Bivariate Cointegrating Time Series with Non-Gaussian Errors

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

Nimitha John

DEPARTMENT OF STATISTICS
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN-682 022

May 2018

CERTIFICATE

This is to certify that the thesis entitled **"Bivariate Cointegrating Time Series with Non-Gaussian Errors "** is a bonafide record of work done by Ms.Nimitha John 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

May 2018

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 have been incorporated in the thesis.

Kochi- 22	**Dr. N. Balakrishna**
May 2018	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 **Nimitha John**

May 2018

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.

It is with all sincerity and high regards that I express my deep sense of gratitude to my supervising guide Dr. N. Balakrishna, Professor and Formerly Head, Department of Statistics, Cochin University of Science and Technology (CUSAT), for his meticulous guidance, consistent encouragement and valuable suggestions throughout my research period. He has always been patient towards my shortcomings and kept encouraging me to work in a better way. This feat was possible only because of the unconditional support provided by him.

I take this opportunity to record my sincere respect and heartiest gratitude to Dr. Asha Gopalakrishnan, Professor and Head, Dr. P. G. Sankaran, Pro-Vice Chancellor and Formerly Head, Dr. K. C. James, Professor and Formerly Head, Prof. S. M. Sunoj, Dr, K.G.Geetha, and Dr, G.Rajesh, Department of Statistics, CUSAT for their whole hearted support extended to me during the research period. I remember with deep gratefulness all my former teachers who gave me light in life through education. I extend my sincere thanks to all non-teaching staff of the Department of Statistics, CUSAT for their kind cooperation.

Words fail to express my gratitude to my best friends, who were giving me encouragement and stood behind me in ups and downs and gave inspiration and moral support for the completion of this task. I owe a lot to my friends and research scholars, Department of Statistics, CUSAT who helped, inspired an encouraged me whenever it needs.

I would also like to thank the Kerala State Council for Science, Technology and Environment, for providing me financial support to carry out this work under KCSTE fellowship.

I am deeply indebted to my brother for his encouragement, prayers and blessings given to me. I am also indebted to my mother, husband, and all my loving family members for their tolerance and constant encouragement for the timely completion of the thesis. I owe everything to them.

Besides this, several people have knowingly or unknowingly helped me in the successful completion of this project. I extend my sincere gratitude to all of them.

Above all, I bow before the grace of the Almighty

Nimitha John.

To My Father

Contents

List of Tables

List of Figures

Chapter 1

Introduction to Linear and Non Linear Time Series

1.1 Motivation

Non stationarity is a property common to many applied time series. An important problem associated with non stationarity is that the statistical inference associated with stationary processes is no longer valid if time series are realisations of non stationary processes. The classical statistical methods used in building and testing large simultaneous equation models, such as Ordinary Least Squares (OLS), were based on the assumption that the variables involved are stationary. Hence, if the time series are non stationary, it is not possible to use OLS to estimate their long-run linear relationships because it would lead to spurious regression. However, this is no longer the case after introducing the concept of cointegration by Granger(1983) (as stated in Engle & Granger (1987)). According to which, models containing non stationary stochastic variables can be constructed in such a way that the derived results are both statistically and economically meaningful.

Further, classical time series methods are based on the assumption that a particular stochastic model generates the observed data. The most commonly used assumption is that the data is a realisation of stationary Gaussian process. However most of the series we come across in practical situations are far from Gaussian series and hence it is of interest to study the non Gaussian type of models. So we extend our study of modelling cointegration in the presence of non Gaussian innovations.

Our objective of the present study is to explore the possibility of employing some non Gaussian error distributions to model cointegration. Cointegration in the presence of hetroskedastic non Gaussian errors has also been studied to investigate the long run equilibrium relationship between the financial variables. This lead us to work on the related problem on statistical inference for cointegrating models with non Gaussian innovations, which is the main contribution of the thesis.

1.2 Introduction

A time series typically consists of set of observations on a variable, taken at equally spaced intervals over time. Some examples includes, the daily maximum temperature, the price series of gold, the daily exchange rate, etc. Time series occur in a variety of fields such as business and economics, agriculture, medical sciences, engineering etc. One of the intrinsic features of time series analysis is that successive observations are always dependent. So the time series analysis is concerned with the techniques involved for this dependence.

The aim of analysis is to summarise the properties of a series and to characterise its salient features. Time series analysis starts with selection of a suitable mathematical model (or class of models) for the data. The main objective for modelling of a time series is to enable forecasting of its future values. In time series analysis, it is natural to suppose that each observation x_t is a realisation of certain random variable X_t. In particular, it has been found that certain time series models are very useful in modelling and analysing of economic and financial data. The sequence of observations representing the prices or price indices are referred to as financial time series.

There are two main objectives in investigating financial time series. First, it is important to know the behaviour of a price series over a period of time. The variance of a price series is particularly relevant to understand the presence of hetroskedasticity in the series. The prices of tomorrow is uncertain and it must therefore be described by a suitable probability distribution. This implies that statistical methods are the natural way to investigate the price behaviour of a time series. The second objective is to use our knowledge of price behaviour in order to reduce risk or to take better decisions. The models of time series can be used for forecasting, option pricing and risk management. This motivates the econometricians and statisticians to devote themselves to the development of new time series models and methods.

The classical set up of time series, 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), Brockwell & Davis (1987)). Their approach of modelling time series is based on the assumption that the time series is a realisation from a Gaussian sequence and the value at time point t is a linear function of past observations. Box et al. (1994) discussed 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 Sections 1.5 to 1.8.

1.3 Useful Characteristics of Time Series

We assume that observed time series is a realisation of $\{X_t, t = 0, \pm 1, \pm 2, ...\}$ which is a discrete time, continuous state-space stochastic process. Let us define some of the commonly used characteristics of time series. The mean, $m(t)$, variance, $V(t)$, auto covariance $\gamma(t,s)$ and auto correlation functions $\rho(t,s)$ of $\{X_t\}$ are respectively defined by,

$$m(t) = E(X_t)$$

$$V(t) = E(X_t - m(t))^2$$

$$\gamma(t,s) = E\{(X_t - m_t)(X_s - m_s)\}$$

and

$$\rho(t,s) = \gamma(t,s) \Big/ \sqrt{V(t)V(s)}.$$

1.3.1 Stationarity

Time series may be stationary or non-stationary. A stationary process is important for time series analysis. A time series $\{X_t\}$ is said to be strictly stationary if the joint probability distribution of $(X_t, X_{t+1}, ..., X_{t+n})$ is exactly same as the joint probability distribution of $(X_{t+h}, X_{t+h+1}, ..., X_{t+h+n})$ for every point t, $t+1$,..$t+n$, h in the time space ((Brooks (2014))). The process $\{X_t\}$ is said to be weakly stationary if it has a constant mean, finite variance and its autocovariance function $\gamma(t,s)$ depends only on the time lag $|t-s|$. A strictly stationary stochastic process with finite variance is always weakly stationary.

1.3.2 White noise

A time series $\{X_t\}$ is called a white noise if $\{X_t\}$ is a sequence of uncorrelated random variables with zero mean and constant variance. In particular, if for every t, $\{X_t\}$ is normally distributed with mean zero and variance σ^2, the series $\{X_t\}$ is called a Gaussian white noise.

1.4 Linear Time Series Models

Time series models are designed to capture various characteristics present in time series data. These models have been widely used in many disciplines such as science, economics, finance etc. This motivated econometricians and statisticians to develop more and more new (or refined) time series models and methods. 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 (ARIMA), etc.

1.4.1 Autoregressive models

A stochastic model that can be extremely useful in the representation of certain practically occurring time series is the autoregressive model. An Autoregressive model of order p, AR(p), is given by

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots\cdots + \phi_p X_{t-p} + a_t.$$

Or equivalently, $\Phi(L)X_t = a_t$ with $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$, where L is the back shift operator, defined by $LX_t = X_{t-1}$, p is a non negative integer and $\{a_t\}$ is a white noise. This model says that the past p values X_{t-i} ,i=1,2,..p jointly determine the conditional expectation of X_t given the past data. Further a_t is uncorrelated with X_{t-i} for every $i > 0$.

1.4.2 Properties of AR models

The mean of stationary AR(p) series is

$$E(X_t) = \frac{\phi_0}{1 - \phi_1 - \cdots\cdots - \phi_p}.$$

The associated polynomial equation of the model is

$$1 - \phi_1 L - \cdots\cdots - \phi_p L^p = 0,$$

which is referred to as the characteristic equation of the model. The resulting process $\{X_t\}$ is weakly stationary, if all the roots of the associated polynomial $\Phi(L) = 0$ lie outside the unit circle. For a stationary AR(p) series, the ACF satisfies the difference equation

$$(1 - \phi_1 L - \cdots\cdots - \phi_p L^p)\rho_h = 0, h > 0,$$

known as Yule-Walker equations. A plot of the 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.

1.4.3 Partial Autocorrelation Function(PACF)

The PACF of a stationary time series is a function of its ACF and is a useful tool for determining the order p of an AR model. For an AR(p) model, PACF of lag more than p vanishes. We make use of this property to determine the order p.

1.4.4 Moving Average Models

Another model of great practical importance in the representation of observed time series is the finite moving average process. The model can be treated as a simple extension of white noise series. In this model, the observation at time t, is expressed as a linear function of the present and past shocks. Specifically an MA(q) model is given by

$$X_t = \mu + a_t - \theta_1 a_{t-1} - \cdots\cdots - \theta_q a_{t-q}.$$

Or, $X_t = \Theta(L)a_t$, where $\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q$ is the characteristic polynomial associated with the MA(q) model, where θ_1's are constants, $\{a_t\}$ is a white noise sequence.

1.4.5 Properties

MA models are always weakly stationary because they are finite linear combinations of a white noise sequence for which the first two moments are time invariant. That is,

$$E(X_t) = \mu.$$

$$V(X_t) = (1 + {\theta_1}^2 + \cdots\cdots + {\theta_q}^2)\sigma_a^2,$$

and the ACF is,

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

Hence, for a MA(q) model, its ACF vanishes after lag q. In particular, the PACF of MA(q) process tails off after lag q. The ACF is useful in identifying the order of MA(q) model.

1.4.6 Autoregressive Moving Average Models

In some applications, the AR or MA models discussed in the previous sections become cumbersome because one may need a higher-order model with many parameters to adequately describe the dynamic structure of the data. To overcome this difficulty, the autoregressive moving-average (ARMA) models are introduced; see Box et al. (1994). A general ARMA(p, q) model is in the form

$$X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + a_t + \sum_{i=1}^{q} \theta_i a_{t-i},$$

here $\{a_t\}$ is a white noise series and p and q are non-negative integers. The model is stationary, if AR(p) component is stationary and invertible if MA(q) component is so. The ACF and PACF are not informative in determining the order of an ARMA model. Tsay & Tiao (1984) propose a new approach that uses the extended autocorrelation function (EACF) to specify the order of an ARMA process.

1.5 Model Identification

The most widely used tools for model identification are the plots of autocorrelation and the partial autocorrelation functions. The behaviour of sample autocorrelation and the sample partial autocorrelation plots are compared to the corresponding theoretical behaviour of these plots when the order is known. ACF of an AR(p) process tails off and PACF has a cut off after lag p. On the other hand, the ACF of moving average process cuts off after lag q, while its

PACF tails off after lag q. And, if both ACF and PACF tail off, then a mixed process is suggested. Furthermore, the ACF for an ARMA 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 PACF for a mixed process is dominated by a mixture of exponential and damped sine waves after the first $|q - p|$ lags.

1.6 Parameter Estimation

Estimating the model parameters is an important aspect of time series analysis. The main approaches followed to fit Box-Jenkins models are the 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. The maximum likelihood estimator (MLE) maximizes the (exact or approximate) log-likelihood function associated with the specified model. 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. When maximum likelihood method becomes difficult for a given model, which is common in the case of stochastic volatility models, we adopt GMM to estimate the parameters.

1.7 Diagnosis Method

When a model has been fitted to a time series, it is advisable to check that the model really does provide an adequate description of the data. As with most statistical models, this is usually done by looking at the residuals, which are generally defined by, residual = observed value - fitted value.

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 measure of dependence is 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. One of the most commonly used model checking methods used in time series analysis is the Portmanteau test.

1.7.1 Portmanteau Test

It is often required to test jointly that several autocorrelations of $\{X_t\}$ are zero. Box & Pierce (1970) propose the Portmanteau statistic,

$$Q_m^* = T \sum_{h=1}^{m} \hat{\rho}_h^2,$$

as a test statistic for the null hypothesis $H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0$ against the alternative hypothesis $H_a : \rho_i \neq 0$ for some $i \in \{1, ..., m\}$, where T is the sample size. Under the assumption that $\{X_t\}$ is an iid sequence with certain moment conditions, Q_m^* asymptotically follows a chi-squared distribution with m degrees of freedom. Ljung & Box (1978) modify the Q_m^* statistic as below to increase the power of the test in finite samples:

$$Q(m) = T(T+2) \sum_{h=1}^{m} \frac{\hat{\rho}_h^2}{T-h}.$$

The decision rule is to reject H_0 if $Q(m) > \chi_\alpha{}^2$, at given significance level α.

1.8 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 using the fitted model. The method of Minimum Mean Square Error (MMSE) forecasting is widely used when the time series follows a linear model. In this case an l-step ahead forecast at time t becomes the conditional expectation,

$$E(X_{t+l} | X_t, X_{t-1}, \cdots).$$

In the study of financial time series, our goal is to forecast the volatility and we have to deal with non-linear models.

1.9 Financial Time Series

Financial time series is concerned with the theory and practice of asset valuation over time. One of the objectives of financial time series is to model the stochastic volatility and forecast its future values. The volatility is measured in terms of the conditional variance of the random variable involved. Although volatility is not directly observable it has some characteristics that are commonly seen in asset return. First, there exist volatility clusters- that is, volatility may be high for certain time periods and low for other periods. 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 is often stationary. Fourth, volatility seems to react differently to a big price increase or a big price drop, referred to as the leverage point.

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

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

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. The log-return series defined by,

$$Y_t = \log\left(P_t/P_{t-1}\right), t = 1, 2, \cdots .$$

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. We refer

$$\{Y_t = \log\left(P_t/P_{t-1}\right), t = 1, 2, ...\}$$

as financial time series. Empirical studies on financial time series (See Mandelbrot (1963) and Fama (1965)) show that the series $\{Y_t\}$ defined above is characterized by the properties below.

1. Absence of autocorrelation in $\{Y_t\}$.

2. Significant autocorrelation in $\{Y_t^2\}$.

3. The marginal distribution $\{Y_t\}$ is symmetric and heavy-tailed.

4. Conditional variance of $\{Y_t\}$ given the past is not constant.

The linear time series models such as ARIMA or ARMA are not suitable for describing the series $\{Y_t\}$ and hence new class of models need to be introduced. While introducing new class of models for such series, we have to see that the model takes care of the special characteristics listed above. Mainly there are two class of models available for analysing the financial time series, namely, **observation driven** and **parameter driven**. In observation driven, volatility is assumed to be a function of past observations, where as in parameter driven case, the conditional variances are generated by some latent models.

1.10 Observation Driven Models

The observation driven models assume that the conditional variances are the functions of past values of the series. The famous autoregressive conditional hetroscedastic (ARCH) model introduced by Engle (1982) is an example of such models. The simplest form of ARCH model assumes that the conditional variance of Y_t given the past is a linear function of the squares of the past data.

1.10.1 Autoregressive Conditional Hetroscedastic model

The autoregressive conditional hetroscedastic(ARCH) model introduced by Engle (1982) was the first attempt in econometrics to capture volatility clustering in time series data. In particular, Engle (1982) used conditional variance to characterise volatility and postulate a dynamic model for conditional variance. 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

$$Y_t = \sqrt{h_t}\varepsilon_t, h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2,$$

where $\{\epsilon_t\}$ is a sequence of independent and identically distributed (iid) random variables with mean zero and variance 1, $\alpha_0 > 0$, and $\alpha_i \geq 0$ for $i > 0$. If $\{\epsilon_t\}$ has standardised Gaussian distribution, Y_t is conditionally normal with mean 0 and variance h_t. The coefficients α_i must satisfy some regularity conditions to ensure that the unconditional variance of y_t to be finite. From the structure of the model, it is seen that large past squared shocks imply a large conditional variance. This means that, under the ARCH framework, large shocks tend to be followed by another large shock. This feature is similar to the volatility clusterings observed in asset returns.

1.10.2 Estimation

The most commonly used estimation procedure for ARCH models has been the maximum likelihood approach. Under the normality assumption, the likelihood function of an ARCH(p) model is

$$L(\alpha|y_1, y_2, \cdots y_n) = \prod_{t=p+1}^{n} \frac{1}{\sqrt{2\pi h_t}} \exp(-\frac{y_t^2}{2h_t}) f(y_1, y_2, \cdots, y_p|\alpha),$$

where $\alpha = (\alpha_0, \alpha_1, \cdots, \alpha_p)^{'}$ and $f(y_1, y_2, \cdots, y_p|\alpha)$ is the joint probability density function of $y_1, y_2, \cdots, y_p$. Since the exact form of $f(y_1, y_2, \cdots, y_p|\alpha)$ 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

$$L(\alpha|y_1, y_2, \cdots y_n) = \prod_{t=p+1}^{n} \frac{1}{\sqrt{2\pi h_t}} \exp(-\frac{y_t^2}{2h_t}).$$

Maximising the conditional-likelihood function is equivalent to maximising its logarithm, which is easier to handle. A variety of alternative estimation methods can be also considered. Least squares and quasi maximum likelihood estimations in ARCH models were considered in the seminal paper by Engle (1982).

1.10.3 Volatility 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 & 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.

1.11 Weakness of ARCH models

The ARCH model has some drawbacks:

1. The model assumes that positive and negative shocks have the same effects on volatility because it depends on the square of the previous shocks. In practice, it is well known that price of a financial asset responds differently to positive and negative shocks.

2. The ARCH model is rather restrictive. For instance, α_1^2 of an ARCH(1) model must be in the interval $[0, 1/3]$ if the series has to have a finite fourth moment. The constraint becomes complicated for higher order ARCH models. In practice, it limits the ability of ARCH models with Gaussian innovations to capture excess kurtosis.

3. The ARCH model does not provide any new insight for understanding the source of variations of financial time series. It merely provides a mechanical way to describe the behaviour of conditional variance. It gives no indication about what causes such behaviour to occur.

4. ARCH models are likely to over predict the volatility because they respond slowly to large isolated shocks to the return series.

1.12 Generalised Observation Driven Models

Although the ARCH model is simple, it often requires many parameters to adequately describe the volatility process of an asset return. Sometimes an ARCH(p) model, where p is of higher order may be needed for the volatility process. So an alternative model must be sought. Bollerslev (1986) proposes a useful extension known as the generalized ARCH (GARCH) model. That is, the Generalized ARCH (GARCH) model is an extension 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. Thus GARCH model is an alternative to ARCH model. The GARCH(p,q) is defined by

$$Y_t = \sqrt{h_t}\varepsilon_t, h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j},$$

where ϵ_t is ae of iid random variables with mean 0 and variance 1, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ and $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_j) < 1$. As before ϵ_t is often assumed to be a standard normal or standardized Student-t distribution or generalised error distribution.

The GARCH model has several extensions like log-GARCH, Exponential GARCH, Integrated GARCH, Fractionally Integrated GARCH, Threshold Garch etc (See Tsay (2005) for more details) .

1.13 Parameter Driven Models

The parameter driven models assume that the volatilities are generated by some latent models. The log-normal stochastic volatility (SV) model by Taylor (1986) is the simplest and the best known example: $Y_t|h_t \sim N(0, \exp(h_t))$, where h_t represents the log-volatility, which is unobserved but can be estimated using the observations. With respect to the previous class, these models are driven by two types of shock, one of which influences the volatility.

1.13.1 Stochastic Volatility Models

For these models the volatility depends on some unobserved components or a latent structure. The most popular of these parameter-driven driven stochastic volatility models, is one by Taylor(1986):

$$Y_t = \varepsilon_t \exp(h_t/2), h_t = \alpha + \beta h_{t-1} + \eta_t,$$

where ϵ_t and η_t are two independent Gaussian white noises, with variances 1 and σ_v^2, respectively. Due to the Gaussianity of η_t, this model is called a log-normal SV model.

1.14 Outline of the Thesis

Many time series exhibit trend or non stationary behaviour. If the time series appears to be non stationary, the standard econometric analysis and distribution theories cannot be applied. Previously, data are differenced in order to make the non stationary series as stationary. Although, this method can be used in large samples, but it may give rise to misleading inferences or spurious regressions in small sample situations. So it is necessary to develop new classes of models to deal with two or more non stationary time series. If two or more series are themselves non stationary, and a linear combination of those series becomes stationary, then the series are said to be cointegrated (William & Wei (2006)).

During the last decade, several estimation methods and test procedures for cointegration among non stationary time series have appeared in the literature. One of the efficient methods for cointegration analysis is the maximum likelihood approach, suggested by Johansen (1988). This method starts from a vector autoregressive (VAR) model representation for a set of variables with Gaussian errors. Some of the test procedures for cointegration in the literature include the Dickey-Fuller unit root test, Engle and Granger two step estimator, Johansen likelihood ratio test etc. Engle & Granger (1987) suggest an efficient estimation technique of the error correction model with the assumption of Gaussianity of errors. All the above mentioned theories and studies are based on the assumption that the possibly cointegrated VAR or error correction model (ECM) has normally distributed errors and, hence, they have the same likelihood function as the classical Johansen method. In this thesis, we study the modelling of cointegration in the presence of non Gaussian innovations.

Chapter 2 briefly discusses an introduction to multivariate time series and cointegration models. Our objective in this study is to identify some non Gaussian time series models among bivariate time series and study their suitability for modelling cointegration.

Cointegration model with logistic innovations is introduced in Chapter 3. In this chapter, a unit root process and cointegration model of first order for $I(1)$ processes which allows for logistic innovation is defined. We propose the maximum likelihood estimator of the cointegrating vector from a first order vector autoregressive process. Then we develop a likelihood ratio test for unit root and cointegration associated with two time series. Monte Carlo simulations are performed to verify the finite sample properties of the estimator and the test statistic. To account for the distortions caused by the specific sample, a bootstrap test based on MLE is performed. Rubber consumption and export data are analysed to illustrate the applications of the proposed model.

In chapter 4, we study the properties of a cointegration model with the errors are generated by a bivariate Student's t distribution. The maximum likelihood estimation for the error correction model and its testing procedure is also discussed. We have also developed a unit root test procedure when the error distribution follows a univariate Student's t distribution with a fixed degrees of freedom. Applications of the model is illustrated using some financial variables.

In Chapter 5, we discuss the cointegration modelling with some non Gaussian GARCH innovations. This chapter presents the estimation procedures for a bivariate cointegration model when the errors are generated by a constant conditional correlation model. In particular, the method of maximum likelihood is discussed when the errors follow Generalised Autoregressive Conditional Hetroskedastic (GARCH) models with Gaussian and some non Gaussian innovations. The method of estimation is illustrated using simulated observations. Since the model is effective in modelling financial data, the descriptive ability of the model is illustrated for a set of data on prices of Oil, Diesel, Palm oil and Soya bean Oil.

Chapter 6 focusses on a bivariate cointegrating model with non-normal errors using a copula. In particular, we propose a bivariate error distribution constructed using two non-identical marginals through a copula. The model parameters are estimated using the method of inference functions for margins and maximum likelihood. Applications of the proposed model is illustrated through real life examples.

Finally, we present the conclusions of the study in Chapter 7.

Chapter 2

Multivariate time series and Cointegration

2.1 Introduction

In modern times, the collection of data became such an easy process that we are able to gather data as frequently as we want, as well as on any number of variables. Since the availability of information is not a big concern nowadays, it only makes sense to analyse all related variables simultaneously to gain more insight on a specific variable. Thus instead of observing a single time series, we rather observe several related time series. That is, a multivariate time series consist of multiple single series referred to as components. If each time series observation is a vector series, then we can model them using a multivariate form of Box-Jenkins model. In particular, the techniques of multivariate time series is used, when we want to analyse and explain the interaction and co movements among a group of time series variables.

The application is wide-spread from, for example, the medical field where the relationship between exercise and blood glucose can be modeled (Crabtree et al. (1990)) to the engineering field where the process control effectiveness can be evaluated (DeVries & Wu (1978)). Whittle (1953) derived the least square estimation equations for a non-deterministic stationary multiple process, while Bartlett

& Rajalakshman (1953) were concerned with the goodness of fit of simultaneous autoregressive series. Akaike (1969), Hannan (1970), up to the more recent Hamilton (1994), Reinsel (2003), Tsay (2013) are just some of the many that have studied and made contributions to the field of multivariate time series analysis. Multivariate time series analysis introduced a way to observe the relationship of variables over time, thus making use of all possible information. In the case of univariate time series, one investigated the influence of all the past values of a single time series on the future values of that specific time series. Now we can extend this to also look at the influence of other variables across time periods. This will ultimately improve the accuracy of the forecasts of an individual time series.

The concepts of vector and matrix are useful in understanding multivariate time series analysis. This chapter serves as an introduction to some of the concepts, namely covariance stationarity, cross correlation matrix, integration, cointegration, and the vector models used in multivariate time series analysis. The aim of this chapter is to study the econometric models for analysing the multivariate process $\{\mathbf{X_t}\}$. Many of the methods and models in univariate time series can be generalised directly to the multivariate case. But there arise situations in which the generalisation requires some attention. We may need new models and methods to handle the complicated relationships between multiple time series. In this chapter, we discuss the general methods and models of multiple time series with emphasis on intuitions and applications. For statistical theory of multivariate time series analysis, one can refer to Liitkepohl (1991), Reinsel (2003).

2.2 Notations

Let $\{\mathbf{X_t}\}$ be a vector of time series with k-components, say $X_{1t}, X_{2t}, \ldots, X_{kt}$. That is,

$$\mathbf{X_t} = \begin{pmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ \vdots \\ X_{kt} \end{pmatrix}. \tag{2.1}$$

For example, an investor holding stocks of IBM, Microsoft and General Motors may consider the three dimensional series of these companies. Here X_{1t} denotes the daily IBM stock series, X_{2t} is that of Microsoft and X_{3t} denotes the series of General Motors. Now, if k time series are observed for a specific time period, say t=1 to T, then the series may be represented in a $k \times T$ matrix form as,

$$\mathbf{X_t} = \begin{pmatrix} X_{11} & \ldots & X_{1T} \\ \vdots & \ddots & \vdots \\ X_{k1} & \cdots & X_{kT} \end{pmatrix}, \tag{2.2}$$

where each row represents a univariate time series and each column represents the observed measurements made on k variables at a specific point in time.

2.3 Weak (Covariance) Stationarity

The vector time series $\{\mathbf{X_t}\}$ is said to be weak or covariance stationary, if its first and second moments are time invariant. In particular, the mean vector and covariance matrix of a covariance stationary time series are constant over time t. For a covariance stationary process $\{\mathbf{X_t}\}$, we define its mean vector and covariance matrix as

$$E(\mathbf{X_t}) = \boldsymbol{\mu}, \boldsymbol{\Gamma_0} = E[(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})^{'}],$$

where the expectation is taken element by element over the joint distribution of $\{\mathbf{X_t}\}$.

The mean vector $\boldsymbol{\mu}$ is a k dimensional vector consisting of the unconditional expectations of the components of $\{\mathbf{X_t}\}$ given by, $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} E(X_{1t}) \\ E(X_{2t}) \\ \vdots \\ \vdots \\ E(X_{kt}) \end{pmatrix}$.

The covariance matrix $\boldsymbol{\Gamma_0}$ is a $k \times k$ matrix. The i th diagonal element of $\boldsymbol{\Gamma_0}$ is the variance of X_{it}, whereas the (i,j)th element of $\boldsymbol{\Gamma_0}$ is the covariance between X_{it} and X_{jt}.

2.4 Cross-Correlation Matrix

In chapter 1, we have seen that, the auto covariance of a univariate time series $\{X_t\}$ is a function that gives the covariance of the process with itself at pairs of time points. If the process $\{X_t\}$ has the mean function μ_t, then the lag-l auto covariance is given by

$$cov(X_t, X_{t-l}) = \Gamma(l) = [E(X_t - \mu_t)(X_{t-l} - \mu_{t-l})].$$

For a vector of time series $\{\mathbf{X_t}\}$ with k-components, say $X_{1t}, X_{2t}, \ldots, X_{kt}$, the concept of cross covariance is used to define the covariance between more than one time series. The cross-covariance is a function that gives the covariance of one process with other, at pairs of time points.

Let $\Gamma_{ij}(h) = Cov(X_{it}, X_{j,t-h})$ and let $\mathbf{D}$ be a $k \times k$ diagonal matrix consisting of the standard deviations of the individual series of $\{X_{it}\}$. That is, $\mathbf{D} = diag\left\{\sqrt{\Gamma_{11}(0)}, \ldots, \sqrt{\Gamma_{kk}(0)}\right\}$. Then, the lag 0 cross correlation matrix of $\{\mathbf{X_t}\}$ is defined as

$$\boldsymbol{\rho}_0 = [\rho_{ij}(0)] = \mathbf{D^{-1}}\boldsymbol{\Gamma}_0\mathbf{D^{-1}},$$

where $\boldsymbol{\Gamma_0}$ is the covariance matrix of $\{\mathbf{X_t}\}$ at lag 0. Hence the cross-covariance and cross-correlation between the i-th and j-th components of the vector $\mathbf{X_t}$ at lag 0 is given by

$$\Gamma_{ij}(0) = \mathrm{cov}(X_{it}, X_{j,t}) = E(X_{it} - \mu_i)(X_{jt} - \mu_j), \tag{2.3}$$

where the (i,j)th element of ρ_0 is given by,

$$\rho_{ij}(0) = \frac{\mathrm{cov}(X_{it}, X_{j,t})}{std(X_{it})std(X_{j,t})} = \frac{\Gamma_{ij}(0)}{\left(\Gamma_{ii}(0)\Gamma_{jj}(0)\right)^{1/2}}, \tag{2.4}$$

where $\Gamma_{ii}(0) = Var(X_{it})$. In time series context, such a correlation coefficient is referred to as a concurrent, or contemporaneous correlation coefficient because it is the correlation of two time series at a time point t. One of the important topics in multivariate time series analysis is the lead lag relationships between component series. To this end, the cross correlation matrices are used to measure the strength of linear dependence between time series. The lag-l cross covariance matrix of $\{\mathbf{X_t}\}$ is defined as

$$\mathbf{\Gamma}_l = [\Gamma_{ij}(l)] = E[(\mathbf{X}_t - \mu)(\mathbf{X}_{t-l} - \mu)^{'}].$$

And, the lag-l cross-correlation matrix of $\{\mathbf{X_t}\}$ is the correlation between X_{it} and $X_{j,t-l}$, which is defined as

$$\boldsymbol{\rho}_l = [\rho_{ij}(l)] = D^{-1}\Gamma_l D^{-1},$$

where $\rho_{ij}(l)$ is given by,

$$\rho_{ij}(l) = \frac{\operatorname{cov}(X_{it}, X_{j,t-l})}{std(X_{it})std(X_{j,t-l})} = \frac{\Gamma_{ij}(l)}{\left(\Gamma_{ii}(l)\Gamma_{jj}(l)\right)^{1/2}}. \tag{2.5}$$

If $l > 0$, then $\rho_{ij}(l)$ measures the linear dependence of X_{it} on $X_{j,t-l}$, which occurred prior to time t. Equation (2.5) also shows that the diagonal element $\rho_{ii}(l)$ is simply the lag-l autocorrelation coefficient of $\{X_{it}\}$. In general, the correlation between i-th variable at time t and j-th variable at time $t - l$, is not same as the correlation between the j-th variable at time t and i-th variable at time $t - l$.
i.e; $\rho_{ij}(l) \neq \rho_{ji}(l)$.

2.4.1 Sample Cross-correlation Matrices

Given the data $\{\mathbf{x_t}, t = 1, 2...n\}$, the cross-covariance matrix $\mathbf{\Gamma}_l$ can be estimated by

$$\hat{\mathbf{\Gamma}}_l = \frac{1}{n}\sum_{t=l+1}^{n}(\mathbf{x_t} - \bar{\mathbf{x}})(\mathbf{x_{t-l}} - \bar{\mathbf{x}})', l \geq 0,$$

where $\bar{\mathbf{x}} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{x_t}$, is the vector of sample means. The cross-correlation matrix $\boldsymbol{\rho}_l$ is estimated by

$$\hat{\rho}_l = \hat{\mathbf{D}}^{-1}\hat{\mathbf{\Gamma}}(l)\hat{\mathbf{D}}^{-1}, l \geq 0,$$

where $\hat{\mathbf{D}}$ is the $k \times k$ diagonal matrix of the sample standard deviations of the component series.

Similar to the univariate case, asymptotic properties of the sample cross correlation matrix $\hat{\boldsymbol{\rho}}_l$ have been investigated under various assumptions; see for instance, Fuller (1976). The estimate is consistent but is biased in a finite sample.

2.5 Multivariate Portmanteau Tests

The univariate Ljung Box Statistics has been generalised to the multivariate case by Hosking (1980) and Li & McLeod (1981). For a multivariate series, the null hypothesis of interest is $H_0 : \boldsymbol{\rho}_1 = \boldsymbol{\rho}_2 = \cdots\cdots = \boldsymbol{\rho}_m = \mathbf{0}$ and the alternate hypothesis $H_1 : \boldsymbol{\rho}_i \neq \mathbf{0}$ for some $i \in \{1, 2, ...m\}$, where $\boldsymbol{\rho}_i$ denotes the lag-i cross correlation matrix of the time series $\{\mathbf{X_t}\}$. The test statistics is given by

$$Q_k(m) = T^2 \sum_{l=1}^{m} \frac{1}{T-l} tr\left(\quad \hat{\boldsymbol{\Gamma}}_l' \quad \hat{\boldsymbol{\Gamma}}_0^{-1} \quad \hat{\boldsymbol{\Gamma}}_l \quad \hat{\boldsymbol{\Gamma}}_0^{-1} \quad \right), \tag{2.6}$$

where T is the sample size, k is the dimension of $\mathbf{X_t}$ and tr$\left(\quad \hat{\boldsymbol{\Gamma}}_l' \quad \hat{\boldsymbol{\Gamma}}_0^{-1} \quad \hat{\boldsymbol{\Gamma}}_l \quad \hat{\boldsymbol{\Gamma}}_0^{-1} \quad \right)$ is the trace of the matrix $\left(\quad \hat{\boldsymbol{\Gamma}}_l' \quad \hat{\boldsymbol{\Gamma}}_0^{-1} \quad \hat{\boldsymbol{\Gamma}}_l \quad \hat{\boldsymbol{\Gamma}}_0^{-1} \quad \right)$, which is the sum of diagonal elements of the matrix (Tsay (2005), page 347). Under the null hypothesis, $Q_k(m)$ follows asymptotically a chi squared distribution with $k^2 m$ degrees of freedom (Hosking (1980), page 605).

2.6 Vector Autoregressive (VAR) Model

The vector autoregressive model generalizes the univariate autoregressive model by allowing for more than one evolving variable. Vector auto regression is a mechanism that is used to link several or multiple stationary time series together. The structure is that, each variable is a linear function of past lags of itself and past lags of the other variables. A p-th order VAR, denoted as VAR(p), is

$$\mathbf{X_t} = \varphi_{\mathbf{0}} + \varphi_1 \mathbf{X_{t-1}} + \varphi_2 \mathbf{X_{t-2}} + \cdots\cdots + \varphi_p \mathbf{X_{t-p}} + \boldsymbol{a}_t,$$

or in a lag operator form

$$(I_k - \varphi_1 L - \cdots\cdots \varphi_p L^p)\mathbf{X_t} = \varphi_0 + \boldsymbol{a}_t, \tag{2.7}$$

where

$L^j \mathbf{X_t} = \mathbf{X_{t-j}}$.

$\mathbf{X_t}$: $k \times 1$ random vector.

$\boldsymbol{\varphi_0}$: $k \times 1$ vector of constant terms.

$\boldsymbol{\varphi_i}$: $k \times k$ autoregressive coefficient matrix.

$\boldsymbol{a}_t$: $k \times 1$ white noise process, satisfying:

$$E\left(\boldsymbol{a}_t\right) = 0 \text{ and } E\left(\boldsymbol{a}_t \boldsymbol{a}_t{}'\right) = E \begin{pmatrix} E({a_{1t}}^2) & E(a_{1t}a_{2t)} & \cdots & E(a_{1t}a_{kt}) \\ E(a_{1t}a_{2t}) & E({a_{2t}}^2) & \cdots & E(a_{2t}a_{kt}) \\ \vdots & \vdots & \vdots & \vdots \\ E(a_{1t}a_{kt}) & E(a_{2t}a_{kt}) & \cdots & E(a_{kt}^2) \end{pmatrix} = \Sigma_a,$$

which is a $k \times k$ symmetric, positive, definite matrix, called the white noise covariance matrix and

$$E\left(\boldsymbol{a}_t \boldsymbol{a}_s{}'\right) = 0,$$

for $t \neq s$, therefore uncorrelated across time.

By replacing the vectors and matrices with scalars will give the definition of an AR(p) process. The process $\{\mathbf{X_t}\}$ is stationary if the zeros of determinant

$$\left| I_k - \boldsymbol{\varphi}_1 L - \cdots\cdots \boldsymbol{\varphi}_p L^p \right|$$

are all greater than one, or in other words, if the zeros of determinant

$$\left| I_k - \boldsymbol{\varphi}_1 L - \cdots\cdots \boldsymbol{\varphi}_p L^p \right|$$

lie outside the complex unit circle (have modulus greater than one).

2.6.1 Bivariate AR(1) model

The simplest VAR is the first order bivariate VAR model which can be expressed as,

$$\mathbf{X_t} = \boldsymbol{\varphi_0} + \boldsymbol{\varphi}_1\mathbf{X_{t-1}} + \boldsymbol{a}_t. \tag{2.8}$$

That is,

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{10} \\ \phi_{20} \end{pmatrix} + \begin{pmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}.$$

Therefore,

$$\begin{aligned} X_{1t} &= \phi_{10} + \phi_{11,1}X_{1,t-1} + \phi_{12,1}X_{2,t-1} + a_{1t}, \\ X_{2t} &= \phi_{20} + \phi_{21,1}X_{1,t-1} + \phi_{22,1}X_{2,t-1} + a_{2t}. \end{aligned}$$

It can be seen that each element of $\{\mathbf{X_t}\}$ is a function of each element of $\{\mathbf{X_{t-1}}\}$. VAR(1) model is always invertible.

2.6.2 Stationarity

The VAR(1) process is covariance stationary, if the eigen values of $\boldsymbol{\varphi}_1$ are less than one in modulus. In the univariate case, this is equivalent to the condition $|\phi_1| < 1$.

2.7 Vector Moving Average (VMA) Model

A vector moving average of order q, or VMA(q), is in the form

$$\mathbf{X_t} = \mathbf{\Theta}_0 + \boldsymbol{a}_t - \mathbf{\Theta}_1 \boldsymbol{a}_{t-1} - \cdots\cdots - \mathbf{\Theta}_q \boldsymbol{a}_{t-q}, \tag{2.9}$$

or

$$\boldsymbol{X}_t = \mathbf{\Theta}\left(L\right)\boldsymbol{a}_t,$$

where $\mathbf{\Theta}_0$ is a k-dimensional vector, $\mathbf{\Theta}_i$ are $k \times k$ matrices and $\mathbf{\Theta}(L) = \left(I - \mathbf{\Theta}_1 L - \cdots\cdots - \mathbf{\Theta}_q L^q\right)$ is the MA matrix polynomial in lag operator L. Similar to univariate case, VMA(q) process are weakly stationary provided that the covariance of $\{\boldsymbol{a}_t\}$ exists. A VMA(q) process is invertible if the zeros of determinant

$$\left|I - \mathbf{\Theta}_1 L - \cdots\cdots - \mathbf{\Theta}_q L^q\right|$$

are all greater than one, or in other words, the zeros of determinant

$$\left|I - \mathbf{\Theta}_1 L - \cdots\cdots - \mathbf{\Theta}_q L^q\right|$$

lie outside the complex unit circle.

2.7.1 Bivariate MA(1) model

To better understand the VMA process, let us consider the bivariate VMA(1) model given by

$$\mathbf{X_t} = \mathbf{\Theta}_0 + \boldsymbol{a}_t - \mathbf{\Theta}_1 \boldsymbol{a}_{t-1}.$$

That is,

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{10} \\ \theta_{20} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix} - \begin{pmatrix} \theta_{11,1} & \theta_{12,1} \\ \theta_{21,1} & \theta_{22,1} \end{pmatrix} \begin{pmatrix} a_{1,t-1} \\ a_{2,t-1} \end{pmatrix},$$

which gives,

$$X_{1t} = \theta_{10} + a_{1t} - \theta_{11,1}a_{1,t-1} + \theta_{12,1}a_{2,t-1}.$$

$$X_{2t} = \theta_{20} + a_{2t} - \theta_{21,1}a_{1,t-1} + \theta_{22,1}a_{2,t-1}.$$

The model says that the current series $\{\mathbf{X_t}\}$ only depends on the current and past shocks. The vector MA(1) process is clearly stationary. For the process $\{\mathbf{X_t}\}$ to be invertible, the eigen values of $\mathbf{\Theta}_1$ should be less than one in absolute.

