right)$$

where $f_t = \Omega_{t\,|\,t-1} + \frac{\pi^2}{2}$.

Then the residuals are calculated by the equation

$$\hat{\varepsilon}_t = r_t \exp\left(-\frac{\bar{h}_t}{2}\right) \tag{5.12}$$

and use this sequence for the residual analysis. The system is initialized at the unconditional values, $\Omega_0 = \frac{\pi^2\sigma^2}{6}$ and $h_0 = \mu - \sigma\,\gamma$. The residual analysis is carried out using this prediction error. The correlograms of the residuals given in Figure 5.5 below suggest that the model performs quite well. The parameters μ, σ and α in the above system are replaced by their respective estimates which are given in Table 5.4.

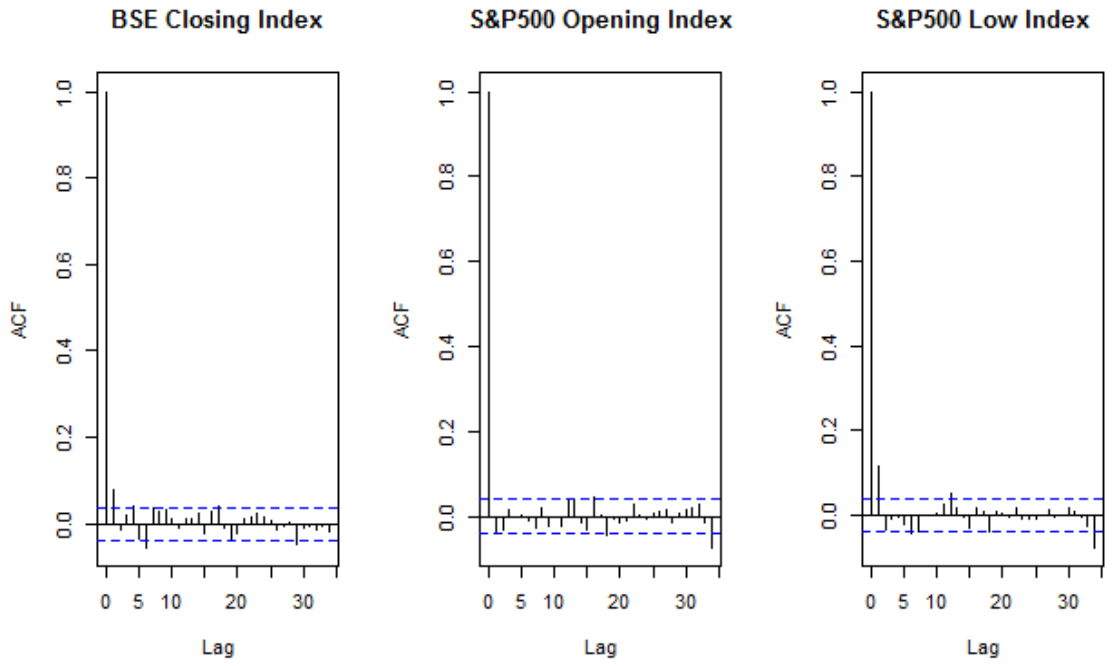

FIGURE 5.5: ACF of the residuals

Further, we also checked the significance of ACF in the residuals by computing the Ljung-Box statistic for the series $\{\hat{\varepsilon}_t\}$ and $\{\hat{\varepsilon}_t^2\}$, which are summarized in the Table 5.5. All these values are less than the 5% Chi-square critical value (10.117) at degrees of freedom 20. Hence we conclude that there is no significant serial dependence among the residuals and the squared residuals.

TABLE 5.5: Ljung-Box Statistic for the residuals and squared residuals

Data	Ljung-Box statistic	
	Residuals	Squared Residuals
BSE Closing Index	3.1938	1.4479
S&P500 Opening Index	0.0827	4.6594
S&P500 Low Index	0.2401	0.0412

In the following Figure 5.6, we superimpose the histogram of the residuals with standard normal density to check whether the series follows standard normal distribution. The figure clearly shows that the standard normal distribution is a good fit for the residuals in all the three cases. The results of this chapter are published in Balakrishna and Shiji (2014b).

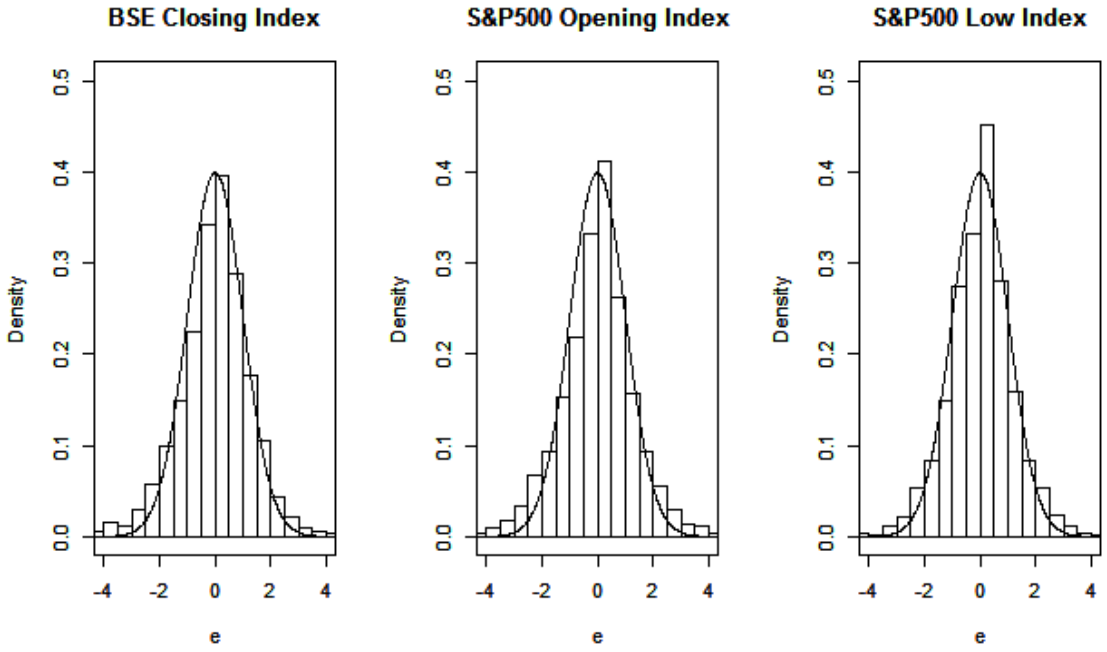

FIGURE 5.6: Histogram of residuals with superimposed standard normal density

Chapter 6

Bivariate Exponential Model with Product Structure

6.1 Introduction

This chapter is a by-product of our studies on product autoregressive models for non-negative rvs. The univariate exponential distribution is very popular among researchers working in the areas such as reliability analysis, life testing, survival analysis, etc. When there are two or more variables affecting the system, in most of the cases the analysis carried out by assuming that they are statistically independent. However, the assumption of independence does not hold in practice. As a consequence, several bivariate models have been introduced in the literature. A detailed up-to-date survey of bivariate distributions and the methods of their

constructions along with their applications may be found in a recent book by Balakrishnan and Lai (2009). Such distributions are important candidates for joint modelling of non-negative variables such as lifetime and repair time of equipments, wind velocity and wave heights, etc. Some well known bivariate exponential distributions are those by Gumbel (1960), Freund (1961), Marshall and Olkin (1967), Block and Basu (1974), Downton (1970) and so on.

Several bivariate exponential distributions were discussed in the literature on non-Gaussian time series (cf, Lawrance and Lewis (1985), Gaver and Lewis (1980), Iyer et al. (2002)) where the component random variables were linked through a linear relation. One of the drawbacks of these models is that they are not absolutely continuous and possess singular components. Further, they have restriction on the range of correlation coefficient. Iyer et al. (2002) considered the detailed analysis of a bivariate exponential distribution constructed using the first order exponential autoregressive model of Gaver and Lewis (1980). Their bivariate distribution is expressed in terms of the Laplace Transform as there is no explicit form for its probability density function. Further there is mass concentration on the line $y = a\,x$, which makes the joint distribution singular.

In this chapter, we introduce a class of absolutely continuous bivariate exponential distributions where the components are linked through a product structure. McKenzie (1982) proposed product form of an autoregressive model of order one (PAR(1)) to generate a gamma sequence and proved a characterizing autocorrelation property. Abraham and Balakrishna (2012) and Balakrishna and Lawrence (2012) discussed several aspects of the PAR(1) models and proposed methods of

estimating their parameters. In this chapter, we use the PAR(1) structure to define an absolutely continuous bivariate exponential distribution.

The chapter is organized as follows: in Section 6.2 and 6.3, we discuss the properties of the bivariate exponential distribution. Bivariate exponential distribution with negative correlation is defined in Section 6.4. Section 6.5 describes the statistical inference for the Bivariate distribution. Simulation studies are conducted in Section 6.6. The applications of the bivariate exponential distribution are given in Section 6.7.

6.2 Models with Product Structure

Let (X, Y) be a random vector defined on non-negative quadrant of R^2 with survival function

$$\bar{F}(x, y) = P(X > x, Y > y),\ x \geq 0,\ y \geq 0.$$

We say that (X, Y) has a specific bivariate distribution D if both the marginal random variables X and Y have the univariate distributions D. Let $F_X(.)$ and $F_Y(.)$ be the distribution functions of X and Y respectively and define

$$Y = X^{\alpha} Z,\ \ 0 \leq \alpha < 1, \tag{6.1}$$

where Z is a non-negative rv, independent of X, such that the equality in (6.1) holds in distribution. One of our tasks here is to identify distribution of Z, for which (X, Y) has a bivariate exponential distribution when the marginal rvs are

tied together using (6.1). The joint survival function of X and Y is obtained as

$$\begin{aligned}\bar{F}(x,y) &= P(X>x, Y>y)\\ &= \int_x^\infty P[Y>y|X=u]dF_X(u).\end{aligned}$$

Using (6.1), we can write the above expression as,

$$\begin{aligned}&= \int_x^\infty P[X^\alpha Z > y|X=u]\, dF_X(u)\\ &= \int_x^\infty P\left[Z > \frac{y}{u^\alpha}\right] dF_X(u)\\ &= \int_x^\infty \bar{F}_Z\left(\frac{y}{u^\alpha}\right)\, dF_X(u).\end{aligned}$$

Applying the integration by parts, we get

$$\begin{aligned}&= \bar{F}_Z\left(\frac{y}{u^\alpha}\right)\int_x^\infty dF_X(u) - \int_x^\infty\left(\int_x^\infty dF_X(u)\right) d\bar{F}_Z\left(\frac{y}{u^\alpha}\right)\\ &= \bar{F}_Z\left(\frac{y}{x^\alpha}\right)\bar{F}_X(x) - \int_x^\infty \bar{F}_X(u)\, dF_Z\left(\frac{y}{u^\alpha}\right)\\ &= \bar{F}_Z\left(\frac{y}{x^\alpha}\right)\bar{F}_X(x) - \int_{y/x^\alpha}^0 \bar{F}_X\left((y/z)^{1/\alpha}\right) d\bar{F}_Z(z)\\ &= \int_0^{y/x^\alpha} \bar{F}_X\left((y/z)^{1/\alpha}\right) d\bar{F}_Z(z) + \bar{F}_X(x)\bar{F}_Z\left((y/x^\alpha)\right). \qquad (6.2)\end{aligned}$$

We need to identify $F_Z(.)$ which provides a meaningful distribution for (X,Y). If there exist density functions for the rvs involved then the joint probability density function may be expressed as

$$f(x,y) = f_X(x) f_Z\left(\frac{y}{x^\alpha}\right)\frac{1}{x^\alpha},\ x>0, y>0. \qquad (6.3)$$

In the following sections, we study the detailed properties of the proposed bivariate exponential distribution.

6.3 Bivariate Exponential Models

Balakrishna and Lawrence (2012) developed a product form of the first order autoregressive time series model for exponential rvs. The exponential product autoregressive model is defined as

$$\vartheta_t = \vartheta_{t-1}^{\alpha} V_t\,, \quad t = 1, 2, \ldots \;\; 0 < \alpha < 1,$$

where $\{V_t\}$ is assumed to be a sequence of independent, identically and non-negatively distributed innovation rvs. The marginal distribution of $\{\vartheta_t\}$ is assumed to be exponential with survival function

$$\bar{F}_{\vartheta_t}(\theta) = e^{-\lambda\,\theta}, \;\; \theta \geq 0,\, \lambda > 0.$$

Corresponding to this marginal distribution, the innovation rv can be expressed as

$$V_t \stackrel{L}{=} \lambda^{-1}(\lambda/U)^{\alpha},$$

where U is the positive stable random variable with Laplace transform,

$$\varphi_U(s) = E\left(e^{-sU}\right) = \exp\left(-s^{\alpha}\right), 0 < \alpha < 1, \tag{6.4}$$

and the pdf of U is given by

$$f_U(u;\alpha) = \frac{1}{\pi u}\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha+1)}{\Gamma(k+1)}(-1)^k\frac{1}{u^{\alpha k}}\sin(-k\pi\alpha), \quad u>0. \tag{6.5}$$

We also have the pdf of $W = U^{-\alpha}$ which may be expressed as

$$f_W(v;\alpha) = \frac{1}{\pi}\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha)}{\Gamma(k)}(-v)^{k-1}\sin(k\pi\alpha), \quad v>0. \tag{6.6}$$

If we choose Z in (6.1) as a suitable function of a positive stable rv U, we can get the required bivariate exponential distribution for (X, Y) with the product structure defined in (6.1).

