TIME SERIES FORECASTING OF POST-WAR TOURISM PROSPECTS FOR SRI LANKA

Sebastian Reyalt Gnanapragasam

(148377 K)

Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science in Operational Research

Department of Mathematics

University of Moratuwa

Sri Lanka

June 2016

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 22-06-2016

The above candidate has carried out research for the Masters Dissertation under my supervision.

Name of the supervisor: Mr. T. M. J. A. Cooray

UOM Verified Signature

Signature of the supervisor: Date: 23/06/2016

ABSTRACT

Tourism plays a big role in the development of a country in terms of economics as it is one of the biggest and fastest-growing economic sectors in the world. It accounts for a large part of Gross Domestic Product of any country through Foreign Exchange. This study focused on international tourist arrivals to Sri Lanka. In the past, nearly three decades, Sri Lanka had to face conflict within the country. Tourists had less interest of visiting Sri Lanka, mainly due to the uncertainty of security. Nevertheless, the internal conflict is over and tourist arrivals have dramatically increased over last six years.

The aim of this study is to investigate the impact of internal conflict in Sri Lanka for tourist arrivals by splitting the entire time frame by *before* and *after* the conflict as two windows. Further this study discusses the factors which are influenced by tourism in Sri Lanka. The data for the study is extracted from the annual reports of the Sri Lanka Tourism Development Authority.

Time series models are developed in two separate time windows by using the methods: Holt-Winters' Exponential Smoothing, Seasonal Autoregressive Integrated Moving Average (ARIMA) modeling, State Space modeling and Dynamic Transfer Function modeling. All necessary tests are carried out for model development, diagnostic checking and forecast.

In the empirical study, behavior of arrivals with its trend and seasonal patterns are analyzed, best models are developed based on the accuracy of fitted models in terms of Mean Absolute Percentage Error (MAPE) values and the impact of the factors influenced by tourism are deeply discussed. MAPE values for the recommended models for *after* the conflict are less than 7%. In both windows, Seasonal ARIMA method performs the best. Moreover it is estimated by ex-post forecast that, 2.085 million international tourist arrivals can be expected in the year 2016.

Key words: ARIMA, Dynamic Transfer Function, State Space, Tourist Arrivals

Dedicated to my Father

ACKNOWLEDGEMENT

Behind the success of my research work, there are many people who deserve my wholehearted gratitude.

To begin with I extend my sincere thanks to my advisor Mr. T. M. J. A. Cooray, Senior Lecturer in the Department of Mathematics, University of Moratuwa, who stood like a shadow, for his scholarly guidance and encouragement. His advice and narration provided an excellent environment for me to conduct my research successfully.

I would like to be grateful to the Head of the Department of Mathematics, Programme Coordinator and the Teaching Panel of the M. Sc in Operational Research degree programme at the University of Moratuwa.

My sincere thanks also goes to the Vice Chancellor of the Open University of Sri Lanka, Dean of the Faculty of Natural Sciences, Head of the Department of Mathematics and Computer Science for granting permission to follow this Master degree programme. Furthermore, I must thank all the staff members in the Department of Mathematics and Computer Science for their support in numerous ways to complete this post graduate study on time.

It would not be possible for me to do this research work without the support of my family. Therefore, I take this opportunity to thank my dearest wife S. Parimala Devi, who was there to render help whenever I felt so downhearted and always being there for me.

Finally, I offer my sincere gratitude for everyone else of whose name I might have forgotten to mention.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
ACF	Auto Correlation Function
AD	Anderson Darling
ADF	Augmented Dickey Fuller
AIC	Akaike information criterion
AR	Autoregressive
ARMA	Autoregressive and Moving Average
ARIMA	Autoregressive Integrated Moving Average
D1Y	1st difference of Y
D12D1Y	12th differences of D1Y
DES	Double Exponential Smoothing
DTF	Dynamic Transfer Function
DW	Durbin Watson
FE	Foreign Exchange
GDP	Gross Domestic Product
HW	Holt Winters
KS	Kolmogorov Smirnov

LM	Lagrange's Multiplier
LOG	Logarithm
MA	Moving Averages
MAPE	Mean Absolute Percentage Error
PACF	Partial Auto Correlation Function
SAS	Statistical Analysis Software
SBC	Schwartz's Bayesian Criterion
SES	Single Exponential Smoothing
SLTDA	Sri Lanka Tourism Development Authority
SQRT	Square root
SS	State Space
SSE	Sums of Squares of Residuals
VAR	Vector Autoregressive

LIST OF APPENDICES

1. INTRODUCTION

1.1 International Tourism

Over the past six decades, tourism has experienced continued expansion and diversification, to become one of the largest and fastest-growing economic sectors in the world. Also tourism is one of the biggest industries in the world. It plays a big role in the development of a country in terms of economics as it is one of the largest economic sectors in the world. Tourism basically contains two types as domestic and international. This study is mainly focused on the international tourist arrivals to Sri Lanka. International tourism can be defined as an activity of visitors who make temporary visits across international borders and remains for more than 24 hours. The purpose of visits can be visiting relatives and friends; leisure, business meeting or conventions, education, healthcare and sports.