2.8 Vector Autoregressive Moving Average (VARMA) Model

The Vector autoregressive moving average model, VARMA(p,q) of order p and q is a combination of the VAR(p) and VMA(q) processes. The VARMA(p,q) model can be written as

$$\mathbf{X_t} = \varphi_0 + \varphi_1 \mathbf{X_{t-1}} + \varphi_2 \mathbf{X_{t-2}} + \cdots\cdots + \varphi_p \mathbf{X_{t-p}} + \boldsymbol{a}_t - \boldsymbol{\Theta}_1 \boldsymbol{a}_{t-1} - \cdots\cdots - \boldsymbol{\Theta}_q \boldsymbol{a}_{t-q}, \tag{2.10}$$

or in the lag operator form

$$(I_k - \varphi_1 L - \cdots\cdots \varphi_p L^p)\mathbf{X_t} = \varphi_0 + \left(I_k - \boldsymbol{\Theta}_1 L - \cdots\cdots - \boldsymbol{\Theta}_q L^q\right) a_t, \tag{2.11}$$

or

$$\varphi\left(L\right) X_t = \varphi_0 + \boldsymbol{\Theta}(L)\boldsymbol{a}_t,$$

where

$\mathbf{X_t}$: $k \times 1$ random vector.

φ_i : $k \times k$ autoregressive coeficient matrix.

$\boldsymbol{\Theta}_i$: $k \times k$ moving average matrix.

φ_0 : $k \times 1$ vector of constant terms.

$\boldsymbol{a}_t$: $k \times 1$ white noise process, which is defines as follows:

$E\left(\boldsymbol{a}_t\right) = 0$ and $E(\boldsymbol{a}_t \boldsymbol{a}_t{}^{'}) = \Sigma_a$. The process is stationary if the zeros of determinant

$$\left|I_k - \varphi_1 L - \cdots\cdots \varphi_p L^p\right|$$

are all greater than one and invertible if the zeros of determinant

$$\left|I_k - \boldsymbol{\Theta}_1 L - \cdots\cdots - \boldsymbol{\Theta}_q L^q\right|$$

are all greater than one.

2.8.1 Bivariate ARMA(1,1) model

The bivariate VARMA(1,1) model is given by

$$(I_2 - \varphi_1 L)\mathbf{X_t} = \varphi_0 + (I_2 - \mathbf{\Theta}_1 L)\, a_t, \tag{2.12}$$

or,

$$\begin{pmatrix} \varphi_{11}(L) & \varphi_{12}(L) \\ \varphi_{21}(L) & \varphi_{22}(L) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varphi_{10} \\ \varphi_{20} \end{pmatrix} + \begin{pmatrix} \theta_{11}(L) & \theta_{12}(L) \\ \theta_{21}(L) & \theta_{22}(L) \end{pmatrix} \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}.$$

The model is stationary if the zeros of the determinant polynomial $|1 - \varphi_1 B|$ are outside the unit circle or, if all the eigen values of φ_1 are inside the unit circle.

2.9 Unit root Non Stationarity and Cointegration

A sequence that contains one or more roots of its characteristic polynomial that are equal to one is called a unit root process. The simplest model that may contain a unit root is the autoregressive model of order one,

$$X_t = \phi X_{t-1} + a_t, \tag{2.13}$$

where a_t denotes a serially uncorrelated white noise with 0 mean and constant variance.

If ϕ= 1, equation (2.13) becomes a random walk without drift model, that is, a non-stationary process. Dickey & Fuller (1979) developed a test procedure to determine whether a variable has a unit root, or equivalently, the variable follows a random walk model. If, $|\phi| < 1$, then the series $\{X_t\}$ is stationary. A process which is not stationary in levels, but stationary in differences is called an integrated series. More generally, a univariate time series $\{X_t\}$ with no deterministic component which has a stationary, invertible, ARMA representation after differencing d times, is said to be integrated of order d, denoted $X_t \sim I(d)$. Thus for d=0, $\{X_t\}$ will be stationary and for d=1, the change, $\Delta X_t = X_t - X_{t-1}$ is stationary (Engle & Granger (1987)). But differencing the vector of time series $\mathbf{X_t}$, is more complicated and should be handled carefully. Over differencing the series may often lead to complications in model fitting. More recently, an alternative way to handle non-stationarity has become popular. Cointegration builds on this structure by defining relationship across time series which transform $I(1)$ series in to $I(0)$. So, when modelling several unit root non stationary time series jointly, we may encounter the case of cointegration.

2.9.1 Cointegration

Even though each individual series under consideration may be I (1), we may still be able to find a linear combination of these series which is I (0). If this is the case, then the series are said to be cointegrated (of order 1). Examples might be short and long term interest rates, capital appropriations and expenditures, household income and expenditures, and prices of the same commodity in different markets. Cointegration may also characterise two or more variables. For

example, the existence of money demand function implies that a linear combination of log series of real money stock, the log aggregate income and nominal interest rate may be stationary even though each of the three variables is $I(1)$. The remedy for problematic regressions with non stationary variables is, to test for cointegration and to estimate a vector error-correction model to distinguish between short-run and long-run responses, since cointegration provides more powerful tools when the data sets are of limited length. An individual economic variable, viewed as a time series, can wander extensively and yet some pairs of series may be expected to move so that they do not drift too apart. Typically economic theory will propose forces which tend to keep such series together. Thus, cointegration is an econometric concept which mimics the existence of a long-run equilibrium among economic time series.
The general definition of cointegration is given below, (cf. Engle & Granger (1987)).

Definition 2.1. The components of the vector $\{\mathbf{X_t}\}$ are said to be co-integrated of order d, b, denoted $\mathbf{X_t} \sim CI(d, b)$, if (i) all components of $\{\mathbf{X_t}\}$ are I(d); (ii) there exists a vector $\boldsymbol{\alpha} \neq 0$ so that $z_t = \boldsymbol{\alpha}'\mathbf{X_t} \sim I(d-b)$, b>0.

Some observations (cf. Engle & Granger (1987)).

- The vector $\boldsymbol{\alpha}$ is called cointegating vector.
- The cointegration vector is not unique. If $\boldsymbol{\alpha}$ is a cointegrating vector, then so is $\lambda\boldsymbol{\alpha}$ for any $\lambda \neq 0$

- Economically, a cointegrating relationship may be interpreted as a long-run equilibrium, in the sense that variables have a tendency to revert to this relationship in the long run. However, cointegration is primarily a statistical concept and whether or not it has meaning in an economic sense depends a lot on the specific application.
- The above definition can be extended to allow for constant term.
- For n integrated series, $n-1$ independent cointegrating vector may exist.
- Differencing the data has the disadvantage that information contained in the data is lost. Explicitly taking the non-stationarity into account has the advantage that the information about the long-run behaviour of the data is used in the estimation.

Since we are mainly focussing on bivariate cointegration in our entire thesis, here we state the definition for bivariate cointegration.

Definition 2.2. (Bivariate Cointegration). Let us assume that $\{X_{1t}\}$ and $\{X_{2t}\}$ be two $I(1)$ series. These series are said to be cointegrated, if there exist a vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^{'}$ with both elements non zero such that,

$$\boldsymbol{\alpha}^{'}[X_{1t}, X_{2t}]^{'} = \alpha_1 X_{1t} - \alpha_2 X_{2t} \sim I(0).$$

2.10 Error Correction model

A principal feature of cointegrated series is that their time paths are influenced by the extent of any kind of deviation from long-run equilibrium. After all, if the system is to return to the long-run equilibrium, the movements of atleast some of the variables must respond to the magnitude of the disequilibrium. For example, the theories of the term "structure of interest rates" imply a long-run relationship between long and short-term interest rates. If the gap between the long and short-term rates is "large" relative to the long-run relationship, the short term rate must ultimately rise relative to the long-term rate. This gap can be closed by (1) an increase in the short-term rate and/or a decrease in the long term rate, (2) an increase in the long-term rate but a larger rise in the short-term rate, (3) a fall in the long-term rate but a smaller fall in the short-term rate. Hence, without a full dynamic model specification of the model, we cannot determine which of the possibilities will occur. Nevertheless, the short-run dynamics must be influenced by the deviation from the long-run relationship.

The dynamic model implied by the above discussion is one of **error correction**. If we difference the $I(1)$ data, we may loose the long run information and could estimate only the short run model. Because, with the differenced data, we can only know the effect of change on one variable on another, not the level effect. An alternative is to use the error correction model, which estimates both short and long run relationships jointly, if variables are cointegrated. Error Correction models have been widely used in economics. Engle & Granger (1987) discuss an error correction representation for a co-integrated system that overcomes the difficulty of estimating non invertible VARMA models.

The Granger representation theorem states that: if the variables $\{X_{it}\}, t = 1, 2, \cdots N$ are cointegrated, then there exists an error correction representation for $X_t = (X_{1t}, X_{2t}, \cdots, X_{Nt})'$ (See Engle & Granger (1987)). For the better understanding of cointegration, we focus on VAR models for their simplicity in estimation. The simplest form of the VAR model is with the bivariate case and is given by,

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix} = \begin{pmatrix} \mathrm{P}_{11} & \mathrm{P}_{12} \\ \mathrm{P}_{21} & \mathrm{P}_{22} \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}. \tag{2.14}$$

Equation (2.14) states that the changes in $\{X_{1t}\}$ and $\{X_{2t}\}$ are related to the levels of $\{X_{1t}\}$ and $\{X_{2t}\}$ through a cointegration matrix P. But since the variables $\{X_{1t}\}$ and $\{X_{2t}\}$ are cointegrated, there exist an α such that $X_{1t} - \alpha X_{2t}$ is $I(0)$. Hence, substituting this relation in equation (2.14) leads to,

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \begin{pmatrix} 1 & -\alpha \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}$$

Hence the short run dynamics or the error correction form for the VAR(1) process $\{\mathbf{X_t}\}$ takes the form:

$$\Delta X_{1t} = \beta_1(X_{1,t-1} - \alpha X_{2,t-1}) + a_{1t}.$$
$$\Delta X_{2t} = \beta_2(X_{1,t-1} - \alpha X_{2,t-1}) + a_{2t}.$$

The term $X_{1,t-1} - \alpha X_{2,t-1}$ represents the deviation from the long run trend (equilibrium correction term) and β_1 and β_2 are the speed adjustment parameters.

Example 2.1. *Consider a simple bivariate cointegrated VAR(1) model given by,*

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}. \tag{2.15}$$

In order to transform this VAR in to error correction model, we begin by subtracting $\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}$ *from both the sides of equation (2.15),*

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} - \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} - \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}.$$

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix} = \left(\begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}.$$

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix} = \begin{pmatrix} -0.2 & 0.2 \\ 0.2 & -0.2 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}.$$

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix} = \begin{pmatrix} -0.2 \\ 0.2 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}.$$

For this example, we can see that the speed of adjustment parameter is -0.2 for ΔX_{1t} and 0.2 for ΔX_{2t}, and the cointegrating vector is $\begin{pmatrix} 1 & -1 \end{pmatrix}$.

Remark 2.1. It is possible to transform a bivariate cointegrated VAR process in to an error correction model by recursive substitution. For that, we consider the cointegrated VAR(3) process given by,

$$\mathbf{X_t} = \varphi_1 \mathbf{X_{t-1}} + \varphi_2 \mathbf{X_{t-2}} + \varphi_3 \mathbf{X_{t-3}} + \mathbf{a}_t.$$

To begin the transformation, we add and subtract $\varphi_3\mathbf{X_{t-2}}$ to the right side,

$$\begin{aligned}\mathbf{X_t} &= \varphi_1\mathbf{X_{t-1}} + \varphi_2\mathbf{X_{t-2}} + \varphi_3\mathbf{X_{t-2}} - \varphi_3\mathbf{X_{t-2}} + \varphi_3\mathbf{X_{t-3}} + \mathbf{a}_t \\ &= \varphi_1\mathbf{X_{t-1}} + \varphi_2\mathbf{X_{t-2}} + \varphi_3\mathbf{X_{t-2}} - \varphi_3\Delta\mathbf{X_{t-2}} + \mathbf{a}_t \\ &= \varphi_1\mathbf{X_{t-1}} + (\varphi_2 + \varphi_3)\mathbf{X_{t-2}} - \varphi_3\Delta\mathbf{X_{t-2}} + \mathbf{a}_t.\end{aligned}$$

Now we add and subtract $(\varphi_2 + \varphi_3)\mathbf{X_{t-1}}$ to the right side gives,

$$\begin{aligned}\mathbf{X_t} &= \varphi_1\mathbf{X_{t-1}} + (\varphi_2 + \varphi_3)\mathbf{X_{t-1}} - (\varphi_2 + \varphi_3)\mathbf{X_{t-1}} + (\varphi_2 + \varphi_3)\mathbf{X_{t-2}} - \varphi_3\Delta\mathbf{X_{t-2}} + \mathbf{a}_t \\ &= \varphi_1\mathbf{X_{t-1}} + (\varphi_2 + \varphi_3)\mathbf{X_{t-1}} - (\varphi_2 + \varphi_3)\Delta\mathbf{X_{t-1}} - \varphi_3\Delta\mathbf{x_{t-2}} + \boldsymbol{a}_t \\ &= (\varphi_1 + \varphi_2 + \varphi_3)\mathbf{X_{t-1}} - (\varphi_2 + \varphi_3)\Delta\mathbf{X_{t-1}} - \varphi_3\Delta\mathbf{X_{t-2}} + \mathbf{a}_t.\end{aligned}$$

Finally, on subtracting $\mathbf{X_{t-1}}$ from both the sides,

$$\begin{aligned}\Delta\mathbf{X}_t &= (\varphi_1 + \varphi_2 + \varphi_3 - \mathbf{I}_2)\mathbf{X_{t-1}} - (\varphi_2 + \varphi_3)\Delta\mathbf{X_{t-1}} - \varphi_3\Delta\mathbf{X_{t-2}} + \mathbf{a}_t \\ &= \mathbf{\Pi X_{t-1}} - \mathbf{\Pi_1 \Delta X_{t-1}} - \mathbf{\Pi_2 \Delta X_{t-2}} + \mathbf{a_t} \qquad (2.16) \\ &= \boldsymbol{\beta\alpha}'\mathbf{X_{t-1}} - \mathbf{\Pi_1}\Delta\mathbf{X_{t-1}} - \mathbf{\Pi_2}\Delta\mathbf{X_{t-2}} + \mathbf{a}_t, \qquad (2.17)\end{aligned}$$

where $\boldsymbol{\beta}$ contains the speed of adjustment parameters, $\boldsymbol{\alpha}$ contains the cointegration vectors, $\mathbf{\Pi} = (\varphi_1 + \varphi_2 + \varphi_3 - \mathbf{I}_2)$, $\mathbf{\Pi_1} = (\varphi_2 + \varphi_3)$ and $\mathbf{\Pi_2} = \varphi_3$. Proceeding like this, we can transform any cointegrated VAR(p) process in to its error correction representation. Hence recursively doing, the error correction representation associated with a VAR(p) process is given by

$$\Delta\mathbf{X}_t = \mathbf{\Pi X_{t-1}} + \mathbf{\Pi_1 \Delta X_{t-1}} + \mathbf{\Pi_2 \Delta X_{t-2}} + \ldots\ldots + \mathbf{\Pi_{p-1} \Delta X_{t-p+1}} + \mathbf{a_t}, \qquad (2.18)$$

where $\mathbf{\Pi} = -\mathbf{I}_2 + \sum_{i=1}^{p} \varphi_i$ and $\mathbf{\Pi_j} = -\sum_{i=j+1}^{p} \varphi_i, j = 1, 2, \cdots, p-1$. (Details of this section can be seen in Tsay (2005), Chapter 8) .

2.11 Test for Cointegration in the presence of Gaussian Errors

Once the variables have been classified as integrated of order I(0), I(1), I(2) etc, we can set up models that lead to stationary relations among the variables, and where standard inference is possible. As one or more integrated variables may have a stationary cointegrating relationship, it is of interest to test for the existence of such a cointegrating relationship. Cointegration is an essential step to check if our modelling has empirically meaningful relationships. If variables have different trends processes, they cannot stay in fixed long-run relation to each other, implying that we cannot model the long-run, and there is usually no valid base for inference based on standard distributions. If we do not find co-integration, it is necessary to continue to work with variables in differences instead. There are a number of cointegration tests in literature, and here we briefly discuss the Engle and Granger two step procedure and the Johansen's procedure, which have the assumption of Gaussian distribution for the errors $\{\mathbf{a_t}\}$ in the error correction equation given by

$$\Delta \mathbf{X}_t = \mathbf{\Pi X_{t-1}} + \mathbf{\Pi_1 \Delta X_{t-1}} + \mathbf{\Pi_2 \Delta X_{t-2}} + \ldots\ldots + \mathbf{\Pi_{p-1} \Delta X_{t-p+1}} + \mathbf{a_t}.$$

2.11.1 Engle-Granger method

Among a number of alternative methods, the Engle-Granger Method, originally suggested by Engle & Granger (1987), has received a great deal of attention in recent years. The main benefit is that the long-run equilibrium relationship (i.e. the co-integrating regression) can be modeled by a straight forward regression involving the levels of the variables. If each component variable of an observed time series vector is subjected to unit root analysis and it is found that all the variables are integrated of order one, I(1), then they contain a unit root. There is a possibility that the regression can still be meaningful (ie; not spurious) provided that the variables co-integrate. In order to find out whether the variables co-integrate, the least squares regression equation is estimated and the residuals (the error term) of the regression equation are subjected to unit root analysis. If the residuals are stationary, that is I (0), it means that the variables under study co-integrate and have a long-term or equilibrium relationship. Accordingly, the steps for determining whether two integrated variable co-integrate of the same order are the following:

- Pre test each variable to determine its order of integration and,
- Estimate the error correction model.

In the two-step estimation procedure, Engle-Granger considered the problem of testing the null hypothesis of no co-integration between a set of variables by estimating the coefficient of a static relationship between economic variables using the OLS and applying well-known unit root tests to the residuals for stationarity.

If the integrated variables are found to be integrated of same order, then it must be tested whether these variables are co-integrated (Johansen (1988)). Rejecting the null hypothesis of a unit root is the evidence in favour of co-integration.

2.12 Testing for Co-integration Using Johansens Methodology

Here we discuss the maximum likelihood estimator of the cointegration space and the likelihood ratio test for testing the rank of $\mathbf{\Pi}$ matrix in the model:

$$\Delta \mathbf{X}_t = \mathbf{\Pi X_{t-1}} + \mathbf{\Pi_1 \Delta X_{t-1}} + \mathbf{\Pi_2 \Delta X_{t-2}} + \ldots\ldots + \mathbf{\Pi_{p-1} \Delta X_{t-p+1}} + \mathbf{a_t}, \tag{2.19}$$

where $\{\mathbf{a}_t\}$ is assumed to be iid Gaussian random variables with mean $\mathbf{0}$ and covariance matrix Λ. Here $\mathbf{\Pi} = \boldsymbol{\beta}\boldsymbol{\alpha}'$, then we shall assume that, although $\Delta\mathbf{X_t}$ is stationary and $\{\mathbf{X_t}\}$ is non stationary as a vector process, the linear combination given by $\boldsymbol{\beta}\boldsymbol{\alpha}'\mathbf{X_{t-1}}$ is stationary. This means that the vector process $\{\mathbf{X_t}\}$ is cointegrated with cointegration vectors $\boldsymbol{\alpha}$. The space spanned by $\boldsymbol{\alpha}$ is the space spanned by the rows of the matrix $\mathbf{\Pi}$, called the cointegration space. The estimation of $\boldsymbol{\alpha}$ is performed by regressing $\Delta\mathbf{X_t}$ and $\mathbf{X_{t-k}}$ on the lagged differences. From the residuals of these regressions, we can calculate the matrix of product moments. The estimates of $\boldsymbol{\alpha}$ is the empirical canonical variates of $\mathbf{X_{t-k}}$ with respect to $\Delta\mathbf{X_t}$ corrected for the lagged differences (See Johansen (1988)). For any $m \leq 2$, the hypothesis of interest is H_0:Rank$(\mathbf{\Pi}) \leq m$. Here we cannot estimate the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, since they form an over parametrisation of the model, but one can estimate the space spanned by $\boldsymbol{\alpha}$. The main result about the estimation of sp($\boldsymbol{\alpha}$) is stated below, whose proof may be found in Tsay (2005), Chapter 8.

Theorem 2.3. *The maximum likelihood estimator of the space spanned by $\boldsymbol{\alpha}$ is the space spanned by the m canonical variates corresponding to the m largest squared canonnical correlations between the residuals of $\mathbf{X_{t-k}}$ and $\Delta\mathbf{X_t}$ corrected for the lagged differences of the process.*

For the testing purpose, let $H(m)$ be the null hypothesis which states that the rank of $\boldsymbol{\Pi}$ is m. Under $H(0)$, Rank($\boldsymbol{\Pi}$)=0, so that $\boldsymbol{\Pi} = \mathbf{0}$, and hence there is no cointegration. The cointegrating rank determines the number of linearly independent cointegrating vectors. Johansen (1988) proposes two different likelihood ratio tests to perform the test, namely, the trace test and maximum eigen value test.

2.12.1 The trace Test

The hypothesis of interest is

$$H_0 : Rank(\boldsymbol{\Pi}) = m \text{ versus } H_a : Rank(\boldsymbol{\Pi}) > m.$$

The test statistic is given by

$$J_{trace} = -(T-p)\sum_{i=m+1}^{k} \ln(1-\hat{\lambda}_i), \tag{2.20}$$

where T is the sample size and $\hat{\lambda}_i$ is the i-th largest canonical correlation. If Rank($\mathbf{\Pi}$)=m, then $\hat{\lambda}_i$ should be small for $i > m$ and hence J_{trace} should be small. Due to the presence of unit roots, the asymptotic distribution of the test statistic is not chi-squared, but it is a function of standard Brownian motions. Thus, the critical values of J_{trace} must be obtained via simulation technique.

2.12.2 Maximum Eigen Value Test

The maximum eigen value test, on the other hand tests the null hypothesis of m co-integrating vectors against the alternative hypothesis of $m+1$ co-integrating vectors. The hypothesis of interest is

$$H_0 : Rank(\mathbf{\Pi}) = m \text{ versus } H_a : Rank(\mathbf{\Pi}) = m + 1.$$

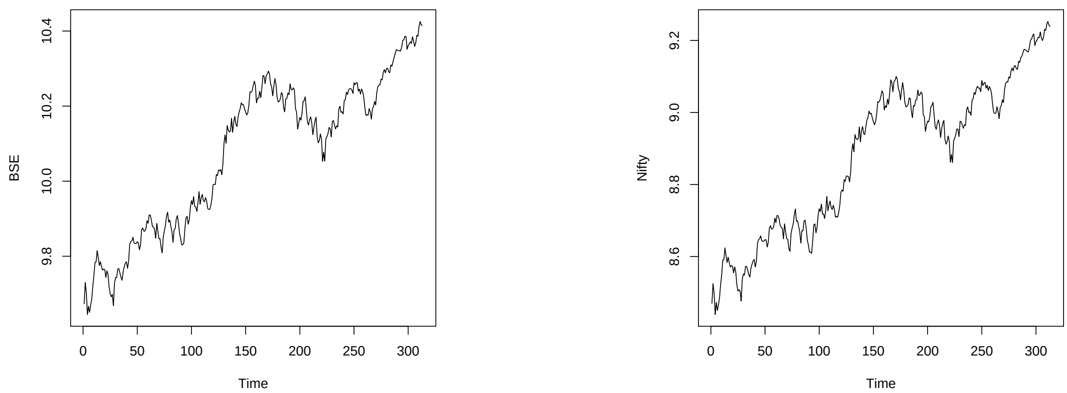
The test statistic is given by

$$J_{\max} = -(T-p)\ln(1 - \hat{\lambda}_{m+1}). \tag{2.21}$$

Neither of these test statistics will follow a chi square distribution in general; assymptotic critical values can be found in Johansen & Juselius (1990). Since the critical values used for the maximum eigenvalue and trace test statistics are based on a pure unit-root assumption, they will no longer be correct when the variables in the system are near-unit-root processes. This method assumes that the co-integrating vector remains constant during the period of study. In reality, it is possible that the long-run relationships between the underlying variables change. The reason for this might be technological progress, economic crisis, changes in people's preferences and behaviour accordingly, policy or regime alteration and institutional development. This is especially the case if the sample period is long.

2.13 Example for Engle Granger Cointegration methodology

This section illustrates the concepts and ideas of cointegration methodology through a financial example. We consider weekly series of BSE sensex and Nifty for the period 2011 to 2017 to see whether there exist any long run relationship between the variables. We transformed both the variables in to their natural logarithm. The time series data are plotted in Fig 2.1. A descriptive summary



(A) Log BSE series (B) Log Nifty series

FIGURE 2.1: Weekly stock price data

of the data is given in Table 2.1.

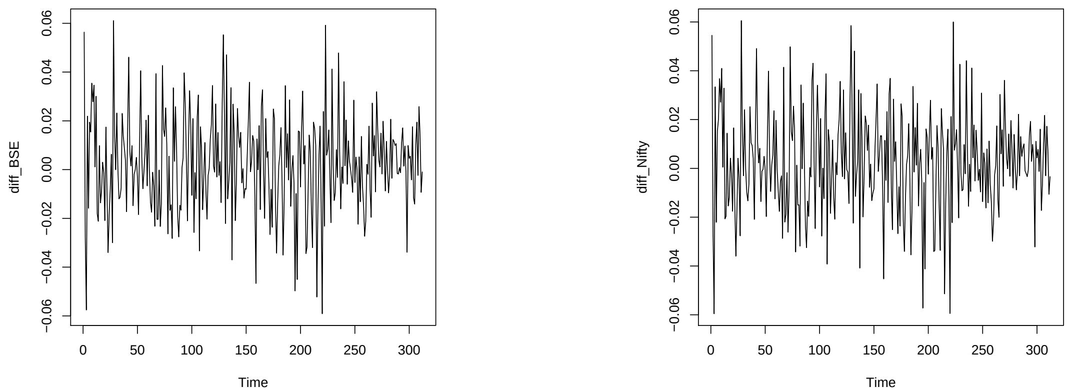
TABLE 2.1: Summary Statistics of log series

	BSE	Nifty
Min	9.650	8.439
1st Qu.:	9.879	8.681
Median	10.148	8.954
Mean	10.072	8.879
3rd Qu.:	10.237	9.049
Max	10.426	9.253

To begin with, we need to pre test each variable in order to find the order of integration. This is needed because cointegration necessitates the variables to be integrated of same order. This has been tested by Augmented Dickey Fuller test. The p values obtained for the series are 0.1474 and 0.1523. Since both the p values are obtained to be large, we cannot reject the null hypothesis of unit root for the variables. This means that the variables contain unit roots and are non stationary.

Next we need to test whether the series is $I(2)$, that is whether it is possible for second order of integration. Fig 2.2 is the plot of the differenced series and it can been seen that the first differenced series is stationary. Second order integration is also tested by using augmented Dickey Fuller test. Both the p values are obtained to be less than 2.2e-16, which also gives the indication that the differenced series are stationary. Now we can test for the existence of the long run relationship between the variables, that is, cointegration.

To implement the Engle Granger methodology, we begin by regressing the S & P BSE sensex and Nifty 50 series on each other and then assess the model fit. If a cointegrating relationship exists among the variables, then the OLS regression yields a super consistent estimator of the cointegrating parameters. That is,



(A) BSE series (B) Nifty price

FIGURE 2.2: Differenced series of log-stock price data

there exist a very strong relationship between the estimated parameters. Taking the S & P BSE sensex series as the dependent variable and Nifty 50 series as the independent variable, gives the following regression equation

$$S\&P = 1.5146 + 0.9631 * Nifty50.$$

TABLE 2.2: Estimates of regression coefficient for BSE series

Coefficients				
	Estimates	Std.Error	t value	P-value
Intercept	1.5146	0.0279	54.12	<2e-16
Nifty	0.9637	0.0031	305.9	<2e-16
		Multiple R-squared: 0.9967		

It is seen that the p value of the Nifty 50 series is very small (<2e-16), hence the regression is statistically significant. Also, the R-Squared value is 0.9967, means that 99.67% of the variation in S&P Series is explained by Nifty 50 series. The regression equation corresponds to NIfty 50 series is

$$Nifty50 = -1.5368 + 1.0341 * S\&P.$$

Next we test the residuals from the regression relationship in order to deter-

TABLE 2.3: Estimates of regression coefficient for Nifty series

Coefficients				
	Estimates	Std.Error	t value	P-value
Intercept	-1.5368	0.0340	-45.12	<2e-16
S&P	1.0341	0.0033	305.9	<2e-16
		Multiple R-squared: 0.9967		

mine if the variables actually form a cointegration relation. So we perform an ADF test for the residuals of the regression equation. The p value obtained for the ADF test is 0.0029 and 0.00268, implies that the residuals are stationary. Hence we can conclude that the two variables form a stationary cointegrating relationship.

In the next step, we have estimated the error correction model for the cointegrated series. The estimated ECM for the S&P and Nifty series are given by,

$$\Delta \hat{S}\&P = 0.00003 + 0.949\Delta \hat{N}ifty - 0.0340\hat{a}_{1t-1},$$

$$\Delta \hat{N}ifty = 0.00001 + 1.028\Delta \hat{S}\&P - 0.0344\hat{a}_{2t-1}.$$

2.14 ECM for a bivariate Cointegrating system

As pointed out in subsection 2.9.1, our main interest in this thesis is a detailed analysis of bivariate cointegrating models when innovations of their error correction model follow non Gaussian distributions. In this section, we consider the general form of a bivariate cointegrating model and its error correction representation.

Let us consider an example of bivariate model discussed by Engle & Granger (1987), where $\{X_{1t}\}$ and $\{X_{2t}\}$ are two cointegrating time series defined by

$$X_{1t} + \beta X_{2t} = u_{1t}, u_{1t} = u_{1t-1} + a_{1t} \tag{2.22}$$

$$X_{1t} + \alpha X_{2t} = u_{2t}, u_{2t} = \phi u_{2t-1} + a_{2t}, |\phi| < 1, \tag{2.23}$$

where $\{(a_{1t}, a_{2t})'\}$, t=1,2,3,.. is a sequence of iid bivariate random variables. Reason for the above two series become cointegrated is as follows: The reduced form for the process in (2.22) and (2.23) will make the variables X_{1t} and X_{2t} as a linear combination of u_{1t} and u_{2t} and therefore both the series will be nonstationary (Integrated of order 1). From (2.22) and (2.23) we will get,

$$X_{1t} = \left(\frac{\alpha}{\alpha - \beta}\right) u_{1t} - \left(\frac{\beta}{\alpha - \beta}\right) u_{2t} \tag{2.24}$$

and

$$X_{2t} = \left(\frac{1}{\alpha - \beta}\right) u_{2t} - \left(\frac{1}{\alpha - \beta}\right) u_{1t}. \tag{2.25}$$

Hence from the above two equations, it is clear that $\{X_{1t}\}$ and $\{X_{2t}\}$ are non stationary as they are linear combinations of a stationary and a non stationary series. Since $\{X_{1t}\}$ and $\{X_{2t}\}$ are integrated series, equation (2.23) describes a stationary linear combination of the nonstationary variables. Thus the variables X_{1t} and X_{2t} are cointegrated and hence we can say that they have a long run relationship in equilibrium. But if $\phi \to 1$, then the series are uncorrelated random walks and hence they are no longer cointegrated. The model given above has been studied by Engle & Granger (1987) in detail with possibly correlated white noise and the model can be transformed in to the error correction form by subtracting the lagged values from both the sides.

Before studying the properties of the model it is convenient to reparameterise the model in (2.22) and (2.23) by subtracting the lagged values from both sides. Let Δ be a difference operator, on applying Δ operator on X_{1t} and X_{2t} of both sides of equations (2.22) and after some algebra we will get,

$$\begin{aligned}
\Delta X_{1t} &= u_{1t} - \beta X_{2t} - X_{1t-1} \\
&= u_{1t-1} + a_{1t} - \beta X_{2t} - X_{1t-1} \\
&= X_{1t-1} + \beta X_{2t-1} + a_{1t} - \beta X_{2t} - X_{1t-1} \\
&= \beta X_{2t-1} + a_{1t} - \frac{\beta}{\alpha}\left(u_{2t} - X_{1t}\right) \\
&= \beta X_{2t-1} + a_{1t} - \frac{\beta}{\alpha}\left(\phi u_{2t-1} + a_{2t} - X_{1t}\right) \\
&= \beta X_{2t-1} + a_{1t} - \frac{\beta}{\alpha}\left(\phi\left(X_{1t-1} + \alpha X_{2t-1}\right) + a_{2t} - X_{1t}\right).
\end{aligned}$$

That is,

$$\left(1-\frac{\beta}{\alpha}\right)\Delta X_{1t}=\beta X_{2t-1}\left(1-\phi\right)+\frac{\beta}{\alpha}X_{1t-1}\left(1-\phi\right)+a_{1t}-\frac{\beta}{\alpha}a_{2t}.$$

On adjusting the terms, we get

$$\begin{aligned}\Delta X_{1t}&=\frac{(1-\phi)}{\alpha-\beta}\beta\left(X_{1t-1}+\alpha X_{2t-1}\right)+\eta_{1t}\\&=\delta\beta Z_{t-1}+\eta_{1t},\end{aligned}$$

where $Z_{t-1}=X_{1,t-1}+\alpha X_{2,t-1}$, $\eta_{1t}=\frac{\alpha}{(\alpha-\beta)}\left(a_{1t}-\frac{\beta}{\alpha}a_{2t}\right)$, where α is the cointegrating parameter. Similarly by using Equation (2.23), we can obtain the second error correction equation. On subtracting αX_{2t-1} from both sides of equation (2.23),we get

$$X_{1t}+\alpha X_{2t}-\alpha X_{2t-1}=u_{2t}-\alpha X_{2t-1}.$$

That is,

$$\begin{aligned}\alpha\Delta X_{2t}&=\phi u_{2t-1}+a_{2t}-\alpha X_{2t-1}-X_{1t}\\&=\phi(X_{1t-1}+\alpha X_{2t-1})+a_{2t}-\alpha X_{2t-1}-X_{1t}\\&=\phi(X_{1t-1}+\alpha X_{2t-1})+a_{2t}-\alpha X_{2t-1}-(u_{1t}-\beta X_{2t})\\&=\phi(X_{1t-1}+\alpha X_{2t-1})+a_{2t}-\alpha X_{2t-1}-(u_{1t-1}+a_{1t}-\beta X_{2t})\\&=\phi(X_{1t-1}+\alpha X_{2t-1})+a_{2t}-\alpha X_{2t-1}-(X_{1t-1}+\beta X_{2t-1}+a_{1t}-\beta X_{2t}).\end{aligned}$$

Thus on adjusting the terms,

$$(\alpha-\beta)\Delta X_{2t}=X_{1t-1}(\phi-1)+\alpha X_{2t-1}(\phi-1)+a_{2t}-a_{1t}.$$

Hence,

$$\Delta X_{2t} = -\delta(X_{1,t-1} + \alpha X_{2,t-1}) + \eta_{2t},$$

where $\eta_{2t} = \frac{1}{(\alpha-\beta)}(a_{2t} - a_{1t})$ and $\delta = \frac{1-\phi}{\alpha-\beta}$. Hence the Error Correction representation becomes:

$$\Delta X_{1t} = \delta\beta Z_{t-1} + \eta_{1t}, \tag{2.26}$$

$$\Delta X_{2t} = -\delta Z_{t-1} + \eta_{2t}. \tag{2.27}$$

The next four Chapters are devoted to the study of bivariate cointegrating models, when (η_{1t}, η_{2t}) follow some non Gaussian distributions.

Chapter 3

Unit root and cointegration with logistic errors

3.1 Introduction

The methods for analysing time series are developed by assuming that the observed series is a realization of certain discrete parameter stationary stochastic process. However, a time series representing a real situation need not be stationary. Box et al. (2015) argued that certain non-stationary time series can be converted in to stationary one by successive differences. The study on basic linear time series reveals that, if the series becomes stationary ARMA, after $d-$ differences then there are d unit roots in the characteristic polynomial of the underlying autoregressive model. That is, we are assuming that the non stationarity was only due to the presence of unit roots. Granger (1981) pointed out that set of all time series which achieve stationarity after differencing may have linear combinations which are stationary without differencing. Engle & Granger (1987) formalized this idea and introduced the concept of co-integration. Since the problem of cointegration and the unit root are closely related, test for cointegration can be carried out by testing for unit root from the residuals of cointegrating regression series. Kim & Schmidt (1993) considered the finite sample accuracy(size) of the Dickey Fuller unit root test when the errors were conditionally hetroskedastic. Lee & Tse (1996) examined the performance of Johansen likelihood ratio tests for cointegration in the presence of GARCH errors.

To the best of our knowledge, apart from the above cointegration tests based on conditionally hetroskedastic errors, to date, there is no study on cointegration and error correction model when the innovations are non normal in their distributions. There are several standard non normal distributions in literature and each distributions may need independent attention.

In this chapter, we study the properties of two cointegrating time series and model them with independent and identically distributed logistic error variables. We propose an estimation procedure for cointegrating parameters using the method of conditional maximum likelihood estimation and then develop a test procedure for unit root and cointegration when the innovation processes are generated by iid logistic random variables. Since the underlying distribution of the test statistic is well-known (asymptotic Chi-square), a bootstrap method provides a way to account for the distortions caused by the finite sample. To account that, we perform a bootstrap test based on MLE for the likelihood ratio test for cointegration.

Rest of the chapter is organized as follows. In Section 3.2, we define a cointegration model with logistic innovations and study the likelihood based estimation in Section 3.3. In Sections 3.4 and 3.5 we study the problems of testing of hypothesis on unit root and cointegration. A simulation study is conducted in Section 3.6 followed by bootstrap method in Section 3.7. To illustrate the applications of our model, a data analysis is presented in Section 3.8.

A brief discussion of cointegration model in the presence of Gaussian innovation is included in Chapter 2, Section 2.12.

3.2 Cointegrating model with logistic innovations

In this Chapter, we discuss a bivariate cointegrating model when the innovations of the corresponding ECM follow iid logistic distributions. That is, we assume that $\{(\eta_{1t}, \eta_{2t})', t = 1, 2, \cdots\}$ in (2.26) and (2.27) are iid bivariate random variables with independent marginals following symmetric logistic distribution with probability density function of the form

$$f(\eta_{it}) = \frac{e^{-\eta_{it}}}{(1 + e^{-\eta_{it}})^2}, i = 1, 2, -\infty < \eta_{it} < \infty. \tag{3.1}$$

with $E(\eta_{it}) = 0$ and $V(\eta_{it}) = \frac{\pi^2}{3}$. Further details on logistic distribution may be found in Johnson et al. (1994).

The definition of a bivariate cointegrating model and its properties are discussed in Section 2.14. The error correction representation given in (2.26) and (2.27) has three unknown parameters and we estimate them by the method of conditional maximum likelihood.