Let X be an exponential rv with pdf,

$$f(x;\lambda) = \lambda\exp(-\lambda x), \;\; x>0, \tag{6.7}$$

and Z be defined by

$$Z = \beta^{-1}(\lambda/U)^{\alpha}. \tag{6.8}$$

For any α, $0 \le \alpha < 1$ consider the distribution of $Y = X^{\alpha}Z$. That is, consider

$$\begin{aligned} P[Y>y] &= P[X^{\alpha}Z>y] \\ &= P\left[X^{\alpha}\frac{1}{\beta}\left(\frac{\lambda}{U}\right)^{\alpha}>y\right] \\ &= \int_0^{\infty} P\left[X>(\beta y)^{1/\alpha}(u/\lambda)\right]\; dF_U(u), \end{aligned}$$

where $F_U(u)$ is the distribution function of the positive stable rv, U. Since X has $Exponential(\lambda)$ distribution, we can write

$$\begin{aligned} P[Y > y] &= \int_0^\infty \exp\left\{-(y\beta)^{1/\alpha}u\right\} dF_U(u) \\ &= E\left(e^{-(y\beta)^{1/\alpha}U}\right) = \varphi_U((y\beta)^{1/\alpha}), \end{aligned}$$

where $\varphi_U(s)$ is the Laplace transform of the positive stable rv U. Now, using (6.4) on the right hand side, we get the marginal survival function of Y as

$$P[Y > y] = e^{-\beta y}. \tag{6.9}$$

This is again an exponential survival function with new scale parameter β.

We employ this result to construct a bivariate random vector as described in Section 6.2 with exponential marginals.

Next, we derive the joint density function of (X, Y). Let $Z = \beta^{-1}(\lambda/U)^\alpha = \beta^{-1}\lambda^\alpha W$. Using the transformation method, the pdf of Z is given by

$$\begin{aligned} f(z) &= f(w)\left|\frac{dw}{dz}\right| \\ &= f_W\left(\frac{\beta}{\lambda^\alpha}z\right)\left(\frac{\beta}{\lambda^\alpha}\right). \end{aligned}$$

Since the rvs involved admit densities, from (6.3) we get the joint density function of (X, Y) as

$$\begin{aligned} f(x,y) &= f(x)\, f(y|x) \\ &= f_X(x)\, f_Z\left(\frac{y}{x^\alpha}\right) \frac{1}{x^\alpha} \\ &= \lambda \exp(-\lambda x) f_W\left(\frac{\beta y}{x^\alpha \lambda^\alpha}\right) \frac{\beta}{\lambda^\alpha} \frac{1}{x^\alpha} \\ &= (\beta \lambda^{1-\alpha} x^{-\alpha}) \exp(-\lambda x) f_W\left(\frac{\beta y}{x^\alpha \lambda^\alpha}\right). \end{aligned} \tag{6.10}$$

The following figures, Figure 6.1-6.3, represents the plot of the joint density function (6.10) of (X, Y) for the selected values of the parameters.

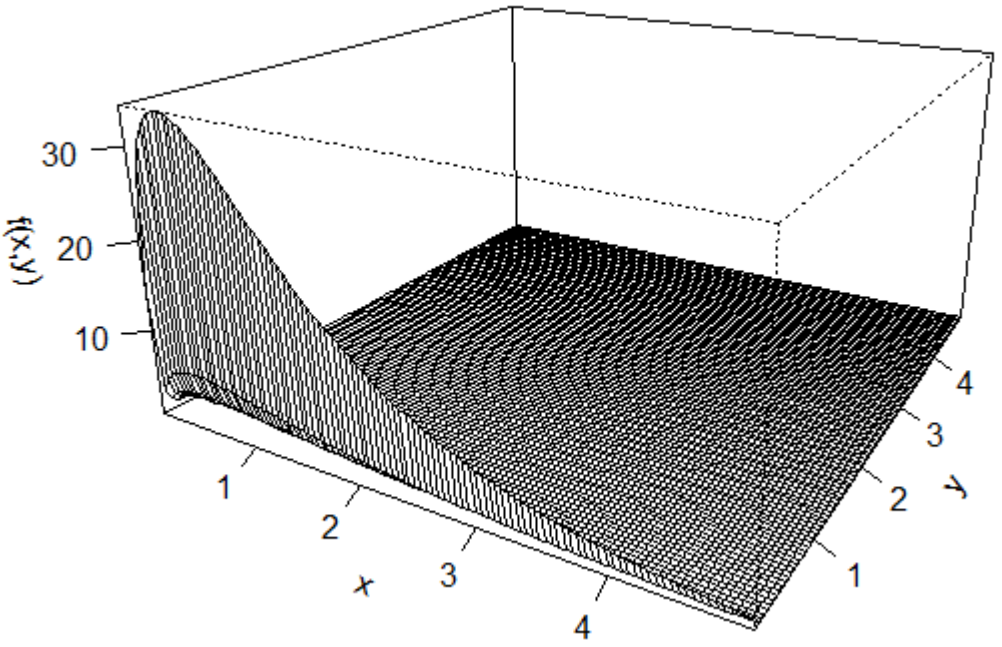

FIGURE 6.1: Plot of the joint density function (6.10) of (X, Y) for $\alpha = 0.8, \beta = 1, \lambda = 1$.

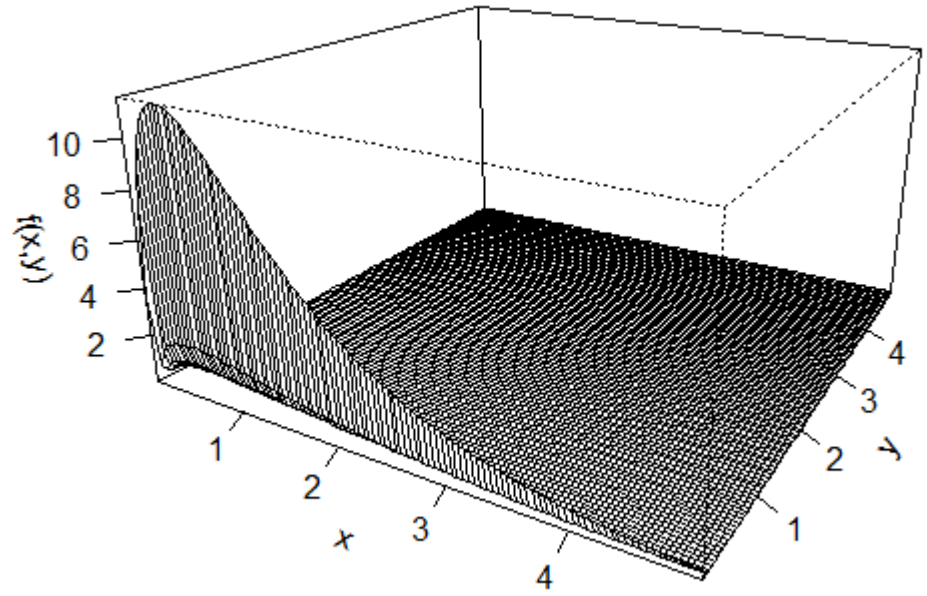

FIGURE 6.2: Plot of the joint density function (6.10) of (X, Y) for $\alpha = 0.6, \beta = 0.5, \lambda = 1$.

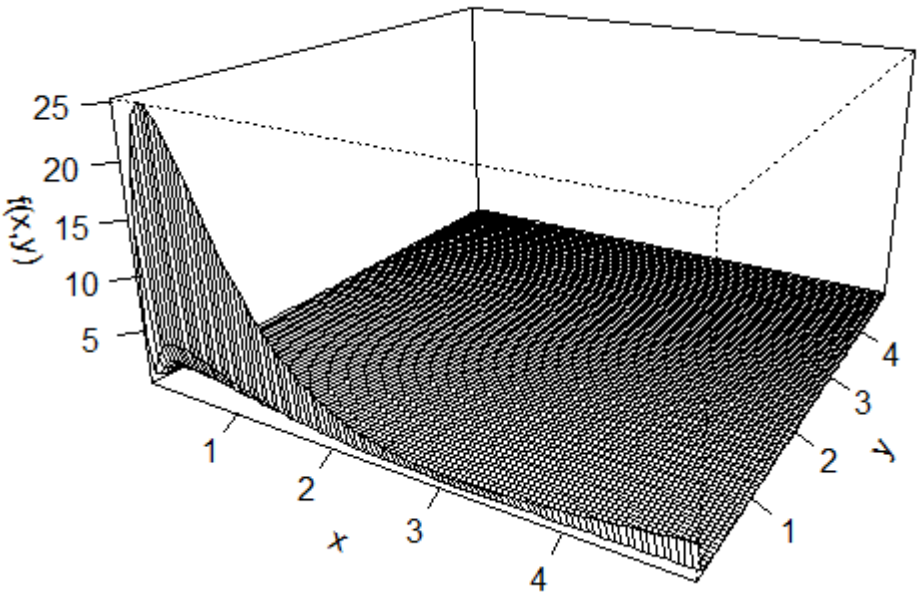

FIGURE 6.3: Plot of the joint density function (6.10) of (X, Y) for $\alpha = 0.3, \beta = 2, \lambda = 2$.

From (6.3) and (6.9) it is readily verified that X and Y have marginal $Exponential(\lambda)$ and $Exponential(\beta)$ distributions respectively.

In terms of the usual notations the conditional probability density functions are given by

$$f(y|x) = f(x,y)/f(x) = (\beta/\lambda^{\alpha}x^{\alpha})\, f_W\left(\beta y/\lambda^{\alpha}x^{\alpha}\right),$$
$$f(x|y) = f(x,y)/f(y) = (\lambda^{1-\alpha}x^{-\alpha})\exp(-\lambda x+\beta y) f_W\left(\beta y/(x\lambda)^{\alpha}\right).$$

The second order properties of the distribution are summarized as follows,

$$E(X) = \lambda^{-1}, E(Y) = \beta^{-1}, E(XY) = (\alpha+1)(\lambda\beta)^{-1}, Corr(X,Y) = \alpha. \qquad (6.11)$$

Theorem 6.1. *The random variables X and Y become independent with exponential marginals as $\alpha \to 0$.*

Proof: If U is a positive stable rv with Laplace transform (6.4), then from Brockwell and Brown (1978), it follows that

$$\Pr\left(U^{-\alpha} \le v\right) \to 1 - e^{-v} \text{ as } \alpha \to 0.$$

Further, it easily follows that as $\alpha \to 0$, $Y = X^{\alpha}\left(\lambda/U\right)^{\alpha}/\beta \xrightarrow{L} (E/\beta)$, where E is unit mean exponential rv, and $\xrightarrow{L}$ stands for convergence in distribution. From the definition of the model, we know that X is independent of U and hence we get the required result.

The regression equation of X on Y is given by

$$\begin{aligned}
E(X|Y=y) &= \int_0^\infty x\, f(x|y)\, dx \\
&= \int_0^\infty x\, (\lambda^{1-\alpha}x^{-\alpha}) \exp(-\lambda x + \beta y) f_W\left(\beta y/(x\lambda)^\alpha\right)\, dx \\
&= \lambda^{1-\alpha}\exp(\beta y) \int_0^\infty x^{1-\alpha}\exp(-\lambda x) f_W\left(\beta y/(x\lambda)^\alpha\right)\, dx \\
&= \lambda^{1-\alpha}\exp(\beta y) \int_0^\infty \left(\frac{(\beta y)^{1/\alpha}u}{\lambda}\right)^{1-\alpha} \exp(-(\beta y)^{1/\alpha}u)\frac{(\beta y)^{1/\alpha}}{\lambda\alpha} u^{1+\alpha} f_U(u)\, du \\
&= \left(\frac{\lambda(\beta y)^{1/\alpha}}{\lambda}\right)^{1-\alpha} \frac{(\beta y)^{1/\alpha}}{\lambda\alpha}\exp(\beta y) \int_0^\infty u^{1-\alpha}\exp(-(\beta y)^{1/\alpha}u) u^{1+\alpha} f_U(u)\, du \\
&= \frac{\beta^{\frac{2}{\alpha}-1} y^{\frac{2}{\alpha}-1}}{\lambda\alpha}\exp(\beta y) \int_0^\infty u^2 \exp(-(\beta y)^{1/\alpha}u) f_U(u)\, du \\
&= \frac{\beta^{\frac{2}{\alpha}-1}}{\alpha\lambda} y^{\frac{2}{\alpha}-1}\exp(\beta y)\, \frac{\partial^2}{\partial s^2}\varphi(s),
\end{aligned}$$

where $s = (\beta y)^{1/\alpha}$ and $\varphi(.) = \exp(-s^\alpha)$ is the Laplace transform of positive stable distribution.

If $\partial^2\varphi(s)/\partial s^2 = \alpha s^{\alpha-2}e^{-s^\alpha}(\alpha s^\alpha - \alpha + 1)$ and $s = (\beta y)^{1/\alpha}$, we get

$$E(X|Y=y) = \ (\alpha\beta/\lambda)y + (1-\alpha)/\lambda.$$

Similarly, the regression equation of Y on X is

$$\begin{aligned} E(Y|X=x) &= \int_0^\infty y\, f(y|x)\, dy \\ &= \int_0^\infty y\, \frac{\beta}{\lambda^\alpha x^\alpha} f_W\left(\frac{\beta y}{\lambda^\alpha x^\alpha}\right) dy \\ &= \frac{\lambda^\alpha x^\alpha}{\beta} E(W) = \frac{\lambda^\alpha x^\alpha}{\beta} E(U^{-\alpha}) \\ &= \frac{\lambda^\alpha x^\alpha}{\beta} E\left(\frac{\beta}{\lambda^\alpha} Z\right) = x^\alpha E(Z) \\ &= x^\alpha \frac{E(Y)}{E(X^\alpha)} = x^\alpha \frac{1/\beta}{\lambda^{-\alpha}\Gamma(1+\alpha)} \\ &= (x\lambda)^\alpha/\beta\,\Gamma(1+\alpha), \end{aligned} \tag{6.12}$$

where $U^{-\alpha} = (\beta/\lambda^\alpha)Z$.