The tourism industry is one of the most important sectors for a successful economy as it accounts for a large part of Gross Domestic Product (GDP), foreign exchange generation and employment figures of any country. International tourist arrivals and international tourist receipts have traditionally been used as benchmark combined series to assess the overall importance of tourism. The twentieth century witnessed a steady increase in tourism all over the world. The level of organization, methods of transportation, and facilities available at destination points have enjoyed an accelerated pace of improvement. Smart entrepreneurs in the field must know how to analyze the available data and interpret them to their advantage.

1.2 Tourism and Economics

Tse (2001) identifies that, tourism in a developing country plays a fundamental role in economic growth and development. Also, it is used in order to make available more employment opportunities. Further, the exchange rate and the price level are important determinants the demand of foreign tourist services. Therefore, tourism has been identified as a key sector of any economy. Countries offer a range of investment incentives that should be targeted at development of tourism to meet the much needs of visitors. Wanhill (1994) examines the nature of investment incentives, their impact, and some methods to evaluate their worth. Further it shows that tourism is a demand-led industry and its influence pervades in different sectors of the economy.

1.3 Importance of Forecast in Tourism

Forecasting plays a major role in tourism planning and predicting future arrivals is a very difficult task. Prediction of tourist arrivals is essential for planning, policy making and budgeting purposes by tourism operators. Policy makers may subsequently be convinced to assist tourism development and further increase profitability from tourism activities. Currently, the governments of developing countries pay much attention to the growth of the number of tourists in their soil.

Thus, forecasting the tourism demand becomes very important. With a correct tourism demand forecasting model that could accurately predict the tourism demand, the government would be able to invest properly and effectively to build tourist infrastructures. Thus proper investment can be done for the airlines, buses, tourist hotels, restaurants, parks, souvenir shops, shopping malls, and etc. A comfortable and enjoyable travel journey may provide tourists with a great travel experience and increase their likelihood of returning. Therefore, a highly accurate forecast of the tourism demand would have a positive influence on the government, the private sectors, and the tourists. The commitment to develop tourism business in other geographical areas would be much easier if it were possible to analyze current and past tourist traffic and predict the nature of changes in tourism demand.

1.4 Tourism in Sri Lanka

The Sri Lanka Tourism Development Authority (SLTDA) was formed as the apex body for Sri Lanka Tourism under section 2 of the Tourism Act (No. 38 of 2005). SLTDA is also formally known as Ceylon Tourist Board, or Sri Lanka Tourist Board and Sri Lanka Tourism Board. The organization is committed towards transforming Sri Lanka to be Asia's foremost tourism destination. The SLTDA will strive to develop diverse, unique and quality tourism services and products that would make Sri Lanka as a unique destination, globally.

1.5 Sri Lankan Tourism and its Development

Sri Lanka entered the international tourism market in the year 1960. Since then, government involvement has been a key factor in tourism development in Sri Lanka. The Ceylon Tourist Board was established in 1966 in order to provide direction and leadership to this promising sector of the developing economy. The tourism sector has been instrumental in generating foreign exchange, employment opportunities and household income for Sri Lankans, as it has in many other developing economies. Therefore, the development of the tourism sector appears to have as important as the development of other sectors of the economy in Sri Lanka.

There were several set-backs during last few decades in the tourism development process such as global economic crisis in 2009, Tsunami in 2004 and civil war from years 1983 to 2009. Study of Kurukulasooriya & Lelwala (2011) says that due to Tsunami affect, tourism in Sri Lanka recorded negative growth with dropping of arrivals by 3% in year 2005. However after year 2009, the Sri Lankan tourism sector has grown significantly, contributing remarkably to GDP. According to Annual Statistical Report of Sri Lanka Tourism Development Authority (2014), international tourist arrivals to Sri Lanka has increased from 18,969 in year 1966 to 1,527,153 in year 2014 exceeding all time high hits in the history which is an increase of 19.8% over the previous year's 1,274,593 arrivals. International tourism receipts also increased from US$ 1.3 to US$

1,245 during the same period, which is an increase of 4.0% compared to the year 2013. Further, this sector's contribution to the direct and indirect employment opportunities increased from 12,078 in year 1970 to 299, 790 in year 2014 which is an increase of 4.0% with year 2013. Although tourism plays a key role in the Sri Lankan economy, a little attention has been paid to this sector in the empirical research.

1.6 Development Planning in Sri Lanka

The tourism sector has been identified as one of the key sectors propelling the country's economic growth. According to the 5 year strategy "*Tourism Development Strategy 2011- 2016*" by the Ministry of Economic Development in Sri Lanka (2011), the ultimate beneficiaries of tourism development strategy should be the people of the country: the farmers who supply rice, vegetables and fruit, the fishermen who deliver the catch of the day, the craftsmen who produce souvenirs, the guides who escort the visitors and the young men and women serving in the industry with the unique Sri Lankan smile and hospitality. It is estimated that by year 2016 the industry is capable of creating 500,000 direct and indirect employments. One of the key objectives of the 5 year strategy is 'Increase tourist arrivals from 650,000 in 2010 to 2.5 Million by 2016'.

It is important that the country moves away from the low cost tourism and focuses on high end tourism. To attract high spending tourists the following are being focused.