3.3 Conditional MLE for Error Correction Model with iid logistic errors

Estimation of model parameters is one of the important problems involved in modelling of Gaussian and non Gaussian time series. Tiku et al. (1999) developed estimation method for a regression model with autocorrelated errors following a shift scaled Student's t distribution. Wong & Bian (2005) extended the work of Tiku et.al to the case, where the underlying distribution is a generalised logistic distribution using the modified maximum likelihood estimators since maximum likelihood estimates are intractable. However, in our model we do not encounter such a problem while estimating the cointegration parameters using logistic innovations and hence we can proceed with the estimation technique using the conditional maximum likelihood method. If an explicit form for the innovation density function is available, then the conditional likelihood based inference is possible for error correction model given in (2.26) and (2.27). To obtain the maximum likelihood estimation of parameters in the error correction model, the innovation random variables are assumed to follow iid logistic distribution with marginal pdf (3.1). The joint probability density function of (η_{1t}, η_{2t}) through the ECM (2.26) and (2.27) is given by

$$f(\eta_{1t}, \eta_{2t}) = \frac{e^{-[(\Delta x_{1t} - \delta\beta z_{t-1}) + (\Delta x_{2t} + \delta z_{t-1})]}}{\left(1 + e^{-(\Delta x_{1t} - \delta\beta z_{t-1})}\right)^2 \left(1 + e^{-(\Delta x_{2t} + \delta z_{t-1})}\right)^2}, \tag{3.2}$$

where $\alpha, \beta, \delta, x_1, x_2$ are all real. The parameter vector to be estimated are the elements of $\theta = (\alpha, \beta, \delta)'$. The conditional log-likelihood function for the ECM for specified values of x_{10}, x_{20} becomes:

$$L_T(\theta) = \sum_{t=1}^{n}\{-[(\Delta x_{1t} - \delta\beta z_{t-1}) + (\Delta x_{2t} + \delta z_{t-1})]$$

$$-2\log(1+e^{-(\Delta x_{1t}-\delta\beta z_{t-1})}) - 2\log(1+e^{-(\Delta x_{2t}+\delta z_{t-1})})\},$$

where $z_0 = x_{1,0} + \alpha x_{2,0}$. The form of the above log likelihood function suggest that we have to maximize it by some numerical methods. Hence on differentiating the log-likelihood function with respect to the parameter vector θ, we will get three equations given by,

$$\begin{aligned}\frac{\partial L_T(\theta|x_{10}, x_{20})}{\partial\delta} &= \sum_{t=1}^{n}\{z_{t-1}(\beta-1) - \frac{2\beta z_{t-1}e^{-(\Delta x_{1t}-\delta\beta z_{t-1})}}{(1+e^{-(\Delta x_{1t}-\delta\beta z_{t-1})})} \\ &\quad + \frac{2z_{t-1}e^{-(\Delta x_{2t}+\delta z_{t-1})}}{(1+e^{-(\Delta x_{2t}+\delta z_{t-1})})}\} = 0. \end{aligned} \tag{3.3}$$

$$\begin{aligned}\frac{\partial L_T(\theta|x_{10}, x_{20})}{\partial\alpha} &= \sum_{t=1}^{n}\{\delta x_{2t-1}(\beta - 1 - \frac{2\beta e^{-(\Delta x_{1t}-\delta\beta z_{t-1})}}{(1+e^{-(\Delta x_{1t}-\delta\beta z_{t-1})})} \\ &\quad + \frac{2e^{-(\Delta x_{2t}+\delta z_{t-1})}}{(1+e^{-(\Delta x_{2t}+\delta z_{t-1})})})\} = 0. \end{aligned} \tag{3.4}$$

$$\frac{\partial L_T(\theta|x_{10}, x_{20})}{\partial\beta} = \sum_{t=1}^{n}\{\delta z_{t-1}(1 - \frac{2e^{-(\Delta x_{1t}-\delta\beta z_{t-1})}}{(1+e^{-(\Delta x_{1t}-\delta\beta z_{t-1})})})\} = 0. \tag{3.5}$$

These equations are solved numerically by using Newton Raphson method and are illustrated using simulated samples in Table 3.3. An algorithm for the simulated sample is summarised in Section 3.6.1.

The study of cointegrating models with logistic errors uses the properties of first order autoregressive models with logistic innovations, which we discuss in the next Section.

3.4 Unit root test for AR(1) model with logistic errors

Dickey & Fuller (1979) developed a unit root test for cointegration among non-stationary time series when the innovations were assumed to follow Gaussian series. Some other authors have examined the size distortions of this test when the errors were conditionally hetroskedastic. In particular, our interest is to analyse the time series in the presence of non-normal innovations, specifically logistic errors. If the time series are integrated of same order and are non stationary, then test for cointegration can be carried out by developing a unit root test for the residual series of either cointegrating regression equation or of the ECM. If the residuals obtained from the error correction model are stationary, then the variables could explain a long run behaviour in the equilibrium and hence they are cointegrated.

Let us consider the first order autoregressive process $\{X_t\}$ defined by,

$$X_t = \phi X_{t-1} + a_t, \tag{3.6}$$

where $X_0 = 0$ and $\{a_t\}$ is a sequence of independent logistic random variables with mean zero. Note that $X_t = a_t + \phi\ a_{t-1} + ... + \phi^{t-1}\ a_1$ and if $|\phi| < 1$, X_t converges to a stationary process as $t \to \infty$ with $E(X_t) = 0$ and $V(X_t) = \pi^2/3(1-\phi^2)$. If a realisation $(X_1, X_2, X_3, ..., X_n)$ of a first order autoregressive time series are given, we are interested in finding an estimator of ϕ and in tests of the null hypothesis that $H_0 : \phi = 1$. Mostly, the alternative hypothesis of interest, $H_1 : \phi < 1$ is that the time series $\{X_t\}$ was generated by $X_t = \phi\ X_{t-1} + a_t$, where $|\phi| < 1$. One can also consider the alternative hypothesis of interest that the time series is generated by $X_t = \beta t + \phi X_{t-1} + a_t$, where $|\phi| < 1$. Our interest is to find $\hat{\phi}$, the maximum likelihood estimator for ϕ in model (3.6) and hence shall obtain the test procedure of unit root under the null hypothesis.

Let us suppose n observations, say x_1,x_2,x_3,.....,x_n are available for the analysis and we shall obtain the likelihood function based on n observations generated by the model (3.6). The joint density function of $(a_1, a_2, a_3, ...a_n)$ is

$$\prod_{t=1}^{n} \frac{e^{-a_t}}{(1 + e^{-a_t})^2}.$$

For the model (3.6), the joint probability density function of $(x_1, x_2, \cdots, x_n | x_0)$ is

$$\prod_{t=1}^{n} \frac{e^{-(x_t - \phi x_{t-1})}}{\left(1 + e^{-(x_t - \phi x_{t-1})}\right)^2}.$$

The log-likelihood function of ϕ conditioned on x_0 is

$$L_T(\phi \mid x_0) = -\sum_{t=1}^{n} \{(x_t - \phi x_{t-1}) + 2\log(1 + e^{-(x_t - \phi x_{t-1})})\}.$$

The critical points of the above log likelihood function can be obtained by setting the first derivative with respect to ϕ equal to zero.

$$\begin{aligned} \frac{\partial L_T(\phi \mid x_0)}{\partial \phi} &= \sum_{t=1}^{n} \left(x_{t-1} - 2x_{t-1} \frac{e^{-(x_t - \phi x_{t-1})}}{(1 + e^{-(x_t - \phi x_{t-1})})} \right) \\ &= \sum_{t=1}^{n} [x_{t-1} - 2x_{t-1}\Gamma], \end{aligned} \tag{3.7}$$

where $\Gamma = \frac{e^{-(x_t - \phi x_{t-1})}}{(1+e^{-(x_t - \phi x_{t-1})})}$. The first order partial derivative equation suggest that the value of ϕ that maximises log likelihood function must satisfy $\sum\limits_{t=1}^{n} [x_{t-1} - 2x_{t-1}\Gamma] = 0$. This equation can be solved by some numerical technique and if any such solution exists specifies a critical point, which is either a maximum or a minimum. It should be noted that if the partial derivative of second order is negative, then the critical point will correspond to a maximum.

$$\begin{aligned}\frac{\partial^2 L_T(\phi \mid x_0)}{\partial\phi\partial\phi'} &= -\sum_{t=1}^{n} x_{t-1}\left(2x_{t-1}\frac{e^{-(x_t-\phi x_{t-1})}}{(1+e^{-(x_t-\phi x_{t-1})})^2}\right)\\ &= \sum_{t=1}^{n} -2x_{t-1}^2\Gamma(1-\Gamma).\end{aligned}$$

We obtain the estimator of ϕ numerically by using the Newton Raphson method. Now let us consider the hypothesis, $H_0 : \phi = 1$ against the alternative $H_1 : |\phi| < 1$, that is the time series $\{X_t\}$ was generated by a stationary model. Under H_0, the maximum value of the likelihood function is $L_0 = e^{-\sum_{t=1}^{n}(x_t-x_{t-1})}\prod_{t=1}^{n}\left(1+e^{-\sum_{t=1}^{n}(x_t-x_{t-1})}\right)$ and under the alternative, the maximum value of likelihood function is, $L_1 = e^{-\sum_{t=1}^{n}(x_t-\hat{\phi}x_{t-1})}\prod_{t=1}^{n}\left(1+e^{-\sum_{t=1}^{n}(x_t-\hat{\phi}x_{t-1})}\right)$.

For $\hat{\phi} \in H_1$, the likelihood ratio test rejects H_0 when $\lambda = e^{-\sum_{t=1}^{n}(X_t(1-\hat{\phi})}\prod_{t=1}^{n}\left(\frac{1+e^{-(x_t-\hat{\phi}x_{t-1})}}{1+e^{-(x_t-x_{t-1})}}\right)$ is small. Wilks (1938) established that under suitable regularity conditions (that is, the MLE exist and is unique), the distribution of $-2\log\lambda$ is asymptotically Chi-square distribution. The regularity conditions are all verified and hence the decision of unit root in the model can be made by comparing the likelihood ratio test statistic

$$-2\log\lambda = -2\sum_{t=1}^{n}\left[x_{t-1}(1-\hat{\phi}) + 2\log\left(\frac{1+e^{-(x_t-\hat{\phi}x_{t-1})}}{1+e^{-(x_t-x_{t-1})}}\right)\right] \tag{3.8}$$

with the corresponding Chi-squared table value at a given level of significance. Accordingly, reject H_0 if $-2\log\lambda > \chi^2_\alpha$.

3.5 Test for Cointegration in an ECM

Modern economic theory often suggests that certain pairs of financial or economic variables should be linked by some long run economic relationship. One of the primary interest concerned with such variables is that to test whether the set of variables are cointegrated. There are several test procedures available for cointegration when the disturbances in vector error correction model are i.i.d Gaussian and some authors have examined the performance of these tests by comparing the sizes and powers of the tests in which the model assumptions are violated (see for example, Kosapattarapim et al. (2013)). If $\phi \to 1$ in the cointegrating equation, then the series will be a random walk and therefore model cannot explain any long run behaviour in the observed series. Hence it is necessary to assure that the variables are all integrated with same order and are non stationary before we test for the presence of cointegration. Then the idea of testing the presence of unit root in the auto regression equation (3.6) can be extended to test the presence of cointegration using a similar approach that considered in Engle & Granger (1987). That is, once the series are identified to be unit root nonstationary with same order of integration, we can extend the test procedure of unit root for testing the presence of cointegration to the residuals of the fitted error correction model. The null hypothesis of unit root $\phi = 1$ can then be identically equal to testing $\delta = 0$ in the error correction model. Note that, unlike the usual cointegration test that applied to the residuals of the cointegrating regression, here we apply the test for the residuals from error correction model. So to test for cointegration, the null hypothesis that has to be taken is no cointegration. That is, $H_0 : \delta = 0$ against the alternative hypothesis of $H_1 : \delta \neq 0$. Once the model parameters are estimated from the data, we test the residuals from the error correction model using the test procedure described below. If the residuals are stationary, (ie; the null hypothesis of no cointegration is rejected) then we can conclude that the variables are cointegrated.

Assuming that the innovations of the ECM (2.26) and (2.27) follow iid logistic distribution, the residuals based on conditional MLE are denoted by,

$\hat{\eta}_{1t} = \Delta x_{1t} - \hat{\beta}\hat{\delta}\hat{z}_{t-1}$ and $\hat{\eta}_{2t} = \Delta x_{2t} + \hat{\delta}\hat{z}_{t-1}$, where $\hat{z}_{t-1} = x_{1,t-1} + \hat{\alpha} x_{2,t-1}$, $\hat{\alpha}$ and $\hat{\beta}$, $\hat{\delta}$ are the estimates of α, δ and β, respectively.

For the model (2.26), under the null hypothesis, the maximum value of the likelihood function is

$L_0 = e^{-\sum\limits_{t=1}^{n} \Delta x_{1t}} \prod\limits_{t=1}^{n} (1 + e^{-\Delta x_{1t}})^{-2}$ and under the alternative hypothesis, the maximum value of the likelihood function is, $L_1 = e^{-\sum\limits_{t=1}^{n} (\Delta x_{1t} - \hat{\beta}\hat{\delta}\hat{z}_{t-1})} \prod\limits_{t=1}^{n} (1 + e^{-(\Delta x_{1t} - \hat{\beta}\hat{\delta}\hat{z}_{t-1})})^{-2}$.

$$
\begin{aligned}
\frac{L_0}{L_1} &= \frac{e^{-\sum\limits_{t=1}^{n} \Delta x_{1t}} \prod\limits_{t=1}^{n} \left(1 + e^{-\Delta x_{1t}}\right)^{-2}}{e^{-\sum\limits_{t=1}^{n} (\Delta x_{1t} - \hat{\beta}\hat{\delta}\hat{z}_{t-1})} \prod\limits_{t=1}^{n} \left(1 + e^{-(\Delta x_{1t} - \hat{\beta}\hat{\delta}\hat{z}_{t-1})}\right)^{-2}} \\
&= e^{-\sum\limits_{t=1}^{n} \hat{\beta}\hat{\delta}\hat{z}_{t-1}} \prod_{t=1}^{n} \frac{\left(1 + e^{-(\Delta x_{1t} - \hat{\beta}\hat{\delta}\hat{z}_{t-1})}\right)^2}{\left(1 + e^{-\Delta x_{1t}}\right)^2}.
\end{aligned}
$$

Hence the log of likelihood ratio is given by,

$$
-2log\left(\frac{L_0}{L_1}\right) = -2\left[-\sum_{t=1}^{n} \hat{\beta}\hat{\delta}\hat{z}_{t-1} + 2\sum_{t=1}^{n} log\frac{(1 + e^{-(\Delta x_{1t} - \hat{\beta}\hat{\delta}\hat{z}_{t-1})})}{(1 + e^{-\Delta x_{1t}})}\right].
$$

For the model (2.27), under the null, the maximum of the likelihood function is $L_0 = e^{-\sum\limits_{t=1}^{n} \Delta x_{2t}} \prod\limits_{t=1}^{n} (1+e^{-\Delta x_{2t}})^{-2}$ and under the alternative, maximum of the likelihood function is $L_1 = e^{-\sum\limits_{t=1}^{n} (\Delta x_{2t}+\hat{\delta}\hat{z}_{t-1})} \prod\limits_{t=1}^{n} (1+e^{-(\Delta x_{2t}+\hat{\delta}\hat{z}_{t-1}})^{-2}$.

Thus,

$$\frac{L_0}{L_1} = \frac{e^{-\sum\limits_{t=1}^{n} \Delta x_{2t}} \prod\limits_{t=1}^{n} \left(1+e^{-\Delta x_{2t}}\right)^{-2}}{e^{-\sum\limits_{t=1}^{n} \left(\Delta x_{2t}+\hat{\delta}\hat{z}_{t-1}\right)} \prod\limits_{t=1}^{n} \left(1+e^{-(\Delta x_{2t}+\hat{\delta}\hat{z}_{t-1}}\right)^{-2}}$$
$$= e^{-\sum\limits_{t=1}^{n} \hat{\delta}\hat{z}_{t-1}} \prod_{t=1}^{n} \frac{\left(1+e^{-(\Delta x_{2t}+\hat{\delta}\hat{z}_{t-1}}\right)^{2}}{(1+e^{-\Delta x_{2t}})^{2}}.$$

Hence the likelihood ratio test statistic is given by,

$$-2log\left(\frac{L_0}{L_1}\right) = -2\left[\sum_{t=1}^{n}\hat{\delta}_1 z_{t-1} + 2\sum_{t=1}^{n}\log\frac{(1+e^{-(\Delta x_{2t}+\hat{\delta}_1\hat{z}_{t-1})})}{(1+e^{-(\Delta x_{2t})})}\right].$$

We reject the null hypothesis of no cointegration if the likelihood ratio test statistic

$$-2\left[\sum_{t=1}^{n}\hat{\delta}_1 z_{t-1} + 2\sum_{t=1}^{n}\log\frac{(1+e^{-(\Delta x_{2t}+\hat{\delta}_1\hat{z}_{t-1})})}{(1+e^{-(\Delta x_{2t})})}\right] \tag{3.9}$$

or

$$2\left[-\sum_{t=1}^{n}\hat{\beta}\hat{\delta}\hat{z}_{t-1} + 2\sum_{t=1}^{n} log\frac{(1+e^{-(\Delta x_{1t}-\hat{\beta}\hat{\delta}\hat{z}_{t-1})})}{(1+e^{-\Delta x_{1t}})}\right] \tag{3.10}$$

is too large or too small based on Chi-square critical value. As our study deals with two time series, we have two error correction models that represent the cointegrating relationship. Though both the ECM have a unique representation for the long run cointegrating relationship, which is represented by the term z_t, the null hypothesis of no cointegration will be rejected if the above test statistic exceeds the Chi-square critical value. If we can reject the null hypothesis of no cointegration, then there exist a long run and a short run relationship between the variables X_{1t} and X_{2t}.

3.6 Simulation Study

As the estimating equations (3.3), (3.4), (3.5) and (3.7) do not admit explicit solutions, we analyse the performance of the above methods by simulation. Hence we carry out a simulation study to understand the performance of the estimator and test statistic described in Sections 3.3 to 3.5 for various sample sizes and for different specified values of the model parameters. We used the method of Newton Raphson to perform the numerical calculations in the simulation study. Simulation experiments are conducted with the help of the software "Mathematica".

3.6.1 Simulation result for logistic errors

For the simulation purpose, we first generate the innovation random variable from a logistic distribution. Then for specified values of the model parameter, we simulated the sequence $\{x_t\}$, t=1,2,...,n using the relation described in (3.6). Based on this sample, we obtain the maximum likelihood estimates of ϕ by solving the score function given in Section 3.4. We used sample autocorrelation as the initial estimate while solving the log likelihood equations numerically. For the given values of the model parameter, we repeated the experiment 100 times for computing the estimates and then averaged them over the repetitions. Next we compute the likelihood ratio test statistic given in equation (3.8) for various sample sizes and for different parameter values. Finally we compute the number of rejections in 500 trials for testing the null hypothesis of interest. The numerical computations are carried out for various value of the model parameter and are summarised in Tables 3.1 and 3.2. Next to evaluate the accuracy of

TABLE 3.1: The average estimates and the corresponding root mean squares errors (RMSE) of the MLE

Sample size	True value ϕ	MLE $\hat{\phi}$	RMSE
100	-0.8	-0.7766	0.0706
	-0.5	-0.5006	0.0883
	-0.3	-0.2951	0.0874
	0.2	0.1797	0.0977
	0.3	0.2781	0.0917
	0.6	0.5699	0.0829
	0.8	0.8832	0.0708
300	-0.8	-0.7921	0.0354
	-0.5	-0.5007	0.0456
	-0.3	-0.2874	0.0547
	0.2	0.2006	0.0492
	0.3	0.3008	0.0571
	0.6	0.6007	0.0425
	0.8	0.7944	0.0341
500	-0.8	-0.7949	0.0477
	-0.5	-0.4995	0.0375
	-0.3	-0.2989	0.0446
	0.2	0.1967	0.0394
	0.3	0.3031	0.0404
	0.6	0.5977	0.0357
	0.8	0.7929	0.0271

the estimation and testing procedure of the error correction model, a simulation study is carried out for different sample sizes and for different values of the model parameters. For the study, we generate the error correction model using (2.26) and (2.27). Then we obtained the MLE of the parameters by solving the likelihood equations in (3.3), (3.4) and (3.5). We then repeated the experiment 50

TABLE 3.2: No of rejections in 500 trials of the hypothesis $H_0 : \phi = 1$ against $H_1 : |\phi| < 1$ using the test statistic given in (3.8) for different values of ϕ

Sample size	50				100				250				350			
	0.01	0.05	0.1	0.2	0.01	0.05	0.1	0.2	0.01	0.05	0.1	0.2	0.01	0.05	0.1	0.2
ϕ =-.50	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500
ϕ =-.20	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500
$\phi = .80$	218	327	411	460	466	493	499	500	500	500	500	500	500	500	500	500
$\phi = .85$	122	216	312	389	376	448	482	496	500	500	500	500	500	500	500	500
$\phi = .90$	59	133	197	283	197	311	399	464	500	500	500	500	500	500	500	500
$\phi = .95$	56	103	167	273	53	108	167	277	285	388	439	480	482	496	500	500

times for computing the estimates and then averaged them over the repetitions. After the parameters of ECM being estimated, we test for cointegration using the residuals from the error correction model. We use the test statistic given in (3.9) and (3.10) to compute the number of rejections of the null hypothesis under the various alternatives. In practical situations, we could reject the null hypothesis of no cointegration based on either of the two test statistics. The numerical computations for estimation and testing are carried out for various values of the model parameters and are summarised in Tables 3.3, 3.4 and 3.5. Note that from Table 3.1 and 3.3, for series of length 100 and 300, estimates are reasonably satisfactory and become more accurate with increasing sample size. From Table 3.2, it is evident that as ϕ becomes closer to 1, the number of rejections of the null hypothesis of unit root becomes smaller. For example, in a length of 50 series, the hypothesis $H_0 : \phi = 0.95$ was rejected 56 times at the 0.01 significance level, while it was rejected 218 times when ϕ was 0.8. Hence we claim that the derived test statistic is powerful for testing the presence of unit root in an observed nonstationary time series. From Tables 3.4 and 3.5, it is seen that for large values of ϕ or as ϕ increases to 1, the number of rejections of the null hypothesis in 500 trial decreases.

Summary of the Algorithms used for the simulation results are given below.

Table 3.3: The average estimates and the corresponding root mean squared errors of MLE

Sample Size	True values			MLE		
	δ	β	α	$\hat{\delta}$	$\hat{\beta}$	$\hat{\alpha}$
300	2.6	1.5	1.8	2.7080(0.3125)	1.4965(0.0156)	1.8016(0.0049)
	0.5	2	3	0.5080(0.0851)	1.9600(0.2108)	3.0096(0.0258)
	0.3	2.5	3.5	0.3178(0.0683)	2.4966(0.1028)	3.5290(0.0547)
	0.2	3	4	0.1863(0.0588)	2.9448(0.1667)	4.0270(0.0837)
	0.1	3	4	0.1018(0.0429)	2.7962(0.1550)	4.0290(0.2978)
500	2.6	1.5	1.8	2.6777(0.2319)	1.5009(0.0108)	1.8040(0.0275)
	0.5	2	3	0.5133(0.0624)	1.9915(0.1727)	3.0351(0.0161)
	0.3	2.5	3.5	0.3107(0.0537)	2.4892(0.0507)	3.5116(0.0269)
	0.2	3	4	0.2037(0.0488)	3.0027(0.0643)	4.0179(0.0451)
	0.1	3	4	0.1048(0.0390)	2.9740(0.1498)	4.0560(0.1380)
700	2.6	1.5	1.8	2.6651(0.1182)	1.5003(0.0068)	1.8006(0.0019)
	0.5	2	3	0.5075(0.0522)	2.0208(0.1477)	3.0244(0.0113)
	0.3	2.5	3.5	0.3035(0.0391)	2.4976(0.0418)	3.5072(0.0194)
	0.2	3	4	0.1904(0.0379)	2.9789(0.0619)	4.0079(0.0421)
	0.1	3	4	0.1047(0.0260)	3.0078(0.0948)	4.0182(0.0653)

Table 3.4: No of rejections in 500 trials of the hypothesis $H_0 : \phi = 1$ (δ=0) against the alternative of $|\phi| < 1$($\delta \neq 0$) using the first ECM

ECM-1								
sample size	50				100			
	0.01	0.05	0.1	0.2	0.01	0.05	0.1	0.2
ϕ=0.5	500	500	500	500	500	500	500	500
ϕ=0.8	450	462	472	475	486	489	491	492
ϕ=0.9	413	429	441	452	470	475	480	492
ϕ=.95	220	250	270	285	295	300	325	347

Table 3.5: No of rejections in 500 trials of the hypothesis $H_0 : \phi = 1$ (δ=0) against the alternative of $\phi < 1$($\delta \neq 0$) using the second ECM

ECM-2								
Sample size	50				100			
	0.01	0.05	0.1	0.2	0.01	0.05	0.1	0.2
ϕ=0.5	500	500	500	500	500	500	500	500
ϕ=0.8	477	488	489	491	484	490	491	496
ϕ=0.9	453	465	477	483	476	480	484	490
ϕ=.95	259	270	280	280	370	320	340	363

Algorithm 1: Algorithm for Table 3.1

1. Set x_0 and ϕ, for some sample of size n.
2. Draw samples $\{a_t\}$, from logistic distribution.
3. Generate x_t using $x_t = \phi x_{t-1} + a_t, t = 1, 2, ...n$
4. Choose the initial value of ϕ as ϕ_0 and obtain $\hat{\phi}$, using the method of maximum likelihood.
5. Repeat Steps 2 to 4, say 100 times.
6. Choose the value of $\hat{\phi}$ as the averages of $\hat{\phi}$ obtained in step 4.

Algorithm 2: Algorithm for Table 3.2

1. Set x_0 and ϕ, for some sample of size n.
2. Draw samples $\{a_t\}$, from logistic distribution.
3. Generate x_t using $x_t = \phi x_{t-1} + a_t, t = 1, 2, ...n$
4. Choose the initial value of ϕ as ϕ_0 and obtain $\hat{\phi}$, using the method of maximum likelihood.
5. Obtain the LRT in Equation (3.8) for the simulated data.
6. Repeat Steps 2 to 5, say 500 times.
7. Count and record the number of rejections of unit root hypothesis in 500 trials.

Algorithm 3: Algorithm for Table 3.3

1. Set x_{10}, x_{20}, α, β and δ, for some sample of size n.
2. Draw samples (η_{1t}, η_{1t}) using Logistic distribution.
3. Draw samples from the ECM in Equations (2.26) and (2.27).
4. Choose the initial value as α_0, β_0 and δ_0. Solve Equations (3.3) to (3.5) for obtaining the maximum likelihood estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\delta}$.
5. Repeat Steps 2 to 4 for 50 times.
6. Set the values of $\hat{\alpha}, \hat{\beta}, \hat{\delta}$ as the averages of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\delta}$ obtained in step 4.

Algorithm 4: Algorithm for Tables 3.4 and 3.5

1. Set x_{10}, x_{20}, α, β and δ, for some sample of size n.
2. Draw samples (η_{1t}, η_{1t}) using Logistic distribution.
3. Draw samples from the ECM in Equations (2.26) and (2.27).
4. Choose the initial value as α_0, β_0 and δ_0. Solve Equations (3.3) to (3.5) for obtaining the maximum likelihood estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\delta}$.
5. Obtain the LRT in Equations (3.9) and (3.10) for the simulated data.
6. Repeat Steps 2 to 5 for 500 times.
7. Count and record the number of rejections of the hypothesis $H_0 : \phi = 1 (\delta = 0)$ against the alternative of $|\phi| < 1 (\delta \neq 0)$ in 500 trials.

3.7 Bootstrap Method

In small sample situations, the asymptotic likelihood ratio test discussed in the earlier sections may not be suitable for determining the cointegrating relationship between two or more time series. The theoretical chi-square distribution for likelihood ratio test will provide much better results if the sample size is reasonably large. Hence for finite sample situations, we can use a parametric bootstrap approach in which we constructs the distribution of the likelihood ratio test statistic empirically. So we provide a Monte Carlo simulation to compare the performance of bootstrap testing with the usual method based on an asymptotic approximation of the distribution of the test statistic.

In this section we address the accuracy of a bootstrap algorithm in small samples for testing the presence of cointegration in an ECM. In recent years, there has been an increasing interest in parametric and non parametric bootstrap inference for econometric and financial time series. The technique of parametric bootstrap suggest estimation of the sampling distribution of the statistic using random sampling methods and it may also be used for constructing tests of hypothesis. Here we provide a simulation based parametric bootstrap method that involves simulating data sets using the maximum likelihood estimates and hence computing the likelihood ratio test statistic for each available simulated data set. The method involves 4 steps.

As a starting point, we estimate the parameters of the cointegration model using the conditional maximum likelihood estimation method and then obtain the asymptotic likelihood ratio test statistic for the real data. Secondly, we generate a bootstrap sample using the maximum likelihood estimates and then compute the likelihood ratio test statistic for the bootstrap sample. Thirdly, we repeat the above step 10000 times which yield an estimate of the distribution of the likelihood ratio test statistic. Finally, we compute the empirical quantiles of the test statistic and then take the decision on the null hypothesis of no cointegration by comparing the calculated critical values with the calculated likelihood ratio test value.

A summary of the algorithm used for bootstrap sampling is given below.

Algorithm 5: Parametric Bootstrap

1 Assuming data sets $x_1 = (x_{11}, x_{12}, \cdots\cdots, x_{1n})$, $x_2 = (x_{21}, x_{22}, \cdots\cdots, x_{2n})$ are available.

2 Assume that the innovation random variable's (η_{1t}, η_{2t}) comes from a iid logistic distribution given by the pdf

$$f_\theta(\eta_{1t}, \eta_{2t}) = \frac{e^{-[(\Delta x_{1t} - \delta\beta z_{t-1}) + (\Delta x_{2t} + \delta z_{t-1})]}}{\left(1 + e^{-(\Delta x_{1t} - \delta\beta z_{t-1})}\right)^2 \left(1 + e^{-(\Delta x_{2t} + \delta z_{t-1})}\right)^2},$$

described by a set of parameters $\theta = (\alpha, \beta, \delta)$.

3 Estimate θ by using maximum likelihood, obtaining the estimate $\hat{\theta}$.

4 Obtain the LRT, $-2log\left(\frac{L_0}{L_1}\right)$ in Equation (3.9) and (3.10) for the data.

5 Fix the number of bootstrap re-samples, N=10000.

6 Draw bootstrap data x_1^*, x_2^* set of size n from $f_{\hat{\theta}}(\eta_{1t}, \eta_{2t})$.

7 Estimate θ from x_1^*, x_2^*. Call the estimate $\hat{\theta}^*$.

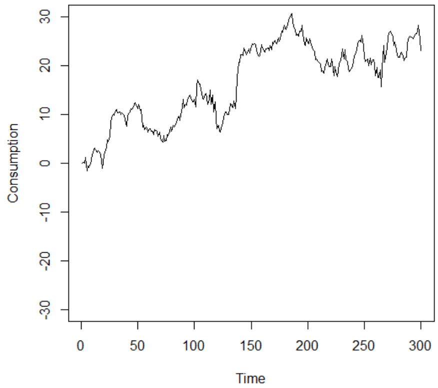
8 Obtain the LRT, $-2log\left(\frac{L_0}{L_1}\right)$ for the simulated data.

9 Repeat steps 6 to 8, 10000 times. Compute the empirical quantiles of the test statistic.

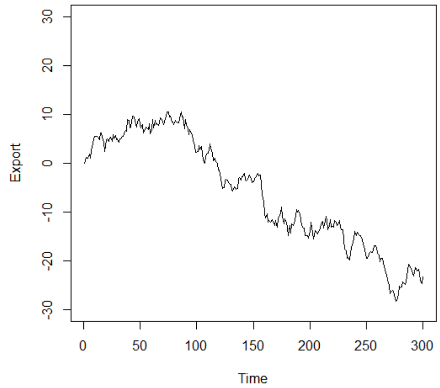
We analyse a real data set for testing the presence of cointegration by our proposed model in section 3.8.

3.8 Data Analysis

In this section, we illustrate the analysis of cointegration with logistic innovations using the real data set. The data set consists of monthly observations on consumption and export of natural rubber (Tonnes) collected from "The Rubber Board", Ministry of Commerce and Industry, Govt. of India, Kottayam. The Figure 3.1 provides the time series plot of the log transformed data and they indicates that the time series are nonstationary.



(A) Time plot of consumption series



(B) Time plot of Export series

FIGURE 3.1: Time Series plots

First we tested the data for cointegration with normally distributed errors using the Johansen test for cointegration. For that, we have tested whether the time series are non stationary or not by using the Augmented Dickey Fuller test. Both the p-values obtained for ADF test are 0.112 and 0.213, implying that the variables are non stationary. In Chapter 2, we have seen that, Johansen's trace test tests the null hypothesis of m cointegrating vectors against the alternative hypothesis of n(>m) cointegrating vectors. If m=0, it means that there is no relationship among the variables, that is stationary. The maximum eigen value test tests the null hypothesis of m cointegrating vectors against the alternative of $(m+1)$ cointegrating vectors. Table 3.6 and 3.7 show the Johansen test of cointegration for normally distributed errors. The values given in brackets are the table values corresponding to 10%, 5% and 1% level of significance.

Table 3.6: Johansen Trace test

	test	10%	5%	1%
m=1	4.9	(7.52)	(9.24)	(12.97)
m=0	32.29	(17.85)	(19.96)	(24.6)

From the tables, it can be seen that in both cases the null hypothesis of one cointegrating vector is not rejected. This implies that, cointegration exist between the rubber consumption and export series.

Table 3.7: Johansen Eigen Value test

	test value	10%	5%	1%
m=1	4.9	(7.52)	(9.24)	(12.97)
m=0	27.39	(13.75)	(15.67)	(20.2)

The parameter estimates are obtained as $\hat{\alpha} = -1.33$, $\hat{\beta} = -0.0059$ and $\hat{\delta} = -0.203$. $x_{1t} - 1.33x_{2t}$ is the estimated cointegrating relationship using the Johansen test. Finally to evaluate the adequacy of the model using normal errors, we checked whether the residual series obtained from the fitted model follows normal distribution. But the assumption of normality is rejected for the residual series, hence we tested for cointegration with errors generated by logistic innovations. Although the plot seems to be nonstationary, it is important to test whether a series is stationary or not before we test for cointegration. Hence we performed a unit root test developed for logistic error variables to the data set in order to test whether the series is stationary or not.

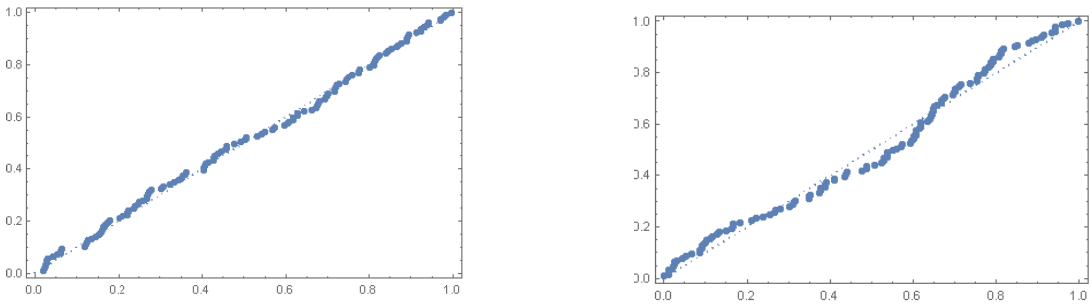
The p-values obtained for testing the unit root for consumption and export series are obtained as 0.9928 and 0.9723 respectively. Since both the p-values obtained are very large, we do no reject the null hypothesis of unit root and hence the series meets cointegration test condition. Next we carry out a maximum likelihood estimation as described in Section 3.3 in order to find the parameter estimates of an error correction model of order 1. The parameter estimates are obtained as $\hat{\alpha} = -1.43$, $\hat{\beta} = -0.0158$ and $\hat{\delta} = -0.1542$. Thus the estimated cointegrating relationship, is $x_{1t} - 1.43x_{2t}$, where $\{x_{1t}\}$ is the month-wise rubber consumption series and $\{x_{2t}\}$ is the month-wise rubber export series. The residuals from the error correction model are obtained as

$$\xi_{1t} = \Delta x_{1t} - \hat{\delta}\hat{\beta}\hat{z}_{t-1}$$

$$\xi_{2t} = \Delta x_{2t} + \hat{\delta}\hat{\beta}\hat{z}_{t-1}, t = 1, 2, 3, ...$$

Using the parameter estimates of the ECM, we tested whether the residuals follow a logistic distribution using Kolmogorov-Smirnov test. The p-values obtained for the series are 0.965 and 0.303 respectively, which indicate that logistic distribution is suitable for the residuals. The probability-probability plots of the residuals are shown in Figure 3.2 also confirm the result.

Finally we performed a bootstrap algorithm for testing cointegration using the error correction model with logistic errors. The bootstrap values and asymptotic



(A) PP- plot of residuals of Consumption series

(B) PP- plot of residuals of Export series

FIGURE 3.2: Monthly Rubber data

values of the two sided hypothesis, $H_0 : \delta = 0$ against $H_1 : \delta \neq 0$ at a given level of significance are given in Table 3.8.

TABLE 3.8: Empirical levels for bootstrap and asymptotic tests for the two sided testing of hypothesis $H_0 : \delta = 0$ against $H_1 : \delta \neq 0$.

Nominal Level	0.2		0.1		0.05	
Bootstrap values	0.024	2.719	0.0005	3.849	-0.095	5.146
Asymptotic values	0.016	2.710	0.0039	3.840	0.009	5.020

The value of the test statistic obtained for the error correction equation of import series is 0.000219. Thus from Table 3.8 we can conclude that we reject the null hypothesis of no cointegration at 10 percent level of significance using a bootstrap test and asymptotic test, implying that the residuals from the ECM are stationary. If the null hypothesis of no cointegration is rejected, then the cointegrating vector parameter estimate provides an estimate of a long run relationship. That is, $x_{1t} - 1.43x_{2t}$ is the cointegrating relationship and the cointegrated vector is $[1, -1.43]'$.

Thus in both situations, that is with normal and logistic errors, the existence of cointegration relationship in the data is identified. But the residual series

obtained from the cointegrating regression using normal errors rejects the assumption of normality. Hence we proceed with the vector autoregression model that allows for logistic innovations to arrive at a right conclusion.
The results of this Chapter are published in Nimitha & Balakrishna (2018b).

Chapter 4

Modelling of Cointegration with bivariate Student t innovation

4.1 Introduction

Our focus in the previous Chapter was modelling of cointegration in the presence of logistic errors. In this Chapter, we focus on the cointegration model when the errors are generated by a bivariate Student t distribution. The Student t distribution can be a useful theoretical tool in the area of applied statistics. The work by Boswijk et al. (1999) considers Student's t distribution with 3 degrees of freedom, a truncated Cauchy distribution, Gaussian mixtures and others. Tiku et al. (2000) discusses an autoregressive models in time series with non normal innovations represented by a member of a wide family of symmetric Student t distributions. Creal et al. (2011) introduced the multivariate Student t generalised autoregressive score model for volatilities and correlations, in which the multivariate normal distribution is a special case. Here, we propose the modelling of two cointegrating time series when the innovations of its ECM are generated from a bivariate student t distribution.

The rest of the chapter is organised as follows. Section 4.2 briefly discusses the cointegration model with bivariate Student t errors. We discuss the conditional maximum likelihood estimation procedure in Section 4.3. Section 4.4 deals with the problems of testing of hypothesis on unit root and cointegration. To evaluate the accuracy of the estimators and test statistic, we carried out a simulation study in Section 4.5. Finally, an application of the proposed model is illustrated in Section 4.6.

4.2 Model Description

Let us assume that $\{X_{1t}\}$ and $\{X_{2t}\}$ be two non stationary cointegrating time series. We use the error correction representation, (2.26) and (2.27) given in Chapter 2 for defining the cointegrating model with Student t distributed errors. The error correction representation with lag 1 is given by :

$$\Delta X_{1t} = \delta\beta Z_{t-1} + \eta_{1t} \tag{4.1}$$

$$\Delta X_{2t} = -\delta Z_{t-1} + \eta_{2t}, \tag{4.2}$$

where $Z_{t-1} = X_{1,t-1} + \alpha X_{2,t-1}$.

Assume that the error variables $\eta_t = (\eta_{1t}, \eta_{2t})^{'}$ in (4.1) and (4.2) follow a bivariate Student t distribution with the probability density function (pdf) given by,

$$f(\eta_{1t}, \eta_{2t}) = \frac{1}{2\pi\sqrt{1-\rho^2}} \left[1 + \frac{1}{v(1-\rho^2)}\left(\eta_{1t}^2 - 2\rho\eta_{1t}\eta_{2t} + \eta_{2t}^2\right)\right]^{-\frac{(\nu+2)}{2}},$$

where $\nu > 0, -1 < \rho < 1, x, y > 0$.

It should be noted that both the marginal distributions are Student t with same degrees of freedom ν (see Balakrishnan & Lai (2009)) with the pdf

$$f\left(\eta_{it}\right)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{\left(\eta_{it}\right)^{2}}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)},-\infty<\eta_{it}<\infty,i=1,2.$$

The above bivariate distribution can be obtained by the transformation,

$$\eta=\left(\frac{\sqrt{s}}{\nu}\right)^{-1}Z+\mu, \tag{4.3}$$

where $\eta=(\eta_{1t},\eta_{1t})$, Z is a bivariate normal random variable with mean **0** and dispersion matrix $\sum=\begin{pmatrix}1&\rho\\\rho&1\end{pmatrix}$, $s\sim\chi^2_{\nu}$. The case with ν=1 reduces to a bivariate Cauchy distribution.

4.3 Conditional MLE for ECM with bivariate student t errors

In this section, we focus on the problem of maximum likelihood estimation of cointegrated model with bivariate Student t distributed errors. When there exist an explicit density function for the innovation random variables, one can use the conditional maximum likelihood estimation for estimating the model parameters. Let the parameter vector to be estimated are the elements of $\theta=(\rho,\alpha,\beta,\delta,\nu)^{'}$. The explicit form of the conditional log likelihood function for $\theta=(\rho,\alpha,\beta,\delta,\nu)^{'}$, given $(x_{10},x_{20})=(0,0)$ based on sample $(x_{1t},x_{2t}),t=1,2,\cdots$

of size n is given by ,

$$
\begin{aligned}
L_T(\theta) = & -n \log\left(1-\rho^2\right)^{1/2} \\
& - \frac{(\nu+2)}{2}\sum_{t=1}^{n}\log\left(1+\frac{1}{\nu\left(1-\rho^2\right)}\left[\left(\eta_{1t}^2-2\rho\eta_{1t}\eta_{2t}+\eta_{2t}^2\right)\right]\right),
\end{aligned}
$$

where (x_{1t}, x_{2t}) are obtained from the equations (4.1) and (4.2). The form of the log likelihood function suggest that we have to maximise the log likelihood function by using some numerical methods. Here we obtain the parameter estimates by using a two step estimation procedure. First we obtain the estimates of $\rho, \alpha, \beta, \delta$ by solving the respective score equations. Then we use the profile likelihood technique for estimating the parameter ν. That is, the estimate of ν is obtained by maximizing the log likelihood function using the maximum likelihood estimates of $\rho, \alpha, \beta, \delta$. The discussion on profile likelihood and its asymptotic properties can be seen in Murphy & Van der Vaart (2000).