We get closed form expressions for the joint density when $\alpha = 1/2$. In this case, the positive stable rv U with Laplace transform $\varphi(s) = \exp(-\sqrt{s})$ has a pdf given by

$$f_U(u) = \frac{1}{2\sqrt{\pi u^3}} \exp\left(-1/4u\right), \quad u \geq 0 \tag{6.13}$$

and the pdf of $W = 1/\sqrt{U}$ is

$$f_W(v) = \frac{1}{\sqrt{\pi}} \exp\left(-v^2/4\right), \quad v \geq 0. \tag{6.14}$$

Noting that $Z = \left(\sqrt{\lambda}/\beta\right) W$ and using the relation (6.3), the joint pdf of X and Y becomes

$$f(x,y) = (\beta\sqrt{\lambda}/\sqrt{\pi x}) \exp\left(-\lambda x - \frac{\beta^2 y^2}{4\lambda x}\right), \quad x > 0, y > 0. \tag{6.15}$$

One can use our bivariate exponential distribution for modelling two-dimensional renewal processes and queuing processes when arrival and service times are dependent. The renewal properties are conveniently discussed with the help of Laplace transforms. Then we have the following theorem.

Theorem 6.2. *The Laplace transform of the bivariate exponential random vector* (X, Y) *with joint pdf* (6.10) *is given by*

$$\varphi_{(X,Y)}(s_1, s_2) = \frac{\lambda\beta}{\lambda + s_1}\left(\beta + s_2\left(\frac{\lambda}{\lambda + s_1}\right)^{\alpha}\right)^{-1}. \tag{6.16}$$

Proof: The Laplace transform of the bivariate exponential random vector (X, Y) can be written as

$$\begin{aligned}
\varphi_{(X,Y)}(s_1, s_2) &= E\left(\exp(-s_1X - s_2Y)\right) \\
&= \int_0^\infty \int_0^\infty \left(\exp(-s_1X - s_2Y)\right) f(x, y)\, dx\, dy \\
&= \int_0^\infty \int_0^\infty \left(\exp(-s_1x - s_2y)\right)\left(\frac{\beta\lambda^{1-\alpha}}{x^\alpha}\right)\exp(-\lambda x)\, f_W\left(\frac{\beta y}{(x\lambda)^\alpha}\right) dxdy \\
&= \int_0^\infty \lambda \exp(-(s_1 + \lambda)x)\, \psi\left(s_2(\lambda x)^\alpha/\beta\right) dx,
\end{aligned}$$

where $\psi(s) = E\left(\exp(-sW)\right) = E\left(\exp(-sU^{-\alpha})\right)$. Now we simplify the expression for $\psi(s)$ in terms of the reciprocal moments of the positive stable random variable U.

We have, $\exp(-sU^{-\alpha}) = \sum_{j=0}^{\infty} \left(-sU^{-\alpha}\right)^j / j!$, then,

$$\begin{aligned}
\psi(s) &= E\left(\exp(-sU^{-\alpha})\right) \\
&= E\left(\sum_{j=0}^{\infty} \left(-sU^{-\alpha}\right)^j / j!\right) \\
&= \sum_{j=0}^{\infty} \left(\frac{(-s)^j}{j!}\right) E(U^{-\alpha j}) \\
&= \sum_{j=0}^{\infty} \left(\frac{(-s)^j}{j!}\right) \frac{\beta^j}{\lambda^{\alpha j}} \frac{E\left(Y^j\right)}{E\left(X^{\alpha j}\right)} \\
&= \sum_{j=0}^{\infty} \left(\frac{(-s)^j}{j!}\right) \frac{\beta^j}{\lambda^{\alpha j}} \frac{\beta^{-j}\Gamma(1+j)}{\lambda^{-\alpha j}\Gamma(1+\alpha j)} \\
&= \sum_{j=0}^{\infty} \frac{(-s)^j}{\Gamma(1+\alpha j)}
\end{aligned}$$

Thus we can write

$$\begin{aligned}
\varphi_{(X,Y)}(s_1, s_2) &= \int_0^{\infty} \lambda \exp(-(s_1+\lambda)x) \left(\sum_{j=0}^{\infty} \frac{(-s_2(\lambda x)^{\alpha}/\beta)^j}{\Gamma(\alpha j+1)}\right) dx \\
&= \lambda \sum_{j=0}^{\infty} (\lambda^{\alpha}/\beta)^j (-s_2)^j \int_0^{\infty} \frac{x^{\alpha j} \exp(-(s_1+\lambda)x)}{\Gamma(\alpha j+1)} dx \\
&= \lambda \sum_{j=0}^{\infty} (\lambda^{\alpha}/\beta)^j \frac{(-s_2)^j}{\Gamma(\alpha j+1)} \int_0^{\infty} x^{\alpha j} \exp(-(s_1+\lambda)x) dx \\
&= \lambda \sum_{j=0}^{\infty} (\lambda^{\alpha}/\beta)^j \frac{(-s_2)^j}{\Gamma(\alpha j+1)} \frac{\Gamma(\alpha j+1)}{(s_1+\lambda)^{\alpha j+1}} \\
&= \lambda \sum_{j=0}^{\infty} (\lambda^{\alpha}/\beta)^j \frac{(-s_2)^j}{(s_1+\lambda)^{\alpha j+1}}
\end{aligned}$$

$$= \frac{\lambda}{\lambda + s_1} \sum_{j=0}^{\infty} \left(- \left(\frac{s_2}{\beta} \right) \left(\frac{\lambda}{\lambda + s_1} \right)^{\alpha} \right)^{j}$$
$$= \frac{\lambda}{\lambda + s_1} \left(1 + \frac{s_2}{\beta} \left(\frac{\lambda}{\lambda + s_1} \right)^{\alpha} \right)^{-1}$$
$$= \frac{\lambda\beta}{\lambda + s_1} \left(1 + s_2 \left(\frac{\lambda}{\lambda + s_1} \right)^{\alpha} \right)^{-1},$$

since $\sum_{j=0}^{\infty}(-q)^j = 1 - q + q^2 + \cdots = (1+q)^{-1}$. The last expression is the Laplace transform of our bivariate exponential distribution. Hence the theorem.

6.4 Bivariate Exponential Distributions with Negative Correlation

For the models discussed in the earlier sections, the correlation coefficient is non-negative. In practice, we may come across situations where the distributions of X and Y may be exponential, but the correlation coefficient between the components may be negative. In order to incorporate such cases, we define negatively correlated bivariate exponential distributions using the method of antithetic variables. To obtain a negative cross correlation between X and Y , we take an approach similar to Iyer et al. (2002).

Let ξ be a Uniform(0,1) rv and define

$$Z_1 = (-(1/\lambda)\log\xi) \quad \text{and} \quad \mathrm{Z}_2 = (-(1/\lambda)\log(1-\xi))\,.$$

Then we know that Z_1 and Z_2 are correlated and have identical $Exponential(\lambda)$ distribution.

Now take $X = Z_1$ and $Y = Z_2^{\alpha}\,(\lambda^{\alpha}/\beta)\,U^{-\alpha}$ in the construction of our bivariate distribution, so that

$$\begin{aligned} P\,(Y > y) &= P\left(Z_2^{\alpha}(\lambda^{\alpha}/\beta)U^{-\alpha} > y\right) \\ &= P\,(Z_2^{\alpha} > y(\beta/\lambda^{\alpha})u^{\alpha}) \\ &= P\left(Z_2 > ((y\beta)^{1/\alpha}/\lambda)u\right) \\ &= \int_0^{\infty} \left\{\exp(-\lambda((y\beta)^{1/\alpha}/\lambda)u)\right\} dF_U(u) \\ &= E\left(\exp(-(y\beta)^{1/\alpha}U)\right) = \phi_U\left((y\beta)^{1/\alpha}\right) = e^{-\beta\, y}. \end{aligned}$$

It follows that (X, Y) will have bivariate exponential distribution with pdf given by (6.10). It is assumed that ξ is independent of U. The correlation between X and Y is defined as,

$$Corr(X,\, Y) = \rho_{X\,Y} = \frac{Cov(X,\, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}.$$

But,

$$\begin{aligned} Cov(X,\, Y) &= E\,(XY) - E(X)E(Y) \\ &= E\left(Z_1 Z_2^{\alpha}(\lambda^{\alpha}/\beta)U^{-\alpha}\right) - E(Z_1)E(Z_2^{\alpha}(\lambda^{\alpha}/\beta)U^{-\alpha}) \\ &= (\lambda^{\alpha}/\beta)E(U^{-\alpha})E\,(Z_1 Z_2^{\alpha}) - (1/\lambda\beta) \\ &= \frac{(\lambda^{\alpha}/\beta)}{\Gamma(\alpha+1)}E\,(Z_1 Z_2^{\alpha}) - (1/\lambda\beta) \end{aligned}$$

$$= \frac{(\lambda^{\alpha}/\beta)}{\Gamma(\alpha+1)} \frac{1}{\lambda^{\alpha+1}} E\left[(-\log\xi)(-\log(1-\xi))^{\alpha}\right] - (1/\lambda\beta)$$
$$= \frac{1}{\lambda\beta}\left\{(1/\Gamma(\alpha+1))E\left[(-\log\xi)(-\log(1-\xi))^{\alpha}\right] - 1\right\}.$$

Now, the correlation coefficient between X and Y becomes

$$\rho_{XY} = \frac{(1/\lambda\beta)\left\{(1/\Gamma(\alpha+1))E\left[(-\log\xi)(-\log(1-\xi))^{\alpha}\right] - 1\right\}}{(1/\lambda\beta)}$$
$$= (1/\Gamma(\alpha+1))E\left[(-\log\xi)(-\log(1-\xi))^{\alpha}\right] - 1$$
$$= (1/\Gamma(\alpha+1))\int_0^1 \left[(-\log\xi)(-\log(1-\xi))^{\alpha}\right] d\xi - 1, \tag{6.17}$$

which can be evaluated numerically. The Figure 6.4 is the plot of the correlation coefficient (6.17), for different values of α.

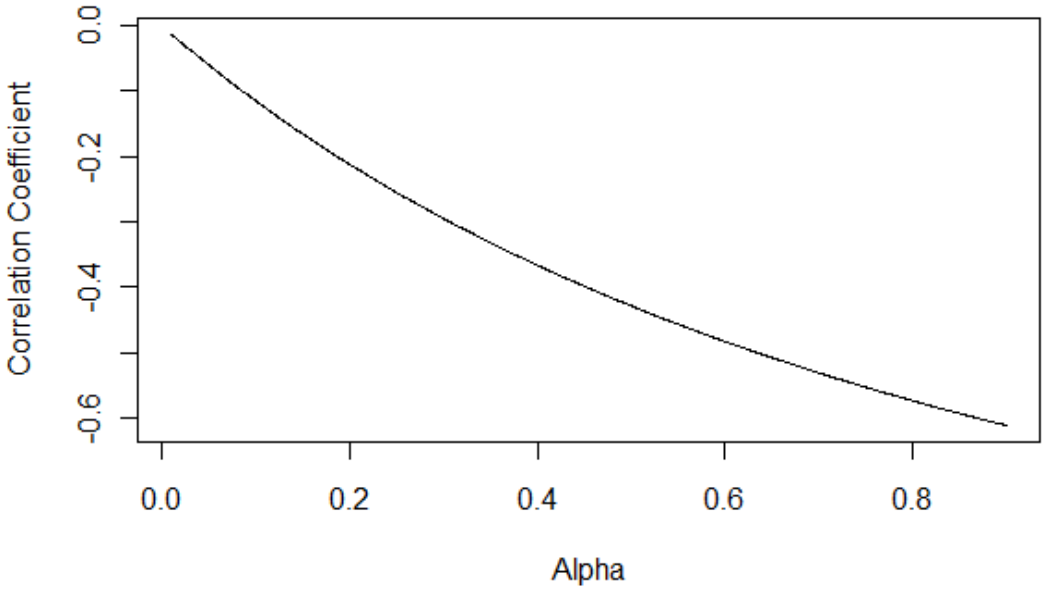

FIGURE 6.4: Plot of the correlation coefficient (6.17), for different values of α

6.5 Statistical Inference for the Product Bivariate Exponential Distribution

In this section, we discuss the problem of estimation for the proposed bivariate exponential distribution. The existence of closed form expressions for all the moments, motivated us to adopt the method of moments to estimate the parameters though they are less efficient. For the bivariate density given by (6.10), the moment estimates are obtained as

$$\widehat{\alpha}= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}, \quad \widehat{\beta}= \frac{1}{\bar{y}}, \quad \widehat{\lambda}= \frac{1}{\bar{x}},$$

where $\bar{x}$ and $\bar{y}$ are the marginal sample means. We will use these estimates as the initial values for obtaining maximum likelihood estimates by numerical methods.

6.5.1 Maximum likelihood estimation

Now we discuss the maximum likelihood method for estimating the parameter vector $\theta = (\alpha,\, \beta,\, \lambda)'$ based on a random sample $\{(x_i,\, y_i), i = 1, 2, ..., n\}$ from the bivariate exponential pdf (6.10). The resulting likelihood function of θ may be expressed as

$$L(\theta|\, x_i,\, y_i,\, i = 1, 2, \ldots, n) = \prod_{i=1}^{n} \frac{\lambda\, e^{-\lambda\, x_i}}{\pi\, y_i}\, R(x_i,\, y_i), \tag{6.18}$$

where $R(x_i,\, y_i) = \sum_{k=1}^{\infty}(-1)^{k-1}\frac{\Gamma(k\,\alpha)}{\Gamma(k)}\left(\frac{\beta\, y_i}{(\lambda\, x_i)^{\alpha}}\right)^k \sin(k\,\pi\,\alpha)$.