- A product that is appealing to the high spenders (business, leisure, shopping wildlife etc.)
- High quality accommodations and service
- Value addition and product segmentation

Furthermore, to reach a high international tourist arrivals level, it may be used in advertising campaigns and also in political discussions to legitimize and emphasize the success of a country in the international community.

1.7 Economic Impact on Tourism in Sri Lanka

Tourism plays an important role in Sri Lankan economy as well. The study of Wickremasinghe & Ihalanayake (2006) reveals its contribution to Gross Domestic Product (GDP). There are some empirical studies which discovered the impact of tourism on economic growth worldwide (see Balaguer & Cantavella (2002), Durbarry (2004) and Dritsakis (2004)). Particularly, King & Gamage (1994) attempts to address some of the issues relevant to the economic impact of tourism in Sri Lanka.

1.8 Social and Cultural aspects of Tourism in Sri Lanka

Some research has examined the impact of tourism on younger residents of communities. Crick (1989 & 1992) in this regard, focuses on the impact of tourism on children. This work expresses concern about the harm that tourism could do to the young. Also it examines how the young themselves see tourism. Further, it reflects the social and cultural aspects, particularly the negative impacts of tourism in Sri Lanka. The impact of political upheaval on tourism in Sri Lanka is discussed in Gamage *et al.* (1997). In the ethnic tourist study of Gamage & King (1999), it is found that ethnic tourists spend more on retail goods and less on hotels and restaurants. The lower spending on hotels by the ethnic tourists is not surprising as they have ethnic relations.

According to the literature, very little study has examined the causal nexus between tourism growth and economic growth in Sri Lanka.

1.9 Aim of the study

The aim is to investigate the impact of internal conflict in Sri Lanka for tourist arrivals by splitting the entire time frame as *before* and *after* the internal conflict and study the behavior of tourist arrivals in these two time spans. To achieve this goal, the following objectives are attained:

1.9.1 Objectives:

- To develop time series models in the two separate time windows by using the following techniques:
    - ➢ Holt- Winters' Exponential Smoothing method
    - ➢ Seasonal ARIMA modeling method
    - ➢ Dynamic Transfer Function modeling method
    - ➢ State Space modeling method
- To identify the influential factors on tourist arrivals

- To recommend the best models based on the accuracy of fitted models in terms of MAPE values (Based on the Ex- ante forecast)

- To forecast (Ex-post forecast) monthly tourist arrivals for the entire year 2016 by comparing with available capacity to cater them and making recommendations.

1.10 Source of data

The data for the study is extracted from the statistical annual reports of Sri Lanka Tourism Development Authority (SLTDA).

1.11 Outline of the Dissertation

The study mainly focuses on investigating impact of internal conflict in Sri Lanka on tourist arrivals. After providing an introduction in the first chapter, the chapters in the rest of the dissertation are organized as follows:

2. **Literature Review**: A detailed review of the previous studies related to this research such as studies relevant to time series model fitting using different methods to forecast the tourist arrivals in different countries and relationship between tourist arrivals and revenue of a country.

3. **Methodology**: In this chapter, almost all the important techniques and tests necessary for time series analysis are explained with their fundamental statistical theories and statistical methodologies.

4. **Preliminary Analysis**: This chapter provides the results of the preliminary analysis prior to the time series model development. The behavior of the series, in different time frames, with their descriptive statistics of tourist arrivals is discussed.

5. **Model Development in Window I**: Detailed discussions about model fitting, in the time span of *before the conflict*, are included in this chapter. Results from all four model fitting methods are discussed. In depth discussion is also carried out by splitting the time frame further into two spans.

6. **Model Development in Window II**: Results obtained from all four time series model fitting methods in the time span of *after the conflict* are discussed in detail followed by a discussion on the other factors influenced by tourism in Sri Lanka.

7. **Factors influenced by tourism in Sri Lanka:** In this chapter, the discussions are based on some factors which are influenced by the tourism in Sri Lanka. Also some highlights relevant to tourism in Sri Lanka are reported at the latter part.

8. **Conclusions**: This chapter presents the conclusions of the study which are based on the results and discussions in the Chapters 4, 5 and 6. In addition, it concludes with the ex-ante forecast for the entire year 2016.

9. **Recommendations**: Based on the study, some recommendations are made to improve the tourism industry in Sri Lanka as well as few suggestions for further research is mentioned.

2. LITERATURE REVIEW

In this chapter, previous studies on international tourist arrivals and its forecasting by using different methods are discussed.

2.1 Previous studies using Exponential Smoothing modeling method

A formulation for the additive Holt–Winters forecasting procedure is provided by Bermudez *et al.* (2007) that simplifies both obtaining maximum likelihood estimates of all unknowns (smoothing parameters and initial conditions) and the computation of forecasts (point forecasts and reliable predictive intervals). By applying Holt-Winters modeling method to UK arrivals by air and comparing the resulting forecasts with those obtained in previous studies show that forecasting accuracy is high in exponential Holt-Winters model. Witt *et al.* (1992) uses Exponential smoothing to domestic tourism in Las Vegas and shows that it obtains a level of accuracy comparable to those of other more sophisticated models.