On differentiating the log-likelihood function with respect to the parameters $\rho, \alpha, \beta, \delta$, we will get four equations given by,

$$
\frac{\partial \log l}{\partial \rho} = 0 \Rightarrow n\frac{\rho}{(1-\rho^2)} - \sum_{t=1}^{n}\frac{(\nu+2)\left[\rho\frac{\left(\eta_{1t}^2-2\rho\eta_{1t}\eta_{2t}+\eta_{2t}^2\right)}{\nu(1-\rho^2)^2} - \frac{\eta_{1t}\eta_{2t}}{\nu(1-\rho^2)}\right]}{\left(1+\frac{1}{\nu(1-\rho^2)}\left[\left(\eta_{1t}^2-2\rho\eta_{1t}\eta_{2t}+\eta_{2t}^2\right)\right]\right)} = 0. \quad (4.4)
$$

$$
\frac{\partial \log l}{\partial \delta} = 0 \Rightarrow \sum_{t=1}^{n}\frac{(\nu+2)\left[2z_{t-1}A(\theta)(\beta+\rho) - 2z_{t-1}B(\theta)(1+\beta\rho)\right]}{2\nu(1-\rho^2)\left[1+\frac{(A(\theta))^2+(B(\theta))-2\rho A(\theta)B(\theta)}{\nu(1-\hat{\rho}^2)}\right]} = 0. \quad (4.5)
$$

$$
\frac{\partial \log l}{\partial \beta} = 0 \Rightarrow \sum_{t=1}^{n}\frac{(\nu+2)\left[2z_{t-1}\delta A(\theta) - 2z_{t-1}\delta\rho B(\theta)\right]}{2\nu(1-\rho^2)\left[1+\frac{(A(\theta))^2+(B(\theta))-2\rho A(\theta)B(\theta)}{\nu(1-\rho^2)}\right]} = 0. \quad (4.6)
$$

$$\frac{\partial \log l}{\partial \alpha} = 0 \Rightarrow \sum_{t=1}^{n} \frac{(\nu+2)\left[2x_{2t-1}\delta(\beta+\rho)A(\theta) + 2x_{2t-1}\delta(1+\beta\rho)B(\theta)\right]}{2\nu(1-\rho^2)\left[1+\frac{(A(\theta))^2+(B(\theta))-2\rho A(\theta)B(\theta)}{\nu(1-\rho^2)}\right]} = 0, \tag{4.7}$$

where $A(\theta) = \Delta x_{1t} - \delta\beta z_{t-1}, B(\theta) = \Delta x_{2t} + \delta z_{t-1}$. Since there does not exist analytically closed form expressions for the estimators, we have to maximize the score functions by using some numerical methods. We have solved these equations by using Newton Raphson method for obtaining the parameter estimates. Let the estimates of ρ, α ,β ,δ be $\hat{\rho}$, $\hat{\alpha}$, $\hat{\beta}$, $\hat{\delta}$ respectively. Once the model parameters are obtained, the profile likelihood function $L\left(\nu, \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\rho}\right)$ is maximized over ν to get $\hat{\nu}$. The profile log likelihood function is given by,

$$\begin{aligned} L_T(\theta) = &- n \log{(1-\hat{\rho}^2)}^{\frac{1}{2}} \\ &- \frac{(\nu+2)}{2}\sum_{t=1}^{n} \log\left(1+\frac{1}{\nu(1-\hat{\rho}^2)}\left(A(\hat{\theta})^2 - 2\rho A(\hat{\theta})B(\hat{\theta}) + B(\hat{\theta})^2\right)\right), \end{aligned} \tag{4.8}$$

where $A(\hat{\theta}) = \Delta x_{1t} - \hat{\delta}\hat{\beta}\hat{z}_{t-1}, B(\hat{\theta}) = \Delta x_{2t} + \hat{\delta}\hat{z}_{t-1}$. Solution of the above equation is obtained numerically, which yields the estimates of the parameter ν.

The properties of first order autoregressive models with students t distributed errors is needed to develop a test procedure for cointegration model in (4.1) and (4.2), which we discuss in Section 4.4.

4.4 Unit root test for AR(1) model with Student t errors

Here, our interest is to analyse the first order autoregressive equation in the presence of Students t errors. We consider the first order autoregressive process $\{X_t\}$ with the errors $\{a_t\}$ is a sequence of Student t random variables.
i.e;

$$X_t = \phi X_{t-1} + a_t, \tag{4.9}$$

where $X_0 = 0$. Our interest is to find the maximum likelihood estimator of ϕ, say $\hat{\phi}$ and the test procedure of unit root under the hypothesis $H_0 : \phi = 1$ against $H_1 : |\phi| < 1$. Suppose n observations are available for the analysis and we shall obtain the likelihood function based on n observations generated by the model (4.9). The joint density function of $(a_1, a_2 \cdots\cdots, a_n)$ is given by,

$$\prod_{t=1}^{n} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{a_t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}.$$

For the model (4.9), the joint probability density function of $(x_1, x_2, \cdots, x_n|x_0)$ is

$$\prod_{t=1}^{n} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(x_t - \phi x_{t-1})^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}.$$

And the conditional likelihood function is,

$$L(\phi \mid x_0) = \left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\right)^n \prod_{t=1}^{n} \left(1 + \frac{(x_t - \phi x_{t-1})^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}.$$

The corresponding log-likelihood function is,

$$L_T(\phi \mid x_0) = -\left(\frac{\nu+1}{2}\right)\sum_{t=1}^{n}\log\left(1+\frac{(x_t-\phi x_{t-1})^2}{\nu}\right) + n\log\Gamma\left(\frac{\nu+1}{2}\right) - n\log\sqrt{\nu\pi} - n\log\Gamma\left(\frac{\nu}{2}\right) \tag{4.10}$$

The critical points of the above log likelihood function can be obtained by solving the score equations given by,

$$\frac{\partial L(\phi \mid x_0)}{\partial \phi} = 0 \Rightarrow \left(\frac{\nu+1}{2}\right)\sum_{t=1}^{T}2x_{t-1}\frac{(x_t-\phi x_{t-1})}{\nu\left(1+\frac{(x_t-\phi x_{t-1})^2}{\nu}\right)} = 0 \tag{4.11}$$

The above equation can be solved by numerically in order to obtain the estimate of ϕ. We obtain the estimator of ϕ numerically by using the Newton Raphson method. Once the estimate of ϕ is obtained, then the profile log likelihood function

$$L_T(\hat{\phi} \mid x_0) = -\left(\frac{\nu+1}{2}\right)\sum_{t=1}^{n}\log\left(1+\frac{(x_t-\hat{\phi} x_{t-1})^2}{\nu}\right) + n\log\Gamma\left(\frac{\nu+1}{2}\right) - n\log\sqrt{\nu\pi} - n\log\Gamma\left(\frac{\nu}{2}\right) \tag{4.12}$$

is maximized over ν to get $\hat{\nu}$. We maximize the above profile log likelihood function by using numerical methods for obtaining the estimate of ν.

Now let us consider the hypothesis $H_0 : \phi = 1$ against the alternative $H_1 : |\phi| < 1$, that is the time series $\{X_t\}$ was generated by a stationary model. Under H_0, the maximum value of the likelihood function is

$$L_0 = \left(\frac{\Gamma\left(\frac{\hat{v}+1}{2}\right)}{\sqrt{\hat{v}\pi}\Gamma\left(\frac{\hat{v}}{2}\right)} \right)^n \prod_{t=1}^{n} \left(1 + \frac{(x_t - x_{t-1})^2}{\hat{v}} \right)^{-\left(\frac{\hat{v}+1}{2}\right)}$$

and under the alternative, the maximum value of likelihood function is,

$$L_1 = \left(\frac{\Gamma\left(\frac{\hat{v}+1}{2}\right)}{\sqrt{\hat{v}\pi}\Gamma\left(\frac{\hat{v}}{2}\right)} \right)^n \prod_{t=1}^{n} \left(1 + \frac{(x_t - \hat{\phi}x_{t-1})^2}{\hat{v}} \right)^{-\left(\frac{\hat{v}+1}{2}\right)}.$$

Hence, for $\hat{\phi} \in H_1$, the likelihood ratio test rejects H_0 when

$$\begin{aligned} -2\log\lambda &= -2\log\prod_{t=1}^{n} \left(\frac{1 + \frac{(x_t - \hat{\phi}x_{t-1})^2}{\hat{v}}}{1 + \frac{(x_t - x_{t-1})^2}{\hat{v}}} \right)^{\left(\frac{\hat{v}+1}{2}\right)} \\ &= -2\left(\frac{\hat{v}+1}{2}\right) \left(\sum_{t=1}^{n} Log\left(1 + \frac{(x_t - \hat{\phi}x_{t-1})^2}{\hat{v}}\right) - \sum_{t=1}^{n} \log\left(1 + \frac{(x_t - x_{t-1})^2}{\hat{v}}\right) \right). \end{aligned} \tag{4.13}$$

We reject H_0 if $-2Log\lambda$ is either too small or too large.

4.4.1 Test for cointegration

If $\phi \to 1$ in (4.9) , then the time series contains a unit root. Here one can extend the idea of testing the presence of unit root for testing the presence of cointegration using the approach discussed in Engle & Granger (1987). First, we have to confirm all the variables are integrated of same order and are non stationary in nature. Once the series are confirmed to be non stationary with same order of integration, we can test for the presence of cointegration by using the residuals from the fitted error correction model. Here the hypothesis of interest for testing the cointegration is, $H_0 : \delta = 0$ against $\delta \neq 0$. Once we obtain the parameter estimates using the data, we tests the residuals from the error correction model using the test procedure described below. We denote the residuals from the error correction model (4.1) and (4.2) by, $\hat{\eta}_{1t} = \Delta x_{1t} - \hat{\beta}\hat{\delta}\hat{z}_{t-1}$ and $\hat{\eta}_{2t} = \Delta x_{2t} + \hat{\delta}\hat{z}_{t-1}$, where $\hat{z}_{t-1} = x_{1,t-1} + \hat{\alpha}x_{2,t-1}$, $\hat{\alpha}$ and $\hat{\beta}$, $\hat{\delta}$ are the parameter estimates of α, δ and β . And, if the residuals obtained from the ECM are stationary, (ie; the null hypothesis of no cointegration is rejected) then we can conclude that the variables will be cointegrated.

Now we will obtain the likelihood ratio test statistic for testing the presence of cointegration in an error correction model. Under the null hypothesis, the maximum value of the likelihood function is,

$$L_0 = \prod_{t=1}^{n} \frac{1}{\sqrt{1-\hat{\rho}_0^2}} \left(1 + \frac{1}{\hat{v}(1-\hat{\rho}_0^2)} \left(\Delta x_{1t}^2 - 2\hat{\rho}_0 \Delta x_{1t} \Delta x_{2t} + \Delta x_{2t}^2\right)\right)^{-\frac{(\hat{v}+2)}{2}},$$

and under the alternative hypothesis, the maximum value of the likelihood function is,

$$L_1 = \prod_{t=1}^{n} \frac{1}{\sqrt{1-\hat{\rho}_1^2}} \left(\frac{1}{\hat{v}(1-\hat{\rho}_1^2)} \left(\left(A_1(\hat{\theta})\right)^2 - 2\hat{\rho}_1 A_1(\hat{\theta}) B_1(\hat{\theta}) + \left(B_1(\hat{\theta})\right)^2\right)\right)^{-\frac{(\hat{v}+2)}{2}}.$$

The likelihood ratio test statistic obtained is given by,

$$n\log\left[\frac{1-\hat{\rho}_0^2}{1-\hat{\rho}_1^2}\right] \\ -(\hat{v}+2)\sum_{t=1}^{n}\log\left(\frac{1+\frac{1}{\hat{v}(1-\hat{\rho}_1^2)}\left(\left(A_1(\hat{\theta})\right)^2-2\hat{\rho}_1A_1(\hat{\theta})B_1(\hat{\theta})+\left(B_1(\hat{\theta})\right)^2\right)}{1+\frac{1}{\hat{v}(1-\hat{\rho}_0^2)}\left(\Delta x_{1t}^2-2\hat{\rho}_0\Delta x_{1t}\Delta x_{2t}+\Delta x_{2t}^2\right)}\right). \tag{4.14}$$

Here we reject the null hypothesis of no cointegration if the likelihood ratio tests statistic is large in absolute.

To evaluate the performance of the estimators and test statistic, a simulation study is carried out, which is illustrated in Section 4.5.

4.5 Simulation study for t distributed errors

First we carry out a simulation study for the first order autoregressive model with Student t distributed errors. For the study, we generate the innovation random variable from a Student t distribution. Then for specified values of the model parameter, we simulated the sequence $\{x_t\}$, t=1,2,...,n using the relation described in (4.9). Based on this sample, we obtain the maximum likelihood estimates by solving the equation (4.11). For the given values of the model parameter, we repeated the experiment 100 times for computing the estimates and then averaged them over the repetitions. Using the estimate of ϕ, we obtained the maximum likelihood estimate of ν by maximizing the profile log likelihood function. Next we compute the likelihood ratio test statistic in equation (4.13) for various sample sizes and for different parameter values. Finally we compute the number of rejections in 500 trials for testing the null hypothesis of interest. The numerical computations are carried out for various value of the model parameter and are summarised in Tables 4.1 to 4.4.

TABLE 4.1: The average estimates and the corresponding root mean squares errors of the MLE for $\nu = 3$

n	ϕ	$\hat{\phi}$	$\hat{\nu}$
300	-0.5	-0.5058(0.0017)	3.0882(0.3451)
	-0.3	-0.2857(0.0038)	3.1144(0.3380)
	0.3	0.8991(0.0026)	3.1082(0.2145)
	0.5	0.4994(0.0022)	3.0703(0.2514)
	0.7	0.6929(0.0020)	3.0560(0.3270)
	0.9	0.8941(0.0007)	3.1036(0.2802)
500	-0.5	-0.4930(0.0009)	3.0564(0.1084)
	-0.3	-0.3002(0.0012)	3.0130(0.1031)
	0.3	0.2992(0.0014)	3.0976(0.1511)
	0.5	0.5011(0.0012)	3.0391(0.1563)
	0.7	0.6987(0.0008)	3.0473(0.1489)
	0.9	0.9020(0.0003)	3.0724(0.1832)

TABLE 4.2: The average estimates and the corresponding root mean squares errors of the MLE for $\nu = 4$

n	ϕ	$\hat{\phi}$	$\hat{\nu}$
300	-0.5	-0.5055(0.0011)	4.0858(0.2152)
	-0.3	-0.3048(0.0037)	4.0941(0.3121)
	0.3	0.3005(0.0013)	4.0702(0.2112)
	0.5	0.4859(0.0021)	4.1221(0.1999)
	0.7	0.6999(0.0024)	4.0927(0.2100)
	0.9	0.8980(0.0006)	4.0854(0.1221)
500	-0.5	-0.4975(0.0015)	4.0521(0.1132)
	-0.3	-0.2973(0.0016)	4.0139(0.1523)
	0.3	0.2932(0.0008)	4.0891(0.1021)
	0.5	0.4926(0.0019)	4.1020 (0.1989)
	0.7	0.7004(0.0013)	4.0698(0.1654)
	0.9	0.8996(0.0003)	4.0541(0.0989)

Next we carried out a simulation study to evaluate the performance of the error correction model based on bivariate Student t distributed errors. For the study, we generate a sample of size, say n, from the bivariate Student t distribution

Table 4.3: No of rejections in 500 trials of the hypothesis $H_0 : \phi = 1$ against $H_1 : \phi < 1$ using the test statistic given in (4.13) for different values of ϕ and $\nu = 3$

n	50			100			250			500		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
ϕ=-0.5	500	500	500	500	500	500	500	500	500	500	500	500
ϕ=-0.2	500	500	500	500	500	500	500	500	500	500	500	500
ϕ=0.5	500	500	500	500	500	500	500	500	500	500	500	500
ϕ=0.8	356	417	453	489	500	500	500	500	500	500	500	500
ϕ=0.9	184	209	324	365	427	461	452	473	492	499	500	500
ϕ=0.95	63	106	170	192	228	333	428	478	493	491	496	498

Table 4.4: No of rejections in 500 trials of the hypothesis $H_0 : \phi = 1$ against $H_1 : \phi < 1$ using the test statistic given in (6.13) for different values of ϕ and $\nu = 4$

n	50			100			250			500		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
ϕ=-0.5	500	500	500	500	500	500	500	500	500	500	500	500
ϕ=-0.2	500	500	500	500	500	500	500	500	500	500	500	500
ϕ=0.5	500	500	500	500	500	500	500	500	500	500	500	500
ϕ=0.8	337	395	417	458	472	477	498	500	500	500	500	500
ϕ=0.9	167	236	265	328	385	396	409	435	438	498	499	500
ϕ=0.95	80	139	169	152	238	266	407	434	439	449	438	463

and then genrerate the ECM in (4.1) and (4.2). Finally we obtained the MLE of the parameters by solving the likelihood equations in (4.4) to (4.7). Using the estimated parameter values of $\rho, \alpha, \beta, \delta$, we maximized the profile log likelihood function in (4.8) and obtained the estimate of ν. We then repeated the experiment 100 times for computing the estimates and then averaged them over the repetitions. Once the parameters of ECM being estimated, we test for cointegration using the residuals from the error correction model. We use the test statistic given in (4.14) to compute the number of rejections of the null hypothesis under the various alternatives.

Table 4.5 and 4.6 corresponds to the parameter estimates of the error correction model associated with t distribution with $\nu = 3$ and $\nu = 4$. Table 4.7 and 44.8

gives the number of rejections of the null hypothesis under the various alternatives. Note that from Tables 4.1 and 4.3, for series of length 300, estimates are

TABLE 4.5: The average estimates and the corresponding root mean squared errors of MLE for $\nu = 3$

n	ρ	α	β	δ	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\nu}$
150	-0.2	2	3	-0.6	-0.1599(0.0244)	2.0010(0.0068)	3.0048(0.0025)	-0.6013(0.0001)	2.6894(0.2216)
	-0.5	3	1	0.2	-0.4369(0.0084)	3.0036(0.0032)	1.0046(0.0044)	0.2016(0.0022)	2.6516(0.3512)
	0.2	1.8	0.5	1	0.1618(0.0088)	1.801(0.0001)	0.4992(0.0009)	1.0031(0.0005)	2.7346(0.2052)
	0.5	1.4	1.5	-7	0.4355(0.0089)	1.3999(0.0006)	1.5000(0.0009)	-6.7570(0.0121)	2.6910(0.2342)
	0.9	0.2	3.5	-0.18	0.8779(0.0007)	0.2038(0.0001)	3.4834(0.0140)	-0.1816(0.0001)	2.6817(0.2878)
300	-0.2	2	3	-0.6	-0.1772(0.0184)	1.9999(0.0031)	2.9997(0.0001)	-0.6001(0.00001)	2.8292(0.1165)
	-0.5	3	1	0.2	-0.4899(0.0053)	2.9917(0.0010)	0.9847(0.0034)	0.1985(0.0001)	2.8032(0.1621)
	0.2	1.8	0.5	1	0.1733(0.0048)	1.7999(0.0042)	0.5029(0.0003)	1.0006(0.0002)	2.8452(0.1123)
	0.5	1.4	1.5	-7	0.4426(0.0055)	1.3999(0.0005)	1.3999(0.0001)	1.4996(0.0002)	2.8012(0.2012)
	0.9	0.2	3.5	-0.18	0.8790(0.0006)	0.2017(0.0001)	3.4966(0.0073)	-0.1811(0.00009)	2.8321(0.1325)

TABLE 4.6: The average estimates and the corresponding root mean squared errors of MLE for $\nu = 4$

n	ρ	α	β	δ	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\nu}$
150	-0.2	2	3	-0.6	-0.1907(0.0061)	2.0017(0.0002)	3.0080(0.0032)	-0.601(0.0001)	4.0538(0.2012)
	-0.5	3	1	0.2	-0.4974(0.0044)	3.0050(0.0020)	0.9813(0.0032)	0.2001(0.0002)	3.9830(0.2136)
	0.2	1.8	0.5	1	0.1933(0.0073)	1.8009(0.00002)	0.4945(0.0007)	0.9993(0.0004)	4.1559(0.2013)
	0.5	1.4	1.5	-7	0.4914(0.0047)	1.3999(0.0005)	1.4998(0.0001)	-7.0004(0.0001)	3.9764(0.1962)
	0.9	0.2	3.5	-0.18	0.8964(0.0002)	0.205(0.0001)	3.5130(0.0014)	-0.1822(0.00002)	4.0261(0.1032)
300	-0.2	2	3	-0.6	-0.2045(0.0028)	2.0013(0.0001)	3.0048(0.0016)	-0.6008(0.00006)	4.0025(0.1145)
	-0.5	3	1	0.2	-0.4940(0.0024)	2.9949(0.0017)	0.9985(0.0009)	0.1999(0.00009)	4.0706(0.1098)
	0.2	1.8	0.5	1	0.1869(0.0038)	1.8004(0.00001)	0.5005(0.0003)	0.9990(0.0001)	4.1018(0.1021)
	0.5	1.4	1.5	-7	0.4982(0.0020)	1.3999(0.00004)	1.5001(0.00005)	-7.0003(0.0060)	4.0778(0.1213)
	0.9	0.2	3.5	-0.18	0.9016(0.0003)	0.2026(0.00009)	3.4999(0.0040)	-0.1817(0.00001)	3.9870(0.1026)

TABLE 4.7: No of rejections in 500 trials of the hypothesis $H_0 : \phi = 1$ (δ=0) against the alternative of $\phi < 1$($\delta \neq 0$) for $\nu = 3$

n	50			100			150		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
ϕ=0.5	500	500	500	500	500	500	500	500	500
ϕ= 0.8	426	456	468	500	500	500	500	500	500
ϕ= 0.9	242	298	342	424	448	465	500	500	500
ϕ= 0.95	101	137	177	223	258	314	361	391	422
ϕ= 0.99	38	42	45	80	85	92	125	130	135

reasonably satisfactory and become more accurate with increasing sample size. From Table 4.3, 4.4, 4.7 and 4.8, it can be seen that as ϕ becomes closer to 1, the

TABLE 4.8: No of rejections in 500 trials of the hypothesis $H_0 : \phi = 1$ (δ=0) against the alternative of $\phi < 1$($\delta \neq 0$) for $\nu = 4$

n	50			100			150		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
ϕ=0.5	500	500	500	500	500	500	500	500	500
ϕ= 0.8	440	452	500	500	500	500	500	500	500
ϕ= 0.9	285	312	350	451	468	481	500	500	500
ϕ= 0.95	111	132	170	218	241	311	350	381	410
ϕ= 0.99	42	46	50	85	98	101	128	140	144

number of rejections of the null hypothesis of unit root and no cointegration becomes smaller. For example, in a length of 50 series in Table 4.3, the hypothesis $H_0 : \phi = 0.95$ was rejected 63 times at the 0.01 significance level, while it was rejected 184 times when ϕ was 0.9. Hence we claim that the derived test statistic is powerful for testing the presence of unit root in an observed non stationary time series.

Algorithm 6: Algorithm for Table 4.1 and 4.2

1 Set x_0, ν and ϕ, for some sample of size n.
2 Draw samples $\{a_t\}$, from Student t distribution.
3 Generate x_t using $x_t = \phi x_{t-1} + a_t, t = 1, 2, ...n$
4 Choose the initial value of ϕ, ν as ϕ_0, ν_0 and obtain $\hat{\phi}$, using the method of maximum likelihood.
5 Using the estimate $\hat{\phi}$, obtain the estimate of ν by maximizing the profile likelihood function.
6 Repeat Steps 2 to 5, say 100 times.
7 Choose the value of $\hat{\phi}$ as the averages of $\hat{\phi}$ obtained in step 4.

Algorithm 7: Algorithm for Table 4.3 and 4.4

1 Set x_0, ν and ϕ, for some sample of size n.
2 Draw samples $\{a_t\}$, from Student t distribution.
3 Generate x_t using $x_t = \phi x_{t-1} + a_t, t = 1, 2, ..., n$
4 Choose the initial value of ϕ, ν as ϕ_0, ν_0 and obtain $\hat{\phi}$, using the method of maximum likelihood.
5 Using the estimate $\hat{\phi}$, obtain the estimate of ν by maximizing the profile likelihood function.
6 Obtain the LRT statistic in Equation (4.13) for the simulated data.
7 Repeat Steps 2 to 6, say 500 times.
8 Count and record the number of rejections of unit root hypothesis in 500 trials.

Algorithm 8: Algorithm for Table 4.5 and 4.6

1 Set $\{(x_{10}, x_{20})\}$, ν, α, β and δ, for some sample of size n.
2 Draw the bivariate samples (η_{1t}, η_{1t}) using the equation given by (4.3).
3 Generate the samples of ECM, using Equations (4.1) and (4.2).
4 Choose the initial value as ν_0, ρ_0, α_0, β_0 and δ_0. Solve Equations (4.4) to (4.7) for obtaining the maximum likelihood estimates $\hat{\rho}_0$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\delta}$.
5 Using the estimates obtained in step 4, obtain the estimate of ν by maximizing the profile likelihood function.
6 Repeat steps 2 to 5 for 100 times.
7 Set the values of $\hat{\rho}, \hat{\alpha}, \hat{\beta}, \hat{\delta}$ as the averages of $\hat{\rho}$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\delta}$ obtained in step 5.

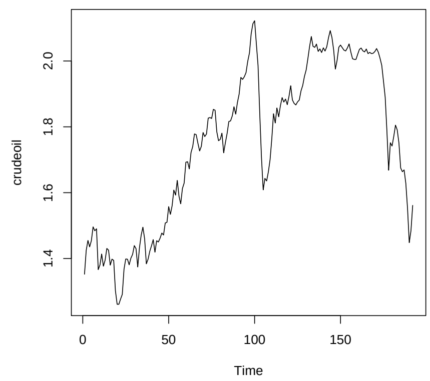
Algorithm 9: Algorithm for Table 4.7 and 4.8

1 Set $\{(x_{10}, x_{20})\}$, ν, α, β and δ, for some sample of size n.
2 Draw the bivariate samples (η_{1t}, η_{1t}) using the equation given by (4.3).
3 Generate the samples of ECM, using Equations (4.1) and (4.2).
4 Choose the initial value as ν_0, ρ_0, α_0, β_0 and δ_0. Solve Equations (4.4) to (4.7) for obtaining the maximum likelihood estimates $\hat{\rho}_0$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\delta}$.
5 Using the estimates obtained in step 4, obtain the estimate of ν by maximizing the profile likelihood function.
6 Obtain the LRT statistic in Equation (4.14) for the simulated data.
7 Repeat Steps 2 to 6 for 500 times.
8 Count and record the number of rejections of the hypothesis $H_0 : \phi = 1$ (δ=0) against the alternative of $|\phi| < 1$ ($\delta \neq 0$) in 500 trials.

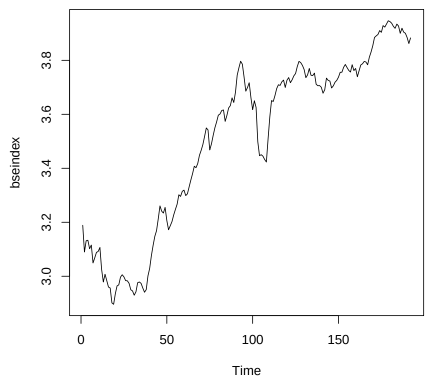
4.6 Data Analysis

Here we illustrate the analysis of cointegration model with bivariate Student t errors with identical marginals using the real data set. The data set consists of 192 monthly observations of crude oil price and Bombay stock exchange index for the period 2000 to 2016. All the variables are transformed in to their natural logarithm. The data set is downloaded from the website of Ministry of Petroleum and Natural gas, Govt. of India (www.ppac.org.in) and the website of Reserve Bank of India (www.rbi.org.in). The time series plots are given in Figure 4.1.

First, we performed a unit root test developed for Student t error variables to



(A) Time plot of crude oil price



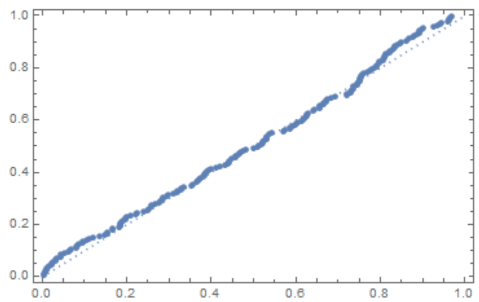
(B) Time plot of BSE index

FIGURE 4.1: Time series plot

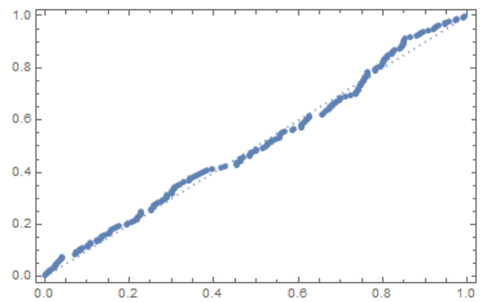
the data set in order to test whether the series is stationary or not. By setting $H_0 : \phi = 1$ against $H_1 : |\phi| < 1$ the test statistic value obtained for crude oil and BSE series are 0.00029 and 0.0015 respectively. These values are much less than the corresponding Chi squared table value. Hence we conclude that the null hypothesis of unit root cannot be rejected at any given level of significance. So we perform a maximum likelihood estimation described in Section 4.3 in order to find the parameter estimates of an error correction model of order 1. The parameter estimates of the model are given by $\hat{\rho} = 0.2425$, $\hat{\alpha} = -2.8930$, $\hat{\beta} = -0.6960$, $\hat{\delta} = 0.00207$ and $\hat{\nu} = 4.54$.

The value of the test statistic obtained for the error correction model is -90 and from the Chi-squared table for two tailed test, we reject the null hypothesis of no cointegration at 5 percent level of significance, implying that the residuals from the ECM are stationary. Thus the cointegrating vector parameter estimate provides an estimate of a long run relationship. That is, $x_{1t} - 2.893x_{2t}$ is the cointegrating relationship and the cointegrated vector is $[1, -2.893]'$. Using the parameter estimates of the ECM, we tested whether the residuals follow a Student t distribution using Kolmogorov-Smirnov test.

The p-values obtained for the series are 0.638 and 0.562 respectively, which indicates that Student t distribution is suitable for the residual series. PP-plot and



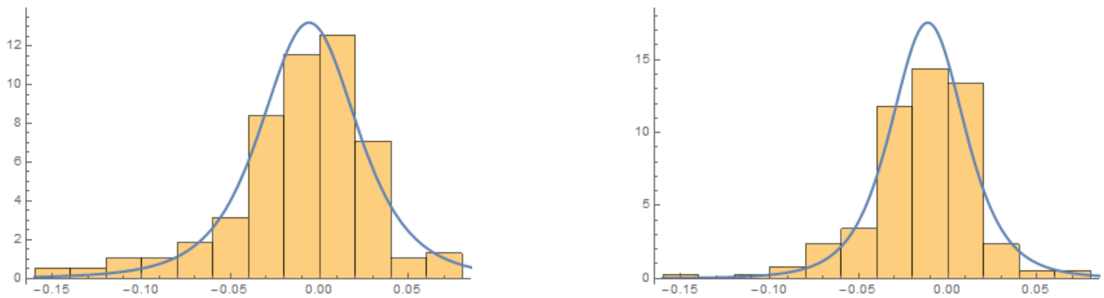
(A) PP- plot of residuals of crude oil series



(B) PP- plot of residuals of BSE series

FIGURE 4.2: Probability- Probability Plots

histogram given in Figure 4.2 and 4.3 also confirms that the chosen distribution is suitable for the residual series.



(A) Histogram of crude oil price with ν=3.69 (B) Histogram of BSE index with ν=4.35

FIGURE 4.3: Residual Histogram

Chapter 5

Cointegration Models with non-Gaussian GARCH innovations

5.1 Introduction

The discussion in the previous chapter largely rely on the assumption that the observed time series fluctuates around changing level with constant variance or homoskedastic in nature. Recent empirical evidences suggest that, financial time series exhibit the presence of heteroskedasticity in the sense that, it possesses non constant conditional variance given the past observations. Conditional heteroskedastic models have been developed to capture this empirical feature in the volatility of financial returns. Engle (1982) introduced the autoregressive conditionally heteroskedastic (ARCH) models and their extension, Generalised autoregressive conditionally heteroskedastic (GARCH) model is due to Bollerslev (1986). For the study of ARCH-type models, see for instance, Bollerslev et al. (1992), Bollerslev et al. (1994), Pagan (1996), Palm (1996), Wong et al. (2005) and Tsay (2005). A time series $\{\varepsilon_t\}$ is said to follow Generalised Autoregressive Conditional Heteroskedastic (GARCH(p,q)) model if

$$\varepsilon_t = a_t\sqrt{h_t},$$

where $\{a_t\}$ is a sequence of independent and identically distributed (iid) random variables with zero mean and unit variance, and

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i}$$

defines the conditional variance. In general, the error term a_t may follow any symmetric distribution like Student t, Generalised error distribution etc. But most of the models in the literature assume that innovations are normally distributed. The structure of the GARCH model makes the likelihood based inference more manageable. Bollerslev (1986) described the maximum likelihood estimation of a GARCH regression model with normally distributed errors. Bollerslev (1987) and Yang & Brorsen (1992) discussed a non normal error GARCH model based on the symmetric, but leptokurtic Student t distribution. Apart from these, Bollerslev (1990) proposes a new class of multivariate GARCH models in which the conditional correlations are constant.

Despite the extensive literature on cointegration and heteroskedastic models, little attention is given to the issue of testing cointegration in the presence of hetroskedastic errors. The available works are mostly based on simulation techniques and hence no formal theories have been developed, see for instance Kim & Schmidt (1993) and Lee & Tse (1996). Perhaps, the work by Li et al. (2001) was the first one to perceive the issues of cointegration and heteroskedasticity together. To our knowledge, there is no systematic studies on cointegration under hetroskedasticity when the innovations are generated from non Gaussian distributions. So, the main objective of the present Chapter is to explore the possibility of modelling cointegrating time series when the errors are generated by non-Gaussian GARCH models.

We propose the modelling of two cointegrating time series with GARCH errors, based on the heteroskedastic model proposed by Bollerslev (1990). Since most of the empirical distributions of financial time series show deviations from normality, a joint modelling of bivariate cointegration in the presence of heteroskedasticity using non Gaussian innovations is also called for. The model applicability is then discussed with the modelling of several price series.

The rest of the chapter is organised as follows. In Section 5.2, we introduce the cointegration GARCH model using Engle-Granger error correction representation and Phillip's triangular representation. Section 5.3 deals with the inference procedures of the cointegration GARCH model under Gaussian and non Gaussian set up. We present the simulation study in Section 5.4. In Section 5.5, we illustrate the application of the proposed model by analysing certain commodity price series.

5.2 Cointegrating Models with GARCH innovations

5.2.1 Using Engle-Granger Error Correction representation

Let $\{X_{1t}\}$ and $\{X_{2t}\}$ be two cointegrating time series and both are integrated of order one. For the formulation of Error Correction representation, we assume that the error term $\mathbf{a_t} = (a_{1t}, a_{2t})'$, t=1,2,3,... follow GARCH with iid innovations. In the following discussion, we obtain an ECM involving two lags, which is required for analysing the model in the present Chapter.

Following Engle & Granger (1987), let us write

$$X_{1t} + \beta X_{2t} = u_{1t}, u_{1t} = u_{1t-1} + a_{1t} \tag{5.1}$$

$$X_{1t} + \alpha X_{2t} = u_{2t}, u_{2t} = \phi u_{2t-1} + a_{2t}, |\phi| < 1, \tag{5.2}$$

where $\{u_{1t}\}$ follows a random walk model and $\{u_{2t}\}$ is a stationary sequence. The Error Correction representation takes of the form:

$$\Delta X_{1t} = \delta\beta Z_{t-1} + \eta_{1t} \tag{5.3}$$

$$\Delta X_{2t} = -\delta Z_{t-1} + \eta_{2t}, \tag{5.4}$$

where $Z_{t-1} = X_{1t-1} + \alpha X_{2t-1}$.

Equivalently, $\Delta\mathbf{X_t} = \boldsymbol{A}Z_{t-1} + \boldsymbol{\eta_t}$, with $\Delta\mathbf{X_t} = \begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix}$, $\mathbf{A} = \begin{pmatrix} \delta\beta \\ -\beta \end{pmatrix}$ and $\boldsymbol{\eta_t} = (\eta_{1t}, \eta_{2t})^{'}$. The above equation relates to the cointegrated vector auto regression model with lag one, that is CVAR(1) model. Similarly, we can obtain CVAR(2) model by adding one more lag in Equations (5.1), (5.2). Let us assume that,

$$\begin{aligned} &(1-\theta_1 L)(1-\theta_2 L)u_{1t} = a_{1t} \\ &\Rightarrow u_{1t} - \theta_2 u_{1t-1} - \theta_1 u_{1t-1} + \theta_1\theta_2 u_{1t-2} = a_{1t} \\ &\Rightarrow u_{1t} = (\theta_1 + \theta_2)u_{1t-1} - \theta_1\theta_2 u_{1t-2} + a_{1t} \\ &\Rightarrow u_{1t} = \phi_1 u_{1t-1} + \phi_2 u_{1t-2} + a_{1t} \end{aligned},$$

where $\theta_1 + \theta_2 = \phi_1$ and $-\theta_1\theta_2 = \phi_2$. If $\theta_2 = 1$, the model has a unit root. That is, $u_{1t} = u_{1t-1} + \theta_1(u_{1t-1} - u_{1t-2}) + a_{1t}$. Now let us redefine equations (5.1) and

(5.2) by

$$X_{1t} + \beta X_{2t} = u_{1t}, u_{1t-1} + \theta_1(u_{1t-1} - u_{1t-2}) + a_{1t} \tag{5.5}$$

$$X_{1t} + \alpha X_{2t} = u_{2t}, u_{2t} = \phi u_{2t-1} + a_{2t}, |\phi| < 1, \tag{5.6}$$

On subtracting with αX_{2t-1} on both sides of Equation (5.6) gives,

$$X_{1t} + \alpha X_{2t} - \alpha X_{2t-1} = u_{2t} - \alpha X_{2t-1}.$$

$$\begin{aligned}
\alpha \Delta X_{2t} &= \phi_1 \left(X_{1t-1} + \alpha X_{2t-1}\right) - \alpha X_{2t-1} - X_{1t} + a_{2t} \\
&= \phi_1 \left(X_{1t-1} + \alpha X_{2t-1}\right) - \alpha X_{2t-1} - \left(u_{1t} - \beta X_{2t}\right) + a_{2t} \\
&= \phi_1 \left(X_{1t-1} + \alpha X_{2t-1}\right) - \alpha X_{2t-1} - \left(u_{1t-1} + \theta_1 \left(u_{1t-1} - u_{1t-2}\right) + a_{1t} - \beta X_{2t}\right) \\
&\quad + a_{2t}.
\end{aligned}$$

On substituting for u_{1t-1} and u_{2t-1},

$$\begin{aligned}
\alpha \Delta X_{2t} = &\left(\phi_1 - 1\right) X_{1t-1} + \alpha X_{2t-1} \left(\phi_1 - 1\right) + \beta \Delta X_{2t} - \theta_1 \left(\Delta X_{1t-1} + \beta \Delta X_{2t-1}\right) \\
&- a_{1t} + a_{2t}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\left(\alpha - \beta\right) \Delta X_{2t} = &\left(\phi_1 - 1\right) X_{1t-1} + \alpha X_{2t-1} \left(\phi_1 - 1\right) - \theta_1 \left(\Delta X_{1t-1} + \beta \Delta X_{2t-1}\right) \\
&- a_{1t} + a_{2t}.
\end{aligned}$$

$$\begin{aligned}\Delta X_{2t} &= -\delta\left(X_{1t-1}+\alpha X_{2t-1}\right)+\delta\left(\Delta X_{1t-1}+\beta\Delta X_{2t-1}\right)+\zeta_{2t}\\ &= -\delta Z_{t-1}+\delta\Delta Z_{t-1}+\zeta_{2t},\end{aligned}$$

where $\delta = \frac{(1-\phi_1)}{\alpha-\beta}$, $Z_{t-1} = X_{1t-1}+\alpha X_{2t-1}$, $\Delta Z_{t-1} = \Delta X_{1t-1}+\beta\Delta X_{2t-1}$ and $\zeta_{2t} = \frac{a_{2t}-a_{1t}}{\alpha-\beta}$.