Now, the MLE, $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda})$ can be obtained by solving the following likelihood equations:

$$\sum_{i=1}^{n} \frac{R_\alpha(x_i,\, y_i)}{R(x_i,\, y_i)} = 0, \;\; \sum_{i=1}^{n} \frac{R_\beta(x_i,\, y_i)}{R(x_i,\, y_i)} = 0 \;\text{ and }\; \frac{\mathrm{n}}{\lambda} - \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i} = 0. \tag{6.19}$$

where

$$R_\alpha(x_i, y_i) \;=\; \sum_{k=1}^{\infty} (-1)^{k-1} \frac{\Gamma(k\,\alpha)}{\Gamma(k)}\, k\, \left(\frac{\beta\, y_i}{(\lambda\, x_i)^\alpha}\right)^k \{[\Psi(k\,\alpha) - \log(\lambda\, x_i)] \sin(k\,\pi\,\alpha) + \pi\, \cos(k\,\pi\,\alpha)\}$$

$$R_\beta(x_i, y_i) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{\Gamma(k\,\alpha)}{\Gamma(k)}\, k\, \left(\frac{\beta\, y_i}{(\lambda\, x_i)^\alpha}\right)^k \sin(k\,\pi\,\alpha)$$

and $\Psi(.)$ is the digamma function defined by $\Psi(k) = d \ln \Gamma(k)/dk = \Gamma'(k)/\Gamma(k)$.

Since MLE of α and β do not have closed forms, they have to be obtained by some numerical methods like Newton-Raphson. We use the moment estimators as the initial guess in the iterative procedure for computing the MLE. The likelihood equations are in terms of infinite series $R(x_i, y_i)$ and we need to truncate them at appropriate finite number of terms. We truncate the series at N terms and denote it by

$$R^N(x_i, y_i) = \sum_{k=1}^{N} (-1)^{k-1} \frac{\Gamma(k\alpha)}{\Gamma(k)} \left(\frac{\beta\, y_i}{(\lambda\, x_i)^\alpha}\right)^k \sin(k\pi\alpha),$$

by choosing N such that, $|R^N(x_i, y_i) - R^{N+1}(x_i, y_i)| < \delta$, where δ is a pre-specified number. We have taken $\delta = 10^{-10}$ while finding MLE based on the simulated data in the next section.

Now, we discuss the asymptotic properties of the MLE. Under the regularity conditions, stated below, Cramér (1946) proved that the MLE of θ is consistent and asymptotically normally distributed with mean zero and covariance I^{-1}, where I is the Fisher information matrix.

Let $\{f_\theta, \theta \in \Theta\}$ be a family of probability density functions, where Θ is an open interval on $\mathcal{R}$. The regularity conditions are:

(i) $\partial \log f_\theta/\partial\theta$, $\partial^2 \log f_\theta/\partial\theta^2$, $\partial^3 \log f_\theta/\partial\theta^3$ exist for all $\theta \in \Theta$ and every x.

Also,

$$\int_{-\infty}^{\infty} \frac{\partial\, f_\theta(x)}{\partial\theta}\, dx = E_\theta \frac{\partial\, f_\theta(X)}{\partial\theta} = 0, \quad \text{for all } \theta \in \Theta.$$

(ii) $\int_{-\infty}^{\infty} \partial^2 f_\theta(x)/\partial\theta^2\, dx = 0$ for all $\theta \in \Theta$.

(iii) $-\infty < \int_{-\infty}^{\infty} (\partial^2 f_\theta(x)/\partial\theta^2)\, f_\theta(x)\, dx < 0$ for all θ.

(iv) There exists a function $H(x)$ such that for all $\theta \in \Theta$,

$$\left|\frac{\partial^3 f_\theta(X)}{\partial\theta^3}\right| < H(x) \text{ and } \int_{-\infty}^{\infty} \mathrm{H(x)\, f_\theta(x)\, dx} = \mathrm{M}(\theta) < \infty.$$

(v) There exists a function $g(\theta)$ that is positive and twice differentiable for every $\theta \in \Theta$ and, a function $H(x)$ such that for all θ

$$\left|\frac{\partial^2}{\partial\theta^2}\left[g(\theta)\frac{\partial \log f_\theta}{\partial\theta}\right]\right| < H(x) \text{ and } \int_{-\infty}^{\infty} \mathrm{H(x)\, f_\theta(x)\, dx} < \infty.$$

Then, we have the following results by Cramér (1946).

Theorem 6.3. *(a) Conditions (i), (iii) and (iv) imply that with probability approaching* 1*, as* $n \to \infty$*, the likelihood equation has a consistent solution.*

(b) Conditions (i) through (iv) imply that a consistent solution $\hat{\theta}$ *of the likelihood equation is asymptotically normal, that is,*$n \to \infty$*,*

$$\sqrt{n}\,(\hat{\theta} - \theta) \xrightarrow{L} N(0, I^{-1}),$$

where $\xrightarrow{L}$ *denotes convergence in distribution and* I *is the Fisher information matrix given by* $I(\theta) = E\left((\partial \log f_\theta(x)/\partial\theta)\,(\partial \log f_\theta(x)/\partial\theta)'\right)$.

The above mentioned regularity conditions are satisfied by the model (6.1) as the likelihood function in (6.18) is differentiable with respect to θ, and all the moments of x and y are finite. Now, the elements of the Fisher information matrix,

$$I = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}$$

are obtained as

$$\begin{aligned} I_{11} &= -E\left(\frac{\partial^2}{\partial\alpha^2}\log\,L\right);\;\; I_{12} = -E\left(\frac{\partial^2}{\partial\alpha\,\partial\beta}\log L\right) = I_{21}; \\ I_{13} &= -E\left(\frac{\partial^2}{\partial\alpha\,\partial\lambda}\log L\right) = I_{31};\;\; I_{22} = -E\left(\frac{\partial^2}{\partial\beta^2}\log L\right); \\ I_{23} &= -E\left(\frac{\partial^2}{\partial\beta\,\partial\lambda}\log L\right) = I_{32};\;\; I_{33} = -E\left(\frac{\partial^2}{\partial\lambda^2}\log L\right). \end{aligned}$$

We evaluate these elements for the simulated samples in the next section.

6.6 Simulation Study

To evaluate the performance of the estimation procedure discussed above, we carried out a simulation study for different sample sizes and for different specified values of the parameters. For the simulation purpose, we first generate the realizations from positive stable rvs $\{U_i\}$ using the procedure discussed in Section 3.4. For specified values of the parameters we simulated independent and identically distributed sequence $\{x_i\}$ from $Exponential(\lambda)$ distribution and then obtain the sequence $\{y_i\}$ using the relation $y_i = x_i^{\alpha}\frac{1}{\beta}\left(\frac{\lambda}{U_i}\right)^{\alpha}, i = 1, 2, ..., n.$ Based on this bivariate sample $\{x_i, y_i\}$, we obtain the MLE of the parameters using the procedures described in the previous section. The estimates are obtained by solving the likelihood equations in (6.19). We used moment estimates as the initial values while solving the likelihood equations by iterative methods. The step-by-step procedure with MATLAB code is given in Appendix D. For each specified value of the parameter, we repeated the experiment 50 times for computing the estimates and then averaged them over the repetitions. It is observed that for smaller values of α, (say $\alpha \leq 0.8$) the computation algorithm works well and we get the MLE. However, for higher values of α, the computation algorithm does not converge and hence we are unable to obtain MLE. In such cases, we propose to use the moment estimates. The final values are entered in Table 6.1 along with the root mean square errors in the parenthesis.

TABLE 6.1: The average estimates and the corresponding root mean squared errors of the MLE

Sample	True values			Maximum Likelihood Estimates		
Size	α	β	λ	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
50	0.80	2.00	2.00	0.8057 (0.0289)	2.0299 (0.3131)	2.0325 (0.2886)
	0.70	2.00	3.00	0.6980 (0.0332)	1.9354 (0.2816)	2.9374 (0.3772)
	0.60	3.00	1.00	0.6004 (0.0583)	2.9989 (0.3331)	1.0154 (0.1410)
	0.50	1.00	2.00	0.5294 (0.1181)	1.0012 (0.1396)	2.0345 (0.3024)
	0.30	2.00	0.50	0.3631 (0.1942)	2.0297 (0.2652)	0.5073 (0.0573)
	0.20	0.50	2.00	0.2393 (0.1153)	0.5080 (0.0712)	2.0514 (0.3142)
100	0.80	2.00	2.00	0.8097 (0.0219)	2.0063 (0.2032)	1.9942 (0.2225)
	0.70	2.00	3.00	0.7042 (0.0298)	2.0061 (0.1930)	2.9747 (0.3343)
	0.60	3.00	1.00	0.6055 (0.0405)	3.0244 (0.2859)	0.9916 (0.1114)
	0.50	1.00	2.00	0.5106 (0.0640)	1.0132 (0.0949)	1.9832 (0.2229)
	0.30	2.00	0.50	0.3253 (0.1102)	2.0266 (0.2314)	0.4958 (0.0557)
	0.20	0.50	2.00	0.2302 (0.0984)	0.5071 (0.0609)	2.0188 (0.1797)
300	0.80	2.00	2.00	0.8019 (0.0177)	2.0173 (0.1145)	2.0035 (0.1228)
	0.70	2.00	3.00	0.7025 (0.0138)	1.9897 (0.1052)	2.9880 (0.1758)
	0.60	3.00	1.00	0.6033 (0.0187)	2.9845 (0.1598)	0.9960 (0.0586)
	0.50	1.00	2.00	0.4991 (0.0234)	1.0041 (0.0605)	2.0037 (0.1047)
	0.30	2.00	0.50	0.3033 (0.0344)	1.9890 (0.1127)	0.4980 (0.0293)
	0.20	0.50	2.00	0.2019 (0.0396)	0.4969 (0.0284)	1.9920 (0.1172)

6.7 Data Analysis

In this section we illustrate the applications of our bivariate exponential distribution by analysing the real data sets. Let us first consider a set of data reported in Hanagal (2011) on bone marrow transplantation for Leukemia patients. The full data consists of observations on 137 patients with three types of diseases which are Acute Myeloctic Leukemia (AML) with low and high risks and Acute Lymphoblastic Leukemia (ALL). In our illustration we considered a group of 38 ALL patients with

data on *Time to death or time on study*(x_i) and the *disease free survival time (time to relapse, death or end of study)*(y_i). Our aim here is to check if the proposed bivariate exponential distribution fits for this data. The moment estimates of the parameters based on the bivariate exponential pdf (6.10) for the data are obtained as $\hat{\alpha} = 0.9430$, $\hat{\beta} = 0.0016$ and $\hat{\lambda} = 0.0015$. We have noted earlier that for large values of α the computation algorithm of MLE does not converge. As the moment estimate of α is 0.9430 in the above example, we are not able to obtain the MLE of the parameters based on this data. Since, the procedure available for testing the goodness-of-fit for a bivariate distribution is complicated, we apply the Kolmogorov-Smirnov (KS) test for the marginal distribution. Based on moment estimates the KS distance between the empirical distribution function and the fitted distribution function along with the associated p-values (in brackets) for the two marginals are respectively 0.1269 (0.5318) and 0.1204 (0.5976). These values clearly indicate that we cannot reject the null hypothesis that both the marginals follow exponential distributions.

Next, we consider another dataset which is obtained from American Football (National Football League) League from the matches on three consecutive weekends in 1986. This data set were first published in 'Washington Post', and they are available in Csörgő and Welsh (1989) and also in Jamalizadeh and Kundu (2012). In this bivariate data set (X, Y), the variable X represents the game time to the first points scored by kicking the ball between goal posts and Y represents the 'game time' by moving the ball into the end zone. The data set was analysed by Csörgő and Welsh (1989) by using the Marshall-Olkin bivariate exponential model. Also, Jamalizadeh and Kundu (2012) analysed the data using Weighted Marshall-Olkin

bivariate exponential model.

The preliminary analysis of this data using hazard function indicated that the hazard function of X is not constant. So we transform the data (X, Y) by $(X^{1.2}, Y^{1.2})$ as suggested by Csörgő and Welsh (1989). The hazard function of the transformed data indicates that both X and Y follow exponential distributions. We estimate the parameters by method of Maximum likelihood. In this case moment estimates of the parameters are $\widehat{\alpha}= 0.7071, \widehat{\beta}= 0.0408, \widehat{\lambda}= 0.0668$. With these as initial values, the MLE of the parameters obtained as $\hat{\alpha} = 0.5288$, $\hat{\beta} = 0.0334$, and $\hat{\lambda} = 0.0668$. The Kolmogorov-Smirnov distances between the empirical marginals and the fitted marginals along with the associated p-values (in brackets) in two cases are 0.1298 (0.4415) and 0.1197 (0.5441) respectively. These results suggest that our bivariate exponential distribution is a good fit for the above data set.

The results of this chapter are published in Balakrishna and Shiji (2014a).

Chapter 7

Conclusions and Future Works

The thesis has covered various aspects of modelling and analysis of financial time series. The main objective of analysing financial time series is to model the volatility and forecast its future values. Time series analysis based on Box and Jenkins methods are the most popular approaches where the models are linear and errors are Gaussian. This is considered to be unrealistic in many areas of economics and finance as the conditional variances are non-Gaussian. In the present thesis, we mainly studied the properties of some non-Gaussian time series models and examined their suitability for modelling stochastic volatility. One of the requirements for suggesting new models for stochastic volatility is that the class of model for generating non-negative sequences of dependent random variables for generating volatilities.

We have proposed an Extreme value autoregressive model, in view of its application in modelling stochastic volatility. Major problems associated with the non-Gaussian

time series models are those of statistical inference. The non-standard form of the error distribution in the GEVAR(1) model motivated us to try the methods of estimation such as conditional least squares and quasi maximum likelihood. However, its link with positive stable distribution helped us to obtain the maximum likelihood estimates of the parameters. Simulation experiments and data analysis are performed to illustrate the applications of these models.