In Lim & McAleer (2001) study, various exponential smoothing models are estimated to forecast quarterly tourist arrivals to Australia from Hong Kong, Malaysia, and Singapore. The Holt–Winters Additive and Multiplicative Seasonal models outperform the Single, Double, and the Holt–Winters Non-Seasonal Exponential Smoothing models in forecasting. Akuno *et al.* (2015) attempts to forecast tourist arrival using double exponential smoothing and the Autoregressive Integrated Moving Average (ARIMA). When the forecasts from these models are validated, Double Exponential Smoothing model has performed better than the ARIMA model. The general theme of these two papers is that exponential smoothing methods perform quite well although the choice of exponential smoothing methods is very subjective.

2.2 Previous studies using ARIMA modeling method

Cho (2001) examines techniques such as, exponential smoothing, univariate Autoregressive Integrated Moving Average (ARIMA), and adjusted ARIMA to forecast the number of tourists from different countries to Hong Kong. According to the analysis, adjusted ARIMA with economic indicators seems to be the best forecasting method for Japan, whereas univariate ARIMA is the best predictor for the United States and United Kingdom. Univariate ARIMA and adjusted ARIMA behave similarly for countries like Taiwan, Singapore, and Korea. Among the three forecasting methods, exponential smoothing is the least accurate. This shows that univariate ARIMA and adjusted ARIMA are more suitable and can be applied to forecast the fluctuating series of tourists arrivals.

Six forecasting time series approaches such as Naïve I, Naïve II, Linear Trend, Sine Wave, Holt- Winters and ARIMA are used in Chu (1998) and examines empirically using monthly international tourist arrivals data in ten countries, which are; Taiwan, Japan, Hong Kong, South Korea, Singapore, the Philippines, Indonesia, Thailand, New Zealand and Australia. The results show that the accuracy of the forecasts differs depending on the country, but that the seasonal-nonseasonal ARIMA model is overall the most accurate method for forecasting international tourist arrivals.

To generate the forecast of international tourism demand for Malaysia, Loganathan & Yahaya (2010) applies ARIMA model. Seasonal ARIMA approaches have been suggested through this study and the forecasting process is based on this combination. Moreover, this study finds that the ARIMA model has offered valuable insights and has provided reliable forecasts of tourism demand for Malaysia. Forecasting model of tourists to Taiwan is built by Chang *et al.* (2011) using ARIMA, artificial neural networks, and multivariate adaptive regression splines. Analytic results demonstrate that ARIMA outperformed other two approaches and provides effective alternatives for forecasting tourism demand.

A short-term forecast on tourist arrivals in Greece is generated by Dimitrios *et al.* (2012). It is found that ARIMA model outperforms exponential smoothing models in forecasting the direction of one year out of sample forecasts. However, this does not interpret into point forecasting accuracy. A model has been fitted by Saayman & Saayman (2010) to forecast tourism to South Africa from the country's main intercontinental tourism markets. These include Great Britain, Germany, the Netherlands, the United States of America and France. The results show that seasonal ARIMA models deliver the most accurate predictions of arrivals. Prasert *et al.* (2008) focuses on forecasting methods to forecast international tourism arrivals to Thailand. Some of the forecasting methods are employed in this study namely Seasonal ARIMA, ARIMA, Holt-Winter-Additive, Holt-Winter-Multiplicative, Holt-Winter-No seasonal, Neural network. The results confirm that the best forecasting is Seasonal ARIMA model and it predicts that international tourism arrivals to Thailand have positive growth rate.

2.3 Previous studies using State Space modeling method

A model has been fit by Athanasopoulos & Hyndman (2006) to forecast Australian domestic tourism demand using a regression framework to estimate important economic relationships for domestic tourism demand. To explore the time series nature of the data, they use innovation state space models to forecast the domestic tourism demand. They build innovation state space models with exogenous variables. These models are able to capture the time series dynamics in the data, as well as economic and other relationships. They show that these models outperform alternative approaches for short-term forecasting and also produce sensible long-term forecasts.

2.4 Previous studies using Dynamic Transfer Function modeling method

The transfer function model is fit to the Natural Rubber production in India by Arumugam & Anithakumari (2013). This model has been used to identify a model and estimate parameters for forecasting of rubber production.

2.5 Previous studies using Neural Network modeling method

Traditional tourism demand forecasting techniques concentrate mainly on univariate time-series models. Law (2000) extends the applicability of neural networks in tourism demand forecasting by incorporating the back-propagation learning process into a non-linearly separable tourism demand data. Empirical results indicate that utilizing a back-propagation neural network outperforms regression models, time-series models, and feed-forward neural networks in terms of forecasting accuracy.

Burger *et al.* (2001) compares a variety of time-series forecasting methods to predict tourism demand for a certain region. A variety of techniques are employed in this survey, namely moving average, decomposition, single exponential smoothing, ARIMA, multiple regression, genetic regression and neural networks with the latter two methods being non-traditional techniques. The actual and predicted number of visitors is then compared. The survey shows that the neural network method performs the best.

Cho (2003) investigates the application of three time-series forecasting techniques, namely exponential smoothing, univariate ARIMA, and Elman's Model of Artificial Neural Networks, to predict travel demand from different countries to Hong Kong. Neural Networks seems to be the best method for forecasting visitor arrivals, especially those series without obvious pattern.