On subtracting X_{1t-1} from both the sides of Equation (5.5) gives,

$$\begin{aligned}\Delta X_{1t} &= u_{1t}-X_{1t-1}-\beta X_{2t}\\ &= u_{1t-1}+\theta_1\left(u_{1t-1}-u_{1t-2}\right)+a_{1t}-X_{1t-1}-\beta X_{2t}.\end{aligned}$$

Substituting for u_{1t-1} and u_{2t-1},

$$\begin{aligned}\Delta X_{1t} &= \beta X_{2t-1}+\theta_1\left(\Delta X_{1t-1}+\beta\Delta X_{2t-1}\right)+a_{1t}-\frac{\beta}{\alpha}\left(u_{2t}-X_{1t}\right)\\ &= \beta X_{2t-1}+\theta_1\left(\Delta X_{1t-1}+\beta\Delta X_{2t-1}\right)+a_{1t}-\frac{\beta}{\alpha}\left(X_{1t-1}+\alpha X_{2t-1}-X_{1t}\right)-\frac{\beta}{\alpha}a_{2t}.\end{aligned}$$

Thus,

$$\begin{aligned}\Delta X_{1t}\left(1-\frac{\beta}{\alpha}\right) &= \beta\left(1-\phi_1\right)X_{2t-1}+\left(\frac{\beta}{\alpha-\beta}\right)\left(1-\phi_1\right)X_{1t-1}\\ &\quad+\left(\frac{\alpha}{\alpha-\beta}\right)\left(1-\phi_1\right)\Delta Z_{t-1}+a_{1t}-\frac{\beta}{\alpha}a_{2t}.\end{aligned}$$

$$\begin{aligned}\Delta X_{1t} &= \frac{\alpha\beta}{\alpha-\beta}\left(1-\phi_1\right)X_{2t-1}+\frac{\alpha\beta}{\alpha-\beta}\left(1-\phi_1\right)X_{1t-1}+\left(\frac{\alpha}{\alpha-\beta}\right)\left(1-\phi_1\right)\Delta Z_{t-1}+\zeta_{1t}\\ &= \beta\delta Z_{t-1}+\alpha\delta\Delta Z_{t-1}+\zeta_{1t},\end{aligned}$$

where $\zeta_{1t} = \frac{\alpha}{\alpha-\beta}\left(a_{1t} - \frac{\beta}{\alpha}a_{2t}\right)$. Thus the CVAR(2) model thus takes of the form,

$$\Delta X_{1t} = \beta\delta Z_{t-1} + \alpha\delta\Delta Z_{t-1} + \zeta_{1t}$$
$$\Delta X_{2t} = -\delta Z_{t-1} + \delta\Delta Z_{t-1} + \zeta_{2t}.$$

This model is needed and defined for the purpose of obtaining the correct lag order for the cointegrating equation.

Now we define the bivariate CVAR(1)-GARCH model as,

$$\Delta \mathbf{X_t} = \mathbf{A} Z_{t-1} + \boldsymbol{\eta}_\mathbf{t}$$

$$\boldsymbol{\eta}_\mathbf{t} = \mathbf{H_t^{1/2} a_t}$$

$$\mathbf{a_t} \mid I_{t-1} \sim N_2(\mathbf{0}, \mathbf{J_2}), \tag{5.7}$$

with $E\left(\boldsymbol{\eta}_t \mid I_{t-1}\right) = \mathbf{0}$ and $V\left(\boldsymbol{\eta}_t \mid I_{t-1}\right) = \mathbf{H_t^{1/2} J_2 H_t^{1/2}}$, where I_{t-1} is the information available up to time t-1, $\mathbf{J_2}$ is a 2×2 identity matrix and $\mathbf{H_t}$ is the time varying conditional covariance matrix which is almost surely positive definite for all t and the conditional correlation is given by $\rho_{ijt} = h_{ijt}/\sqrt{(h_{iit}h_{jjt})}$, h_{ijt} denote the ij^{th} element of $\mathbf{H_t}$. However, this conditional covariance will be time varying as $\mathbf{H_t}$ varies over time. But, in some applications, the time varying conditional covariances can be proportional to the square root of the product of the two conditional variances given by,

$$h_{ijt} = \rho_{ij}\sqrt{(h_{iit}h_{jjt})},$$

leaving the conditional correlations constant through time.

5.2.2 Using Phillip's triangular representation

Another convenient representation of cointegrated system, is the triangular representation, introduced by Phillips (1991). Writing of the same model in (5.3) and (5.4) in the usual triangular representation (see Hayashi (2000), Chapter 10), reduces the number of parameters (δ is eliminated) and the estimation procedure becomes simpler. In particular, it simplifies the parameter vector of cointegration.

Phillip's triangular representation is used for the purpose of estimating cointegrating vectors. This representation is valid for any cointegrating rank, but often it is assumed that the cointegrating rank is 1. The representation is called triangular, because the parameter matrices are in a triangular form with fewer number of parameters. So, on following Hayashi (2000), we use the Phillip's triangular representation for the purpose of defining cointegration model with GARCH innovations. Suppose that $\boldsymbol{X_t} = (X_{1t}\ X_{2t})'$ is a vector of non stationary series with order of integration one, and let the cointegrating vector $\boldsymbol{\alpha}$ be

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 \\ -\gamma \end{pmatrix}.$$

We focus on vector moving average model of order 1 and further we assume that the rank of cointegration is one. So on choosing the appropriate coefficient matrix with rank one, the triangular representation in bivariate case consists of two equations, the cointegrating regression equation given by,

$$X_{1t} = \mu + \gamma X_{2t} + (\eta_{1t} - \gamma\eta_{2t}), \tag{5.8}$$

where $\mu = (X_{1,0} - \gamma X_{2,0}) - (\eta_{1,0} - \gamma \eta_{2,0})$ and the second row of

$$\underset{(2\times1)}{\Delta \boldsymbol{X}_t} = \underset{(2\times1)}{\boldsymbol{\delta}} + \underset{(2\times2)}{\boldsymbol{\psi}(L)} \underset{(2\times1)}{\varepsilon_t}, \boldsymbol{\psi}(L) = \boldsymbol{I}_2 + \boldsymbol{\psi}_1 L + \boldsymbol{\psi}_2 L^2 + \cdots \tag{5.9}$$

We choose the coefficient matrix in such a way that the rank of cointegration is one. So on choosing the appropriate coefficient matrix with rank one, the above equation becomes,

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} + \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{1t-1} \\ \eta_{2t-1} \end{bmatrix}. \tag{5.10}$$

Hence the second equation of the triangular representation is:

$$\Delta X_{2t} = \delta_2 + \eta_{2t}. \tag{5.11}$$

The above equation relates to the triangular representation of cointegration model with lag one, that is cointegrated vector moving average model of order 1 (CVMA(1) model), see Hayashi (2000). The cointegrated vector moving average model with higher lags may be needed for obtaining the correct lag length of the cointegration while modelling a real data set. In this case, equation (5.10) become,

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} + \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{1t-1} \\ \eta_{2t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \eta_{1t-2} \\ \eta_{2t-2} \end{bmatrix}. \tag{5.12}$$

Hence the triangular representation associated with CVMA(2) model becomes:

$$X_{1t} = \mu + \gamma X_{2t} + (\eta_{1t} - \gamma \eta_{2t})$$

$$\Delta X_{2t} = \delta_2 + \eta_{2t} - \eta_{2t-2}.$$

Now we define the bivariate CVMA(1)- GARCH model as,

$$X_{1t} = \mu + \gamma X_{2t} + (\eta_{1t} - \gamma \eta_{2t})$$

$$\Delta X_{2t} = \delta_2 + \eta_{2t}$$

$$\boldsymbol{\eta}_t = \mathbf{H_t^{1/2} a_t}$$

$$\mathbf{a_t} \mid I_{t-1} \sim N_2(\mathbf{0}, \mathbf{J_2}), \tag{5.13}$$

where $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t})'$ with $E(\boldsymbol{\eta}_t \mid I_{t-1}) = \mathbf{0}$ and $V(\boldsymbol{\eta}_t \mid I_{t-1}) = \mathbf{H_t^{1/2} J_2 H_t^{1/2}}$. In both representations (5.7) and (5.13), we assume that the error variable η_t follow a constant conditional correlation model proposed by Bollerslev (1990). Since the vector $\mathbf{a_t} \mid I_{t-1}$ follows a conditional bivariate normal distribution, $\boldsymbol{\eta}_t \sim N_2(\mathbf{0}, \mathbf{H_t})$. Further, following Bollerslev (1990), we assume that the conditional covariance matrix $\mathbf{H_t}$ is of the type,

$$\mathbf{H_t} = \mathbf{D_t \Gamma D_t},$$

where $\mathbf{D_t}$, denotes 2×2 diagonal matrix given by $\mathbf{D_t} = diag(\sigma_{1t}, \sigma_{2t})$, the conditional variances given I_{t-1} takes the form,

$$\sigma_{it}^2 = c_i + b_i \eta_{it-1}^2 + g_i \sigma_{it-1}^2 \tag{5.14}$$

for i=1,2 and t=1,2,...N with $c_i > 0$, $b_i \geq 0$, $g_i \geq 0$, and the correlation matrix is assumed to be, $\mathbf{\Gamma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where ρ is the correlation parameter. Hence the conditional correlation matrix $\mathbf{H_t}$ becomes,

$$\begin{aligned} \mathbf{H_t} &= \begin{pmatrix} \sigma_{1t} & 0 \\ 0 & \sigma_{2t} \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sigma_{1t} & 0 \\ 0 & \sigma_{2t} \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{1t}^2 & \rho(\sigma_{1t}\sigma_{2t}) \\ \rho(\sigma_{1t}\sigma_{2t}) & {\sigma_{2t}}^2 \end{pmatrix}, \end{aligned} \tag{5.15}$$

with $\det(\mathbf{H_t}) = \sigma_{1t}^2 \sigma_{2t}^2 (1 - \rho^2)$. Here each of the conditional variances is obtained as $h_{ijt} = \rho(\sigma_{1t}\sigma_{2t})$ and $h_{iit} = \sigma_{it}^2, i, j = 1, 2$. From (5.15) it follows that $\mathbf{H_t}$ will be almost surely positive definite for all t.

In the following section, we discuss the classical approach for the parameter estimation of the joint modelling of cointegration and GARCH model using Gaussian and some non-Gaussian error distribution.

5.3 Inference for a given model

The parameter estimation of heteroskedastic models is usually carried out by some numerical maximum likelihood technique such as method of scoring or direct likelihood optimization, as the likelihood function is highly non linear in parameters; see, for instance, Mak (1993), Mak et al. (1997) and Berndt et al. (1974). Vrontos et al. (2003) obtained the parameter estimates of a full factor GARCH model by using Fisher scoring algorithm and Bayesian approach. Bollerslev (1990) discuss the estimation procedures for the multivariate time series model with time varying conditional variances and covariances by using a numerical iterative maximisation technique. In this section, we discuss the problem of estimation for the bivariate cointegration model (5.7) and (5.13) with GARCH errors when the errors $\mathbf{a_t} \sim N_2(\mathbf{0}, \mathbf{J_2})$ and $\mathbf{a_t} \sim GED$ with shape parameter $v = 2$ (cf. (5.27)). We obtain the maximum likelihood estimates of the parameters of cointegration model with GARCH errors that have normal errors by using Fisher scoring algorithm, and by using numerical optimization technique for the cointegration GARCH model using GED errors. The motivations for using Fisher scoring algorithm in our research comes from the results of the above mentioned works. The kth iteration of the algorithm takes the form,

$$\hat{\theta}^{(k)} = \hat{\theta}^{(k-1)} + \left\{ -E\left(\frac{\partial^2 L_T}{\partial\theta\partial\theta'} \right) \right\}^{-1} \frac{\partial L_T}{\partial\theta}, \tag{5.16}$$

where $\hat{\theta}^{(k-1)}$ denotes the estimate of the parameter vector obtained after k-1 iterations, L_T denotes the conditional log likelihood function; which is conditioned on past observations, $\left\{ -E\left(\frac{\partial^2 L_T}{\partial\theta\partial\theta'} \right) \right\}^{-1}$ denotes the expected information matrix computed at $\hat{\theta}^{(k-1)}$ and $\frac{\partial L_T}{\partial\theta}$ denote the gradient computed at $\hat{\theta}^{(k-1)}$. An advantage of the model (5.7) and (5.13) is that, under the assumption of bivariate normality, a closed form expression is readily available for the gradient and expected information matrix. Besides, as the covariance matrix is guaranteed to be positive definite, the model parameters are easily estimated and can be extended to higher dimensions (See Vrontos et al. (2003)).

5.3.1 MLE using Gaussian GARCH errors

Now we discuss the maximum likelihood estimation of the parameters involved in model (5.7) and(5.13) based on the sample $\mathbf{x}_t = \{x_{1t}, x_{2t}\}, t = 1, 2, ...T$. Under the assumption of a bivariate normal distribution for the vector $\mathbf{a_t}$, the conditional likelihood function given $(\eta_{10}, \eta_{20}, \sigma_{10}, \sigma_{20})$ for the cointegration GARCH model can be written as,

$$L\left(\theta | \, x_t\right) = \prod_{t=1}^{T} (1/2\pi) \left|\mathbf{H_t}\right|^{-1/2} Exp\left[-\frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{\eta}_t^{'}\mathbf{H_t^{-1}}\boldsymbol{\eta}_\mathbf{t}\right)\right], \tag{5.17}$$

where θ denotes the vector of unknown parameters to be estimated. For the ease of the computational procedures, we will assume that η_{i0} and σ_{i0} $(\mathrm{i} = 1, 2)$ are known. The log likelihood function for the above cointegration GARCH model is given by:

$$\begin{aligned} L_T\left(\theta \,|x_t\right) &= -1/2\sum_{t=1}^{T}\log\left(\sigma_{1t}^2\sigma_{2t}^2(1-\rho^2)\right) - 1/2\sum_{t=1}^{T}\left(\boldsymbol{\eta}_t^{'}(\boldsymbol{D}_t\boldsymbol{\Gamma}\boldsymbol{D}_t)^{-\mathbf{1}}\boldsymbol{\eta}_t\right) \\ &= -1/2\sum_{t=1}^{T}\log\left(1-\rho^2\right) - 1/2\sum_{t=1}^{T}\left(\sum_{i=1}^{2}\log\left(\sigma_{it}^2\right)\right) - 1/2\sum_{t=1}^{T}\left(\boldsymbol{\varepsilon}_t^{'}\boldsymbol{\Gamma}^{-\mathbf{1}}\boldsymbol{\varepsilon}_t\right), \end{aligned} \tag{5.18}$$

where $\boldsymbol{\varepsilon_t} = \boldsymbol{D}_t^{-1}\boldsymbol{\eta_t} = \left(\frac{\eta_{1t}}{\sigma_{1t}}, \frac{\eta_{2t}}{\sigma_{2t}}\right)^{'}$ and $\boldsymbol{\varepsilon}_t^{'}\Gamma^{-1}\boldsymbol{\varepsilon_t} = \frac{1}{1-\rho^2}\left(\frac{\eta_{1t}^2}{\sigma_{1t}^2} + \frac{\eta_{2t}^2}{\sigma_{2t}^2} - \frac{2\rho\eta_{1t}\eta_{2t}}{\sigma_{1t}\sigma_{2t}}\right)$,
so that $L_T\left(\theta\,|x_t\right) = -1/2\sum_{t=1}^{T}\log\left(1-\rho^2\right) - 1/2\sum_{t=1}^{T}\left(\sum_{i=1}^{2}\log\left(\sigma_{it}^2\right)\right)$

$$-1/2\sum_{t=1}^{T}\frac{1}{1-\rho^2}\left(\frac{\eta_{1t}^2}{\sigma_{1t}^2} + \frac{\eta_{2t}^2}{\sigma_{2t}^2} - \frac{2\rho\eta_{1t}\eta_{2t}}{\sigma_{1t}\sigma_{2t}}\right).$$

Here the parameters to be estimated are the elements of θ, where $\theta = (\rho, \gamma, \delta_2, c_1, c_2, b_1, g_1, b_2, g_2)^{'}$. The form of the log likelihood function indicates that we have to maximize it by some numerical methods. Following Vrontos et al. (2003), we transformed all our positive parameters in to its logarithmic transformation in order to avoid the positivity restrictions built in the GARCH model. The transformations of the positive parameters leads to the variance equation :

$$\sigma_{it}^2 = e^{c_i^*} + e^{b_i^*}\eta_{it-1}^2 + e^{g_i^*}\sigma_{it-1}^2, i = 1, 2. \tag{5.19}$$

Several GARCH models can be seen in the literature and they have been analysed by many researchers. The model proposed by Bollerslev (1990) and Engle (2002) does not impose any positivity conditions for the GARCH model, hence it may lead to difficulty in estimation of large set of parameters simultaneously. Among those, the assumptions stated here are very easy to verify and it reduces the difficulty in parameter estimation as it avoids the positivity restrictions in the model; see Vrontos et al. (2003) and Engle & Kroner (1995) for a similar discussion for a multivariate GARCH model. Now for obtaining the estimates, the parameter vector is partitioned in to three blocks. The first block consists of the correlation parameter, $\theta_1 = \rho$, second block consists of the cointegration parameters, $\theta_2 = (\alpha, \beta, \delta)^{'}$ (for the model (5.7)), $\theta_2 = (\gamma, \delta_2)^{'}$ (for the model (5.13)) and the final block contains the GARCH parameters, $\theta_3 = (c_1, c_2, b_1, g_1, b_2, g_2)^{'}$. Since the GARCH model is dynamic, some restrictions are needed for the initial values of the variance term σ_{it}^2. That is we assume $\sigma_{i0}^2 = 0, i = 1, 2$. Hence differentiating the log likelihood function with respect to the correlation parameter $\theta_1 = \rho$ yields,

$$\frac{\partial L_T}{\partial \rho} = \sum_{t=1}^{T} \left(\frac{\rho}{(1-\rho^2)} \right) - \sum_{t=1}^{T} \left[\frac{\rho}{(1-\rho^2)} \left(\frac{\eta_{1t}^2}{\sigma_{1t}^2} + \frac{\eta_{2t}^2}{\sigma_{2t}^2} \right) \right] + \sum_{t=1}^{T} \left[\frac{\eta_{1t}\eta_{2t}}{\sigma_{1t}\sigma_{2t}} \left(\frac{1+\rho^2}{(1-\rho^2)^2} \right) \right]. \tag{5.20}$$

And,

$$\frac{\partial^2 L_T}{\partial \rho^2} = \sum_{t=1}^{T} \left(\frac{1+\rho^2}{(1-\rho^2)^2} \right) - \sum_{t=1}^{T} \left(\frac{\eta_{1t}^2}{\sigma_{1t}^2} + \frac{\eta_{2t}^2}{\sigma_{2t}^2} \right) \left(\frac{1+2\rho^2-3\rho^4}{(1-\rho^2)^4} \right)$$
$$+ \sum_{t=1}^{T} \left(\frac{\eta_{1t}\eta_{2t}}{\sigma_{1t}\sigma_{2t}} \right) \left(\frac{6\rho^2-4\rho^4-2\rho^6}{(1-\rho^2)^4} \right). \tag{5.21}$$

For obtaining the expected Fisher Information term $-E\left(\frac{\partial^2 L_T}{\partial \rho^2}\right)$, we need to calculate the covariance between η_{1t} and η_{2t} which is given by,

$$Cov\left(\eta_{1t}, \eta_{2t}\right) = \rho\sigma_{1t}\sigma_{2t}.$$

Hence,

$$-E\left(\frac{\partial^2 L_T}{\partial \rho^2}\right) = -\sum_{t=1}^{T} \left(\frac{1+\rho^2}{(1-\rho^2)^2} \right) + 2\sum_{t=1}^{T} \left(\frac{1+2\rho^2-3\rho^4}{(1-\rho^2)^4} \right) - \sum_{t=1}^{T} \left(\frac{6\rho^2-4\rho^4-2\rho^6}{(1-\rho^2)^4} \right). \tag{5.22}$$

Differentiating L_T with respect to the cointegration parameters, $\theta_2 = (\gamma, \delta_2)^{'}$ yields,

$$\frac{\partial L_T}{\partial \theta_2} = \sum_{t=1}^{T} \left[\sum_{i=1}^{2} \frac{1}{2\sigma_{it}^2} \frac{\partial \sigma_{it}^2}{\partial \theta_2} \left(\frac{\eta_{it}^2}{(1-\rho^2)\,\sigma_{it}^2} - 1 \right) + \frac{\eta_{it}}{(1-\rho^2)\,\sigma_{it}^2} \frac{\partial \eta_{it}}{\partial \theta_2} \right] +$$

$$\sum_{t=1}^{T} \frac{\rho}{(1-\rho^2)} \left[\frac{\eta_{1t}\eta_{2t}}{2\sigma_{1t}^3\sigma_{2t}} \frac{\partial \sigma_{1t}^2}{\partial \theta_2} + \frac{\eta_{1t}\eta_{2t}}{2\sigma_{2t}^3\sigma_{1t}} \frac{\partial \sigma_{2t}^2}{\partial \theta_2} \right]$$

$$+ \frac{\rho}{(1-\rho^2)} \sum_{t=1}^{T} \left[\eta_{1t} \frac{\partial \eta_{2t}}{\partial \theta_2} + \eta_{2t} \frac{\partial \eta_{1t}}{\partial \theta_2} \right] \tag{5.23}$$

And the expected information matrix for the second block is given by,

$$-E\left(\frac{\partial^2 L_T}{\partial \theta_2 \partial \theta_2^{'}} \right) = \sum_{t=1}^{T} \left[\sum_{i=1}^{2} \frac{1}{\sigma_{it}^4} \frac{\partial \sigma_{it}^2}{\partial \theta_2} \frac{\partial \sigma_{it}^2}{\partial \theta_2^{'}} \left(\frac{1}{(1-\rho^2)} - \frac{1}{2} \right) \right] - \frac{1}{(1-\rho^2)} \sum_{t=1}^{T} \left[\sum_{i=1}^{2} \frac{1}{\sigma_{it}^2} \frac{\partial \eta_{it}}{\partial \theta_2} \frac{\partial \eta_{it}}{\partial \theta_2^{'}} \right]$$

$$- \sum_{t=1}^{T} \frac{\rho}{(1-\rho^2)} \frac{\left[\frac{\partial \eta_{2t}}{\partial \theta_2} \frac{\partial \eta_{1t}}{\partial \theta_2^{'}} + \frac{\partial \eta_{1t}}{\partial \theta_2} \frac{\partial \eta_{2t}}{\partial \theta_2^{'}} \right]}{\sigma_{1t}\sigma_{2t}}$$

$$- \sum_{t=1}^{T} \frac{\rho}{(1-\rho^2)} \left[\frac{\rho}{2\sigma_{1t}^5\sigma_{2t}} \left[\frac{3}{2}\sigma_{1t}\sigma_{2t} \frac{\partial \sigma_{1t}^2}{\partial \theta_2} \frac{\partial \sigma_{1t}^2}{\partial \theta_2^{'}} + \frac{1}{2}\sigma_{2t}^{-1}\sigma_{1t}^3 \frac{\partial \sigma_{2t}^2}{\partial \theta_2} \frac{\partial \sigma_{2t}^2}{\partial \theta_2^{'}} \right] \right]$$

$$- \sum_{t=1}^{T} \frac{\rho}{(1-\rho^2)} \left[\frac{\rho}{2\sigma_{2t}^5\sigma_{1t}} \left[\frac{3}{2}\sigma_{2t}\sigma_{1t} \frac{\partial \sigma_{2t}^2}{\partial \theta_2} \frac{\partial \sigma_{2t}^2}{\partial \theta_2^{'}} + \frac{1}{2}\sigma_{1t}^{-1}\sigma_{2t}^3 \frac{\partial \sigma_{1t}^2}{\partial \theta_2} \frac{\partial \sigma_{1t}^2}{\partial \theta_2^{'}} \right] \right], \tag{5.24}$$

where $\frac{\partial \sigma_{it}^2}{\partial \theta_2} = 2e^{b_i^*}\eta_{it-1}\frac{\partial \eta_{it-1}}{\partial \theta_2} + e^{g_i^*}\frac{\partial \sigma_{it-1}^2}{\partial \theta_2}, i = 1, 2.$ and the derivatives of η_{it} with respect to the cointegration parameters θ_2 are given by, $\frac{\partial \eta_{1t}}{\partial \theta_2} = (x_{2,0} - \eta_{2,0} - x_{2t} + \eta_{2t}, -\gamma)^{'}$ and $\frac{\partial \eta_{2t}}{\partial \theta_2} = (0, -1)^{'}$.

Differentiating L_T with respect to the GARCH parameters, $\theta_3 = (c_1, c_2, b_1, g_1, b_2, g_2)^{'}$ yields,

$$\frac{\partial L_T}{\partial \theta_3} = -\frac{1}{2} \sum_{t=1}^{T} \left[\sum_{i=1}^{2} \frac{1}{\sigma_{it}^2} \left(\frac{\eta_{it}^2}{\sigma_{it}^2(1-\rho^2)} - 1 \right) \frac{\partial \sigma_{it}^2}{\partial \theta_3} \right] - \frac{\rho}{(1-\rho^2)}$$

$$\sum_{t=1}^{T} \frac{\eta_{1t}\eta_{2t}}{\sigma_{1t}^2\sigma_{2t}^2} \left[\frac{1}{2}\sigma_{1t}^{-1}\sigma_{2t}\frac{\partial\sigma_{1t}^2}{\partial\theta_3} + \sigma_{1t}\sigma_{2t}^{-1}\frac{\partial\sigma_{1t}^2}{\partial\theta_3} \right]. \tag{5.25}$$

And the expected information matrix for the third block becomes,

$$J3 = -E\left(\frac{\partial^2 L_T}{\partial\theta_3\partial\theta_3'}\right) = \sum_{t=1}^{T}\left[\sum_{i=1}^{2}\frac{1}{\sigma_{it}^4}\frac{\partial\sigma_{it}^2}{\partial\theta_3}\frac{\partial\sigma_{it}^2}{\partial\theta_3'}\left(\frac{1}{(1-\rho^2)}-\frac{1}{2}\right)\right]$$

$$-\frac{\rho^2}{(1-\rho^2)}\sum_{t=1}^{T}\left[\sum_{i=1}^{2}\left[\frac{3}{2}\sigma_{1t}^2\sigma_{2t}^2\frac{\partial\sigma_{1t}^2}{\partial\theta_3}\frac{\partial\sigma_{1t}^2}{\partial\theta_3'}\right]+\sum_{i=1}^{2}\left[\frac{1}{2}\sigma_{1t}\sigma_{2t}^4\frac{\partial\sigma_{1t}^2}{\partial\theta_3}\frac{\partial\sigma_{2t}^2}{\partial\theta_3}+\frac{3}{2}\sigma_{1t}^2\sigma_{2t}^2\frac{\partial\sigma_{2t}^2}{\partial\theta_3}\frac{\partial\sigma_{2t}^2}{\partial\theta_3'}\right]\right], \tag{5.26}$$

where $\frac{\partial\sigma_{it}^2}{\partial\theta_3} = r_{i,t} + e^{g_i^*}\frac{\partial\sigma_{it-1}^2}{\partial\theta_3}, i = 1,2$ and the vectors $r_{1,t} = \left(e^{c_1^*}, 0, e^{b_1^*}\eta_{1t-1}^2, 0, e^{g_1{}^*}\sigma_{1t-1}^2, 0\right)'$ and $r_{2,t} = \left(0, e^{c_2^*}, 0, e^{b_2^*}\eta_{2t-1}^2, 0, e^{g_2^*}\sigma_{2t-1}^2\right)'$. Once the expected information matrix and the gradients of three blocks are obtained, we can find the maximum likelihood estimators by using Fisher scoring algorithm given in (5.16). A simulation study to illustrate the computations is discussed in Section 5.4.

5.3.2 MLE using some non-Gaussian GARCH errors

In this section, we discuss the estimation procedure for the cointegration GARCH model using a generalised error distribution(GED). The probability density function of a univariate generalised error distribution is defined by,

$$f(x;\mu,\phi,v) = \frac{1}{\phi\Gamma\left(1+\frac{1}{2v}\right)2^{1+\frac{1}{2v}}} exp\left\{-\frac{1}{2}\left|\frac{x-\mu}{\phi}\right|^{2v}\right\}, -\infty < x < \infty,$$

$\mu \in R$ and $\phi, \nu > 0$. The multivariate generalisation of the above model will be useful to model the multidimensional random phenomena that have heavy or thin tails than those of normal distribution. Gómez et al. (1998) introduced a multivariate generalisation of the generalised error distributions for modelling multidimensional random phenomena. Their definition is as follows: A random variable $\mathbf{X} = (X_1, X_2 \cdots\cdots X_n)'$ with $n \geq 1$ has an n-dimensional generalised error distribution with parameters $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_n)$, $\boldsymbol{\Sigma}$, v if its density has the form

$$f(X; \boldsymbol{\mu}, \boldsymbol{\Sigma}, v) = k|\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left[(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right]^{v}\right\},$$

where $k = \frac{n\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{n}{2}}\Gamma\left(1+\frac{n}{2v}\right)2^{1+\frac{n}{2v}}}$. In our case, we assume that $\mathbf{a_t}$ follows the above distribution with $v = 2$, $n = 2$ and pdf:

$$f(\mathbf{a_t}; \mu, \boldsymbol{\Sigma}, \beta) = \frac{2}{\pi\Gamma\left(\frac{3}{2}\right)2^{\frac{3}{2}}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left[(\mathbf{a_t})'\boldsymbol{\Sigma}^{-1}(\mathbf{a_t})\right]^{2}\right\}. \tag{5.27}$$

Further $E(\mathbf{a}_t|I_{t-1}) = \mathbf{0}$ and $V(\mathbf{a}_t|I_{t-1}) = \frac{1}{\sqrt{2\pi}}\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is assumed to be an identity matrix. Now we redefine the bivariate cointegration GARCH model in (5.7) and (5.13) as,

$$\Delta\mathbf{X_t} = \mathbf{A}Z_{t-1} + \boldsymbol{\eta}_\mathbf{t}$$

$$\boldsymbol{\eta}_\mathbf{t} = \mathbf{H}_\mathbf{t}^{1/2}\mathbf{a_t}$$

$$\mathbf{a_t} \mid I_{t-1} \sim GED \tag{5.28}$$

and

$$X_{1t} = \mu + \gamma X_{2t} + (\eta_{1t} - \gamma\eta_{2t})$$

$$\Delta X_{2t} = \delta_2 + \eta_{2t}$$

$$\boldsymbol{\eta_t} = \mathbf{H_t^{1/2} a_t}$$

$$\mathbf{a_t} \mid I_{t-1} \sim GED, \tag{5.29}$$

with shape parameter $v = 2$. Here $E\left(\boldsymbol{\eta}_t | I_{t-1}\right) = \mathbf{0}$, $Cov\left(\boldsymbol{\eta}_t | I_{t-1}\right) = \frac{1}{\sqrt{2\pi}}\mathbf{H_t}$ and as previously, $\mathbf{H_t}$ is defined as,

$$\mathbf{H_t} = \mathbf{D_t \Gamma D_t},$$

where $\mathbf{D_t}$ denotes 2×2 diagonal matrix given by $\mathbf{D_t} = diag(\sigma_{1t}, \sigma_{2t})$ with σ_{it}^2 following a GARCH (1,0) model defined by,

$$\sigma_{it}^2 = c_i + b_i \eta_{it-1}^2, i = 1, 2.$$

As mentioned in Section 5.3.1, the transformations of the positive parameters lead to the variance equation σ_{it}^2, i=1,2 given by

$$\sigma_{it}^2 = e^{c_i^*} + e^{b_i^*} \eta_{it-1}^2, i = 1, 2, t = 1, 2, 3, ... \tag{5.30}$$

To avoid computational difficulties, we assume that the correlation matrix is an identity matrix. Under the assumption of a bivariate generalised error distribution, the conditional likelihood function given (η_{10}, η_{20}) is given by,

$$L(\theta \,| x_t) = \prod_{t=1}^{T} \frac{2}{\pi \Gamma\left(\frac{3}{2}\right) 2^{\frac{3}{2}}} |\mathbf{H_t}|^{-\frac{1}{2}} Exp\left\{-\frac{1}{2}\left(\boldsymbol{\eta_t'} \boldsymbol{H_t^{-1}} \boldsymbol{\eta_t}\right)^2\right\}.$$

And the corresponding log likelihood function becomes

$$L_T(\theta\,|x_t) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{2}\log\left(\sigma_{it}^2\right) - \frac{1}{2}\sum_{t=1}^{T}\left[\sum_{i=1}^{2}\frac{\eta_{it}^2}{\sigma_{it}^2}\right]^2. \tag{5.31}$$

Since a closed form expression for the expected information matrix is not readily available for the chosen distribution, we obtain the parameter estimates by directly optimizing the log likelihood function. In the next section, we simulate observations from the cointegration GARCH model described in (5.28), (5.29) and obtain the estimates using numerical optimization techniques. We used the method of Newton Raphson to solve the equations numerically.

5.4 Simulation study

We carried out a simulation study for computing the estimates and to examine the finite sample performance of the model proposed in previous sections. We used the method of Newton Raphson to obtain the parameter estimates by solving the log likelihood equation for the case of non normal error distribution.

5.4.1 Simulation study for the model (5.7) with Normal errors

Here, we simulate the bivariate data (x_{1t}, x_{2t}) from the model specified in equation (5.7). The maximum likelihood estimates are obtained by solving the gradients and expected information matrix given in Section 5.3.1

We repeat the experiment 500 times for computing the maximum likelihood estimates and then averaged over the repetitions. Computational time for obtaining the estimates is 24 and 48 minutes for the sample sizes 150 and 300 respectively. Tables 5.1 to 5.4 give the average estimates and MSE based on simulated observations. Tables 5.1, 5.3 reports the parameter estimates of cointegration with GARCH parameters b=0.2, g=0.3 and b=0.1, g=0.85 respectively. The corresponding estimates of GARCH parameters are reported in Tables 5.2 and 5.4.

TABLE 5.1: The average estimates of the MLE and MSE (in parenthesis) of cointegration and correlation parameters with GARCH parameters, c1=10, c2=15,b=0.2,g=0.3.

n	ρ	α	β	δ	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$
150	0.1	0.8	1	-4.95	0.1222(0.0251)	0.8003(0.0515)	1.002(0.0032)	-4.974(0.0323)
		2	2.5	-1.4	0.1644(0.0647)	1.9978(0.0027)	2.348(0.1648)	-1.428(0.0366)
		4	4.5	-0.6	0.1232(0.0257)	4.0098(0.0108)	4.3581(0.1010)	-0.6150(0.0196)
	0.8	0.8	1	-4.95	0.8471(0.0510)	0.8010(0.0007)	1.0002(0.0012)	-4.9490(0.0181)
		2	2.5	-1.4	0.9114(0.0517)	1.9990(0.0001)	2.5019(0.0034)	-1.4001(0.0061)
		4	4.5	-0.6	0.7974(0.0284)	4.00003(0.008)	4.5018(0.004)	-0.6005(0.0069)
300	0.1	0.8	1	-4.95	0.1025(0.0143)	0.8004(0.0226)	1.0010(0.0012)	-4.9604(0.0225)
		2	2.5	-1.4	0.1452(0.0456)	2.0021(0.0013)	2.4932(0.0388)	-1.4020(0.0161)
		4	4.5	-0.6	0.1345(0.0121)	4.0001(0.0012)	4.5123(0.0412)	-0.6005(0.0012)
	0.8	0.8	1	-4.95	0.8461(0.0482)	0.8005(0.0003)	0.9999(0.0002)	-4.9510(0.0015)
		2	2.5	-1.4	0.9110(0.0112)	2.0002(0.00003)	2.4997(0.0018)	-1.3990(0.0009)
		4	4.5	-0.6	0.7972(0.0115)	4.0002(0.0040)	4.4923(0.0009)	-0.6002(0.0031)

From Tables 5.1 and 5.3, we can see that the estimates of ρ are slightly biased, compared to the other estimates. But when the sample size increases, the estimates perform reasonably well and there is a significant reduction in MSE. Further we can see that, the estimates of δ, β, α perform well and the mean square error decreases when the sample size increases. Table 5.2, 5.4 give the average values and MSE of the GARCH parameters for different choice of GARCH parameters. It can be seen that mean square errors of the estimates decreases as the value of g increases.

TABLE 5.2: Average values of the MLE and MSE (in paranthesis) of GARCH parameters c1=10, c2=15,b=0.2,g=0.3.

n	ρ	$\hat{c_1}$	$\hat{c_2}$	$\hat{b}$	$\hat{g}$
150	0.1	10.4361(0.0815)	15.5505(0.0688)	0.1906(0.0626)	0.3001(0.0031)
		10.4051(0.0491)	15.4914(0.0441)	0.1904(0.0646)	0.3028(0.0104)
		9.9912(0.0029)	15.0108(0.0087)	0.200(0.0093)	0.3001(0.0006)
	0.8	9.9916(0.0018)	15.0153(0.0088)	0.2000(0.0093)	0.3001(0.0005)
		9.9916(0.0018)	15.0115(0.0088)	0.20002(0.0093)	0.3001(0.0051)
		9.9912(0.00178)	15.0108(0.0087)	0.2000(0.0093)	0.3001(0.0068)
300	0.1	10.4142(0.0488)	15.5440(0.0281)	0.1914(0.0585)	0.300(0.0012)
		10.4570(0.0320)	15.582(0.0106)	0.1909(0.0144)	0.2988(0.0045)
		9.9920(0.0018)	15.0119(0.0088)	0.2001(0.0053)	0.3001(0.0004)
	0.8	9.9910(0.0017)	15.0111(0.0060)	0.2000(0.0093)	0.30016(0.0006)
		9.9916(0.0018)	15.0110(0.0042)	0.2000(0.0061)	0.3010(0.0021)
		10.0020(0.0002)	15.0006(0.0022)	0.1990(0.0029)	0.3010(0.0015)

TABLE 5.3: The average estimates of the MLE and MSE (in parenthesis) of cointegration and correlation parameters with GARCH parameters, c1=10, c2=15,b=0.1, g=0.85.

n	ρ	α	β	δ	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$
150	0.1	0.8	1	-4.95	0.1807(0.0807)	0.8002(0.0026)	0.9999(0.0002)	-4.9490(0.0612)
		2	2.5	-1.4	0.1905(0.0905)	2.0003(0.00048)	2.5078(0.0101)	-1.3906(0.0129)
		4	4.5	-0.6	0.18110(0.0811)	4.00074(0.0008)	4.5149(0.0172)	-0.5960(0.0044)
	0.8	0.8	1	-4.95	0.9472(0.1473)	0.8001(0.0008)	1.0002(0.0004)	-4.95002(0.0004)
		2	2.5	-1.4	0.9617(0.1610)	2.000(0.0007)	2.5003(0.0003)	-1.399(0.0002)
		4	4.5	-0.6	0.9366(0.1368)	4.0001(0.0002)	4.5003(0.0004)	-0.5999(0.006)
300	0.1	0.8	1	-4.95	0.1811(0.0411)	0.8003(0.0012)	1.0001(0.0001)	-4.948(0.0018)
		2	2.5	-1.4	0.1906(0.0407)	2.0007(0.0021)	2.5063(0.0011)	-1.3921(0.0086)
		4	4.5	-0.6	0.1100(0.0413)	4.0005(0.0006)	4.5109(0.0012)	-0.5971(0.0034)
	0.8	0.8	1	-4.95	0.9450(0.0176)	0.7999(0.0005)	1.0002(0.0021)	-4.9499(0.0024)
		2	2.5	-1.4	0.9613(0.0512)	1.9999(0.0030)	2.4999(0.0001)	-1.3998(0.0001)
		4	4.5	-0.6	0.9377(0.0519)	4.00001(0.0001)	4.5007(0.0001)	-0.5999(0.0008)

Table 5.4: The average estimates of the MLE and MSE (in parenthesis) of GARCH parameters c1=10, c2=15,b=0.1, g=0.85.

n	ρ	$\hat{c_1}$	$\hat{c_2}$	$\hat{b}$	$\hat{g}$
150	0.1	10.1015(0.0027)	14.9910(0.0081)	0.0999(0.0029)	0.8508(0.0001)
		10.1010(0.0026)	14.9981(0.0080)	0.0998(0.0028)	0.8650(0.0001)
		10.2015(0.0027)	14.9792(0.0080)	0.0997(0.0029)	0.8601(0.0027)
	0.8	9.9950(0.0025)	15.1012(0.0080)	0.100(0.0025)	0.8508(0.0001)
		9.9990(0.0025)	15.1101(0.0081)	0.1001(0.0025)	0.8512(0.0001)
		9.9990(0.0025)	15.0011(0.0082)	0.1001(0.0025)	0.8511(0.0001)
300	0.1	9.9989(0.0022)	15.0001(0.0051)	0.1001(0.0012)	0.8503(0.00001)
		9.9991(0.0013)	15.0001(0.0011)	0.1001(0.0011)	0.8502(0.0001)
		10.0013(0.0017)	14.9990(0.0047)	0.0999(0.0019)	0.8503(0.00002)
	0.8	10.0002(0.0015)	15.0030(0.0042)	0.1005(0.0015)	0.8499(0.0016)
		9.9998(0.0025)	15.0020(0.0039)	0.1001(0.0020)	0.8502(0.0001)
		9.9990(0.0015)	15.0010(0.0040)	0.1002(0.0017)	0.8500(0.00001)

Algorithm 10: Algorithm for Table 5.1 to 5.4

1 Set x_{10}, x_{20}, σ_{10}, σ_{20}, θ_1, θ_2, θ_3 for a sample of size n.
2 Generate the bivariate samples $\{\mathrm{x}_{1t}, x_{2t}\}$ using (5.7).
3 Compute the gradients and the expected information matrix in Equation (5.20) to (5.26).
4 Compute the the estimates $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$ by using the Fisher scoring equation in (5.16)
5 Repeat steps 2 to 4 say 500 times.
6 Choose the value of $\hat{\theta}_i$ as the averages of $\hat{\theta}_i$ obtained in step 6 for $i = 1, 2$.