In the context of developing models for non-negative random variables, we have proposed Weibull product autoregressive model, anticipating that they could be used for generating variances for SV models. The PAR models have several advantages over the corresponding linear AR models for non-negative rvs. We have studied properties of the first order PAR model with Weibull marginal distribution. The estimation problem is really challenging here as the likelihood functions do not have closed form expressions. As a preliminary step we have proposed the method of an approximate maximum likelihood approach. These methods need to be improved and we are currently working on them. Further, there is a wide scope for developing SV models generated by stationary PAR(1) models and we have already made some progress in this direction.

We proposed a SV model generated by first order extreme value autoregressive process as an alternative to normal-lognormal SV model. The model parameters are estimated using the method of moments as the likelihood function is intractable. But we have to come up with more efficient method of estimation and diagnosis procedures for effective use of these models. One practical approach in this context is to develop Bayesian inference procedures. Numerical estimation methods such as Gibbs sampler and Markov chain Monte Carlo procedures will be handier here.

The bivariate exponential distribution introduced in the thesis may serve as a life time model in the context of a repairable system, where could be survival and repair time of an equipment. We have discussed the properties of this distribution and also proposed inference procedures. The computation algorithm for MLE works well for smaller values of the correlation coefficient. If the correlation coefficient exceeds 0.8, the algorithm does not converge. In such cases we may use the moment estimates. Two data sets are analysed to illustrate the applications of the new distribution. The method of constructing bivariate exponential distributions introduced in this chapter can be extended to obtain other classes of distributions such as bivariate Weibull, bivariate gamma, bivariate Pareto, etc. The bivariate Pareto distribution will have a singular component. Further details on these aspects need to be worked out.

We conclude this thesis with a note that we have several unsolved problems which are more challenging than the problems we have already solved. Some of these problems can be solved under the Bayesian frame work. The problems related to volatility forecasting and model selection are yet be discussed. The non-parametric and semi-parametric approaches are potential alternatives to the already established parametric approaches to deal with financial time series. These methods will work better when there are no closed form expressions for likelihood functions. Even though we have focused on discrete time space in our studies so far, the events such as changes in price, temperature, etc. take place continuously. So it is more appropriate to study such problems in continuous time space, which requires the knowledge of stochastic calculus.

Appendix A

Estimation of parameters for GEVAR(1) Model

```
1. Conditional Least Squares Estimation
---------------------------------------
 clear;
 N=600;
 alpha=input('enter alpha');
 mu=input('enter mu');
 sig=input('enter sigma');
 ahat(50)=0;mhat(50)=0;sighat(50)=0;
 for z=1:50

 Generation of positive stable rvs
```

```
-----------------------------------

E=exprnd(1,N,1);

U=unifrnd(0,pi,N,1);

kk1=power(E,(-(1-alpha)/alpha));

kk2=power((sin(U)),(-1/alpha));

kk3=sin(alpha*U);

kk4=power((sin((1-alpha)*U)),((1-alpha)/alpha));

y=kk1.*kk2.*kk3.*kk4;

y1=(1./y);

v=power(y1,alpha);

v1=log(v);

ab=((1-alpha)*mu)-(sig.*v1);

x1(N)=0;

x1(1)=ab(1);

for p=2:N

    x1(p)=alpha*x1(p-1)+ab(p);

end

x=x1(101:N);

n=length(x);

s1=0;s2=0;s3=0;s4=0;

for t=2:n

    s1=s1+(x(t-1)*x(t));

    s2=s2+x(t);

    s3=s3+x(t-1);
```

```
    s4=s4+(x(t-1)*x(t-1));
end
computation of CLS estimates
------------------------------


aa=(s1-((1/n)*s2*s3))/(s4-((1/n)*s3\^2));
gam=0.577216;
sg=(1/pi)*sqrt(6*var(x));
mu1=((s2-aa*s3)/(n*(1-aa)))-(sg*gam);
ahat(z)=aa;
mhat(z)=mu1;
sighat(z)=sg;
end
est=[mean(ahat) mean(mhat) mean(sighat)]
stder=[sqrt(var(ahat)) sqrt(var(mhat)) sqrt(var(sighat))]
------------------------------------------------------------

------------------------------------------------------------
```

```
2. Quasi Maximum Likelihood Estimation

-----------------------------------

  clear;

  N=600;

  alpha=input('enter alpha');

  mu=input('enter mu');

  sigm=input('enter sigm');

  e1(50)=0;e2(50)=0;e3(50)=0;

  for p=1:50


  Generation of positive stable rvs

  ---------------------------------

  E=exprnd(1,N,1);

  U=unifrnd(0,pi,N,1);

  k1=power(E,(-(1-alpha)/alpha));

  k2=power((sin(U)),(-1/alpha));

  k3=sin(alpha*U);

  k4=power((sin((1-alpha)*U)),((1-alpha)/alpha));

  y=k1.*k2.*k3.*k4;

  y1=(1./y);

  v=power(y1,alpha);

  v1=log(v);

  ab=((1-alpha)*mu)-(sigm.*v1);

  x1(N)=0;
```

```
 x1(1)=ab(1)};
 for q=2:N
     x1(q)=alpha*x1(q-1)+ab(q);
 end
 x=x1(101:N);
 n=length(x);
 sum1=0;sum2=0;sum3=0;sum4=0;sum5=0;
 for h=2:n
     sum1=sum1+x(h);
     sum2=sum2+x(h-1);
     sum3=sum3+(x(h)*x(h));
     sum4=sum4+(x(h-1)*x(h-1));
     sum5=sum5+(x(h)*x(h-1));
 end
 x2=autocorr(x);

QML Estimates
------------------

a(10)=0;
a(1)=x2(2);
for j=1:10
falp=(((n-1)/n)*ff1(n,a(j))*power(ff2(n,a(j),x),2))
      +(((n-1)/n)*a(j)*(1+a(j))*ff3(n,a(j),x))
```

```
      +(((a(j)-1)/(2*a(j)+n-n*a(j)))*(sum1+sum2)*ff2(n,a(j),x))
   -((1/(1-a(j)))*((a(j)*(sum3+sum4))
   -((1+(a(j)*a(j)))*sum5)));
falp1=-((2*((n-1)/n))*ff4(n,a(j))*ff2(n,a(j),x)*ff2(n,a(j),x))
 +((2*((n-1)/n))*ff1(n,a(j))*(x(1)-sum2)*ff2(n,a(j),x))
 +(((n-1)/n)*(1+(2*a(j)))*ff3(n,a(j),x))
 +(((2*(n-1)*a(j)*(1+a(j)))/(n*(1-(a(j)*a(j)))*(1-(a(j)*a(j)))
 ))*(a(j)*(sum3+sum4)-(1+(a(j)*a(j)))*sum5))
 +(((sum1+sum2)/((2*a(j)+n-n*a(j))\^2))*(((2*a(j)+n-n*a(j))*
 (2*a(j)*x(1)+sum1+(1-2*a(j))*sum2))-((a(j)-1)*(2-n)*ff2(n,a(j
 ),x))))-((1/((1-a(j))\^2))*(sum3+sum4-((1+2*a(j)-a(j)*a(j)
 *sum5)));
 a(j+1)=a(j)-(falp/falp1);
 if a(j+1)-a(j)<0.0001
       ah=a(j+1);
 end
 end
 sum6=0;sum7=0;
 for i=2:n
     sum6=sum6+(x(i)-ah*x(i-1));
     sum7=sum7+((x(i)-ah*x(i-1))*(x(i)-ah*x(i-1)));
 end
 sig=sqrt((6/(pi*pi*n))*((x(1)*x(1))+((1/(1-ah*ah))*sum7)
    -(((((1+ah)*x(1))+sum6)\^2)/((2*ah+n-n*ah)*(1+ah)))));
```

```
muh=((((1+ah)*x(1))+sum6)/(2*ah+n-n*ah))-(sig*0.5772);
e1(p)=ah;
e2(p)=muh;
e3(p)=sig;
end
est=[mean(e1) mean(e2) mean(e3)]
st=[sqrt(var(e1)) sqrt(var(e2)) sqrt(var(e3))]
-------------------------


function f1=ff1(n,ap)
f1=((n*(ap-1)*(ap-1))-(2*ap*ap))/((2*ap+n-n*ap)\^2);
function f2=ff2(n,ap,x)
s=0;
for t=2:n
    s=s+(x(t)-ap*x(t-1));
end
f2=((1+ap)*x(1))+s;
function f3=ff3(n,ap,x)
s1=0;
for t=2:n
    s1=s1+((x(t)-ap*x(t-1))*(x(t)-ap*x(t-1)));
end
f3=(x(1)*x(1))+((1/(1-ap*ap))*s1);
function f4=ff4(n,ap)
```

```
f4=(((2*ap+n-n*ap)\^2)+((2-n)*((n*(ap-1)*(ap-1))-(2*ap*ap))))
/((2*ap+n-n*ap)\^3);
------------------------------------------------------------
------------------------------------------------------------
```

```
3. Maximum Likelihood Estimation
-----------------------------------
 clear;
 N=200;
 alpha=input('enter alpha');
 sig=input('enter sigma');
 mu=input('enter mu');
 ah(50)=0;mh(50)=0;shh(50)=0;
 for z=1:50
     tic

 Generation of positive stable rvs
 ---------------------------------
 E=exprnd(1,N,1);
 U=unifrnd(0,pi,N,1);
 kk1=power(E,(-(1-alpha)/alpha));
 kk2=power((sin(U)),(-1/alpha));
 kk3=sin(alpha*U);
 kk4=power((sin((1-alpha)*U)),((1-alpha)/alpha));
 y=kk1.*kk2.*kk3.*kk4;
 y1=(1./y);
 v=power(y1,alpha);
 v1=log(v);
 ab=((1-alpha)*mu)-(sig.*v1);
```

```
x1(N)=0;
x1(1)=ab(1);
for p=2:N
    x1(p)=alpha*x1(p-1)+ab(p);
end
x=x1(101:N);
n=length(x);
s1=0;s2=0;s3=0;s4=0;
for t=2:n
    s1=s1+(x(t-1)*x(t));
    s2=s2+x(t);
    s3=s3+x(t-1);
    s4=s4+(x(t-1)*x(t-1));
end
aa=(s1-((1/n)*s2*s3))/(s4-((1/n)*s3\^2));
sg=(1/pi)*sqrt(6*var(x));
gam=0.577216;
mu1=((s2-aa*s3)/(n*(1-aa)))-(sg*gam);

calculation of mle
---------------------
ap(10)=0;m(10)=0;sh(10)=0;
ap(1)=aa;
m(1)=mu1;
```

```
sh(1)=sg;


newton raphson for sigma

-------------------------

sigm(5)=0;

sigm(1)=sg;

for r=1:10

for j=1:5

b(n)=0;

for h=2:n

sumR(1)=stable\_R(1,ap(r),sigm(j),m(r),x,h);

    sumR(2)=stable\_R(1,ap(r),sigm(j),m(r),x,h)

     +stable\_R(2,ap(r),sigm(j),m(r),x,h);

    i=2;

    while abs(sumR(i)-sumR(i-1))>.0001

        sumR(i+1)=sumR(i)+stable\_R(i+1,ap(r),sigm(j),m(r),x,h);

        i=i+1;

    end

    b(h)=i;

end

b1(n)=0;

for h=2:n

    sumL(1)=stable\_L(1,ap(r),sigm(j),m(r),x,h);

    sumL(2)=stable\_L(1,ap(r),sigm(j),m(r),x,h)
```

```
        +stable\_L(2,ap(r),sigm(j),m(r),x,h);
      i1=2;
      while abs(sumL(i1)-sumL(i1-1))>.0001
          sumL(i1+1)=sumL(i1)+stable\_L(i1+1,ap(r),sigm(j),m(r),x,h);
          i1=i1+1;
      end
      b1(h)=i1;
  end
  b2(n)=0;
  for h=2:n
      sumL1(1)=stable\_L1(1,ap(r),sigm(j),m(r),x,h);
      sumL1(2)=stable\_L1(1,ap(r),sigm(j),m(r),x,h)
       +stable\_L1(2,ap(r),sigm(j),m(r),x,h);
      i2=2;
      while abs(sumL1(i2)-sumL1(i2-1))>.0001
          sumL1(i2+1)=sumL1(i2)
           +stable\_L1(i2+1,ap(r),sigm(j),m(r),x,h);
          i2=i2+1;
      end
      b2(h)=i2;
  end
  sum1=0;sum2=0;
  for h=2:n
      sL=0;sL1=0;sR=0;
```

```
      for q=1:b(h)
          sR=sR+stable\_R(q,ap(r),sigm(j),m(r),x,h);
      end
      for q1=1:b1(h)
          sL=sL+stable\_L(q1,ap(r),sigm(j),m(r),x,h);
      end
      for q2=1:b2(h)
          sL1=sL1+stable\_L1(q2,ap(r),sigm(j),m(r),x,h);
      end
      sum1=sum1+((x(h)-ap(r)*x(h-1)-(1-ap(r))*m(r))*(sL/sR));
      sum2=sum2+(power(((x(h)-ap(r)*x(h-1)
       -(1-ap(r))*m(r))/sigm(j)),2)*((sR*sL1-sL*sL)/(sR*sR)));
 end
fsig=x(1)-m(r)-n*sigm(j)-((x(1)-m(r))*exp(-(x(1)-m(r))/sigm(j)))+sum1;
fsig1=-n-(power((x(1)-m(r))/sigm(j),2)*exp(-(x(1)-m(r))/sigm(j)))+sum2;
 sigm(j+1)=sigm(j)-(fsig/fsig1);
 if sigm(j+1)-sigm(j)<0.0001
     sigh=sigm(j+1);
 end
 end

newton raphson for mu
-----------------------
 mm(5)=0;
```

```
mm(1)=mu1;
for j=1:5
d(n)=0;
for h=2:n
 sumR1(1)=stable\_R(1,ap(r),sigh,mm(j),x,h);
 sumR1(2)=stable\_R(1,ap(r),sigh,mm(j),x,h)
  +stable\_R(2,ap(r),sigh,mm(j),x,h);
  j1=2;
  while abs(sumR1(j1)-sumR1(j1-1))>.0001
    sumR1(j1+1)=sumR1(j1)+stable\_R(j1+1,ap(r),sigh,mm(j),x,h);
      j1=j1+1;
  end
  d(h)=j1;
end
d1(n)=0;
for h=2:n
    sumLL(1)=stable\_L(1,ap(r),sigh,mm(j),x,h);
    sumLL(2)=stable\_L(1,ap(r),sigh,mm(j),x,h)
     +stable\_L(2,ap(r),sigh,mm(j),x,h);
    j2=2;
    while abs(sumLL(j2)-sumLL(j2-1))>.0001
        sumLL(j2+1)=sumLL(j2)+stable\_L(j2+1,ap(r),sigh,mm(j),x,h);
        j2=j2+1;
    end
```

```
        d1(h)=j2;
    end
    d2(n)=0;
    for h=2:n
        sumLL1(1)=stable\_L1(1,ap(r),sigh,mm(j),x,h);
        sumLL1(2)=stable\_L1(1,ap(r),sigh,mm(j),x,h)
         +stable\_L1(2,ap(r),sigh,mm(j),x,h);
        j3=2;
        while abs(sumLL1(j3)-sumLL1(j3-1))>.0001
            sumLL1(j3+1)=sumLL1(j3)+stable\_L1(j3+1,ap(r),sigh,mm(j),x,h);
            j3=j3+1;
        end
        d2(h)=j3;
    end
    sum3=0;sum4=0;
    for h=2:n
        sLL=0;sLL1=0;sR1=0;
        for q3=1:d(h)
            sR1=sR1+stable\_R(q3,ap(r),sigh,mm(j),x,h);
        end
        for q4=1:d1(h)
            sLL=sLL+stable\_L(q4,ap(r),sigh,mm(j),x,h);
        end
        for q5=1:d2(h)
```

```
          sLL1=sLL1+stable\_L1(q5,ap(r),sigh,mm(j),x,h);
      end
      sum3=sum3+(sLL/sR1);
      sum4=sum4+((sR1*sLL1-sLL*sLL)/(sR1*sR1));
 end
 fmu=1-exp(-(x(1)-mm(j))/sigh)+(1-ap(r))*sum3;
 fmu1=-(exp(-(x(1)-mm(j))/sigh)/sigh)+(power((1-ap(r)),2)/sigh)*sum4;
 mm(j+1)=mm(j)-(fmu/fmu1);
 if mm(j+1)-mm(j)<0.0001
     muh=mm(j+1);
 end
 end

newton raphson for alpha
------------------------
 alp(5)=0;
 alp(1)=aa;
 for j=1:5
 k1(n)=0;
 for h=2:n
     sumR2(1)=stable\_R(1,alp(j),sigh,muh,x,h);
     sumR2(2)=stable\_R(1,alp(j),sigh,muh,x,h)
      +stable\_R(2,alp(j),sigh,muh,x,h);
     j4=2;
```

```
    while abs(sumR2(j4)-sumR2(j4-1))>.0001
        sumR2(j4+1)=sumR2(j4)+stable\_R(j4+1,alp(j),sigh,muh,x,h);
        j4=j4+1;
    end
    k1(h)=j4;
end
k2(n)=0;
for h=2:n
    sumalp(1)=stable\_R1(1,alp(j),sigh,muh,x,h);
    sumalp(2)=stable\_R1(1,alp(j),sigh,muh,x,h)
     +stable\_R1(2,alp(j),sigh,muh,x,h);
    j5=2;
    while abs(sumalp(j5)-sumalp(j5-1))>.0001
        sumalp(j5+1)=sumalp(j5)+stable\_R1(j5+1,alp(j),sigh,muh,x,h);
        j5=j5+1;
    end
    k2(h)=j5;
end
k3(n)=0;
for h=2:n
    sumalp1(1)=stable\_R2(1,alp(j),sigh,muh,x,h);
    sumalp1(2)=stable\_R2(1,alp(j),sigh,muh,x,h)
     +stable\_R2(2,alp(j),sigh,muh,x,h);
    j6=2;
```

```
    while abs(sumalp1(j6)-sumalp1(j6-1))>.0001
        sumalp1(j6+1)=sumalp1(j6)+stable\_R2(j6+1,alp(j),sigh,muh,x,h);
        j6=j6+1;
    end
    k3(h)=j6;
end
sum5=0;sum6=0;
for h=2:n
    salp=0;salp1=0;sR2=0;
    for q6=1:k1(h)
        sR2=sR2+stable\_R(q6,alp(j),sigh,muh,x,h);
    end
    for q7=1:k2(h)
        salp=salp+stable\_R1(q7,alp(j),sigh,muh,x,h);
    end
    for q8=1:k3(h)
        salp1=salp1+stable\_R2(q8,alp(j),sigh,muh,x,h);
    end
    sum5=sum5+(salp/sR2);
    sum6=sum6+((sR2*salp1-salp*salp)/(sR2*sR2));
end
falp=sum5;
falp1=sum6;
alp(j+1)=alp(j)-(falp/falp1);
```

```
if alp(j+1)-alp(j)<0.0001
    alph=alp(j+1);
end
end
ap(r+1)=alph;
m(r+1)=muh;
sh(r+1)=sigh;
if ap(r+1)-ap(r)<0.001
    aha=ap(r+1);
end
if m(r+1)-m(r)<0.001
    mha=m(r+1);
end
if sh(r+1)-sh(r)<0.001
    sha=sh(r+1);
end
end
ah(z)=aha;
mh(z)=mha;
shh(z)=sha;
toc
end
est=[mean(ah) mean(shh) mean(mh)]
stdrr=[sqrt(var(ah)) sqrt(var(shh)) sqrt(var(mh))]
```

```
-----------------------------------------------------

  function fl=stable\_L(k,alpha,sig,mu,x,h)

  fl=(power(-1,k-1))*(gamma(k*alpha))*k*(exp(-(k*(x(h)-alpha*x(h-1)

   -(1-alpha)*mu)/sig)))*(sin(k*pi*alpha))/factorial(k-1);


 function fl1=stable\_L1(k,alpha,sig,mu,x,h)

  fl1=(power(-1,k-1))*(gamma(k*alpha))*k*k*(exp(-(k*(x(h)-alpha*x(h-1)

  -(1-alpha)*mu)/sig)))*(sin(k*pi*alpha))/factorial(k-1);


 function f=stable\_R(k,alpha,sig,mu,x,h)

  f=(power(-1,k-1))*(gamma(k*alpha))*(exp(-(k*(x(h)-alpha*x(h-1)

  -(1-alpha)*mu)/sig)))*(sin(k*pi*alpha))/factorial(k-1);


 function fR=stable\_R1(k,alpha,sig,mu,x,h)

  fR=(power(-1,k-1))*k*(gamma(k*alpha))*(exp(-(k*(x(h)-alpha*x(h-1)

  -(1-alpha)*mu)/sig)))*(((((x(h-1)-mu)/sig)+psi(0,k*alpha))*sin(k*pi*alpha))

  +(pi*cos(k*pi*alpha)))/factorial(k-1);


  function fR1=stable\_R2(k,alpha,sig,mu,x,h)

  fR1=(power(-1,k-1))*k*k*(gamma(k*alpha))*(exp(-(k*(x(h)-alpha*x(h-1)

  -(1-alpha)*mu)/sig)))*((power((((x(h-1)-mu)/sig)+psi(0,k*alpha)),2)

  +psi(1,k*alpha)-pi*pi)*sin(k*pi*alpha)+pi*(((x(h-1)-mu)/sig)

  +2*psi(0,k*alpha))*cos(k*pi*alpha))/factorial(k-1);
```