2.6 Previous studies on forecasting tourists in Sri Lanka

A comprehensive study of time series behavior of the postwar international tourist arrivals to Sri Lanka has been carried out by Kurukulasooriya & Lelwala (2014). The empirical study is based on tourist arrivals from all origins that create a demand for tourism in Sri Lanka. In the modeling exercise, classical time series decomposition approach is employed. It demonstrates that the linear trend component and seasonal fluctuations are the two prominent components whereas multiplicative model is comparatively the most accurate model in forecasting.

Study of Kurukulasooriya & Lelwala (2011) is to select the most accurate time series model for forecasting international tourist arrivals to Sri Lanka and generate short – term forecasts. It aims to analyze different forecasting models; exponential smoothing models, on monthly international tourist arrivals to Sri Lanka. From several types of models best performing models are selected. Empirical results show that Holt – Winter's exponential smoothing model with multiplicative seasonality is the best performing model for international tourist arrivals.

2.7 Comprehensive Reviews

Song & Li (2008) provides comprehensive reviews of published studies on tourism demand modeling and forecasting since 2000. One of the key findings of this review is that the methods used in analyzing and forecasting the demand for tourism has been more varied. In addition to the most popular time-series and econometric models, a number of new techniques have emerged in the literature.

2.8 Synopsis

According to the literature, as far as the forecasting accuracy is concerned, the studies show that **there is no single model that consistently outperforms other models in all situations**. Also there is no study that has carried out to forecast international tourist arrivals to Sri Lanka using State Space (SS) modeling method and Dynamic Transfer Function (DTF) modeling method.

Therefore this study fits models to forecast international tourist arrivals to Sri Lanka by using four forecasting techniques, namely, Holt- Winters' Smoothing method, Seasonal ARIMA method, Dynamic Transfer Function method and State Space method. Ultimately, the recommendations are made to the best models based on the least Mean Absolute Percentage Error (MAPE) values.

3. METHODOLOGY

3.1 Preliminary Analysis

At preliminary stage prior to model fittings, the following techniques are carried out.

3.1.1 Time Series Plot

The plot of time series is generally used to get an idea about the data and its behaviour. After undertaking the background of research regarding the series of interest, it is first necessary to visually examine the data. The plot is to inspect it for extreme observations, missing data, or elements of non-stationary such as trend or seasonality or cyclic pattern or irregular variations.

3.1.2 Augmented Dickey- Fuller test

Augmented Dickey- Fuller (ADF) test is used to test whether the series has a unit root. It is to confirm, statistically, that the stationary of series in terms of trend availability.

Test statistic for the model $Y_t = \rho Y_{t-1} + u_t$ and $-1 < \rho < 1$, is $DF = \frac{\hat{\rho}}{SE(\hat{\rho})} \sim t_{n-1}$

where u_t is the white noise and n is the number of observations.

Hypothesis: H_0: series is non-stationary $(|\rho| = 1)$ versus H_1: series is stationary $(|\rho| < 1)$.

3.1.3 Kruskal- Wallis Test

Kruskal- Wallis test is used to confirm the seasonality in the series. The hypothesis to be tested in this test is, H_0: series has no seasanality versus H_1: series has seasonality. The test statistic of Kruskal- Wallis test is defined as:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{n_i} \frac{R_i^2}{n_i} - 3(N+1) \sim \chi^2_{L-1},$$ where N is the total number of rankings,

R_i is the sum of the rankings in a specific season, n_i is the number of rankings in a specific season and L is the length of season.

3.1.4 Transformation to Stationary Series

There are usually three methods to transform non-stationary to stationary series, which are: regular differencing, seasonal differencing and variance stabilizing methods.

3.1.4.1 Regular Differencing method

If the series has a trend then by taking first difference (or at most 2 differences) the trend can be eliminated from the series and it is defined as follows:

$W_t = Y_t - Y_{t-L}$, where Y_t is the response variable at time t and $L = 1$ or 2.

3.1.4.2 Seasonal Differencing method

In this method differences are taken at seasonal lags. If the peaks appear seasonally in the autocorrelation function at particular lags, then it can be assumed that there is a seasonal pattern in the series. In this case, the difference is taken with the particular lag (such as lags 3, 4, 6 or 12) to remove the seasonality from the series. It is defined as:

$W_t = Y_t - Y_{t-L}$, where L is the length of season.

3.1.4.3 Variance Stabilizing method

If the series has fluctuation with extreme level, then it has to be transformed using Box-Cox transformation. Based on the estimated value suggested by Box-Cox transformation, the series can be converted to a stationary series.

3.1.5 Autoregressive process

Autoregressive (AR) process is a regression process with lagged values of the dependent variable in the independent variable positions; hence it is named autoregressive process. The expression of a p th -order autoregressive process, AR(p), is :

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t ,$$

where

Y_t = response variable at time t, Y_{t-k} = observation (predictor variable) at time $t-k$,
ϕ_i = regression coefficients to be estimated and ε_t = error term at time t

3.1.6 Moving Averages process

Moving Averages (MA) process is also a regression process with the dependent variable, Y_t, depending on previous values of the errors rather than on the variable itself. The expression of a q th -order moving average process, MA(q), is :

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

where Y_t = response variable at time t, μ = constant mean of the process,

θ_i = regression coefficients to be estimated and ε_{t-k} = error in time period t - k

3.1.7 Autocorrelation function and partial auto correlation function

In time series analysis, a process of examining the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) is to determine the nature of the process under consideration. Graphs of ACF and PACF are obtained to see the stationary condition as well as to guess the AR terms and MA terms which will involve in the autoregressive and moving average processes.