5.4.2 Simulation Study for the model (5.28) with GED errors

In this case, the bivariate data (x_{1t}, x_{2t}) is simulated from the model given in (5.28). Then we maximize the log likelihood function in (5.31) using likelihood numerical optimization and hence obtained the parameter estimates.

We repeat the experiment 500 times for computing the estimates and then averaged them over the repetitions. Computational time for obtaining the estimates is 20 and 40 minutes for sample sizes 150 and 300 respectively. Tables 5.5, 5.6

give the average estimates and MSE based on simulated observations. From the tables, we observe that, the estimates perform reasonably well with the increase of sample size and there is a significant reduction in MSE. Also, the estimates of δ, β, α perform well as sample size increases. For instance, if $\delta = -1$, $\beta = 3.5$, $\alpha = 3$, the average $\hat{\delta} = -0.9672$, $\hat{\beta} = 3.4875$, $\hat{\alpha} = 2.999$, for n=150. When the sample size is increased to 300, the estimates of δ, β and α are respectively -1.005, 3.50 and 3.0003. Further the mean square error of the estimates decrease when the sample size increases. The estimates of GARCH parameters also behave in a similar way.

Table 5.5: Average values of the MLE and MSE (in parenthesis) of Cointegration parameters based on simulated observations of sample sizes n=150,300.

n	δ	α	β	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\beta}$
150	-4.95	0.8	1	-4.9570(0.0428)	0.80004(0.0007)	1.0080(0.0032)
	-1.8	2	2.5	-1.8009(0.0156)	2.0001(0.0015)	2.5102(0.0481)
	-1	3	3.5	-0.9672(0.0142)	2.999(0.00004)	3.4875(0.3027)
300	-4.95	0.8	1	-4.948(0.0211)	0.7999(0.0003)	0.9998(0.0016)
	-1.8	2	2.5	-1.8001(0.0079)	1.999(0.00007)	2.5224(0.0286)
	-1	3	3.5	-1.005(0.0072)	3.0003(0.00146)	3.5065(0.1420)

Table 5.6: Average values of the MLE and MSE (in paranthesis) of GARCH parameters

n	c_1	c_1	b	$\hat{c_1}$	$\hat{c_2}$	$\hat{b}$
150	6	12	0.4	5.9930(0.0200)	12.016(0.0201)	0.4192(0.0510)
	10	16	0.6	9.9721(0.0230)	15.9486(0.0224)	0.5920(0.0602)
	18	22	0.7	1.5232(0.0269)	21.9321(0.0275)	0.7100(0.0311)
300	6	12	0.4	6.0580(0.0109)	12.1931(0.0124)	0.4116(0.0391)
	10	16	0.6	10.0703(0.0113)	16.4862(0.0116)	0.6109(0.0487)
	18	22	0.7	18.2952(0.0111)	22.4604(0.0114)	0.7082(0.0271)

Algorithm 11: Algorithm for Table 5.5 to 5.6

1 Set x_{10}, x_{20}, σ_{10}, σ_{20}, α, β, δ, c_i, b_i, $i = 1, 2$ for a sample of size n.
2 Draw GARCH samples $\boldsymbol{\eta}_t$ in Equation (5.28).
3 Generate the bivariate samples $\{x_{1t}, x_{2t}\}$ using (5.28)
4 Choose the initial values α_0, β_0, δ_0, c_{i0}, b_{i0} and obtain the estimates, $\hat{\alpha}$, $\hat{\beta}$, $\hat{\delta}$, $\hat{c}_i$, $\hat{b}_i$ by solving Equation (5.31).
5 Repeat Steps 2 to 4 say 500 times.
6 Choose the value of estimates as the averages of estimates obtained in step 6.

5.4.3 Simulation study for the model (5.13) with Normal errors

Here, we demonstrate the estimation procedure described in Section 5.1 using a simulated sample. For the simulation study, we first generate the innovation random variables from a Normal distribution. Then for specified values of model parameters, we obtained the bivariate observations from the model given by (5.13). Based on this sample, we obtained the maximum likelihood estimates by solving the gradients and expected information matrix given in Section 5.3.1. The experiment was repeated 500 times for computing the maximum likelihood estimates and then averaged over the repetitions. Computational time for obtaining the estimates of the well specified model is 4 and 9 minutes for the sample sizes 150 and 300 respectively. All the computations in section 5.3.1 are carried out using the software Mathematica with an i3 CPU and 2.93 G.Hz. Tables 5.7 to 5.10 give the average estimates and mean squared error(MSE) based on simulated observations.

In order to better describe the performance of the proposed model, we compare the estimation results on the simulated data obtained both by the proposed model (well specified) and by a traditional co-integration model without

GARCH error (miss-specified). The evaluation of the performance obtained by the well specified and the misspecified model can help to understand how the error-misspecification affect the parameter estimation and it will provide further motivation to use the proposed model. Mis-specified model means, a model that has an incorrect functional form. Misspecification can arise either because of the omission of a variable specified by truth, or because of inclusion of a variabe not specified by truth. Tables 5.7, 5.9 reports the estimates of the cointegration parameters with well specified (Model specified in Section 5.2) and miss-specified innovations. The corresponding estimates of the GARCH parameters for the proposed model with well specified errors are reported in Tables 5.8 and 5.10. $\hat{\gamma}_{miss.spec}$ and $\hat{\delta}_{miss.spec}$ correspond the parameter estimates of a miss-specified innovations. From the tables it can be seen that, when the sample size increases, the estimates perform reasonably well and there is a significant reduction in MSE. From Tables 5.7 and 5.9, it can be seen that, the parameter estimates corresponds to the miss-specified innovations are little biased compare to the estimates of the well-specified innovations.

Table 5.7: The average estimates of the MLE and MSE (in parenthesis) of cointegration and correlation parameters with GARCH parameters, c1=3, c2=4, b1=0.15, b2=0.1, g1=0.8, g2=0.85

n	ρ	γ	δ_2	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\delta}_2$	$\hat{\gamma}_{miss.spec}$	$\hat{\delta}_{miss.spec}$
150	0.8	-1	3	0.7801(0.0192)	-0.9870(0.0031)	3.1702(0.1198)	-1.5120(0.1121)	2.6120(0.2771)
		-2	6	0.7601(0.0192)	-2.1091(0.0090)	6.1495(0.1760)	-2.5120(0.1321)	5.5212(0.5650)
		3	2	0.7800(0.0200)	2.8991(0.0043)	2.1190(0.1301)	3.411(0.1102)	1.6012(0.5521)
300	0.8	-1	3	0.7902(0.0091)	-1.1021(0.0003)	2.9783(0.0270)	-1.4951(0.0690)	2.7094(0.1122)
		-2	6	0.7801(0.0080)	-2.1022(0.0041)	6.1062(0.0811)	-2.4651(0.0780)	5.4191(0.2565)
		3	2	0.8010(0.0091)	3.0201(0.0011)	2.0443(0.0741)	3.3011(0.0662)	1.6992(0.3742)

TABLE 5.8: Average values of the MLE and MSE (in paranthesis) of GARCH parameters c1=3, c2=4, b1=0.15, b2=0.1, g1=0.8, g2=0.85

n	$\hat{c}_1$	$\hat{c}_2$	$\hat{b}_1$	$\hat{b}_2$	$\hat{g}_1$	$\hat{g}_2$
150	2.9711(0.0223)	4.0524(0.0131)	0.1401(0.0040)	0.1101(0.0053)	0.8194(0.006)	0.8692(0.0044)
	2.9112(0.0084)	3.9321(0.0176)	0.1396(0.0094)	0.0897(0.0055)	0.8210(0.0045)	0.8651(0.0042)
	2.9913(0.0083)	4.1094(0.0173)	0.1675(0.0064)	0.0954(0.0055)	0.7985(0.0025)	0.8665(0.0045)
300	2.9950(0.0061)	4.0104(0.0022)	0.1453(0.0031)	0.1091(0.0035)	0.8092(0.0043)	0.8591(0.0023)
	2.9913(0.0051)	4.0250(0.0024)	0.1581 (0.0045)	0.1156(0.0031)	0.8101(0.0035)	0.8596(0.0022)
	2.9942(0.0045)	4.0144(0.0015)	0.1559(0.0047)	0.1168(0.0021)	0.7991(0.0011)	0.8601(0.0033)

TABLE 5.9: The average estimates of the MLE and MSE (in parenthesis) of cointegration and correlation parameters with GARCH parameters, c1=5, c2=7, b1=0.1, b2=0.08, g1=0.85, g2=0.9, ρ=0.3.

n	ρ	γ	δ_2	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\delta}_2$	$\hat{\gamma}_{miss.spec}$	$\hat{\delta}_{miss.spec}$
150	0.3	-1	3	0.3342(0.0481)	-0.9855(0.0034)	3.1746(0.2566)	-1.6164(0.1015)	3.4212(0.3124)
		-2	6	0.3345(0.0644)	-1.9122(0.0053)	6.2302(0.1046)	-2.5125(0.1105)	6.5124(0.2125)
		3	2	0.3291(0.0493)	3.1021(0.0092)	2.2022(0.1041)	3.4211(0.1112)	2.5120(0.2211)
300	0.3	-1	3	0.3110(0.0213)	-1.0252(0.0021)	2.9846(0.1412)	-1.5112(0.0612)	3.3002(0.1122)
		-2	6	0.3312(0.0212)	-1.9993(0.0031)	6.1101(0.0602)	-2.4127(0.0508)	6.4218(0.1016)
		3	2	0.3046(0.0217)	2.9984(0.0046)	1.9952(0.0662)	3.3012(0.0600)	2.4102(0.1122)

TABLE 5.10: Average values of the MLE and MSE (in paranthesis) of GARCH parameters c1=5, c2=7, b1=0.1, b2=0.08, g1=0.85, g2=0.9, ρ=0.3.

n	$\hat{c}_1$	$\hat{c}_2$	$\hat{b}_1$	$\hat{b}_2$	$\hat{g}_1$	$\hat{g}_2$
150	4.9859(0.0034)	6.8901(0.0083)	0.0983(0.0095)	0.0774(0.0053)	0.8351(0.0053)	0.8861(0.0023)
	4.9862(0.0031)	6.9512(0.0092)	0.1021 (0.0062)	0.0754(0.0032)	0.8643(0.0064)	0.9212(0.008)
	5.1215(0.0054)	6.9120(0.0072)	0.1172(0.0073)	0.0854(0.0052)	0.8424(0.0054)	0.9126(0.0034)
300	4.9990(0.0011)	6.9952(0.0054)	0.0994(0.0055)	0.0800(0.0024)	0.8495(0.0035)	0.9104(0.0014)
	5.0116(0.0024)	7.0214(0.0056)	0.0915 (0.0044)	0.0814(0.0015)	0.8594(0.0035)	0.9014(0.0042)
	4.9992(0.0031)	6.9911(0.0031)	0.0991(0.0035)	0.0802(0.0022)	0.8511(0.0022)	0.9013(0.0013)

5.4.4 Simulation Study for the model (5.29) with GED errors

In this case, the bivariate data (x_{1t}, x_{2t}) is simulated from the model given in (5.29). Then we maximize the log likelihood function in (5.31) using likelihood numerical optimization and hence obtained the parameter estimates.

Algorithm 12: Algorithm for Table 5.7 to 5.10

1 Set x_{10}, x_{20}, σ_{10}, σ_{20}, θ_1, θ_2, θ_3 for a sample of size n.
2 Draw GARCH samples $\boldsymbol{\eta}_t$ in Equation (5.13).
3 Generate the bivariate samples $\{\mathrm{x}_{1t}, x_{2t}\}$ using (5.13)
4 Compute the gradients and the expected information matrix in Equation (5.20) to (5.26).
5 Compute the Fisher Scoring Algorithm in Equation(5.16) to get the estimates $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$.
6 Repeat steps 2 to 6 say 500 times.
7 Choose the value of $\hat{\theta}_i$ as the averages of $\hat{\theta}_i$ obtained in step 6, $i = 1, 2$.

We repeat the experiment 500 times for computing the estimates and then averaged them over the repetitions. Computational times for obtaining the estimates of the well specified model is 6 and 13 minutes for sample sizes 150 and 300 respectively. Tables 5.11, 5.12 give the average estimates and MSE based on simulated observations. From the tables, we observe that, the estimates perform reasonably well with the increase of sample size. Further the mean square error of the estimates decrease when the sample size increases. From Table 5.11, it can be seen that the parameter estimates corresponds to the miss-specified model is little biased compared to the proposed model with GARCH innovations.

TABLE 5.11: Average values of the MLE and MSE (in parenthesis) of Cointegration parameters based on simulated observations of sample sizes n=150,300.

n	γ	δ_2	$\hat{\gamma}$	$\hat{\delta}_2$	$\hat{\gamma}_{miss.spec}$	$\hat{\delta}_{miss.spec}$
150	1	3	1.1122(0.0062)	3.1020(0.0510)	1.3101(0.1210)	3.5612(0.1112)
	2	1	1.9882(0.0034)	0.9795(0.0520)	2.4131(0.1020)	1.4013(0.1215)
	3	0.5	3.1115(0.0020)	0.5203(0.0122)	3.4002(0.2134)	0.8124(0.1105)
300	1	3	0.9984(0.0022)	3.0012(0.0354)	1.3013(0.0623)	3.3213(0.0515)
	2	1	2.0197(0.0015)	1.0056(0.0376)	2.3908(0.6108)	1.3905(0.0721)
	3	0.5	2.9925(0.0016)	0.4998(0.0086)	3.3115(0.1103)	0.7015(0.0411)

TABLE 5.12: Average values of the MLE and MSE (in paranthesis) of GARCH parameters

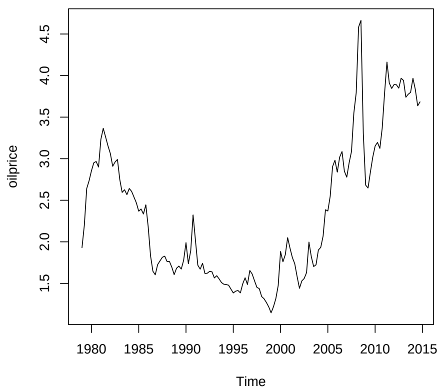
n	c_1	c_2	b_1	b_2	$\hat{c}_1$	$\hat{c}_2$	$\hat{b}_1$	$\hat{b}_2$
150	2	3	0.3	0.4	1.9631(0.1022)	2.9162(0.1413)	0.3105(0.1002)	0.3920(0.0962)
	5	6	0.5	0.7	4.9653(0.1262)	5.9231(0.1322)	0.4702(0.1203)	0.6863(0.1125)
	0.5	0.6	0.8	0.9	0.5015(0.0543)	0.6153(0.1251)	0.7622(0.1435)	0.8473(0.1434)
300	2	3	0.3	0.4	1.9901(0.0623)	2.9912(0.0512)	0.3053(0.0721)	0.4002(0.0782)
	5	6	0.5	0.7	4.9893(0.0653)	6.0200(0.0752)	0.4842(0.0402)	0.7053(0.0663)
	0.5	0.6	0.8	0.9	0.4932(0.0301)	0.5992(0.0083)	0.7881(0.0632)	0.8931(0.0753)

5.5 Data Analysis

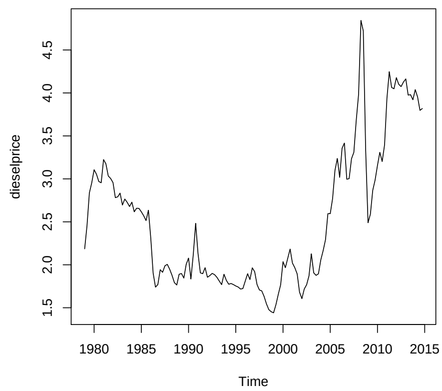
We now analyse two data sets to illustrate the applications of cointegrated GARCH model described in Section 5.3.1 and 5.3.2.

Example 1- Oil and Diesel Price series

The data set consists of 144 quarterly real value price series of heating Oil and Diesel from 1979 to 2014. Units are measured in US dollars/gallon. The data set is downloaded from the website of data market (www.datamarket.com). Seasonally adjusted values of the price series are considered for the further analysis. Figure 5.1 provides the time series plot of the data and it indicates that the time series are non stationary. We have confirmed the non stationarity of time series by using Augmented Dickey Fuller test. Here we test the null hypothesis of non stationary time series against the alternative hypothesis of stationarity about a linear time trend.



(A) Oil Price



(B) Diesel price

FIGURE 5.1: Quarterly price data

The p-values for Augmented Dickey Fuller test of the oil and diesel price is obtained as 0.215 and 0.290 respectively, and the corresponding lag order for both the series is obtained to be 2. The p-values indicate that both the series are non stationary. The summary statistics of the price series are reported in Table 5.13.

TABLE 5.13: Summary statistics of the price series ($ | gallon)

Statistics	Oil price	Diesel Price
Sample size	144	144
Std. Dev	0.861	0.818
Mean	2.359	2.527
Minimum	1.146	1.441
Maximum	4.663	4.843

Next we perform the maximum likelihood estimation method described in Section 5.3.1 for the model (5.13), to find the parameter estimates of the model. The Akaike Information Criterion(AIC) for the model with one and two lags are 51525 and 62349 respectively.

It is found that AIC is minimum for the CVMA(1) model compared to CVMA(2), and hence we use the model with one lag for the given data. In Table 5.14, we present the parameter estimates of model (5.13) with well specified and miss-specified errors.

TABLE 5.14: MLE of Cointegration parameter under well specifed and misspecified errors

Parameters	Estimates	Std.Error
ρ	0.983	0.0004
γ	-0.996	0.0004
δ	-0.008	0.0202
$\gamma_{miss.spec}$	-2.239	0.8816
$\delta_{miss.spec}$	-1.431	1.3250

From the table it can be seen that, the parameter estimates of the model with the error misspecification is biased and also the standard error of the estimates has increased. Therefore it is evident that the proposed cointegration model with GARCH errors using normal innovations are suitable for the given data. The estimated parameter values of GARCH model and standard errors are given in Table 5.15. Next we have to test the adequacy of the model by checking the

TABLE 5.15: MLE of GARCH parameters

Parameters	Estimates	Std.Error
c1	0.297	0.006
c2	0.002	0.0001
b1	0.030	0.012
b2	0.041	0.041
g1	0.961	0.028
g2	0.930	0.006

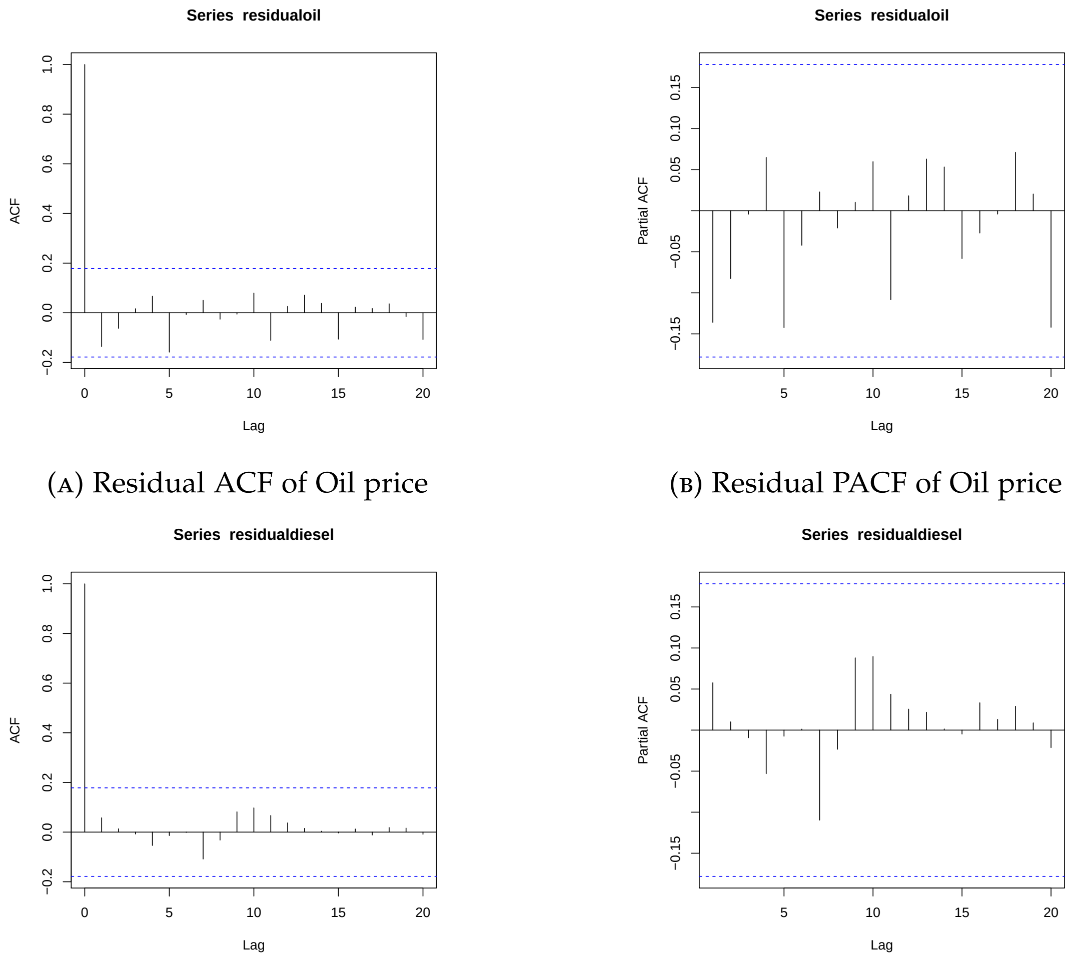
validity of the assumptions imposed on errors. In Fig 5.3, we superimpose the histogram of the residuals with normal density to check whether the series follows normal distribution. We have also validated the assumption of normality of errors using CramerVonMises test. The p-values obtained are 0.231 and 0.552 respectively for oil and diesel price series. The p-values indicate that the null hypothesis of normality is not rejected at 10 percent level of significance.

For a visual check, we plot the ACF of the fitted residuals of the GARCH series in Figure 5.2. It is observed that the ACF of the resulting residual series is negligible and hence there is no significant serial correlation among the residuals. The ACF, PACF and histogram of the residual series suggest that the cointegration GARCH model with normal errors is a good fit for the above data sets.

Example 2- Palm oil and Soya bean oil Price series

The data set consists of 147 quarterly real price series of Palm oil and Soya bean Oil with reference to US market from 1980 to 2016. The data set is downloaded from the website of data market (www.datamarket.com). The prices are seasonally adjusted, which is with respect to the average of the global quotations and the units are measured in US dollars. Alias & Othman (1997) present a cointegration approach to ascertain whether there exist a long run relationship between palm oil price and Soya bean oil price under the normality of errors. Their study concluded that the time series are cointegrated and hence there exists a long run equilibrium relationships between the variables.

For our study, we transformed the variables to their natural logarithm. The time series plots of the log transformed data is given in Figure 5.4 and it indicates that the time series is non stationary. The p-values for ADF test are obtained as



(A) Residual ACF of Oil price

(B) Residual PACF of Oil price

(C) Residual ACF of Diesel price

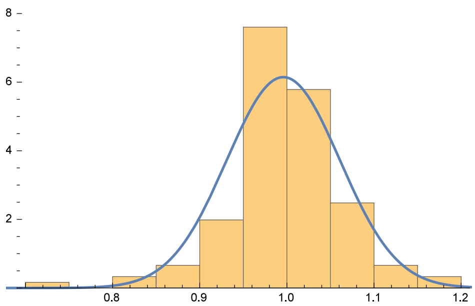
(D) Residual PACF of Diesel price

FIGURE 5.2: ACF and PACF from fitted GARCH series

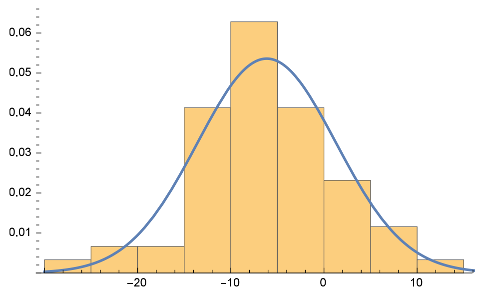
0.077 and 0.183 respectively, indicates that both the variables are non stationary with respect to the linear time trend and the corresponding lag order is obtained to be 2. The summary statistics for the price series are given in Table 5.16.

TABLE 5.16: Summary statistics of the price series ($ | gallon)

Statistics	Palm Oil	Soyabean Oil
Sample size	147	147
Std. Dev	0.428	0.346
Mean	6.106	6.341
Minimum	5.183	5.767
Maximum	7.074	7.173

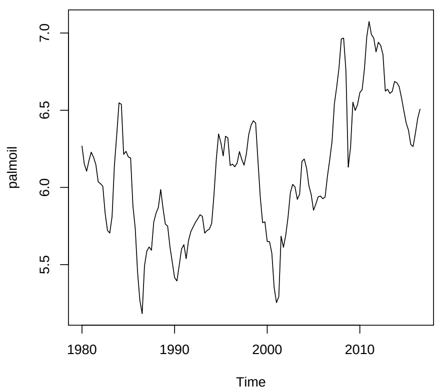


(A) Normal Histogram of oil price

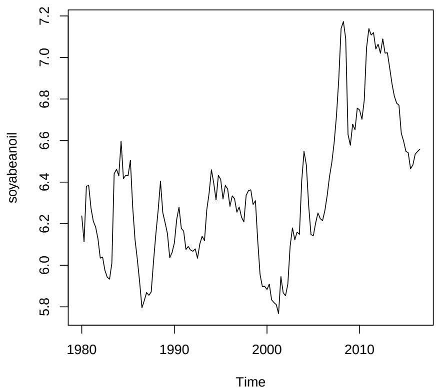


(B) Normal Histogram of diesel price

FIGURE 5.3: Histogram of the residuals of fitted GARCH series



(A) Palm oil price



(B) Soyabean Oil price

FIGURE 5.4: Quarterly Oil price data

The AIC for the model with one and two lags are 31218 and 36775 respectively. Since the CVMA(1) model results with the minimum AIC, we proceed with the model (5.29) for fitting the cointegration GARCH model.

We fit a cointegration GARCH model described in Section 5.3.2 for the given data set. The maximum likelihood estimates described in Section 5.2 of model (5.29) are obtained and reported in Tables 5.17 and 5.18.

The parameter estimates of the model with the error misspecification is biased and also the standard error of the estimates have increased. Therefore the

TABLE 5.17: MLE of Cointegration parameters under well specified and mis-specified errors

Parameters	Estimates	Std.Error
δ	-0.755	0.062
γ	0.976	0.052
$\delta_{miss.spec}$	-2.612	0.511
$\gamma_{miss.spec}$	0.653	0.131

proposed model with GARCH errors is more suitable for the given data.

TABLE 5.18: MLE of GARCH parameters

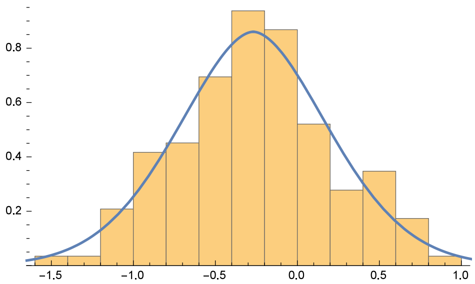
Parameters	Estimates	Std.Error
c1	0.090	0.021
c2	0.434	0.066
b1	0.845	0.016
b2	0.907	0.002

Next we have to test the adequacy of the model by checking the validity of the assumptions imposed on errors. In Figure 5.5, we superimpose the histogram of the residuals with GED density to check whether the series follows Generalised Error distribution. We have also validated the assumption of errors using CramerVonMises test.

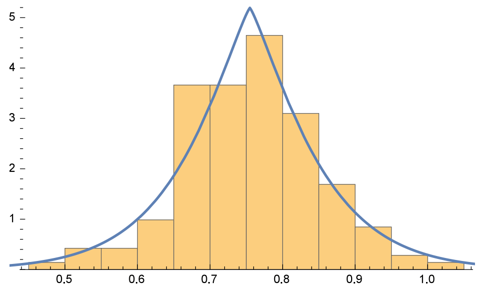
The p-values obtained are 0.960 and 0.889 respectively for Palm oil and Soya bean oil price series. The p-values indicate that the null hypothesis of GED errors is not rejected. The squared ACF and PACF of the fitted residuals of the GARCH series is shown in Figure 5.6.

The residual ACF plots and residual histogram suggest that the cointegration GARCH model with GED errors is a good fit for the above data set.

The results of this Chapter are summarized in Nimitha & Balakrishna (2018a).

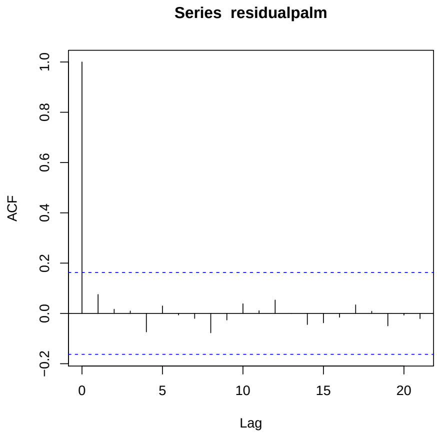


(A) Histogram of Palm Oil price with shape paraameter $v = 1.79$

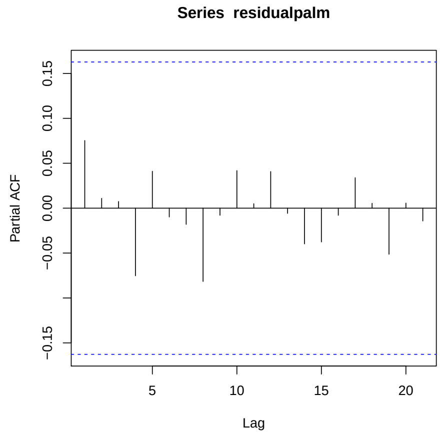


(B) Histogram of Soyabean Oil price with shape parameter $v = 1.85$

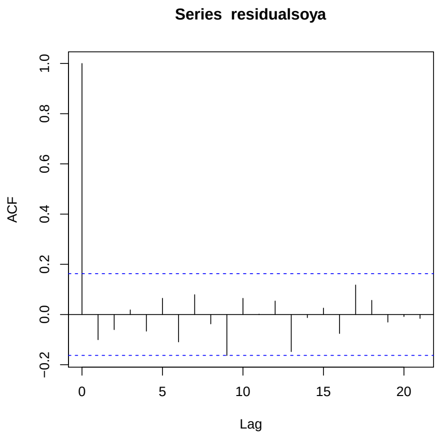
FIGURE 5.5: Histogram of the Residual series



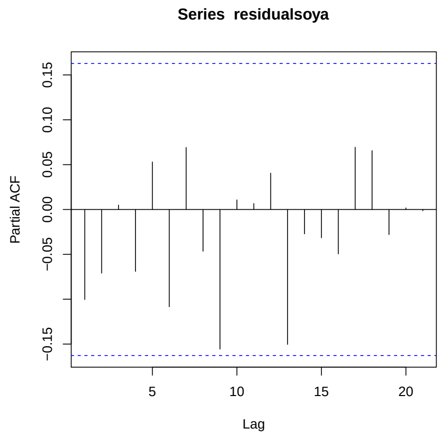
(A) Residual ACF of Palm Oil price



(B) Residual PACF of Palm Oil price



(C) Residual ACF of Soyabean Oil price



(D) Residual PACF of Soyabean Oil price

FIGURE 5.6: Residual ACF and PACF from fitted GARCH series

Chapter 6

Copula based bivariate cointegration model

6.1 Introduction

There is a vast literature on modelling and analysis of financial time series using copula, but the use of copula in cointegration modelling is virtually nil. Some of the most influential papers on copula in finance is Cherubini et al. (2004), which addressed the mathematics of copula functions illustrated using financial applications, Chen & Fan (2006), which developed the estimation of copula based semi parametric stationary models, Lee & Long (2009), developed a copula based multivariate GARCH model and Scaillet & Fermanian (2002), discussed a non parametric method of estimation of copulas in time series model. Sklar (1959), justified the importance of modelling the distribution of a multivariate random variable through a copula. By coupling different marginal distributions with different copula functions, it is possible to construct a wide variety of copula based time series models and it helps one to model the dependence structure and the marginal behaviour separately. Due to this flexibility, copulas have gained much attention in finance and economics in the past few years. Joe & Xu (2016) proposed a method for estimating the parameters separately by maximizing the marginal likelihoods and then estimating the dependence parameter from the joint likelihood function. Choroś et al. (2010) provided a survey of estimation methods including both parametric and non-parametric, on copula based time series models. While modelling the dependence structure between the variables using a copula, it is also of interest to find the long run relationship between the financial variables, that can be captured through cointegration.

A famous theorem that describes the relationship between copula function and the joint distribution is due to Sklar (1959). The theorem is stated as follows:

Theorem 6.1. *Let $F_{(X_1,X_2)}$ be a joint distribution function with margins F_{X_1} and F_{X_2}. Then there exists a copula C such that for all X_1,X_2 in $\bar{R}$,*

$$F_{(X_1,X_2)}\left(x_1,x_2\right)=C\left(F_{X_1}\left(x_1\right),F_{X_2}\left(x_2\right)\right).$$

If F_{X_1} and F_{X_2} are continuous, then C is unique; otherwise C is uniquely determined on RanF $\times$ RanG, where RanF is the range of F. Conversely, if C is a copula and F_{X_1} and F_{X_2} are distribution functions, then the function $F_{(X_1,X_2)}$ defined above is a joint distribution functions with margins F_{X_1} and F_{X_2}.

To the best of our knowledge, modelling and estimation on copula based cointegration model has not yet been studied in literature. Hence this Chapter takes the advantage of modelling copulas and cointegration jointly. Clearly this is important, since most of the financial multivariate series exhibit dependence structures between themselves and hence the dependence can be modelled with the help of a suitable copula function, and at the same time the marginal distributions need not be necessarily identical. Unlike the usual assumption of identical distributions for all the set of marginals, we obtain the joint distribution function by considering different marginal distributions for the error terms. The rest of the Chapter is organised as follows. In Section 6.2, we discuss the cointegrated model with the errors generated using a Clayton copula with different marginals. In Section 6.3, we discuss the inference procedures of the model. Algorithm for random number simulation for the obtained bivariate density is given in Section 6.4. Section 6.5 deals with the simulation studies and results. We then apply the model to a real data set in Section 6.6.

6.2 Model and Properties

Our focus here is the cointegration model generated by copula errors in which the marginal distributions and copula function is completely specified. The explicit form of a bivariate cointegration model can be written as,

$$\Delta X_{1t} = \delta\beta Z_{t-1} + \eta_{1t} \tag{6.1}$$

$$\Delta X_{2t} = -\delta Z_{t-1} + \eta_{2t}, t = 1, 2, ... \tag{6.2}$$

where $Z_{t-1} = X_{1t-1} + \alpha X_{2t-1}$, see Engle & Granger (1987) for more details. Equations (6.1) and (6.2) correspond to the cointegrated vector error correction model with lag one. The cointegrated vector error correction model with higher lags may be needed for obtaining the correct lag length of the cointegration while modelling a real set of data. We have obtained the cointegrated vector error correction model with lags two by following the method described in

Chapter 4, and is given by

$$\Delta X_{1t} = \beta\delta Z_{t-1} + \alpha\delta\Delta Z_{t-1} + \eta_{1t}$$
$$\Delta X_{2t} = -\delta Z_{t-1} + \delta\Delta Z_{t-1} + \eta_{2t},$$

where $\Delta Z_{t-1} = \Delta X_{1t-1} + \beta\Delta X_{2t-1}$. In a similar way, we can obtain the expression for the cointegration model with lags three, four and so on. Now we assume that the innovation random variables η_{1t} and η_{2t} defined in (6.1) and (6.2) follow non-identical marginals, specifically Logistic and Normal distributions with the respective densities given by,

$$f(\eta_{1t}) = \frac{e^{\frac{-\eta_{1t}}{s}}}{s(1+e^{\frac{-\eta_{1t}}{s}})^2}, -\infty < \eta_{1t} < \infty, s > 0$$

$$f(\eta_{2t}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\eta_{2t}^2}{2\sigma^2}}, -\infty < \eta_{2t} < \infty, \sigma > 0.$$

It is possible to construct a bivariate distribution or density function, having the desired marginal distribution along with a chosen dependence structure, that is a copula. Conversely with every bivariate distribution functions $F_{(X_1,X_2)}$ with corresponding marginal distributions F_{X_1} and F_{X_2}, there associates a unique function, $C : [0,1] \times [0,1] \rightarrow [0,1]$ called a copula such that,

$$F_{(X_1,X_2)}(x_1, x_2) = C\left(F_{X_1}(x_1), F_{X_2}(x_2)\right).$$

One may refer Sklar (1959) for details on copula. Analogously, for a given copula C, there exist a unique survival copula $\hat{C}$ such that,

$$\hat{C}(u,v) = u + v - 1 + C(1-u, 1-v).$$

Hence the advantage of the copula approach is that we have the freedom to choose the copula function and margins separately. Let us consider the Clayton survival copula given by:

$$\hat{C}_\theta\left(u,v\right) = \left(u^{-\frac{1}{\theta}} + v^{-\frac{1}{\theta}} - 1\right)^{-\theta}, \theta > 0, \tag{6.3}$$

with the corresponding conditional copula densities are:

$$\hat{C}_{v|u}(u,v) = \frac{\partial}{\partial u}\hat{C}\left(u,v\right) = u^{-1-\frac{1}{\theta}}\left(-1 + u^{-\frac{1}{\theta}} + v^{-\frac{1}{\theta}}\right)^{-1-\theta},$$
$$\hat{C}_{u|v}(u,v) = \frac{\partial}{\partial v}\hat{C}\left(u,v\right) = v^{-1-\frac{1}{\theta}}\left(-1 + u^{-1/\theta} + v^{-1/\theta}\right)^{-1-\theta}.$$

Note that the above copula is a member of one parameter families of Archimedean copula which is useful for modelling positive dependent data as can be seen in Remark 6.1. The generator of the Clayton survival copula given in (6.3) is,

$$\varphi(t) = \left(t^{-\frac{1}{\theta}} - 1\right).$$

This generator is used to find the cumulative distribution of the copula which is useful while checking the goodness of fit of the data. Recall from Sklar (1959), for a given marginal survival function,

$$
\begin{aligned}
S\left(\eta_{1t}\right) &= \int_{\eta_{1t}}^{\infty} f\left(\eta_{1t}\right) d\eta_{1t} \\
&= \int_{\eta_{1t}}^{\infty} \frac{e^{\frac{-\eta_{1t}}{s}}}{s(1+e^{\frac{-\eta_{1t}}{s}})^{2}} d\eta_{1t} \\
&= \frac{1}{(1+e^{\frac{\eta_{1t}}{s}})}. \\
S\left(\eta_{2t}\right) &= \int_{\eta_{2t}}^{\infty} f\left(\eta_{2t}\right) d\eta_{2t} \\
&= \int_{\eta_{2t}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\eta_{2t}^{2}}{2\sigma^{2}}} d\eta_{2t} \\
&= \left[\frac{1}{2} Erfc\left(\frac{\eta_{2t}}{\sqrt{2}\sigma}\right)\right]
\end{aligned}
$$

and for the Clayton survival copula defined by (6.3), the joint survival function of η_{1t} and η_{2t} can be written as:

$$
\begin{aligned}
\overline{H}\left(\eta_{1t}, \eta_{2t}\right) &= \hat{C}_{\theta}\left(S\left(\eta_{1t}\right), S\left(\eta_{2t}\right)\right) \\
&= \left(\left(\frac{1}{\left(1+e^{\frac{\eta_{1t}}{s}}\right)}\right)^{-\frac{1}{\theta}} + \left(\frac{1}{2} Erfc\left(\frac{\eta_{2t}}{\sqrt{2}\sigma}\right)\right)^{-\frac{1}{\theta}} - 1\right)^{-\theta},
\end{aligned}
$$

where $Erfc$ represents the complimentary error function which takes the form,

$$
Erfc(z) = \frac{2}{\sqrt{\pi}} \int_{z}^{\infty} e^{-t^{2}} dt.
$$

It can be shown that the survival copula has uniform marginals given by,

$$\hat{C}_\theta(u,1) = \left(u^{-\frac{1}{\theta}}\right)^{-\theta} = u = \frac{1}{(1+e^{\frac{\eta_{1t}}{s}})}$$

and

$$\hat{C}_\theta(1,v) = \left(v^{-\frac{1}{\theta}}\right)^{-\theta} = v = \left[\frac{1}{2}Erfc\left(\frac{\eta_{2t}}{\sqrt{2}\sigma}\right)\right].$$

Also, the margins of $\bar{H}(\cdot,\cdot)$ are the functions $\bar{H}(\eta_{1t},\infty)$ and $\bar{H}(-\infty,\eta_{2t})$, which are the univariate survival functions. That is,

$$\bar{H}(\eta_{1t},\infty) = \frac{1}{(1+e^{\frac{\eta_{1t}}{s}})} = S_1(\eta_{1t})$$

and,

$$\bar{H}(-\infty,\eta_{2t}) = \left[\frac{1}{2}Erfc\left(\frac{\eta_{2t}}{\sqrt{2}\sigma}\right)\right] = S_2(\eta_{2t}).$$

The cumulative distribution function of the Clayton's copula $W^* = C(u,v)$ is given by,

$$\begin{aligned}
K_c(w) &= w - \frac{\phi(w)}{\phi'(w)} \\
&= w - \frac{w^{-\frac{1}{\theta}} - 1}{-\frac{w^{-1-\frac{1}{\theta}}}{\theta}} \\
&= \frac{\frac{1}{\theta}w^{-\frac{1}{\theta}} + w^{-\frac{1}{\theta}} - 1}{w^{-1-\frac{1}{\theta}}} \\
&= \frac{1}{\theta}w + w - w^{1+\frac{1}{\theta}} \\
&= w + \theta w - \theta w^{1+\frac{1}{\theta}} \\
&= w\left(1+\theta - w^{\frac{1}{\theta}}\theta\right); 0 \le w \le 1,
\end{aligned}$$

where $\phi(w)$ is the generator of the defined copula. The above function $K_c(w)$ can be used to check whether the proposed copula model is a good fit for the data by plotting the cumulative distribution, $K_c(w)$ and the empirical distribution function, $k_n(w)$ from the data, as can be seen in Section 6.6.