Appendix B

Estimation of Parameters for Weibull PAR(1) Model

```
Maximum Likelihood Estimation
-----------------------------
clear;
al=input('initial alpha');
th=input('initial theta');
la=input('initial lamda');
gam=0.577216;
n=2000;
be(10)=0;
be(1)=0.3;
for i=1:10
```

```
    LL=((gamma((1/be(i))+1)*gamma((1/be(i))+1))/gamma((2/be(i))+1))
    -(gamma((1/th)+1)*gamma((1/th)+1)*gamma((2*al/th)+1))/(gamma((2/th)+1)
    *gamma((al/th)+1)*gamma((al/th)+1));

    LL1=-((2*gamma((1/be(i))+1)*gamma((1/be(i))+1)
    *psi(0,1+(1/be(i))))/(be(i)*be(i)*gamma((2/be(i))+1)))
    +((2*gamma((1/be(i))+1)*gamma((1/be(i))+1)
    *psi(0,1+(2/be(i))))/(be(i)*be(i)*gamma((2/be(i))+1)));

    be(i+1)=be(i)-(LL/LL1);
    if be(i+1)-be(i)<0.001
        be1=be(i+1);
    end
end
c1=power(la,-1/th)*gamma((1/th)+1);
c2=power(la,-al/th)*gamma((al/th)+1)*gamma((1/be1)+1);
dee=c1/c2;
de=power(dee,-be1);
mom=[be1 de]
aa(100)=0;tt(100)=0;ll(100)=0;
for z=1:100
u=rand(1,n);
u1=-(log(1-u)/de);
u2=1/be1;
```

```
v=u1.\^{u2}
y(n)=0;
y(1)=v(1);
for t=2:n
    y(t)=power(y(t-1),al)*v(t);
end
m=10;
bet(m)=0;alp(m)=0;alph(m)=0;bet2(m)=0;
alp(1)=al;
bet(1)=be1;
alph(1)=al;
bet2(1)=be1;
for j=1:m
for i=1:m
    L1=((n-1)/bet(i))+fp3(alph(j),y,n)-(n-1)
    *(fp4(alph(j),bet(i),y,n)/fp(alph(j),bet(i),y,n));

    L2=-((n-1)/power(bet(i),2))-(n-1)*((fp(alph(j),bet(i),y,n)
    *fp6(alph(j),bet(i),y,n))-power(fp4(alph(j),bet(i),y,n),2))
    /power(fp(alph(j),bet(i),y,n),2);

    bet(i+1)=bet(i)-(L1/L2);
    if bet(i+1)-bet(i)<0.001
        bet1=bet(i+1);
```

```
    end
end
for i=1:m
    L3=(n-1)*(fp1(alp(i),bet1,y,n)/fp(alp(i),bet1,y,n))-fp2(y,n);

    L4=((n-1)*bet1)*(power(fp1(alp(i),bet1,y,n),2)
    -(fp(alp(i),bet1,y,n)*fp5(alp(i),bet1,y,n)))
    /power(fp(alp(i),bet1,y,n),2);

    alp(i+1)=alp(i)-(L3/L4);
    if alp(i+1)-alp(i)<0.001
        alp1=alp(i+1);
    end
end
alph(j+1)=alp1;
bet2(j+1)=bet1;
if alph(j+1)-alph(j)<0.001
    alpha=alph(j+1);
end
if bet2(j+1)-bet2(j)<0.001
    beta=bet2(j+1);
end
end
delta=(n-1)/fp(alpha,beta,y,n);
```

```
tht(10)=0;

tht(10)=0.4;

for i=1:10

  H1=((gamma(1+(1/tht(i)))*gamma(1+(1/tht(i)))

    *gamma(1+(2*alpha/tht(i))))/(gamma(1+(2/tht(i)))

    *gamma(1+(alpha/tht(i)))*gamma(1+(alpha/tht(i)))))

    -((gamma(1+(1/beta))*gamma(1+(1/beta)))/gamma(1+(2/beta)));


  H2=(2*gamma((1/tht(i))+1)*gamma((1/tht(i))+1)

    *gamma((2*alpha/tht(i))+1)*psi(0,(1/tht(i))+1))/(tht(i)*tht(i)

    *gamma((2/tht(i))+1)*gamma((alpha/tht(i))+1)*gamma((alpha/tht(i))+1));


  H3=(2*gamma((1/tht(i))+1)*gamma((1/tht(i))+1)

    *gamma((2*alpha/tht(i))+1)*psi(0,(2/tht(i))+1))/(tht(i)

    *tht(i)*gamma((2/tht(i))+1)*gamma((alpha/tht(i))+1)

    *gamma((alpha/tht(i))+1));


  H4=(2*alpha*gamma((1/tht(i))+1)*gamma((1/tht(i))+1)

    *gamma((2*alpha/tht(i))+1)*psi(0,(alpha/tht(i))+1))/(tht(i)

    *tht(i)*gamma((2/tht(i))+1)*gamma((alpha/tht(i))+1)

    *gamma((alpha/tht(i))+1));


  H5=(2*alpha*gamma((1/tht(i))+1)*gamma((1/tht(i))+1)

    *gamma((2*alpha/tht(i))+1)*psi(0,(2*alpha/tht(i))+1))/(tht(i)
```

```
    *tht(i)*gamma((2/tht(i))+1)*gamma((alpha/tht(i))+1)
    *gamma((alpha/tht(i))+1));

  H=-H2+H3+H4-H5;
  tht(i+1)=tht(i)-(H1/H);
  if tht(i+1)-tht(i)<0.001
        theta=tht(i+1);
  end
end
lam=(gamma((1/theta)+1)*power(delta,(1/beta)))/(gamma((alpha/theta)+1)
*gamma((1/beta)+1));
po=theta/(1-alpha);
lamda=power(lam,po);
tt(z)=theta;
aa(z)=alpha;
ll(z)=lamda;
end
est=[mean(aa) mean(tt) mean(ll)]
sds=[sqrt(var(aa)) sqrt(var(tt)) sqrt(var(ll))]

-------------------------------------------------

function f=fp(a,b,y,n)
s=0;
```

```
for t=2:n
    s=s+power((y(t)/power(y(t-1),a)),b);
end
f=s;

function f1=fp1(a,b,y,n)
s1=0;
for t=2:n
    s1=s1+power((y(t)/power(y(t-1),a)),b)*log(y(t-1));
end
f1=s1;

function f2=fp2(y,n)
s2=0;
for t=2:n
    s2=s2+log(y(t-1));
end
f2=s2;

function f3=fp3(a,y,n)
s3=0;
for t=2:n
    s3=s3+log(y(t)/power(y(t-1),a));
end
```

```
f3=s3;


function f4=fp4(a,b,y,n)
s4=0;
for t=2:n
    s4=s4+power((y(t)/power(y(t-1),a)),b)*log(y(t)/power(y(t-1),a));
end
f4=s4;


function f5=fp5(a,b,y,n)
s5=0;
for t=2:n
    s5=s5+power((y(t)/power(y(t-1),a)),b)*power(log(y(t-1)),2);
end
f5=s5;


function f6=fp6(a,b,y,n)
s6=0;
for t=2:n
    s6=s6+power((y(t)/power(y(t-1),a)),b)
    *power(log(y(t)/power(y(t-1),a)),2);
end
f6=s6;
```