3.1.7.1 Autocorrelation function

Autocorrelation function (ACF) at lag k is defined by

$$\rho_k = \frac{\operatorname{cov}\left[\left(Y_t - \hat{Y}_t\right)\left(Y_{t+k} - \hat{Y}_{t+k}\right)\right]}{\sqrt{\operatorname{var}\left(Y_t - \hat{Y}_t\right)\operatorname{var}\left(Y_{t+k} - \hat{Y}_{t+k}\right)}}$$

The first several autocorrelations are persistently large in the graph of ACF and trailed off to zero rather slowly, it can be assumed that a trend exists and the time series is non-stationary. If the series is stationary, then ACF graph must decay exponentially. ACF is an excellent tool in identifying the order of a MA(q) process.

3.1.7.2 Partial autocorrelation function

Partial autocorrelation function between Y_t and Y_{t+k} is the conditional correlation between Y_t and Y_{t+k} and defined as follows:

$$\phi_{kk} = corr\left(Y_t,\ Y_{t+k} \mid Y_{t+1}, Y_{t+2},..., Y_{t+k-1}\right)$$

The PACF between Y_t and Y_{t+k} is the autocorrelation between Y_t and Y_{t+k} after adjusting for Y_{t-1}, Y_{t-2},..., Y_{t-k+1}. Hence for an $\text{AR}(p)$ model the PACF between Y_t and Y_{t+k} for $k > p$ should be equal to zero. PACF is an excellent tool in identifying the order of a $\text{AR}(p)$ process.

In this study, *MINITAB* is used to get the above results and relevant graphs for preliminary analysis.

3.2 Model Development and Forecasting using exponential smoothing method

Exponential smoothing is a method of forecasting that induces historical patterns such as trends and seasonal patterns into the future. It is a procedure for continually revising a forecast in the light of more recent experience. It assigns exponentially decreasing weights as the observations get older. In other words, recent observations are given relatively more weights in forecasting than the older observations. An exponentially weighted average refers to a weighted average of the data in which the weights decay exponentially.

In more general terms, the smoothing equation can be written as:

next period forecast = weight(present period observation) + (1-weight)(present period forecast), where 0 < weight < 1.

3.2.1 Single Exponential Smoothing

Single Exponential Smoothing (SES) is used for short-term forecasting, usually just one period into the future. The model assumes that the data fluctuates around a reasonably stable mean (no trend or consistent pattern of growth). The SES method produces forecasts that are a level line for any period in the future, but it is not appropriate for projecting trending data or patterns that are more complex.

3.2.2 Double Exponential Smoothing

Double Exponential Smoothing (DES) method is used when the data shows a trend. This method with a trend works much like single smoothing except that two components (level and trend) must be updated each period. The level is a smoothed estimate of the value of the data at the end of each period. The trend is a smoothed estimate of average growth at the end of each period. DES model containing linear trend is represented as: $Y_t = a_t + b_t(t) + \varepsilon_t$ where a_t and b_t are the intercept and slope parameters respectively, Y_t is the response variable at time t and ε_t is the error term at time t.

3.2.3 Holt-Winter's Seasonal Exponential Smoothing

Holt-Winter's (HW) Seasonal Exponential Smoothing method is used when the data shows trend and seasonality. To handle seasonality, a third parameter has to be added. By introducing a third equation to take care of seasonality, the resulting set of equations is called the "Holt-Winters" (HW) method. There are two main HW models, depending on the type of seasonality, which are: Multiplicative Seasonal Model and Additive Seasonal Model.

3.2.3.1 Multiplicative Holt-Winter's seasonal model

The multiplicative HW model containing linear trend is represented as $Y_t = \left[a_t + b_t(t)\right] S_t \varepsilon_t$, where a_t, b_t and S_t are the level, slope and seasonal component parameters respectively, Y_t is the response variable at time t, ε_t is the error term at time t.

3.2.3.2 Estimates of model parameters of multiplicative Holt-Winter's model

The HW method can be used to update the parameters found in the multiplicative decomposition process. Using optimal values for level, slope and seasonal component, parameters can be updated as follows.

3.2.3.3 Updating level parameter of multiplicative Holt-Winter's model

$a_t = \alpha\left[\frac{Y_t}{S_t(t-L)}\right] + (1-\alpha)\left[a_{t-1} + b_{t-1}\right]$, where $0 < \alpha \leq 1$ is the first smoothing constant.

3.2.3.4 Updating trend parameter of multiplicative Holt-Winter's model

$b_t = \beta\left[a_t - a_{t-1}\right] + (1-\beta)b_{t-1}$, where $0 < \beta \leq 1$ is the second smoothing constant.

3.2.3.5 Updating seasonal parameter of multiplicative Holt-Winter's model

$S_{t+L} = \gamma\left[\frac{Y_t}{a_t}\right] + (1-\gamma)\left[S_t(t-L)\right]$, where $0 < \gamma \leq 1$ is the third smoothing constant.

Note that the best estimate of the seasonal factor for this time period in the season is used, which is last updated L periods ago.