For a defined survival Copula function, the bivariate density associated can be obtained by,

$$h\left(\eta_{1t}, \eta_{2t}\right) = (-1)^2 \frac{\partial^2 \bar{H}\left(\eta_{1t}, \eta_{2t}\right)}{\partial \eta_{1t} \partial \eta_{2t}},$$

where $\frac{\partial^2 \bar{H}(\eta_{1t}, \eta_{2t})}{\partial \eta_{1t} \partial \eta_{2t}}$ can be further decomposed in to,

$$\begin{aligned} \frac{\partial^2 \bar{H}\left(\eta_{1t}, \eta_{2t}\right)}{\partial \eta_{1t} \partial \eta_{2t}} &= \frac{\partial^2}{\partial \eta_{1t} \partial \eta_{2t}} \hat{C}_\theta \left(S\left(\eta_{1t}\right), S\left(\eta_{2t}\right)\right) \\ &= \frac{\partial^2 C_\theta\left(u, v\right)}{\partial u \partial v} \frac{\partial u}{\partial \eta_{1t}} \frac{\partial v}{\partial \eta_{2t}} \\ &= \hat{c}_\theta\left(u, v\right) . f\left(\eta_{1t}\right) f\left(\eta_{2t}\right). \end{aligned}$$

Thus the bivariate density function constructed using Logistic and Normal marginals is of the form,

$$\begin{aligned} h\left(\eta_{1t}, \eta_{2t}\right) = & \frac{-2^{\frac{1}{2}+\frac{1}{\theta}}}{\sqrt{\pi} s \theta \sigma} e^{-\frac{\eta_{2t}{}^2}{2\sigma^2}+\frac{\eta_{1t}}{s}} \left(1+e^{\frac{\eta_{2t}}{s}}\right)^{-1+\frac{1}{\theta}} \left(-1-\theta\right) Erfc\left(\frac{\eta_{2t}}{\sqrt{2}\sigma}\right)^{-\frac{1}{\theta}-1} \\ & . \left(-1+\left(\frac{1}{\left(1+e^{\frac{\eta_{1t}}{s}}\right)}\right)^{-\frac{1}{\theta}} + 2^{\frac{1}{\theta}} Erfc\left(\frac{\eta_{2t}}{\sqrt{2}\sigma}\right)^{-\frac{1}{\theta}}\right)^{-2-\theta}. \end{aligned}$$

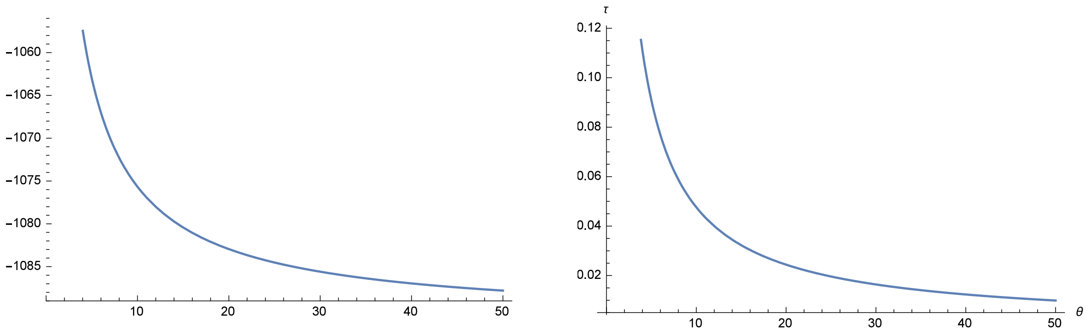
As discussed above, the bivariate density can be written in terms of the product of univariate marginals and a copula function, which describes the dependence structure between the variables. That is,

$$h\left(\eta_{1t},\eta_{2t}\right)=\left(-1+\left(\frac{1}{\left(1+e^{\frac{\eta_{1t}}{s}}\right)}\right)^{-\frac{1}{\theta}}+2^{\frac{1}{\theta}}Erfc\left(\frac{\eta_{2t}}{\sqrt{2}\sigma}\right)^{-\frac{1}{\theta}}\right)^{-2-\theta}$$
$$.\frac{-2^{1+\frac{1}{\theta}}}{\theta}\left(\frac{1}{1+e^{\frac{\eta_{1t}}{s}}}\right)^{-1-\frac{1}{\theta}}\left(-1-\theta\right)Erfc\left(\frac{\eta_{2t}}{\sqrt{2}\sigma}\right)^{-\frac{1}{\theta}-1}f\left(\eta_{1t}\right)f\left(\eta_{2t}\right). \quad (6.4)$$

For a sample of T observations, the conditional log-likelihood function given (x_{10},x_{20}) is given by,

$$L\left(\Theta\right)=\sum_{t=1}^{T}\log\left[\hat{C}_{\theta}\left(S\left(\eta_{1t}\right),S\left(\eta_{2t}\right)\right).f\left(\eta_{1t}\right)f\left(\eta_{2t}\right)\right], \quad (6.5)$$

where Θ is the set of all parameters in the model. Fig 6.1(A) shows the plot



(A) Plot of Log likelihood function

(B) Plot of Kendall's correlation

FIGURE 6.1: Plot of log likelihood function & Kendall's correlation

of log likelihood function for various values of θ and for a fixed set of density parameters α=1, β=0.5, $\phi = 0.5$ σ=1 and s=1. It can be seen that the function is concave up and decreasing, as the value of θ increases.

It is of interest to obtain the measure of dependence for non elliptical distributions, for which the linear correlation coefficient is inappropriate and misleading. One of the alternatives to the linear correlation coefficient is the Kendalls

Tau measure of dependence, for more details refer Kruskal (1958).

Remark 6.1. For a given copula function $\hat{C}_\theta(u,v) = \left(u^{-\frac{1}{\theta}} + v^{-\frac{1}{\theta}} - 1\right)^{-\theta}, \theta > 0,$ the population version of Kendall's tau is obtained as

$$\begin{aligned}
\tau_{\eta_{1t},\eta_{2t}} = \tau_{\hat{C}} &= 4\int_0^1\int_0^1 \hat{C}(u,v)\, d\hat{C}(u,v) - 1 \\
&= \frac{(-1-\theta)}{\theta}\int_0^1\int_0^1 -u^{-1-\frac{1}{\theta}}v^{-1-\frac{1}{\theta}}\left(-1+u^{-1/\theta}+v^{-1/\theta}\right)^{-2-2\theta} dudv - 1 \\
&= -\frac{(-1-\theta)}{\theta(2+\frac{1}{\theta})}\int_0^1 u^{-1-\frac{1}{\theta}}\left(u^{-\frac{1}{\theta}}\right)^{-(2+\frac{1}{\theta})\theta} du - 1 \\
&= \frac{1}{(1+2\theta)}.
\end{aligned}$$

Remark 6.2. The diagonal section of copula is the function δ_c from I to I defined as $\delta_c(t) = C(t,t)$ which is non decreasing and uniformly continuous on I, where $I = [0,1]$. The diagonal section of the copula is given by,

$$\begin{aligned}
\delta_c(t) &= C(t,t) \\
&= \left(t^{\frac{-1}{\theta}} + t^{\frac{-1}{\theta}} - 1\right)^{\theta}, 0 \le t \le 1.
\end{aligned}$$

Plot of the diagonal section of the copula is given in Figure 6.2. It is confirmed that the diagonal section of the copula is non decreasing and uniformly continuous on I. From Figure 6.1(B), it is clear that the defined copula is a positive dependent, since the dependency measure τ is always positive for all values of θ. Figure 6.3 shows the plot of density functions and contours with the chosen marginals for different values of copula parameters and it is clear that the new distribution is uni-modal for different choice of parameters.

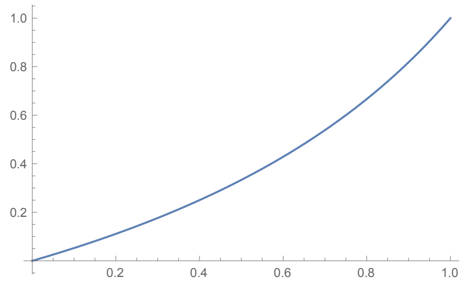
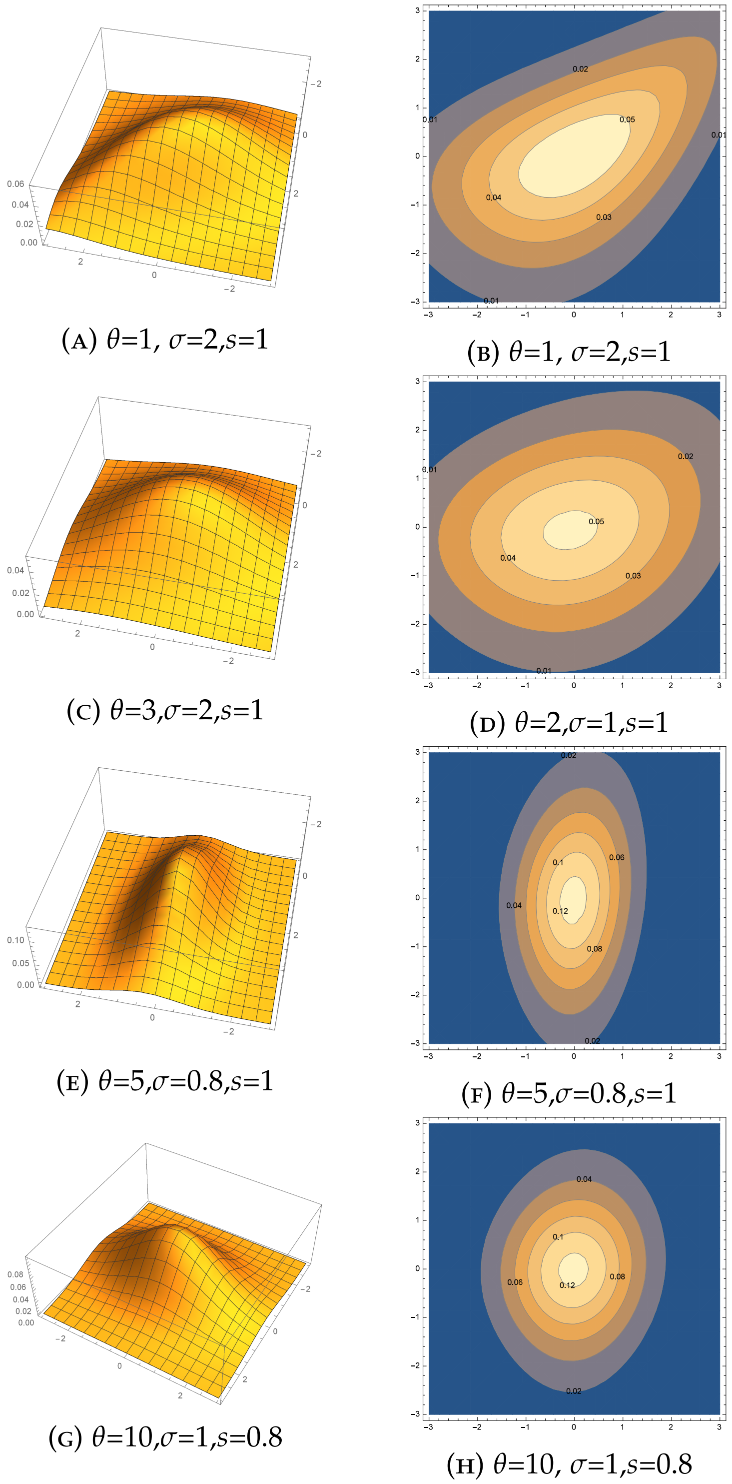


FIGURE 6.2: Plot of diagonal section of the copula



(A) θ=1, σ=2,s=1

(B) θ=1, σ=2,s=1

(C) θ=3,σ=2,s=1

(D) θ=2,σ=1,s=1

(E) θ=5,σ=0.8,s=1

(F) θ=5,σ=0.8,s=1

(G) θ=10,σ=1,s=0.8

(H) θ=10, σ=1,s=0.8

FIGURE 6.3: Bivariate Density Plots and Contours with Normal and Logistic marginals

6.3 Estimation of Copula based cointegration model

The method of maximum likelihood estimation (MLE) in copula modelling is computationally intensive, especially for high dimensional cases, as it needs estimation of the marginal and copula parameters jointly. Joe & Xu (2016) proposed a two step estimation procedure for obtaining the parameters of a copula based model. Their method is computationally simpler than estimating all parameters simultaneously from the log likelihood equation in (6.5). This section deals with the estimation of a copula based bivariate cointegration model using the method of Inference functions for margins. The efficiency of this estimator is compared with the MLE (cf. Xu (1996),Barnard (1991)).

6.3.1 Inference Functions for Margins (IFM) method

This approach consists of obtaining the maximum likelihood estimates of model parameters based on marginal likelihood function and maximum likelihood estimate of dependency parameter based on the whole likelihood function.
Let us assume that we have a set of T observations from two financial series. The conditional log likelihood functions based on the univariate marginals based on iid random variables $\{\eta_{1t}\}$ and $\{\eta_{2t}\}$, $t = 1, 2, \cdots, T$ given (x_{10}, x_{20}) are respectively

$$L_1(\Theta_1) = \sum_{t=1}^{T} \log f(\eta_{1t}; \Theta_1) \tag{6.6}$$

and

$$L_2(\Theta_2) = \sum_{t=1}^{T} \log f(\eta_{2t}; \Theta_2), \tag{6.7}$$

where $\Theta_1 = (s, \alpha, \beta, \delta)^{'}$ and $\Theta_2 = (\sigma, \alpha, \delta)^{'}$. The corresponding log likelihood function based on the joint distribution function is,

$$\begin{aligned} L(\theta, \alpha, \beta, \delta, \sigma, s) &= \sum_{t=1}^{T} \log h(\eta_{it}; \theta, \alpha, \beta, \delta, \sigma, s) \\ &= \sum_{t=1}^{T} \log \left[\hat{C}_\theta \left(S(\eta_{1t}), S(\eta_{2t})\right) . f(\eta_{1t}) f(\eta_{2t})\right]. \end{aligned}$$

The procedure of IFM works as follows: We first obtain the estimates of Θ_1 and Θ_2 by maximizing the log likelihood functions given in (6.6) and (6.7). The estimates of the parameters common to both margins are obtained by taking the averages of the individual estimators. If $\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\sigma}, \hat{s}$ are the estimators obtained using first step, in the second stage we obatin the estimate of θ by maximizing,

$$L\left(\theta, \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\sigma}, \hat{s}\right) = \sum_{t=1}^{T} \log h\left(\eta_{it}; \theta, \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\sigma}, \hat{s}\right)$$

with respect to θ.

Now we apply this procedure to the model (6.1) and (6.2). The conditional log likelihood function of Θ_1 based on (x_{1t}, x_{2t}), $t = 1, 2, \cdots, T$ given (x_{10}, x_{20}) when the errors have logistic distribution is given by,

$$\begin{aligned} L_1(\Theta_1) &= \sum_{t=1}^{T} \log \left(\frac{e^{\frac{-\eta_{1t}}{s}}}{s(1+e^{\frac{-\eta_{1t}}{s}})^2} \right) \\ &= \sum_{t=1}^{T} \left(-\frac{\eta_{1t}}{s} - \log(s) - 2\log\left(1+e^{\frac{-\eta_{1t}}{s}}\right) \right) \\ &= \sum_{t=1}^{T} \left(-\left(\frac{\Delta x_{1t} - \delta\beta z_{t-1}}{s} \right) - \log(s) - 2\log\left(1+e^{-\left(\frac{\Delta x_{1t} - \delta\beta z_{t-1}}{s}\right)}\right) \right). \end{aligned}$$

The likelihood equations for the parameters δ, β, α and s are:

$$\frac{\partial L_1\left(\Theta_1\right)}{\partial \delta}=0 \Rightarrow \sum_{t=1}^{T} \frac{\beta z_{t-1}}{s}\left[1-\frac{2e^{-\left(\frac{\Delta x_{1t}-\delta\beta z_{t-1}}{s}\right)}}{\left(1+e^{-\left(\frac{\Delta x_{1t}-\delta\beta z_{t-1}}{s}\right)}\right)}\right]=0. \tag{6.8}$$

$$\frac{\partial L_1\left(\Theta_1\right)}{\partial \beta}=0 \Rightarrow \sum_{t=1}^{T} \frac{\hat{\delta}_1 z_{t-1}}{s}\left[1-\frac{2e^{-\left(\frac{\hat{\delta}_1 x_{1t}-\hat{\delta}_1\beta z_{t-1}}{s}\right)}}{\left(1+e^{-\left(\frac{\Delta x_{1t}-\hat{\delta}_1\beta z_{t-1}}{s}\right)}\right)}\right]=0. \tag{6.9}$$

$$\frac{\partial L_1\left(\Theta_1\right)}{\partial \alpha}=0 \Rightarrow \sum_{t=1}^{T} \frac{\hat{\delta}_1\hat{\beta}_1 x_{2t-1}}{s}\left[1-\frac{2e^{-\left(\frac{\Delta x_{1t}-\hat{\delta}_1\hat{\beta}_1 z_{t-1}}{s}\right)}}{\left(1+e^{-\left(\frac{\Delta x_{1t}-\hat{\delta}_1\hat{\beta}_1 z_{t-1}}{s}\right)}\right)}\right]=0. \tag{6.10}$$

$$\frac{\partial L_1\left(\Theta_1\right)}{\partial s}=0 \Rightarrow \sum_{t=1}^{T}\left(\left(\frac{\Delta x_{1t}-\hat{\delta}_1\hat{\beta}_1\hat{z}_{t-1}}{s^2}\right)\left[1-\frac{2e^{-\left(\frac{\Delta x_{1t}-\hat{\delta}_1\hat{\beta}_1\hat{z}_{t-1}}{s}\right)}}{\left(1+e^{-\left(\frac{\Delta x_{1t}-\hat{\delta}_1\hat{\beta}_1\hat{z}_{t-1}}{s}\right)}\right)}\right]-\frac{1}{s}\right)=0, \tag{6.11}$$

where $\hat{z}_{t-1}=x_{1,t-1}+\hat{\alpha}_1 x_{2,t-1}$. We do not have analytically closed form expressions for the estimators. Thus on solving the likelihood equations numerically we will get the estimates of δ, β, α and s and we call them as $\hat{\delta}_1$, $\hat{\beta}_1$, $\hat{\alpha}_1$ and $\hat{s}$. The conditional log likelihood function on (x_{10}, x_{20}) based on the normally distributed errors is given by,

$$
\begin{aligned}
L_2(\Theta_2) &= \sum_{t=1}^{T}\left(\log\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-\eta_{2t}^2}{2\sigma^2}}\right) \\
&= \sum_{t=1}^{T}\left(\log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2}\frac{\eta_{2t}^2}{\sigma^2}\right) \\
&= \sum_{t=1}^{T}\left(\log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{(\Delta x_{2t} + \delta z_{t-1})^2}{2\sigma^2}\right).
\end{aligned}
$$

On solving the likelihood equations, the MLE's of the parameter δ, α and σ are recorded as δ_2, α_2 and σ and are given by:

$$
\hat{\delta}_2 = \frac{\sum_{t=1}^{T} -\Delta x_{2t} z_{t-1}}{\sum_{t=1}^{T} z_{t-1}^2}. \tag{6.12}
$$

$$
\hat{\alpha}_2 = \frac{-\sum_{t=1}^{T}\left[\Delta x_{2t} x_{2t-1} + \hat{\delta}_2 x_{1t-1} x_{2t-1}\right]}{\sum_{t=1}^{T} \hat{\delta}_2 x_{2t-1}^2}. \tag{6.13}
$$

$$
\hat{\sigma} = \frac{1}{n}\sum_{t=1}^{T}(\Delta x_{2t} - \delta \hat{z}_{1,t-1})^2, \tag{6.14}
$$

where $\hat{z}_{1,t-1} = x_{1,t-1} + \hat{\alpha}_2 x_{2,t-1}$. If a parameter appears in more than one bivariate margin, there are several ways to obtain its estimate. A possible way to obtain the parameter estimates is to average the estimators from the log likelihoods of the margin with the common parameter (cf. Joe (1997)). One can also obtain the estimates by using the higher dimensional log likelihood with the given univariate parameters. In our case, two of the parameters α and δ are common in both the margins. We consider the first method as it will reduce the computational complexity of the estimation procedure. Thus the estimates of the common parameters are given by,

$$\hat{\delta} = \frac{\hat{\delta}_1+\hat{\delta}_2}{2}.$$
$$\hat{\alpha} = \frac{\hat{\alpha}_1+\hat{\alpha}_2}{2}.$$

Once the marginal parameters are obtained , the function $L\left(\theta, \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\sigma}, \hat{s}\right)$ is maximized over θ to get $\hat{\theta}$. On differentiating the log likelihood function over the maximised set of parameters with respect to θ implies,

$$\frac{\partial \log L\left(\Theta_1\right)}{\partial \theta} = 0 \Rightarrow$$
$$\sum_{t=1}^{T} \left(-\frac{1}{-1-\theta} - \frac{1}{\theta} - \frac{\log[2]}{\theta^2} + \frac{\log\left[\frac{A(\hat{\Omega})}{1+A(\hat{\Omega})}\right]}{\theta^2} + \frac{\log\left[Erfc\left[B(\hat{\Omega})\right]\right]}{\theta^2} \right.$$
$$\log\left(-1 + \left(\frac{A(\hat{\Omega})}{1+A(\hat{\Omega})}\right)^{-1/\theta} + 2^{\frac{1}{\theta}} Erfc\left[B(\hat{\Omega})\right]^{-1/\theta}\right)$$
$$+ \frac{(-2-\theta)\,\theta^{-2}\left(\frac{A(\hat{\Omega})}{1+A(\hat{\Omega})}\right)^{-1/\theta} \log\left[\frac{A(\hat{\Omega})}{1+A(\hat{\Omega})}\right]}{-1 + \left(\frac{A(\hat{\Omega})}{1+A(\hat{\Omega})}\right)^{-1/\theta} + 2^{\frac{1}{\theta}} Erfc\left[B(\hat{\Omega})\right]^{-1/\theta}}$$
$$\left. - \frac{\left((-2-\theta)\, 2^{\frac{1}{\theta}} Erfc\left[B(\hat{\Omega})\right]^{-1/\theta} \theta^{-2} \left(\log[2] - \log\left(Erfc(B(\hat{\Omega}))\right)\right)\right)}{-1 + \left(\frac{A(\hat{\Omega})}{1+A(\hat{\Omega})}\right)^{-1/\theta} + 2^{\frac{1}{\theta}} Erfc\left[B(\hat{\Omega})\right]^{-1/\theta}} \right) = 0, \tag{6.15}$$

where $A(\hat{\Omega}) = e^{\frac{\Delta x_{1t} - \hat{\delta}\hat{\beta}\hat{z}_{t-1}}{\hat{s}}}$ and $B(\hat{\Omega}) = \frac{\Delta x_{2t} + \hat{\delta}\hat{z}_{t-1}}{\sqrt{2}\hat{\sigma}}$.

Solution of the above equation is obtained numerically, which yields the estimates of the parameter θ using the method of inference functions for margins.

The two stage estimators obtained above are consistent estimator for the unknown parameters, see Joe & Xu (2016).

6.3.2 Maximum Likelihood Estimation

Here we discuss the maximum likelihood estimation of the parameters involved in the likelihood function based on the sample $\{x_{1t}, x_{2t}\}, t = 1, 2, ...T$. Under the assumption of the bivariate distribution given in (6.4), for a sample of T observations, the conditional likelihood function given (x_{10}, x_{20}) for the copula based cointegration model can be written as,

$$\begin{aligned} L\left(\Theta \mid x_t\right) = &\prod_{t=1}^{T}\left(-1+\left(1+e^{\frac{\eta_{1t}}{s}}\right)^{-1}+2^{\frac{1}{\theta}}Erfc\left(\frac{\eta_{2t}}{\sqrt{2}\sigma}\right)^{-\frac{1}{\theta}}\right)^{-2-\theta} \times \\ &\prod_{t=1}^{T}\left(\frac{-2^{1+\frac{1}{\theta}}}{\theta}\left(\frac{1}{1+e^{\frac{\eta_{1t}}{s}}}\right)^{-1-\frac{1}{\theta}}(-1-\theta)\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{\eta_{2t}^2}{2\sigma^2}}\frac{e^{\frac{-\eta_{1t}}{s}}}{s(1+e^{\frac{-\eta_{1t}}{s}})^2}\right), \end{aligned}$$

where Θ denotes the vector of unknown parameters to be estimated. For the above copula based cointegration model, the log likelihood function is given as follows:

$$\begin{aligned} \log L\left(\Theta \mid x_t\right) = &\sum_{t=1}^{T}\left(\left(1+\frac{1}{\theta}\right)\log(-2)-\log(\theta)-\left(-1-\frac{1}{\theta}\right)\log\left(1+e^{\frac{\eta_{1t}}{s}}\right)^{-1}+\right. \\ &(-2-\theta)\left(\log(-1+\left(1+e^{\frac{\eta_{1t}}{s}}\right)^{-1}+2^{\frac{1}{\theta}}Erfc\left(\frac{\eta_{2t}}{\sqrt{2}\sigma}\right)\right)+\left(\frac{1}{\theta}-1\right) \\ &\left.-\frac{\eta_{2t}^2}{2\sigma^2}-\log\sigma-\frac{\eta_{1t}}{s}+\log\left(-1-\theta\right)-\log s-2\log\left(1+e^{\frac{\eta_{1t}}{s}}\right)\right). \end{aligned}$$

The likelihood equations for the parameters σ, s, β, δ, α and θ are:

$$\frac{\partial \log L\left(\Theta_1\right)}{\partial \sigma}=0$$
$$\Rightarrow \sum_{t=1}^{T}\left(-\frac{1}{\sigma}+\frac{\left(A(\hat{\Omega})\right)^2\sqrt{\frac{2}{\pi}}\left(-1-\frac{1}{\theta}\right)\left(\Delta x_{2t}+\delta z_{t-1}\right)}{\sigma^2 Erfc\left[B(\hat{\Omega})\right]}+\frac{\left(\Delta x_{2t}+\delta z_{t-1}\right)^2}{\sigma^3}\right.$$
$$\left.-\frac{2^{\frac{1}{2}+\frac{1}{\theta}}\left(A(\hat{\Omega})\right)^2\left(-2-\theta\right)Erfc\left[B(\hat{\Omega})\right]^{-1-\frac{1}{\theta}}\left(\Delta x_{2t}+\delta z_{t-1}\right)}{\sqrt{\pi}\theta\sigma^2\left(-1+\left(\frac{A(\hat{\Omega})}{1+A(\hat{\Omega})}\right)^{-1/\theta}+2^{\frac{1}{\theta}}Erfc\left[B(\hat{\Omega})\right]^{-1/\theta}\right)}\right)=0, \quad (6.16)$$

where $A(\hat{\Omega})=e^{\frac{\Delta x_{1t}-\delta\beta z_{t-1}}{s}}$ and $B(\hat{\Omega})=\frac{\Delta x_{2t}+\delta z_{t-1}}{\sqrt{2}\sigma}$.

$$\frac{\partial \log L\left(\Theta_1\right)}{\partial s}=0$$
$$\Rightarrow \sum_{t=1}^{T}\left(-\frac{1}{s}+\frac{\eta_{1t}}{s^2}-\frac{2A(\hat{\Omega})\eta_{1t}}{\left(1+A(\hat{\Omega})\right)s^2}+A(\hat{\Omega})\left(1+A(\hat{\Omega})\right)\left(-1-\frac{1}{\theta}\right)\right.$$
$$\left(\frac{A(\hat{\Omega})\eta_{1t}}{\left(1+A(\hat{\Omega})\right)s^2}-\frac{\mathrm{e}^{-\frac{2\eta_{1t}}{s}}\eta_{1t}}{\left(1+A(\hat{\Omega})\right)^2 s^2}\right)\left(\frac{A(\hat{\Omega})}{1+A(\hat{\Omega})}\right)^{-1-\frac{1}{\theta 1}}$$
$$\left.\left(-2-\theta\right)\frac{\left(-\frac{\mathrm{e}^{-\frac{2\eta_{1t}}{s}}\eta_{1t}}{\left(1+A(\hat{\Omega})\right)^2 s^2}+\frac{A(\hat{\Omega})\eta_{1t}}{\left(1+A(\hat{\Omega})\right)s^2}\right)}{\theta\left(-1+\left(\frac{A(\hat{\Omega})}{1+A(\hat{\Omega})}\right)^{-1/\theta}+2^{\frac{1}{\theta}}Erfc\left[B_1(\Omega)\right]^{-1/\theta}\right)}\right)=0, \quad (6.17)$$

where $B_1(\hat{\Omega})=\frac{\Delta x_{2t}+\delta z_{t-1}}{\sqrt{2}\hat{\sigma}}$.

$$
\begin{aligned}
&\frac{\partial \log L\left(\Theta_{1}\right)}{\partial \beta}=0 \\
&\Rightarrow \sum_{t=1}^{T}\left(\frac{\delta z_{t-1}}{\hat{s}}-\frac{2 A_{1}(\hat{\Omega}) \delta z_{t-1}}{\left(1+A_{1}(\hat{\Omega})\right) \hat{s}}+A_{1}(\hat{\Omega})\left(1+A_{1}(\hat{\Omega})\right) \delta z_{t-1}\left(-1-\frac{1}{\theta}\right)\right. \\
&\qquad\left(-\frac{\mathrm{e}^{-2 \frac{\eta_{1 t}}{\hat{s}}}}{\left(1+A_{1}(\hat{\Omega})\right)^{2} \hat{s}}+\frac{A_{1}(\hat{\Omega})}{\left(1+A_{1}(\hat{\Omega})\right) \hat{s}}\right) \\
&\left.\frac{\left(-\frac{\mathrm{e}^{-2 \frac{\eta_{1 t}}{\hat{s}}}}{\left(1+A_{1}(\hat{\Omega})\right)^{2} \hat{s}}+\frac{A_{1}(\hat{\Omega})}{\left(1+A_{1}(\hat{\Omega})\right) \hat{s}}\right) \theta\left(-1+\left(\frac{A_{1}(\hat{\Omega})}{\left(1+A_{1}(\hat{\Omega})\right)}\right)^{-1 / \theta}+2^{\frac{1}{\theta}} Erfc\left[B_{1}(\hat{\Omega})\right]^{-1 / \theta}\right)}{\left(-\frac{\mathrm{e}^{-\frac{2 \eta_{1 t}}{\hat{s}}} \eta_{1 t}}{\left(1+A_{1}(\hat{\Omega})\right)^{2} \hat{s}^{2}}+\frac{A_{1}(\hat{\Omega}) \eta_{1 t}}{\left(1+A_{1}(\hat{\Omega})\right) \hat{s}^{2}}\right)}\right)=0,
\end{aligned}
\tag{6.18}
$$

where $A_1(\hat{\Omega}) = e^{\frac{\Delta x_{1t}-\delta\beta z_{t-1}}{\hat{s}}}$.

$$
\begin{aligned}
&\frac{\partial \log L\left(\Theta_{1}\right)}{\partial \delta}=0 \\
&\Rightarrow \sum_{t=1}^{T}\left(\frac{\hat{\beta} z_{t-1}}{\hat{s}}-C(\hat{\Omega}) \frac{2 \hat{\beta} z_{t-1}}{\hat{s}}-\frac{\mathrm{e}^{\frac{\left(\Delta x_{1 t}\right)}{\hat{s}}} \hat{\beta}(1+\theta) z_{t-1}}{\left(\mathrm{e}^{\frac{\hat{\beta} \delta z_{t-1}}{\hat{s}}}+\mathrm{e}^{\frac{\left(\Delta x_{1 t}\right)}{\hat{s}}}\right) \hat{s} \theta}+\frac{\mathrm{e}^{-\frac{\eta_{2 t}{ }^{2}}{2 \hat{\sigma}^{2}}} \sqrt{\frac{2}{\pi}}(1+\theta) z_{t-1}}{\theta \sigma Erfc\left[B_{1}(\hat{\Omega})\right]}\right. \\
&\quad-\frac{z_{t-1} \eta_{2 t}}{\sigma^{2}}+(-2-\theta)\left(\frac{-\hat{s}^{-1} \theta^{-1} A_{2}(\hat{\Omega})(C(\hat{\Omega}))^{\frac{-1+\theta 1}{\theta 1}} \hat{\beta} z_{t-1}}{-1+(C(\hat{\Omega}))^{-1 / \theta}+2^{\frac{1}{\theta}} Erfc\left[B_{1}(\hat{\Omega})\right]^{-1 / \theta}}\right) \\
&\left.\quad+(-2-\theta) \frac{\left((\sqrt{\pi} \theta \hat{\sigma})^{-1} 2^{\frac{1}{2}+\frac{1}{\theta}} \mathrm{e}^{-\frac{\eta_{2 t}{ }^{2}}{2 \hat{\sigma}^{2}}} Erfc\left[B_{1}(\hat{\Omega})\right]^{-\frac{1+\theta}{\theta}} z_{t-1}\right)}{-1+(C(\hat{\Omega}))^{-1 / \theta}+2^{\frac{1}{\theta}} Erfc\left[B_{1}(\hat{\Omega})\right]^{-1 / \theta}}\right)=0,
\end{aligned}
\tag{6.19}
$$

where $A_2(\hat{\Omega}) = e^{\frac{\Delta x_{1t}-\delta\hat{\beta} z_{t-1}}{\hat{s}}}$ and $C(\hat{\Omega}) = \frac{\mathrm{e}^{\frac{\hat{\beta}\delta z_{t-1}}{\hat{s}}}}{\mathrm{e}^{\frac{\hat{\beta}\delta z_{t-1}}{\hat{s}}}+\mathrm{e}^{\frac{(\Delta x_{1t})}{\hat{s}}}}$.

$$
\begin{aligned}
&\frac{\partial \log L\left(\Theta_1\right)}{\partial \alpha} = 0 \\
&\Rightarrow \sum_{t=1}^{T} \left(\frac{\hat{\beta}\delta x_{2t-1}}{\hat{s}} - \frac{2A_3(\hat{\Omega})\hat{\beta}\hat{\delta}x_{2t-1}}{\left(A_3(\hat{\Omega}) + \mathrm{e}^{\frac{(\Delta x_{1t})}{\hat{s}}}\right)\hat{s}} - \frac{\mathrm{e}^{\frac{(\Delta x_{1t})}{\hat{s}}}\hat{\beta}\hat{\delta}\left(1+\theta\right)x_{2t-1}}{\left(A_3(\hat{\Omega}) + \mathrm{e}^{\frac{(\Delta x_{1t})}{\hat{s}}}\right)\hat{s}\theta} - \hat{\delta}x_{2t-1}2B_2(\hat{\Omega}) + \right. \\
&\left(\frac{(\hat{s}\theta)^{-1}A_3(\hat{\Omega})\left(\frac{A_3(\hat{\Omega})}{A_3(\hat{\Omega})+\mathrm{e}^{\frac{(\Delta x_{1t})}{\hat{s}}}}\right)^{\frac{-1+\theta}{\theta}}\hat{\beta}\hat{\delta}x_{2t-1}}{1+\left(\frac{A_3(\hat{\Omega})}{A_3(\hat{\Omega})+\mathrm{e}^{\frac{(\Delta x_{1t})}{\hat{s}}}}\right)^{-1/\theta} + 2^{\frac{1}{\theta}}Erfc\left[B_2(\hat{\Omega})\right]^{-1/\theta}} \right) + (-2-\theta) \\
&\left.\left(\frac{\hat{\beta}\hat{\delta}x_{2t-1} + (\sqrt{\pi}\theta\sigma)^{-1}2^{\frac{1}{2}+\frac{1}{\theta}}\mathrm{e}^{-(B_2(\hat{\Omega}))^2}Erfc\left[B_2(\hat{\Omega})\right]^{-\frac{1+\theta}{\theta}}x_{2t-1}}{1+\left(\frac{A_3(\hat{\Omega})}{A_3(\hat{\Omega})+\mathrm{e}^{\frac{(\Delta x_{1t})}{\hat{s}}}}\right)^{-1/\theta} + 2^{\frac{1}{\theta}}Erfc\left[B_2(\hat{\Omega})\right]^{-1/\theta}} \right)\right) = 0, \quad (6.20)
\end{aligned}
$$

where $A_3(\hat{\Omega}) = e^{\frac{\Delta x_{1t} - \hat{\delta}\hat{\beta}z_{t-1}}{\hat{s}}}$ and $B_2(\hat{\Omega}) = \frac{\Delta x_{2t} + \hat{\delta}z_{t-1}}{\sqrt{2}\hat{\sigma}}$.

$$
\begin{aligned}
&\frac{\partial \log L\left(\Theta_1\right)}{\partial \theta} = 0 \\
&\Rightarrow \sum_{t=1}^{T} \Bigg(-\frac{1}{-1-\theta} - \frac{1}{\theta} - \frac{\log[2]}{\theta^2} + \frac{\log\left[\frac{A_4(\hat{\Omega})}{1+A_4(\hat{\Omega})}\right]}{\theta^2} + \frac{\log\left[Erfc\left[B_3(\hat{\Omega})\right]\right]}{\theta^2} \\
&\qquad \log\left(-1+\left(\frac{A_4(\hat{\Omega})}{1+A_4(\hat{\Omega})}\right)^{-1/\theta} + 2^{\frac{1}{\theta}} Erfc\left[B_3(\hat{\Omega})\right]^{-1/\theta}\right) \\
&\qquad + \frac{(-2-\theta)\,\theta^{-2}\left(\frac{A_4(\hat{\Omega})}{1+A_4(\hat{\Omega})}\right)^{-1/\theta}\log\left[\frac{A_4(\hat{\Omega})}{1+A_4(\hat{\Omega})}\right]}{-1+\left(\frac{A_4(\hat{\Omega})}{1+A_4(\hat{\Omega})}\right)^{-1/\theta} + 2^{\frac{1}{\theta}} Erfc\left[B_3(\hat{\Omega})\right]^{-1/\theta}} \\
&\quad - \frac{\left((-2-\theta)\,2^{\frac{1}{\theta}} Erfc\left[B_3(\hat{\Omega})\right]^{-1/\theta}\theta^{-2}\left(\log[2]-\log\left(Erfc(B_3(\hat{\Omega}))\right)\right)\right)}{-1+\left(\frac{A_4(\hat{\Omega})}{1+A_4(\hat{\Omega})}\right)^{-1/\theta} + 2^{\frac{1}{\theta}} Erfc\left[B_3(\hat{\Omega})\right]^{-1/\theta}} \Bigg) = 0,
\end{aligned}
\tag{6.21}
$$

where $A_4(\hat{\Omega}) = e^{\frac{\Delta x_{1t} - \hat{\delta}\hat{\beta}\hat{z}_{t-1}}{\hat{s}}}$ and $B_3(\hat{\Omega}) = \frac{\Delta x_{2t} + \hat{\delta}\hat{z}_{t-1}}{\sqrt{2}\hat{\sigma}}$. Since the score functions are in a complicated form, we will obtain the MLE's by solving the likelihood equations numerically. Here we used the method of Newton Raphson to solve the likelihood equations.