Appendix C

Estimation of parameters for GEV-SV Model

```
GMM Estimation (R Code)
-----------------------
mu=2
sigma=2
alpha=0.2
n1=1500
est1=c();est2=c();est3=c();
for(r in 1:1000){
E=rexp(n1,1)
U=runif(n1,0,pi)
k1=(sin(U))\^(-1/alpha)
```

```
k2=sin(alpha*U)

k3=(sin((1-alpha)*U))\^((1-alpha)/alpha)

k4=E\^(-(1-alpha)/alpha)

Z=k4*k1*k2*k3

v=(Z)\^(-alpha)

v1=-log(v);

eta=((1-alpha)*mu)-(sigma*v1)

h1=c()

h1[1]=eta[1]

for(t in 2:n1){

 h1[t]=alpha*h1[t-1]+eta[t]

}

h=h1[501:n1]

n=length(h)

abl=rnorm(n,0,1)

y=exp(h/2)*abl

s1=0;s2=0;s3=0

for(l in 2:n){

 s1=s1+(y[l]*y[l])

 s2=s2+(y[l]*y[l]*y[l]*y[l])

 s3=s3+(y[l]*y[l]*y[l-1]*y[l-1])

}

m1=s1/n;m2=s2/n;m3=s3/n

sig=c()
```

```
 sig[1]=0.5
 for(i in 1:10){
 L1=((m1\^2)/m2)-((gamma(1+sig[i])*gamma(1+sig[i]))
   /(3*gamma(1+(2*sig[i]))))
 L2=((2*gamma(1+sig[i])*gamma(1+sig[i]))
   /(3*gamma(1+(2*sig[i])))))*(digamma(1+(2*sig[i]))
   -digamma(1+sig[i]))
 sig[i+1]=sig[i]-(L1/L2)
 if((sig[i+1]-sig[i])<0.0001) sigmahat=sig[i+1]
 }
 muhat=log(m1/gamma(sigmahat+1))
 al=c()
 al[1]=0.1
 for(j in 1:10){
 L3=(m3/((exp(2*muhat))*gamma(1+sigmahat)))
   -((gamma(1+((1+al[j])*sigmahat)))/(gamma(1+(al[j]*sigmahat))))
 L4=(sigmahat*(gamma(1+((1+al[j])*sigmahat)))
   /(gamma(1+(al[j]*sigmahat))))*(digamma(1+(al[j]*sigmahat))
   -digamma(1+((1+al[j])*sigmahat)))
 al[j+1]=al[j]-(L3/L4)
 if((al[j+1]-al[j])<0.0001) alphahat=al[j+1]
 }
est1[r]=muhat
est2[r]=sigmahat
```

```
est3[r]=alphahat

}

a1mu=mean(est1)

v1mu=sqrt(var(est1))

a2sigma=mean(est2)

v2sigma=sqrt(var(est2))

a3alpha=mean(est3)

v3alpha=sqrt(var(est3))

a1mu

v1mu

a2sigma

v2sigma

a3alpha

v3alpha
```

Appendix D

Estimation of parameters for Bivariate Exponential Distribution

```
Maximum Likelihood Estimation
-----------------------------
clear;
N=150;
alpha=input('Enter alpha: ');
beta=input('Enter beta: ');
lamda=input('Enter lamda: ');
aa(50)=0;bb(50)=0;ll(50)=0;
for z=1:50
     tic
E=exprnd(1,N,1);
```

```
U=unifrnd(0,pi,N,1);

kk1=power(E,(-(1-alpha)/alpha));

kk2=power((sin(U)),(-1/alpha));

kk3=sin(alpha*U);

kk4=power((sin((1-alpha)*U)),((1-alpha)/alpha));

Y1=kk1.*kk2.*kk3.*kk4;

V=power(Y1,-alpha);

Z=(power(lamda,alpha)/beta).*V;

X=exprnd((1/lamda),N,1);

Y=power(X,alpha).*Z;

x=X(101:N);

y=Y(101:N);

n=length(x);

lha=n/sum(x);


newton raphson for alpha

------------------------

ap(20)=0;be(20)=0;

cc=corrcoef(x,y);

ap(1)=cc(1,2);

be(1)=1/mean(y);

for r=1:20

a(10)=0;  a(1)=ap(1);

for j=1:10
```

```
q(n)=0;
for h=1:n
s(1)=(power(-1,1-1))*(gamma(1*a(j))/gamma(1))
    *power((be(r)*y(h)/(power(lha*x(h),a(j)))),1)
    *(sin(1*pi*a(j)));
s(2)=s(1)+(power(-1,2-1))*(gamma(2*a(j))/gamma(2))
    *power((be(r)*y(h)/(power(lha*x(h),a(j)))),2)
    *(sin(2*pi*a(j)));
   i=2;
while abs(s(i)-s(i-1))>.0000000001
s(i+1)=s(i)+(power(-1,(i+1)-1))
      *(gamma((i+1)*a(j))/gamma((i+1)))
      *power((be(r)*y(h)/(power(lha*x(h),a(j)))),(i+1))
      *(sin((i+1)*pi*a(j)));
     i=i+1;
end
q(h)=i;
end
G(n)=0;
for h=1:n
sum1=0;
for k=1:q(h)
sum1=sum1+(power(-1,k-1))*(gamma(k*a(j))/gamma(k))
    *power((be(r)*y(h)/(power(lha*x(h),a(j)))),k)
```

```
    *(sin(k*pi*a(j)));
end
G(h)=sum1;
end
q1(n)=0;
for h=1:n
s1(1)=(power(-1,1-1))*(gamma(1*a(j))/gamma(1))*1
     *(power((be(r)*y(h)/(power(lha*x(h),a(j)))),1))
     *((pi*cos(1*pi*a(j)))+((psi(0,1*a(j))
     -log(lha*x(h)))*sin(1*pi*a(j))));
s1(2)=s1(1)+(power(-1,2-1))*(gamma(2*a(j))/gamma(2))
     *2*(power((be(r)*y(h)/(power(lha*x(h),a(j)))),2))
     *((pi*cos(2*pi*a(j)))+((psi(0,2*a(j))
     -log(lha*x(h)))*sin(2*pi*a(j))));
   i1=2;
while abs(s1(i1)-s1(i1-1))>.0000000001
s1(i1+1)=s1(i1)+(power(-1,(i1+1)-1))
        *(gamma((i1+1)*a(j))/gamma((i1+1)))*(i1+1)
        *(power((be(r)*y(h)/(power(lha*x(h),a(j)))),(i1+1)))
        *((pi*cos((i1+1)*pi*a(j)))+((psi(0,(i1+1)*a(j))
        -log(lha*x(h)))*sin((i1+1)*pi*a(j))));
      i1=i1+1;
end
q1(h)=i1;
```

```
end
Ga(n)=0;
for h=1:n
sum2=0;
for k1=1:q1(h)
sum2=sum2+(power(-1,k1-1))*(gamma(k1*a(j))/gamma(k1))
    *k1*(power((be(r)*y(h)/(power(lha*x(h),a(j)))),k1))
    *((pi*cos(k1*pi*a(j)))+((psi(0,k1*a(j))
    -log(lha*x(h)))*sin(k1*pi*a(j))));
end
Ga(h)=sum2;
end
q2(n)=0;
for h=1:n
s2(1)=(power(-1,1-1))*(gamma(1*a(j))/gamma(1))*1*1
     *power((be(r)*y(h)/(power(lha*x(h),a(j)))),1)
     *((2*pi*(psi(0,1*a(j))-log(lha*x(h)))*cos(1*pi*a(j)))
     +(power((psi(0,1*a(j))-log(lha*x(h))),2)
     *sin(1*pi*a(j)))+((psi(1,1*a(j))-pi*pi)
     *sin(1*pi*a(j))));
s2(2)=s2(1)+(power(-1,2-1))*(gamma(2*a(j))/gamma(2))
     *2*2*power((be(r)*y(h)/(power(lha*x(h),a(j)))),2)
     *((2*pi*(psi(0,2*a(j))-log(lha*x(h)))*cos(2*pi*a(j)))
     +(power((psi(0,2*a(j))-log(lha*x(h))),2)
```

```
      *sin(2*pi*a(j)))+((psi(1,2*a(j))-pi*pi)
      *sin(2*pi*a(j))));
   i2=2;
while abs(s2(i2)-s2(i2-1))>.0000000001
s2(i2+1)=s2(i2)+(power(-1,(i2+1)-1))*(gamma((i2+1)*a(j))
        /gamma((i2+1)))*(i2+1)*(i2+1)*power((be(r)*y(h)
        /(power(lha*x(h),a(j)))),(i2+1))
        *((2*pi*(psi(0,(i2+1)*a(j))-log(lha*x(h)))
        *cos((i2+1)*pi*a(j)))+(power((psi(0,(i2+1)*a(j))
        -log(lha*x(h))),2)*sin((i2+1)*pi*a(j)))
        +((psi(1,(i2+1)*a(j))-pi*pi)*sin((i2+1)*pi*a(j))));
      i2=i2+1;
end
q2(h)=i2;
end
Gaa(n)=0;
for h=1:n
sum3=0;
for k2=1:q2(h)
sum3=sum3+(power(-1,k2-1))*(gamma(k2*a(j))/gamma(k2))
    *k2*k2*power((be(r)*y(h)/(power(lha*x(h),a(j)))),k2)
    *((2*pi*(psi(0,k2*a(j))-log(lha*x(h)))*cos(k2*pi*a(j)))
    +(power((psi(0,k2*a(j))-log(lha*x(h))),2)*sin(k2*pi*a(j)))
    +((psi(1,k2*a(j))-pi*pi)*sin(k2*pi*a(j))));
```

```
end
Gaa(h)=sum3;
end
falpha=sum(Ga./G);
fdalpha=sum((G.*Gaa-Ga.*Ga)./(G.*G));
a(j+1)=a(j)-(falpha/fdalpha);
if a(j+1)-a(j)<0.0001
  aha=a(j+1);
end
end

newton raphson for beta
-----------------------
b(10)=0;  b(1)=1/mean(y);
for j=1:10
q3(n)=0;
for h=1:n
s3(1)=(power(-1,1-1))*(gamma(1*aha)/gamma(1))
     *power((b(j)*y(h)/(power(lha*x(h),aha))),1)
     *(sin(1*pi*aha));
s3(2)=s3(1)+(power(-1,2-1))*(gamma(2*aha)/gamma(2))
     *power((b(j)*y(h)/(power(lha*x(h),aha))),2)
     *(sin(2*pi*aha));
   i3=2;
```

```
while abs(s3(i3)-s3(i3-1))>.0000000001
s3(i3+1)=s3(i3)+(power(-1,(i3+1)-1))*(gamma((i3+1)*aha)
        /gamma((i3+1)))*power((b(j)*y(h)
        /(power(lha*x(h),aha))),(i3+1))
        *(sin((i3+1)*pi*aha));
      i3=i3+1;
end
q3(h)=i3;
end
bG(n)=0;
for h=1:n
sum4=0;
for k3=1:q3(h)
sum4=sum4+(power(-1,k3-1))*(gamma(k3*aha)/gamma(k3))
    *power((b(j)*y(h)/(power(lha*x(h),aha))),k3)
    *(sin(k3*pi*aha));
end
bG(h)=sum4;
end
q4(n)=0;
for h=1:n
s4(1)=(power(-1,1-1))*(gamma(1*aha)/gamma(1))
     *1*power((b(j)*y(h)/(power(lha*x(h),aha))),1)
     *(sin(1*pi*aha));
```

```
s4(2)=s4(1)+(power(-1,2-1))*(gamma(2*aha)/gamma(2)
     *2*power((b(j)*y(h)/(power(lha*x(h),aha))),2)
     *(sin(2*pi*aha));
   i4=2;
while abs(s4(i4)-s4(i4-1))>.0000000001
s4(i4+1)=s4(i4)+(power(-1,(i4+1)-1))
        *(gamma((i4+1)*aha)/gamma((i4+1)))*(i4+1)
        *power((b(j)*y(h)/(power(lha*x(h),aha))),(i4+1))
        *(sin((i4+1)*pi*aha));
      i4=i4+1;
end
q4(h)=i4;
end
Gb(n)=0;
for h=1:n
sum5=0;
for k4=1:q4(h)
sum5=sum5+(power(-1,k4-1))*(gamma(k4*aha)/gamma(k4))
    *k4*power((b(j)*y(h)/(power(lha*x(h),aha))),k4)
    *(sin(k4*pi*aha));
end
Gb(h)=sum5;
end
q5(n)=0;
```

```
for h=1:n
s5(1)=(power(-1,1-1))*(gamma(1*aha)/gamma(1))
     *1*1*power((b(j)*y(h)/(power(lha*x(h),aha))),1)
     *(sin(1*pi*aha));
s5(2)=s5(1)+(power(-1,2-1))*(gamma(2*aha)/gamma(2))
     *2*2*power((b(j)*y(h)/(power(lha*x(h),aha))),2)
     *(sin(2*pi*aha));
i5=2;
while abs(s5(i5)-s5(i5-1))>.0000000001
s5(i5+1)=s5(i5)+(power(-1,(i5+1)-1))
        *(gamma((i5+1)*aha)/gamma((i5+1)))*(i5+1)*(i5+1)
        *power((b(j)*y(h)/(power(lha*x(h),aha))),(i5+1))
        *(sin((i5+1)*pi*aha));
      i5=i5+1;
end
q5(h)=i5;
end
Gbb(n)=0;
for h=1:n
sum6=0;
for k5=1:q5(h)
sum6=sum6+(power(-1,k5-1))*(gamma(k5*aha)/gamma(k5))
    *k5*k5*power((b(j)*y(h)/(power(lha*x(h),aha))),k5)
    *(sin(k5*pi*aha));
```

```
end
Gbb(h)=sum6;
end
fbeta=sum(Gb./bG);
fdbeta=(1/b(j))*sum((bG.*Gbb-Gb.*Gb)./(bG.*bG));
b(j+1)=b(j)-(fbeta/fdbeta);
if b(j+1)-b(j)<0.0001
   bha1=b(j+1);
end
end
be(r+1)=bha1;  ap(r+1)=aha;
if be(r+1)-be(r)<0.0001
  bha=be(r+1);
end
if ap(r+1)-ap(r)<0.0001
  aph=ap(r+1);
end
end
aa(z)=aph;  bb(z)=bha;  ll(z)=lha;
   toc
end
es=[mean(aa) mean(bb) mean(ll)]
va=[sqrt(var(aa)) sqrt(var(bb)) sqrt(var(ll))]
```