3.2.3.6 Forecast for the next period of multiplicative Holt-Winter's model

forecast=[(level estimates) +(slope estimate)] (seasonal estimates)

The forecast for the next period is given by: $\hat{Y}_t = \left[a_t + b_t(T)\right]S_t$

3.2.3.7 Additive Holt-Winter's seasonal model

The additive HW model containing linear trend is represented as $Y_t = a_t + b_t(t) + S_t + \varepsilon_t$ where a_t, b_t and S_t are the level, slope and seasonal component parameters respectively, Y_t is the response variable at time t and ε_t is the error term at time t.

3.2.3.8 Estimates of model parameters of additive Holt-Winter's model

The HW method can be used to update the parameters found in the additive decomposition process. Using optimal values for level, slope and seasonal component, parameters can be updated as follows.

3.2.3.9 Updating level parameter of additive Holt-Winter's model

$a_t = \alpha\left[Y_t - S_t(t-L)\right] + (1-\alpha)\left[a_{t-1} + b_{t-1}\right]$, where $0 < \alpha \leq 1$ is the first smoothing constant.

3.2.3.10 Updating trend parameter of additive Holt-Winter's model

$b_t = \beta\left[a_t - a_{t-1}\right] + (1-\beta)b_{t-1}$, where $0 < \beta \leq 1$ is the second smoothing constant.

3.2.3.11 Updating seasonal parameter of additive Holt-Winter's model

$S_{t+L} = \gamma(Y_t - a_t) + (1-\gamma)\left[S_t(t-L)\right]$, where $0 < \gamma \leq 1$ is the third smoothing constant.

Note that the best estimate of the seasonal factor for this time period in the season is used, which is last updated L periods ago.

3.2.3.12 Forecast for the next period of Additive HW model

forecast= (level estimates)+ (slope estimate)+(seasonal estimates)

The forecast for the next period is given by: $\hat{Y}_{t+1} = \left[a_t + b_t(T)\right] + S_{t+1}(t+1-L)$

3.2.4 Obtaining the optimal values for smoothing constants

One of the most important parts of any exponential smoothing model is estimating smoothing constant. The common approach is to work with several values of smoothing constants and select the best combination which produces the minimum for the evaluation criteria used. The procedure is time consuming and thus in this study the grid search procedure for the determination of smoothing constants is carried out using the software *STATISTICA*.

3.2.4.1 Parameter Grid Search algorithm

It is recommended by Gardner (1985) that, in practice, to compute forecasts the optimum smoothing parameters must be used. One common method is to perform a grid search of the parameter space. Thus, *STATISTICA* provides each parameter from the minimum (from zero) to maximum (to one) by incrementing step by step. For each combination of parameter values, *STATISTICA* computes the Sums of Squares for the Errors (SSE). It displays ten combination of smoothing constants based on the SSE from minimum.

3.2.4.2 Parameter Auto Search algorithm

Auto search algorithm in *STATISTICA* automatically computes the parameters based on the data entered into the system.

3.2.5 Advantages of Exponential Smoothing method

- It is relatively good short- term accuracy, simplicity, and low cost
- The process is easily implemented on the computer
- No need to have large amount of historical data
- New forecasts are easy to obtain and updating only depends on last data point
- Provides a possible framework for forecasting disaggregate demand patterns
- For short-term planning and control systems, these techniques are extremely valuable and have a more than adequate track record in forecast accuracy
- Each technique employs weights that give more emphasis to data from the recent past than to the older data. As a result of weighting of the past data, a fitted model appears smoother (less volatile) than the original data

3.2.6 Disadvantages of Exponential Smoothing method

- The start-up time required for finding the 'best' parameter estimates along with the need of continuously monitoring of the process and updating the values of parameters
- Does not account for any of the other variables that might influence the forecast
- Exponential smoothing is only valid if it is assumed that the error terms are random
- For demand forecasting, exponential smoothing models are univariate (that is, no explanatory drivers) and are not models designed for use with causal variables, and cannot handle economic business cycles

3.3 Model development and forecasting using ARIMA modelling method

Generally, building non-seasonal or seasonal ARIMA has 3 stages, which are: Model Identification, Parameter Estimation and Residual Analysis.

3.3.1 Model Identification

Tentative model is identified through analysis of historical data. It is useful to look at a plot of the series along with the sample ACF and PACF. According to these plots the tentative model can be identified.

3.3.1.1 Autoregressive Moving Averages models

A model with combinations of autoregressive terms and moving average terms are generally called as Auto Regressive Moving Averages (ARMA) model. The ACF and PACF of an ARMA process are determined by the AR and MA components, respectively. It can therefore be shown that the ACF and PACF of a ARMA both exhibit exponential decay, which makes the identification of the order of the ARMA(p, q) model relatively. A formulation of an ARMA process is given as:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

ARMA(p, q) models can describe a wide variety of behaviors for stationary time series. In practice, the values of p and q each rarely exceed 2.

Table 3.1: Behavior of theoretical ACF and PACF for stationary series

Process	ACF	PACF
AR(p)	Exponential decay/ Die out	Cuts off after lag p
MA(q)	Cuts off after lag q	Exponential decay/ Die out
ARMA(p, q)	Exponential decay/ Die out	Exponential decay/ Die out

Adapted: Douglas *et al.* (2015)

In Table 3.1, "Die out" means "tend to zero gradually" and "Cut off" means "disappear" or "is zero".