6.4 Algorithm for Random Number Generation

The copula $\hat{C}$ of (η_{1t}, η_{2t}) is given by

$$
\hat{C}\left(u, v\right) = \left(u^{-\frac{1}{\theta}} + v^{-\frac{1}{\theta}} - 1\right)^{-\theta},
$$

and so the conditional distribution function $\hat{C}_u$ is given by

$$\begin{aligned}\hat{C}_u(v) &= \frac{\partial}{\partial u}\hat{C}(u,v)\\ &= u^{-1-\frac{1}{\theta}}\left(-1+u^{-\frac{1}{\theta}}+v^{-\frac{1}{\theta}}\right)^{-1-\theta}.\end{aligned}$$

Now the quasi inverse $\hat{C}_u^{-1}$ can be obtained by setting $\hat{C}_u(v) = t$.

$$\Rightarrow u^{-1-\frac{1}{\theta}}\left(-1+u^{-\frac{1}{\theta}}+v^{-\frac{1}{\theta}}\right)^{-1-\theta} = t$$

$$\Rightarrow \left(-1+u^{-\frac{1}{\theta}}+v^{-\frac{1}{\theta}}\right) = \left(tu^{1+\frac{1}{\theta}}\right)^{\frac{1}{-1-\theta}}$$

$$\Rightarrow \hat{C}_u^{-1}(t) = \left(t^{1/(-1-\theta)}u^{\left(1+\frac{1}{\theta}\right)/(-1-\theta)} - u^{-\frac{1}{\theta}}+1\right)^{-\theta}.$$

Thus an algorithm to generate the bivariate data from the given copula is (cf. Nelsen (2007)):

Step 1: Generate two independent uniform random variates u and t.

Step 2: Set v=$\hat{C}_u^{-1}(t)$.

i.e; $v = \left(t^{1/(-1-\theta)}u^{\left(1+\frac{1}{\theta}\right)/(-1-\theta)} - u^{-\frac{1}{\theta}}+1\right)^{-\theta}$.

Step 3: Set $\eta_1 = F_u^{-1}$ and $\eta_2 = F_v^{-1}$.

On solving we will get, $\eta_1 = \log\left(u^{-1}-1\right)$ and $\eta_2 = \sqrt{2}\left(Erfc\left(2v\right)\right)^{-1}$. Now we have the bivariate data from (η_1, η_2).

6.5 Simulation

In this section, we carried out a simulation study to evaluate the efficiency of the estimates obtained by IFM method relative to the MLE method.

For the simulation study, we first generate the bivariate data (η_{1t}, η_{2t}) using the procedure discussed in Section 6.4 for a specified choice of the dependency parameter, θ. Then for different values of the model parameters we simulated the bivariate data (x_{1t}, x_{2t}) using (6.1) and (6.2). Based on this bivariate sample, we obtain the estimates of the parameters through the method of Inference functions for margins and the method of maximum likelihood. The estimates are obtained by solving the equations (6.8) to (6.21). For each specified value of the parameter, we repeated the experiment 500 times and then averaged them over the repetitions. Alternatively, we calculated the dependency parameter θ using the population version of Kendall's tau measure given in Remark 6.1, $\hat{\theta}_\tau = \frac{(1-\tau)}{2\tau}$ The estimates of IFM as well as MLE along with their root mean squared error are reported in Tables 6.1- 6.8.

TABLE 6.1: Average values and MSE of Cointegration and distribution parameters based on IFM for θ=0.5, σ=1 and κ=0.5

n	α	β	δ	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\sigma}$	$\hat{\kappa}$
500	1	0.5	1.8	1.0051(0.0006)	0.5114(0.0235)	1.8316(0.0889)	0.9496(0.0992)	0.4945(0.0489)
	2	1.5	1.4	1.9995(0.0058)	1.533(0.0547)	1.5235(0.1952)	0.8893(0.1516)	0.5260(0.0427)
	3	2.5	1	2.9957(0.0052)	2.5620(0.0670)	1.1547(0.1713)	1.1018(0.1841)	0.4556(0.0582)
	4	3.5	0.6	3.9954(0.0052)	3.5694(0.0754)	0.7032(0.1182)	1.2026(0.2380)	0.6413(0.1211)
	5	4.5	0.2	4.9544(0.0500)	4.2386(0.2862)	0.1455(0.0589)	0.8131(0.223)	0.4517(0.0634)
1000	1	0.5	1.8	1.0021(0.0003)	0.5245(0.0013)	1.8713(0.0625)	0.9616(0.0494)	0.4849(0.0290)
	2	1.5	1.4	1.9993(0.0025)	1.468(0.0349)	1.3190(0.089)	0.9148(0.0964)	0.4901(0.0204)
	3	2.5	1	2.9981(0.0020)	2.4592(0.0479)	0.9321(0.0809)	1.0673(0.0902)	0.6169(0.0612)
	4	3.5	0.6	3.9978(0.0029)	3.4842(0.0255)	0.5846(0.0270)	0.9341(0.1028)	0.5048(0.033)
	5	4.5	0.2	4.9885(0.0136)	4.4170(0.0958)	0.1767(0.0274)	1.0596(0.0904)	0.4325(0.0404)

Note that from Tables 6.1 to 6.8, for series of length 500 estimates are reasonably satisfactory and become more accurate with increasing sample size, n=1000. It is seen that, both methods are efficient with respect to the mean square errors

TABLE 6.2: Average values and MSE of dependency parameter estimates based on IFM ($\hat{\theta}_{IFM}$) and Kendals tau measure ($\hat{\theta}_{\tau}$) and the estimate of the measure of dependence (τ) for θ=0.5, σ=1 and κ=0.5

n	α	β	δ	$\hat{\theta}_{IFM}$	$\hat{\theta}_{\tau}$	$\hat{\tau}$
500	1	0.5	1.8	0.4999(0.0557)	0.5132(0.0724)	0.4958(0.0244)
	2	1.5	1.4	0.5056(0.0522)	0.5119(0.0696)	0.4962(0.0191)
	3	2.5	1	0.5080(0.1189)	0.5070(0.0510)	0.4977(0.0602)
	4	3.5	0.6	0.5163(0.0933)	0.4868(0.0678)	0.5088(0.0549)
	5	4.5	0.2	0.6159(0.2386)	0.5082(0.0656)	0.4978(0.0869)
1000	1	0.5	1.8	0.5080(0.0341)	0.5112(0.0377)	0.4930(0.0143)
	2	1.5	1.4	0.50061(0.0253)	0.4979(0.0346)	0.5016(0.0114)
	3	2.5	1	0.5076(0.0335)	0.4983(0.0325)	0.5013(0.0181)
	4	3.5	0.6	0.5014(0.0516)	0.4955(0.0363)	0.5029(0.0249)
	5	4.5	0.2	0.5142(0.0973)	0.4938(0.028)	0.5034(0.0478)

TABLE 6.3: Average values and MSE of Cointegration and distribution parameters based on IFM for θ=1, σ=3 and κ=2

n	α	β	δ	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\sigma}$	$\hat{\kappa}$
500	1	0.5	1.8	1.0040(0.0054)	0.5807(0.1004)	2.1476(0.4511)	2.9035(0.2208)	1.9358(0.1502)
	2	1.5	1.4	1.9998(0.0004)	1.5708(0.1148)	1.6863(0.4656)	3.0809(0.3128)	2.6816(0.4321)
	3	2.5	1	2.9984(0.0018)	2.5531(0.0717)	1.1299(0.1809)	2.9265(0.2144)	2.3650(0.4645)
	4	3.5	0.6	3.9948(0.0054)	3.5353(0.0554)	0.6580(0.0886)	2.6791(0.3751)	1.9417(0.1841)
	5	4.5	0.2	4.9733(0.0322)	4.4270(0.1485)	0.1910(0.0343)	2.5978(0.6056)	1.8646(0.1992)
	1	0.5	1.8	1.0070(0.0008)	0.5010(0.0177)	1.7793(0.0612)	2.9861(0.1527)	2.1497(0.1735)
1000	2	1.5	1.4	1.9994(0.00074)	1.5081(0.0165)	1.4285(0.0482)	3.0603(0.3010)	1.909(0.1241)
	3	2.5	1	2.998(0.0015)	2.5023(0.0202)	1.0103(0.0412)	3.1478(0.1980)	1.9289(0.1860)
	4	3.5	0.6	3.9948(0.0054)	3.5353(0.055)	0.6580(0.0886)	2.6912(0.3751)	1.9417(0.1841)
	5	4.5	0.2	4.990(0.0117)	4.4361(0.0342)	0.1940(0.0088)	3.002(0.1532)	1.6696(0.1045)

TABLE 6.4: Average values and MSE of dependency parameter estimates based on IFM and Kendals tau measure ($\hat{\theta}_{IFM}$) and Kendals tau measure ($\hat{\theta}_{\tau}$) and the estimate of the measure of dependence (τ) for θ=1, σ=3 and κ=2

n	α	β	δ	$\hat{\theta}_{IFM}$	$\hat{\theta}_{\tau}$	$\hat{\tau}$
500	1	0.5	1.8	0.9992(0.1554)	0.9947(0.1865)	0.3394(0.0261)
	2	1.5	1.4	0.98926(0.1311)	1.0051(0.1598)	0.3358(0.0241)
	3	2.5	1	1.0067(0.1567)	1.0030(0.1946)	0.3380(0.0263)
	4	3.5	0.6	1.042(0.1739)	1.0164(0.1479)	0.3328(0.0397)
	5	4.5	0.2	1.0651(0.1992)	1.0414(0.2598)	0.3284(0.0448)
1000	1	0.5	1.8	1.0253(0.1065)	1.0362(0.0965)	0.3265(0.0166)
	2	1.5	1.4	0.9993(0.0729	0.9899(0.0882)	0.3367(0.0122)
	3	2.5	1	1.081(0.0929)	0.9889(0.0808)	0.3367(0.0142)
	4	3.5	0.6	1.044(0.1032)	1.0104(0.1107)	0.3318(0.0297)
	5	4.5	0.2	1.0412(0.1739)	1.0164(0.1479)	0.3328(0.0397)

TABLE 6.5: Average values and MSE of Cointegration and distribution parameters based on MLE for θ=0.5, σ=1 and κ=0.5

n	α	β	δ	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\sigma}$	$\hat{\kappa}$
500	1	0.5	1.8	0.9996(0.0002)	0.4741(0.0032)	1.7310(0.1031)	0.8891(0.1289)	0.482(0.0373)
	2	1.5	1.4	2.00003(0.0005)	1.5118(0.0149)	1.4347(0.0453)	1.1336(0.1456)	0.5945(0.1627)
	3	2.5	1	2.998(0.0030)	2.517(0.0293)	1.0438(0.0708)	0.9917(0.0717)	0.6977(0.0227)
	4	3.5	0.6	4.0003(0.0004)	3.553(0.0565)	0.6709(0.0762)	1.0632(0.0787)	0.4466(0.0679)
	5	4.5	0.2	4.9981(0.0025)	4.5015(0.0241)	0.2021(0.0096)	1.0471(0.0785)	0.5631(0.0699)
1000	1	0.5	1.8	1.0013(0.0015)	0.5044(0.0080)	1.8126(0.0287)	1.0748(0.0796)	0.4821(0.0221)
	2	1.5	1.4	2.0001(0.0002)	1.5152(0.0200)	1.4419(0.0570)	0.9724(0.0528)	0.5257(0.0429)
	3	2.5	1	2.9991(0.0001)	2.4949(0.0106)	0.9913(0.0194)	0.9430(0.0707)	0.5173(0.022)
	4	3.5	0.6	4.0008(0.0001)	3.5253(0.0305)	0.6313(0.0384)	0.9828(0.0350)	0.4943(0.0102)
	5	4.5	0.2	5.0004(0.0006)	4.4408(0.0066)	0.1798(0.0012)	1.0400(0.0564)	0.5062(0.0180)

TABLE 6.6: Average values and MSE of dependency parameter estimates based on MLE and Kendals tau measure ($\hat{\theta}_{IFM}$) and Kendals tau measure ($\hat{\theta}_{\tau}$) and the estimate of the measure of dependence (τ) for θ=0.5, σ=1 and κ=0.5

n	α	β	δ	$\hat{\theta}_{IFM}$	$\hat{\theta}_{\tau}$	$\hat{\tau}$
500	1	0.5	1.8	0.5011(0.0373)	0.5126(0.0417)	0.4965(0.0252)
	2	1.5	1.4	0.4985(0.0396)	0.5131(0.0678)	0.4954(0.0190)
	3	2.5	1	0.5002(0.0380)	0.5134(0.0673)	0.4954(0.0255)
	4	3.5	0.6	0.4976(0.0679)	0.5142(0.0327)	0.4946(0.0211)
	5	4.5	0.2	0.6126(0.1212)	0.4116(0.1254)	0.4085(0.0249)
1000	1	0.5	1.8	0.4986(0.0398)	0.5024(0.0096)	0.4995(0.0195)
	2	1.5	1.4	0.5032(0.00043)	0.5112(0.0010)	0.4948(0.0016)
	3	2.5	1	0.5173(0.0205)	0.4983(0.0362)	0.5014(0.0130)
	4	3.5	0.6	0.5010(0.0219)	0.5020(0.0353)	0.4996(0.0122)
	5	4.5	0.2	0.4985(0.0189)	0.4960(0.0342)	0.5025(0.0103)

TABLE 6.7: Average values and MSE of Cointegration and distribution parameters based on MLE for θ=1, σ=3 and κ=2

n	α	β	δ	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\sigma}$	$\hat{\kappa}$
500	1	0.5	1.8	0.9988(0.0013)	0.477(0.0538)	1.7386(0.1808)	3.0441(0.3190)	2.5056(0.1160)
	2	1.5	1.4	2.0001(0.0008)	1.4783(0.0467)	1.3549(0.1125)	2.8935(0.1772)	1.7943(0.2955)
	3	2.5	1	3.0001(0.0015)	2.4071(0.0994)	0.8496(0.1608)	3.448(0.1123)	1.7896(0.2657)
	4	3.5	0.6	4.0022(0.0011)	3.5472(0.0718)	0.6739(0.1139)	2.8725(0.3827)	2.4273(0.1254)
	5	4.5	0.2	5.0001(0.0031)	4.634(0.1522)	0.2818(0.1001)	3.1861(0.2625)	2.1664(0.1370)
1000	1	0.5	1.8	1.0002(0.0003)	0.5491(0.0526)	1.9896(0.0119)	2.606(0.2252)	1.8371(0.1826)
	2	1.5	1.4	2.0007(0.0007)	1.5023(0.0125)	1.4071(0.0342)	2.9502(0.0994)	1.7509(0.1090)
	3	2.5	1	3.0002(0.00025)	2.5066(0.0169)	1.0141(0.0357)	3.1933(0.2307)	1.9113(0.1277)
	4	3.5	0.6	3.9997(0.0003)	3.5236(0.0276)	0.6301(0.0360)	3.4008(0.2428)	2.1087(0.1035)
	5	4.5	0.2	4.9998(0.0015)	4.4360(0.035)	0.1785(0.0246)	3.4352(0.1882)	1.8137(0.1036)

TABLE 6.8: Average values and MSE of dependency parameter estimates based on MLE and Kendals tau measure ($\hat{\theta}_{IFM}$) and Kendals tau measure ($\hat{\theta}_{\tau}$) and the estimate of the measure of dependence (τ) θ=1, σ=3 and κ=2

n	α	β	δ	$\hat{\theta}_{IFM}$	$\hat{\theta}_{\tau}$	$\hat{\tau}$
300	1	0.5	1.8	1.0066(0.0952)	1.0209(0.1526)	0.3320(0.0221)
	2	1.5	1.4	0.9862(0.2955)	0.9968(0.1358)	0.3383(0.0222)
	3	2.5	1	0.9932(0.1069)	0.3372(0.1625)	0.3386(0.0208)
	4	3.5	0.6	0.9991(0.1231)	1.0254(0.1832)	0.3322(0.0230)
	5	4.5	0.2	1.0142(0.1037)	1.0451(0.1741)	0.3270(0.0220)
500	1	0.5	1.8	0.9904(0.0442)	1.0014(0.0833)	0.3339(0.0124)
	2	1.5	1.4	0.9968(0.0522)	0.9856(0.0877)	0.3377(0.0136)
	3	2.5	1	0.9966(0.0581)	1.0071(0.0201)	0.3331(0.0126)
	4	3.5	0.6	1.0028(0.0470)	1.0013(0.0804)	0.3339(0.0117)
	5	4.5	0.2	0.9856(0.0578)	0.9950(0.0807)	0.3354(0.0103)

of the estimates. The comparison suggests that ML method of estimation is slightly efficient than IFM method for our proposed model.

Algorithm 13: Algorithm for Table 6.1 to 6.4

1 Set x_{10}, x_{20}, θ, $\Theta_1 = (s, \alpha, \beta, \delta)^{'}$ and $\Theta_2 = (\sigma, \alpha, \delta)^{'}$ for a sample of size n.
2 Generate η_1 and η_2 as described in Section 6.4.
3 Generate x_{1t} and x_{2t} using Equation (6.1) and (6.2).
4 Set the initial values $\alpha_0, \beta_0, \delta_0, \sigma_0$, s_0.
5 Record the estimates of Θ_1 as $\hat{\Theta}_1$ by solving Equations (6.8) to (6.11).
6 Compute $\hat{\Theta}_2$ by using Equations (6.12) to (6.14).
7 Set $\hat{\delta} = \frac{\hat{\delta}_1+\hat{\delta}_2}{2}$, $\hat{\alpha} = \frac{\hat{\alpha}_1+\hat{\alpha}_2}{2}$.
8 Obtain $\hat{\theta}$ by maximising Equation (6.15). Set $\theta = \hat{\theta}$
9 Repeat Steps 2 to 8, say 500 times.
10 Choose the value of estimates as the averages of estimates obtained in step 9.

Algorithm 14: Algorithm for Table 6.5 to 6.8

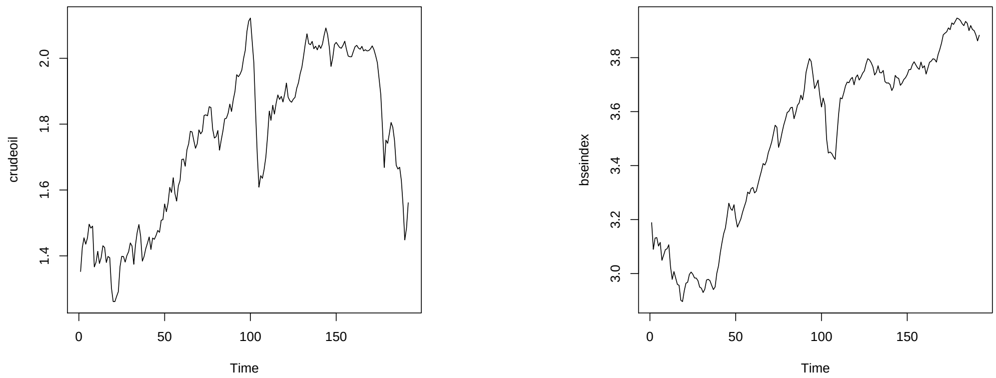
1 Set x_{10}, x_{20}, θ, $\Theta_1 = (s, \alpha, \beta, \delta)^{'}$ and $\Theta_2 = (\sigma, \alpha, \delta)^{'}$ for a sample of size n.
2 Generate η_1 and η_2 as described in Section 6.4.
3 Generate x_{1t} and x_{2t} using Equation (6.1) and (6.2).
4 Set the initial values $\alpha_0, \beta_0, \delta_0, \sigma_0$, s_0, θ_0.
5 Compute the estimates $\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\sigma}$, $\hat{s}$, $\hat{\theta}$ by solving equations (6.16) to (6.21).
6 Repeat Steps 2 to 6, say 500 times.
7 Choose the value of estimates as the averages of estimates obtained..

6.6 Data Analysis

To illustrate the use of the model developed above, we consider some real data sets. The data set consists of 192 monthly observations of crude oil price and Bombay stock exchange index for the period 2000 to 2016. All the variables are transformed in to their natural logarithm. We have used the same data set as given in Chapter 4.

Figure 6.4 provides the time series plot of the log transformed data and it indicates that the time series is non stationary. We also confirmed the non stationarity of time series by using Augmented Dickey Fuller test. The p-values for Augmented Dickey fuller test of the oil and diesel price is obtained as 0.9108 and 0.3208 respectively, indicates that both the variables are non stationary in nature.

Next we perform the estimation procedure of the cointegration model using inference functions for margins to find the parameter estimates of cointegration and the dependence parameter θ. Table 6.9 reports the AIC for the cointegration model with one and two lags. It is found that AIC is minimum for the CVAR(1) model compared to CVAR(2), and hence we use the model with one



(A) Time plot of crude oil price (B) Time plot of BSE index

FIGURE 6.4: Time series plot

TABLE 6.9: Information criterion

Model	AIC
CVAR(1)	-37638.018
CVAR(2)	-2498.425

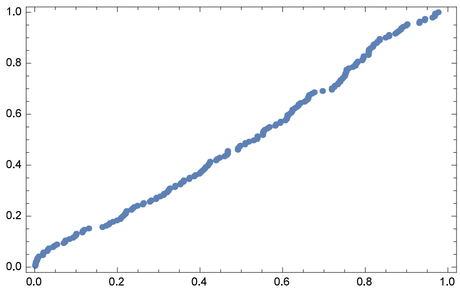
lag for the given data. The parameter estimates of the cointegration parameters and the copula parameters are obtained using the method of Inference functions for margins and are, $\hat{\theta}$=0.29240, $\hat{\alpha}$=-0.70509, $\hat{\beta}$=-0.06485, $\hat{\delta}$=0.02911 and the scale parameters of the distributions are σ=0.02077 and s=0.02888. Alternatively, the dependency parameter is also estimated from Kendall's tau measure of dependence using the formula, $\theta = \frac{(1-\tau)}{2\tau} = 0.3807$, where τ is the sample version of the measure of association (Kendall's Tau). It can be seen that the dependency parameter θ we obtained using the two step procedure is close to the one obtained using the dependency measure of association. The estimated cointegrating relationship is $x_{1t} - 0.70509x_{2t}$.

Now the residuals from the fitted cointegration model is observed and analysed.

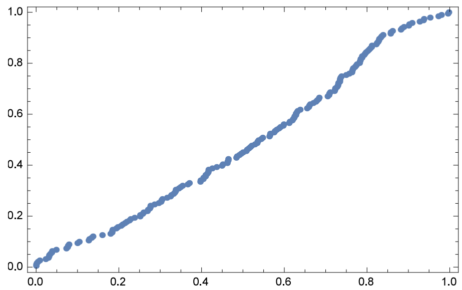
The residuals from the fitted error correction model is obtained as

$$\hat{\eta}_{1t} = \Delta x_{1t} - \hat{\delta}\hat{\beta}\hat{z}_{t-1}$$
$$\hat{\eta}_{2t} = \Delta x_{2t} + \hat{\delta}\hat{\beta}\hat{z}_{t-1}, t = 1, 2, ...$$

Using the above estimates of the ECM, we tested whether the marginal residual series follow a logistic and normal distribution using Kolmogorov-Smirnov test. The marginals of the data can be fitted by Logistic[0.003306, 0.0288] and Normal[-0.0166, 0.0207] with Kolmogorov- Smirnov test statistic values, 0.454 and 0.129 respectively. The probability-probability plots and histograms of the residuals are shown in Figure 6.5 and Figure 6.6. Plots indicate that the residuals follow the chosen distributions.

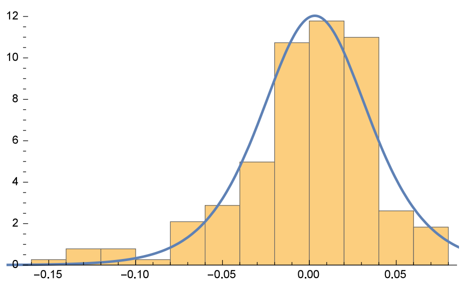


(A) Residual pp-plot of Crude oil price

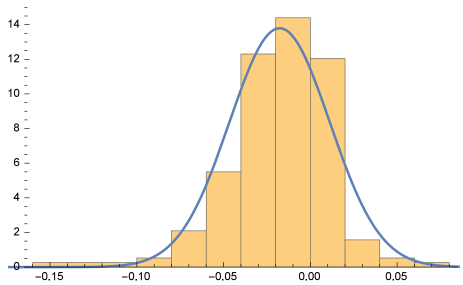


(B) Residual pp-plot of BSE index

FIGURE 6.5: Probability-Probability Plot



(A) Residual Histogram of Crude oil price



(B) Residual histogram of BSE index

FIGURE 6.6: Residual Histogram

Thus the joint density of the errors can be fitted by,

$$f\left(\eta_{1t},\eta_{2t}\right)=C\left(f\left(\Delta x_{1t}-\hat{\delta}\hat{\beta}(x_{1t-1}+\hat{\alpha}x_{2t-1})\right),f\left(\Delta x_{2t}+\hat{\beta}(x_{1t-1}+\hat{\alpha}x_{2t-1})\right);0.29240\right)$$

$$\times f\left(\Delta x_{1t}-\hat{\delta}\hat{\beta}(x_{1t-1}+\hat{\alpha}x_{2t-1})\right),f\left(\Delta x_{2t}+\hat{\beta}(x_{1t-1}+\hat{\alpha}x_{2t-1})\right).$$

Next we evaluate the goodness of fit of the copula model. There are many methods of goodness of fit tests for multivariate models in literature (D'Agostino & Stephens (1986) ,Fang et al. (2000)). We adopt the method of Fang et al. (2000) which is based on the copula technique and independent of the marginal distributions. Suppose that (X, Y) is a bivariate random vector with cdf F(x,y) and copula function C(u,v), then we have, that

$$K(w)=P\left(H(X,Y)\leq w\right)$$

is a univariate distribution function on the interval (0,1).

This K(w) can be used as a statistic to test the goodness of fit of a copula. For the Clayton survival copula, $K(w)=w-\frac{\varphi(w)}{\varphi'(w)}$, where $\varphi(w)$ is the generator of the copula.

Let $(x_1,y_1),(x_2,y_2)\cdots\cdots,(x_n,y_n)$ be a sample drawn from a bivariate distribution H(x,y). Then a non parametric estimate of K(w) is given by $K_n(w)=\sum_{i=1}^{n}\frac{\delta(w-W_i)}{n}$, where

$$W_i=\frac{1}{(n-1)}\#\left\{(x_j,y_j):x_j<x_i,y_j<y_i\right\},1\leq i\leq n,$$

and # represents the cardinality of a set.

The cdf K(w) with θ=0.29240 and the empirical distribution function $K_n(w)$ from the data are plotted in Figure 6.7. Now to test whether the chosen copula is

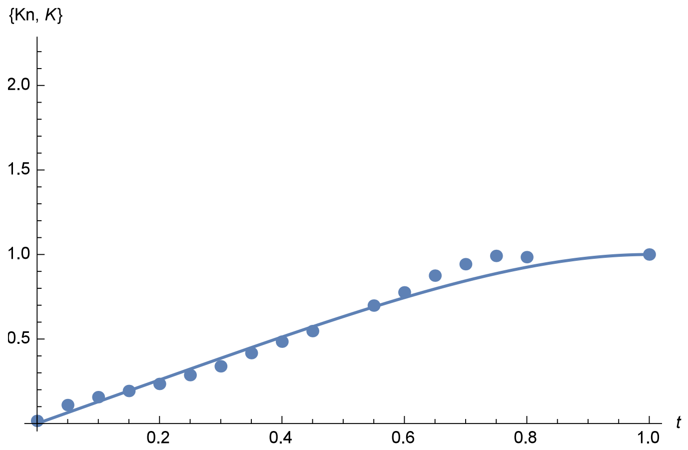


FIGURE 6.7: Plots of $K(w)$ and $K_n(w)$

suitable for the data, we test the hypothesis $H_0 : C(u, v) = C(u, v, \theta)$ with θ=0.29240 using the test statistics given in Franq et.al(2000),

$$D_n = \sqrt{n} \sup_{0 \leq w \leq 1} |K_n(w) - K(w)|.$$

D_n is exactly the Kolmogorov test statistic (Refer Saunders and Laud (1980)). For the the level of test 0.05, L(1-0.05)=1.358, where $L(r)$ is the limit distribution function of the Kolmogorov test statistic. We obtained the value of D_n as D_n=0.16682<1.358, thus we can not reject the null hypothesis and hence the given copula can be fitted by $C(u, v, 0.29240)$. Hence, the proposed distribution is suitable for the data. The estimates obtained using the method of maximum likelihood also lead to the same conclusion, so we omit those results here.

The results of this Chapter are communicated in Nimitha & Balakrishna (2017).

Chapter 7

Conclusions and Future Work

The focus of this thesis is modelling and analysis of bivariate cointegration under various non normal distributions. The method of analysing time series based on the assumption of linear ARMA models with Gaussian errors are found to be unrealistic in many areas of finance and economics as most of the financial time series deviates from Gaussianity. If two or more series under consideration are non stationary, but a linear combination of those series can brought to be a stationary series, then the variables are cointegrated. On taking this in to account, in the present thesis, we studied the modelling of bivariate cointegrating time series and examined their suitability in the presence of non Gaussian errors.

We have proposed a bivariate cointegration model with the errors generated from a iid logistic distribution. We have developed maximum likelihood estimation of the cointegration vector in a first order vector autoregressive model that allows for logistic innovations. Then we developed a likelihood ratio test to detect the presence of cointegration for the residuals of the ECM. All the estimating equations are solved by numerical techniques. From the simulation studies, it is observed that the proposed procedure is powerful for detecting the presence of unit root and cointegration. Along with the usual asymptotic test, a bootstrap test based on MLE is carried out to account for the size distortions caused by the finite samples. Finally the study provides an evidence of the long-run cointegrating relationship of Rubber consumption and Export series. Existence of cointegration with normally distributed errors using the Johansen likelihood ratio test has also been tested. The results indicate that the existence of cointegration relationship in the data set is identified with both normal and logistic errors. But the residuals obtained from the cointegrating regression using normal errors rejects the assumption of normality. Hence we proceeded with the vector autoregression model that allows for logistic innovations. The data analysis confirms that the proposed model detects the presence of cointegration.

We developed a bivariate cointegration model with the errors generated from a bivariate Student t distribution. The estimation procedure and testing for the presence of cointegration with bivariate Student t distributed errors is developed. The model parameters are estimated using the method of maximum likelihood. To evaluate the performance of the estimators and test statistic, a simulation study has been conducted. The simulation study shows that estimates perform reasonably well and become more accurate with increase of sample sizes. Finally, data analysis are conducted to illustrate the applications of the proposed model and found that model is a good fit for the data.

In the context of developing models for financial time series, we investigated the presence of non Gaussian heteroskedasticity in a bivariate cointegration model. We introduced a bivariate Cointegrating time series model with time varying conditional variances and covariances with constant and zero conditional correlations. We have considered two different forms of cointegration representation, namely the Engle and Granger's error correction form and Phillip's triangular representation, for estimating the long run relationship between the variables. We developed the maximum likelihood estimation of the parameter vector for our proposed model with Gaussian and non Gaussian innovations. As the estimating equations do not admit explicit solution, the likelihood equations are all solved by using numerical techniques such as the method of Fisher scoring and likelihood optimization technique. A simulation study has been conducted for estimating the parameters of the model. The simulation experiments show that the estimates perform reasonably well. Finally to illustrate the applications of the proposed model, two data sets are analysed and found that the proposed model is a good fit for the data sets.

We introduced a bivariate cointegration model with errors generated from a survival copula model. Unlike the usual copula density obtained through identical marginals, we obtained the bivariate density by specifying different marginal distributions. The parameters of the marginal distributions and copula function are estimated through simulation technique using the method of inference functions for margins and the method of maximum likelihood estimation. Even though ML estimates are found to be slightly better than IFM, mean square errors suggest that in both cases the estimates are found to be efficient. The applicability of the proposed model using the dependence structure is analysed and is illustrated by using financial time series of crude oil price and Bombay stock exchange index.

We summarise this thesis with the note that there are several unsolved problem that have to be undertaken. The problems related to forecasting of cointegrating time series under non Gaussian error distribution are yet to be discussed. In applied work, it is very common to render a strong seasonal time series in to a stationary series by seasonal differencing. For example, if $\{X_t\}$ is a non stationary quarterly time series, its seasonal difference $\Delta_4 X_t = X_t - X_{t-4}$ may be I(0). And, if two non stationary seasonal series $\{X_t\}$ and $\{Y_t\}$can be made I(0) by seasonal differencing and if there exists a linear combination $Y_t - \beta X_t \sim I(0)$, then the two series are called seasonally co-integrated. So we plan to undertake a study of series which are seasonally cointegrated under non Gaussian error distributions.

List of Accepted/ Communicated Papers

1. Nimitha, John & Balakrishna, N. (2018). Cointegration models with non Gaussian GARCH innovations . *Metron,* 76(1), 83-98 (Published).

2. Nimitha, John & Balakrishna, N. (2018). Unit root and Cointegration with Logistic innovations. *Journal of Indian Society of Agricultural Statistics,* 72(1), 39-48 (Published).

3. Nimitha, John & Balakrishna, N. (2017). Copula based bivariate Cointegration . *Calcutta Statistical Association Bulletin* (Communicated)

Bibliography

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics*, *21*(1), 243–247.

Alias, M. H. & Othman, J. (1997). Cointegration between palm oil price and soyabean oil price: A study on market integration. *Malaysian Oil Science and Technology*, *6*(2), 92–95.

Balakrishnan, N. & Lai, C. D. (2009). Simple forms of the bivariate density function. In *Continuous Bivariate Distributions* (pp. 351–400). Springer.

Barnard, G. (1991). The theory and applications of statistical inference functions. lecture notes in statistics 44-dl mcleish, cg small. *Metrika*, *38*, 249–250.

Bartlett, M. & Rajalakshman, D. (1953). Goodness of fit tests for simultaneous autoregressive series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 107–124.

Berndt, E. R., Hall, B. H., Hall, R. E., & Hausman, J. A. (1974). Estimation and inference in nonlinear structural models. In *Annals of Economic and Social Measurement, Volume 3, number 4* (pp. 653–665). NBER.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

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

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The review of economics and* *statistics*, 498–505.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, *52*(1-2), 5–59.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). Arch models. *Handbook of econometrics*, *4*, 2959–3038.

Boswijk, H. P., Lucas, A., & Taylor, N. (1999). A comparison of parametric, semi-nonparametric, adaptive, and nonparametric cointegration tests.

Box, G., Jenkins, G., & nsel, G. (1994). Time series analysis, englewood cliffs.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Box, G. E. & 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. & Davis, R. A. (1987). Time series theory and methods.

Brooks, C. (2014). *Introductory econometrics for finance*. Cambridge university press.

Chen, X. & Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, *130*(2), 307–335.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula methods in finance*. John Wiley & Sons.

Choroś, B., Ibragimov, R., & Permiakova, E. (2010). Copula estimation, 77–91.

Crabtree, B. F., Ray, S. C., Schmidt, P. M., O'Connor, P. T., & Schmidt, D. D. (1990). The individual over time: time series applications in health care research. *Journal of clinical epidemiology*, *43*(3), 241–260.

Creal, D., Koopman, S. J., & Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics*, *29*(4), 552–563.

D'Agostino, R. & Stephens, M. (1986). Goodness-of-fit techniques marcel dekker new york.

DeVries, W. & Wu, S. (1978). Evaluation of process control effectiveness and diagnosis of variation in paper basis weight via multivariate time-series analysis. *IEEE Transactions on Automatic Control*, *23*(4), 702–708.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427–431.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F. & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251–276.

Engle, R. F. & Kraft, D. (1983). Multiperiod forecast error variances of inflation estimated from arch models. *Applied time series analysis of economic data*, 293–302.

Engle, R. F. & Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric theory*, *11*(01), 122–150.

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

Fang, K.-T., Fang, H.-B., & Rosen, D. v. (2000). A family of bivariate distributions with non-elliptical contours. *Communications in Statistics-Theory and Methods*, *29*(8), 1885–1898.

Fuller, W. (1976). *Introduction to statistical time series, Chapter 6*. New York, NY (USA) Wiley.

Gómez, E., Gomez-Viilegas, M., & Marin, J. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics-Theory and Methods*, *27*(3), 589–600.

Granger, C. W. (1981). Some properties of time series data and their use in econometric model specification. *Journal of econometrics*, *16*(1), 121–130.

Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton university press Princeton.

Hannan, E. J. (1970). *Multiple time series*, volume 38. John Wiley & Sons.

Hayashi, F. (2000). Econometrics. *Princeton University Press Princeton, NJ*, 633–636.

Hosking, J. R. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, *75*(371), 602–608.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Joe, H. & Xu, J. J. (2016). The estimation method of inference functions for margins for multivariate models. *Technical Report No. 166, Department of Statistics, University of British Columbia*.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, *12*(2), 231–254.

Johansen, S. & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration-with applications to the demand for money.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). 14: lognormal distributions. *Continuous univariate distributions*, *1*.

Kim, K. & Schmidt, P. (1993). Unit root tests with conditional heteroskedasticity. *Journal of Econometrics*, *59*(3), 287–300.

Kosapattarapim, C., Lin, Y.-X., & McCrae, M. (2013). Size and power of cointegration tests with non-normal garch error distributions.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, *53*(284), 814–861.

Lee, T.-H. & Long, X. (2009). Copula-based multivariate garch model with uncorrelated dependent errors. *Journal of Econometrics*, *150*(2), 207–218.

Lee, T.-H. & Tse, Y. (1996). Cointegration tests with conditional heteroskedasticity. *Journal of Econometrics*, *73*(2), 401–410.

Li, W., Ling, S., & Wong, H. (2001). Estimation for partially nonstationary multivariate autoregressive models with conditional heteroscedasticity. *Biometrika*, *88*(4), 1135–1152.

Li, W. & McLeod, A. (1981). Distribution of the residual autocorrelations in multivariate arma time series models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 231–239.

Liitkepohl, H. (1991). Introduction to multiple time series analysis. *Berlin et al*.

Ljung, G. M. & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Mak, T., Wong, H., & Li, W. (1997). Estimation of nonlinear time series with conditional heteroscedastic variances by iteratively weighted least squares.

Computational statistics & data analysis, *24*(2), 169–178.

Mak, T. K. (1993). Solving non-linear estimation equations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 945–955.

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

Murphy, S. A. & Van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, *95*(450), 449–465.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Nimitha, J. & Balakrishna, N. (2017). Copula based bivariate cointegration model. *Calcutta Statistical Association Bulletin*.

Nimitha, J. & Balakrishna, N. (2018a). Cointegration models with non gaussian garch innovations. *METRON*, *76*(1), 83–98.

Nimitha, J. & Balakrishna, N. (2018b). Unit root and cointegration with logistic errors. *Journal of the Indian Society of Agricultural Statistics*, *72*(1), 39–48.

Pagan, A. (1996). The econometrics of financial markets. *Journal of empirical finance*, *3*(1), 15–102.

Palm, F. C. (1996). 7 garch models of volatility. *Handbook of statistics*, *14*, 209–240.

Phillips, P. C. (1991). Optimal inference in cointegrated systems. *Econometrica: Journal of the Econometric Society*, 283–306.

Reinsel, G. C. (2003). *Elements of multivariate time series analysis*. Springer Science & Business Media.

Scaillet, O. & Fermanian, J.-D. (2002). Nonparametric estimation of copulas for time series.

Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.

Taylor, S. J. (1986). *Modelling financial time series*. world scientific.

Tiku, M. L., Wong, W.-K., & Bian, G. (1999). Estimating parameters in autoregressive models in non-normal situations: Symmetric innovations. *Communications in Statistics-Theory and Methods*, *28*(2), 315–341.

Tiku, M. L., Wong, W.-K., Vaughan, D. C., & Bian, G. (2000). Time series models in non-normal situations: Symmetric innovations. *Journal of Time Series Analysis*, *21*(5), 571–596.

Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.

Tsay, R. S. (2013). *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons.

Tsay, R. S. & Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary arma models. *Journal of the American Statistical Association*, *79*(385), 84–96.

Vrontos, I. D., Dellaportas, P., & Politis, D. N. (2003). A full-factor multivariate garch model. *The Econometrics Journal*, *6*(2), 312–334.

Whittle, P. (1953). The analysis of multiple stationary time series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 125–139.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, *9*(1), 60–62.

William, W. & Wei, S. (2006). Time series analysis: univariate and multivariate methods. *USA, Pearson Addison Wesley, Segunda edicion. Cap*, *10*, 212–235.

Wong, H., Li, W., & Ling, S. (2005). Joint modeling of cointegration and conditional heteroscedasticity with applications. *Annals of the Institute of Statistical Mathematics*, *57*(1), 83–103.

Wong, W.-K. & Bian, G. (2005). Estimating parameters in autoregressive models with asymmetric innovations. *Statistics & probability letters*, *71*(1), 61–70.

Xu, J. J. (1996). *Statistical modelling and inference for multivariate and longitudinal discrete response data*. PhD thesis, University of British Columbia.

Yang, S.-R. & Brorsen, B. W. (1992). Nonlinear dynamics of daily cash prices. *American Journal of Agricultural Economics*, *74*(3), 706–715.