List of Accepted/Published Papers

1. N. Balakrishna and K. Shiji (2010). Markovian Weibull sequence generated by product autoregressive models and its statistical analysis. *Journal of Indian Society for Probability and Statistics*, 12, 53–67.

2. N. Balakrishna and K. Shiji (2013). Extreme Value Autoregressive Model and its Applications. *Journal of Statistical Theory and Practice*, (To appear).

3. N. Balakrishna and K. Shiji (2014a). On a class of bivariate exponential distributions. *Statistics and Probability Letters*, 85, 153–160.

4. N. Balakrishna and K. Shiji (2014b). Stochastic volatility processes generated by Gumbel Extreme Value Autoregressive model. *Journal of the Indian Statistical Association*, 52(1), 45–64.

Bibliography

Abraham, B. and Balakrishna, N. (2012). Product Autoregressive Models for Non-negative Variables. *Statistics and Probability Letters*, 82(8):1530–1537.

Abraham, B., Balakrishna, N., and Sivakumar, R. (2006). Gamma Stochastic Volatility Models. *Journal of Forecasting*, 25(3):153–171.

Adke, S. R. and Balakrishna, N. (1992). Markovian Chi-square and Gamma Processes. *Statistics and Probability Letters*, 15(5):349–356.

Anderson, T. W. (1954). *An Introduction to Multivariate Statistical Analysis*. Wiley Eastern Private Limited; New Delhi.

Aue, A., Horváth, L., and Steinebach, J. (2006). Estimation in Random Coefficient Autoregressive Models. *Journal of Time Series Analysis*, 27(1):61–76.

Balakrishna, N. and Jacob, T. M. (2003). Parameter Estimation in Minification Processes. *Communications in Statistics, Theory & Methods*, 32(11):2139–2152.

Balakrishna, N. and Lawrence, A. J. (2012). Development of Product Autoregressive Models. *Journal of Indian Statistical Association*, 50:1–20.

Balakrishna, N. and Nampoothiri, K. (2003). A Cauchy Autoregressive Model and its Applications. *Journal of Indian Statistical Association*, 41(2):143–156.

Balakrishna, N. and Shiji, K. (2010). Markovian Weibull Sequence Generated by Product Autoregressive Models and its Statistical Analysis. *Journal of Indian Society for Probability and Statistics*, 12:53–67.

Balakrishna, N. and Shiji, K. (2013). Extreme Value Autoregressive Model and its Applications. *Journal of Statistical Theory and Practice*, (To appear).

Balakrishna, N. and Shiji, K. (2014a). On A Class of Bivariate Exponential Distributions. *Statistics and Probability Letters*, 85:153–160.

Balakrishna, N. and Shiji, K. (2014b). Stochastic Volatility Processes Generated by Gumbel Extreme Value Autoregressive Model. *Journal of the Indian Statistical Association*, 52(1):45–64.

Balakrishnan, N. and Lai, C. D. (2009). *Continuous Bivariate Distributions*. Springer, New York.

Bali, T. G. (2003). An Extreme Value Approach to Estimating Volatility and Value at Risk*. *The Journal of Business*, 76(1):83–108.

Basawa, I. V. and Rao, B. L. S. P. (1980). *Statistical Inference for Stochastic Processes*. Academic press, London.

Bhat, B. R. (1961). On the Asymptotic Distribution of the " PSI-Squared" Goodness-of-Fit Criteria for Markov Chains and Markov Sequences. *The Annals of Mathematical Statistics*, 32(1):49–58.

Billard, L. and Mohamed, F. Y. (1991). Estimation of the Parameters of an EAR(p) Process. *Journal of Time Series Analysis*, 12(3):179–192.

Billingsley, P. (1961). *Statistical Inference for Markov Processes*, volume 2. University of Chicago Press, Chicago.

Block, H. W. and Basu, A. P. (1974). A Continuous, Bivariate Exponential Extension. *Journal of the American Statistical Association*, 69(348):1031–1037.

Blumenfeld, R. and Mandelbrot, B. B. (1997). Lévy dusts, Mittag-Leffler Statistics, Mass Fractal Lacunarity, and Perceived Dimension. *Physical Review E. Statistical Physics, Plasmas, Fluids & Related Interdisciplinary Topics*, 56(1):112–118.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992). ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence. *Journal of Econometrics*, 52(1):5–59.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Englewood Cliffs, NJ: Prentice Hall.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*, 65(332):1509–1526.

Brockwell, P. J. and Brown, B. M. (1978). Expansions for the Positive Stable Laws. *Z. Wahrschein. Ver.*, 45(3):213–224.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods*. Springer-Verlag,New York.

Broto, C. and Ruiz, E. (2004). Estimation Methods for Stochastic Volatility Models: A Survey. *Journal of Economic Surveys*, 18(5):613–649.

Brown, R. G. (1963). *Smoothing, Forecasting and Prediction*. Englewood Cliffs, NJ: Prentice-Hall.

Chavez-Demoulin, V. and Davison, A. C. (2012). Modelling Time Series Extremes. *REVSTAT-Statistical Journal*, 10(1):109–133.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton university press, Princeton.

Csörgő, S. and Welsh, A. H. (1989). Testing for Exponential and Marshall–Olkin Distributions. *Journal of Statistical Planning and Inference*, 23(3):287–300.

Diebold, F. X. (1988). *Empirical Modelling of Exchange Rate Dynamics*. Springer-Verlag, New York.

Downton, F. (1970). Bivariate Exponential Distributions in Reliability Theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32:408–417.

Durbin, J. (1960). Estimation of Parameters in Time Series Regression Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 139–153.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica: Journal of the Econometric Society*, 50:987–1007.

Engle, R. F. and Kraft, D. (1983). Multi-period Forecast Error Variances of Inflation Estimated from ARCH Models. *Applied Time Series Analysis of Economic Data*, pages 293–302.

Fama, E. F. (1965). The Behaviour of Stock-market Prices. *Journal of Business*, 38:34–105.

Freund, J. E. (1961). A Bivariate Extension of the Exponential Distribution. *Journal of the American Statistical Association*, 56(296):971–977.

Gallant, A. R., Hsieh, D., and Tauchen, G. (1997). Estimation of Stochastic Volatility Models with Diagnostics. *Journal of Econometrics*, 81(1):159–192.

Gaver, D. P. and Lewis, P. A. W. (1980). First-order Autoregressive Gamma Sequences and Point Processes. *Advances in Applied Probability*, 12:727–745.

Geweke, J. (1989). Exact Predictive Densities for Linear Models with ARCH Disturbances. *Journal of Econometrics*, 40(1):63–86.

Ghysels, E., Harvey, A., and Renault, E. (1996). *Stochastic Volatility: In G.S. Maddala and C. R. Rao (eds.), Statistical Methods in Finance.* North Holland, Amsterdam.

Granger, C. W. J., White, H., and Kamstra, M. (1989). Interval Forecasting: An Analysis based upon ARCH-Quantile Estimators. *Journal of Econometrics*, 40(1):87–96.

Gumbel, E. J. (1960). Bivariate Exponential Distributions. *Journal of the American Statistical Association*, 55(292):698–707.

Hanagal, D. D. (2011). *Modelling Survival Data using Frailty Models*. Chapman & Hall/CRC.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica: Journal of the Econometric Society*, 50:1029–1054.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate Stochastic Variance Models. *The Review of Economic Studies*, 61(2):247–264.

Heyde, C. C. and Scott, D. J. (1973). Invariance Principles for the Law of the Iterated Logarithm for Martingales and Processes with Stationary Increments. *The Annals of Probability*, 1(3):428–436.

Hughes, G. L., Rao, S. S., and Rao, T. S. (2007). Statistical Analysis and Time-Series Models for Minimum/Maximum Temperatures in the Antarctic Peninsula. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 463(2077):241–259.

Iyer, S. K., Manjunath, D., and Manivasakan, R. (2002). Bivariate Exponential Distributions using Linear Structures. *Sankhya: The Indian Journal of Statistics, Series A*, 64:156–166.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 12(4):371–417.

Jamalizadeh, A. and Kundu, D. (2012). Weighted Marshall–Olkin Bivariate Exponential Distribution. *Statistics*, pages 1–12.

Jayakumar, K. and Pillai, R. N. (1993). The First-order Autoregressive Mittag-Leffler Process. *Journal of Applied Probability*, 30(2):462–466.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Vol. 2*. John Wiley & Sons, New York.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of basic Engineering*, 82(1):35–45.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies*, 65(3):361–393.

Klimko, L. A. and Nelson, P. I. (1978). On Conditional Least Squares Estimation for Stochastic Processes. *The Annals of Statistics*, 6(3):629–642.

Lawrance, A. J. (1991). Directionality and Reversibility in Time Series. *International Statistical Review*, 59(1):67–79.

Lawrance, A. J. and Lewis, P. A. W. (1985). Modelling and Residual Analysis of Non-linear Autoregressive Time Series in Exponential Variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47:165–202.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business*, 36:394–419.

Mann, H. B. and Wald, A. (1943). On the Statistical Treatment of Linear Stochastic Difference Equations. *Econometrica*, 11:173–220.

Marshall, A. W. and Olkin, I. (1967). A Multivariate Exponential Distribution. *Journal of the American Statistical Association*, 62(317):30–44.

Matilla-García, M. and Ruiz Marín, M. (2008). A Non-parametric Independence Test using Permutation Entropy. *Journal of Econometrics*, 144(1):139–155.

McKenzie, E. D. (1982). Product Autoregression: A Time Series Characterization of the Gamma Distribution. *Journal of Applied Probability*, 19:463–468.

McLeish, D. L. (1974). Dependent Central Limit Theorems and Invariance Principles. *the Annals of Probability*, 2:620–628.

Melino, A. and Turnbull, S. M. (1990). Pricing Foreign Currency Options with Stochastic Volatility. *Journal of Econometrics*, 45(1):239–265.

Nakajima, J., Kunihama, T., Omori, Y., and Frühwirth-Schnatter, S. (2012). Generalized Extreme Value Distribution with Time-dependence using the AR and MA Models in State- Space Form. *Computational Statistics & Data Analysis*, 56(11):3241–3259.

Patankar, V. N. (1954). The Goodness-of-Fit of Frequency Distributions Obtained from Stochastic Processes. *Biometrika*, 41(3-4):450–462.

Rachev, S. T., Mittnik, S., Fabozzi, F. J., Focardi, S. M., and Jašić, T. (2007). *Financial Econometrics: From Basics to Advanced Modelling Techniques*, volume 150. John Wiley & Sons.

Ruiz, E. (1994). Quasi-maximum Likelihood Estimation of Stochastic Volatility Models. *Journal of econometrics*, 63(1):289–306.

Shephard, N. (1996). *Statistical Aspects of ARCH and Stochastic Volatility*. Springer, New York.

Sim, C. (1986). Simulation of Weibull and Gamma Autoregressive Stationary Process. *Communications in Statistics-Simulation & Computation*, 15(4):1141–1146.

Sim, C. (1990). First-order Autoregressive Models for Gamma and Exponential Processes. *Journal of Applied Probability*, 27:325–332.

Stout, W. F. (1974). *Almost Sure Convergence*. Academic press, New York.

Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models*. Springer, New York.

Taylor, S. J. (1986). *Modelling Financial Time Series*. John Wiley & Sons, New York.

Taylor, S. J. (1994). Modelling Stochastic Volatility: A Review and Comparative Study. *Mathematical finance*, 4(2):183–204.

Toulemonde, G., Guillou, A., Naveau, P., Vrac, M., and Chevallier, F. (2010). Autoregressive Models for Maxima and their Applications to CH4 and N2O. *Environmetrics*, 21(2):189–207.

Tsay, R. S. (2005). *Analysis of Financial Time Series*. John Wiley & Sons, New York.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50(1):1–25.