3.3.1.2 Autoregressive Integrated Moving Averages models

Autoregressive Integrated Moving Averages (ARIMA) model is a non- stationary model. The first step in model identification is to determine whether the series is stationary. If the series is not stationary, it can often be converted to a stationary series by differencing: the original series is replaced by a series of differences and an ARMA model is then specified for the differenced series. It is generally denoted as ARIMA(p, d, q), where d indicates the amount of differencing. Once a stationary series has been obtained, it is necessary to identify the form of the model to be used through the sample ACF and PACF for the various ARIMA models.

3.3.1.3 Seasonal ARIMA models

In some cases, the series shows a repeating, cyclic behaviour. This pattern or as more commonly called *seasonal pattern* can be very effectively used to further improve the forecasting performance. The most important structural issue to recognize about seasonal time series is that if the season is s periods long, then observations those are s time intervals apart are alike. In addition, for non-stationary seasonal series, an additional seasonal difference is often required.

A Seasonal ARIMA (SARIMA) models or $\text{ARIMA}(p,d,q)(P,D,Q)_s$ usually contains: regular $\text{AR}(p)$ and $\text{MA}(q)$ terms that account for the correlation at low lags Seasonal $\text{AR}(P)$ and Seasonal $\text{MA}(Q)$ terms that account for the correlation at the seasonal lags where d and D indicate the amount of regular and seasonal differencing respectively, S is the seasonality.

3.3.2 Parameter Estimation

The second stage of building non- seasonal ARIMA or seasonal ARIMA is estimating parameters of those models. Once tentative models have been selected, the unknown parameters of those models are estimated using least squares estimates.

3.3.3 Limitation of ARIMA models (by Robert and Monnie (1999))

- If there are not enough data, they may be no better at forecasting than the decomposition or exponential smoothing techniques. Most authors recommend that at least 30 to 50 observations are needed ARIMA modeling
- ARIMA models usually are based on stochastic rather than deterministic or axiomatic processes
- These models are better at formulating incremental rather than structural change

3.3.4 Residual Analysis

The third and the last stage of non- seasonal ARIMA or seasonal ARIMA model building is residual analysis. Before using the model for forecasting, it must be checked for adequacy. Diagnostic checks are performed to determine the adequacy of the model. Basically, a model is adequate if the residuals cannot be used to improve the forecasts. Thus, the residuals should be random and normally distributed with constant variance.

The following tests are carried out for the residual analysis:

3.3.4.1 Normality of Residuals

An assumption of ordinary least squares regression analysis is that the errors of a model are normally distributed. The values of Skewness and Kurtosis are considered to check the normality of the residuals. The skewness closer to 0 and the kurtosis closer to 3 suggests the residuals follow a normal distribution. Histogram is employed for this purpose.

3.3.4.2 Anderson- Darling Test

The Anderson- Darling (AD) test is used to test if a sample of data comes from a population with a specific distribution. It is a modification of Kolmogorov- Smirnov (K-S) test and gives more weight to the tails than does the K-S test. Here the hypotheses are H_0: The data follow normal distribution versus H_1: The data do not follow normal distribution.

The test statistic of AD test is:

$A^2 = -N - \sum_{i=1}^{N} \frac{(2i-1)}{N}\left[\ln F(Y_i) + \ln(1 - F(Y_{N+1-i}))\right]$, where F is the cumulative distribution function of the specified distribution, Y_i are the ordered data and N is the total number of observations.

3.3.4.3 Durbin-Watson statistic

The most important test for detecting serial correlation is that developed by statisticians Durbin and Watson. That means, DW statistic is used to test for randomness of residuals. The test statistic is defined as:

$$d = \frac{\sum_{t=2}^{n}\left(\hat{u}_t - \hat{u}_{t-1}\right)^2}{\sum_{t=1}^{n}\hat{u}_t^2}, \qquad \text{where } u_t \text{ is white noise of a fitted model.}$$

A great advantage of DW statistic is that, it is based on the estimated residuals. The DW closer to 2 reveals that the residuals are randomly scatted.

3.3.4.4 Lagrange's Multiplier test

Lagrange's Multiplier (LM) test is used to test the independency of residuals. It is an alternative test of Durbin Watson test for auto correlation among residuals. The null hypothesis to be tested is that, H_0: there is no serial correlation of any order. The individual residual autocorrelations should be small. Significant residual autocorrelations at low lags or seasonal lags suggest that the model is inadequate.

$$W = nR^2 \sim \chi^2_{df}$$

Where, df is the number of regressors in the auxiliary regression (only linear terms of the dependent variable are in the auxiliary regression), R^2 is the determination of coefficients and n is the number of observations.

3.3.4.5 White's General test

White's General test is used in order to check constant variance of residuals. Accordingly the null hypothesis is H_0: Homoscedasticity against the alternative hypothesis H_1: Heteroscedasticity.

Test statistic of White's General test is:

$W = nR^2 \sim \chi^2_{df}$ Where, df is the number of regressors in the auxiliary regression (squared terms of the dependent variable are also included in addition to terms in the LM test in auxiliary regression), R^2 is the determination of coefficients and n is the number of observations.

3.3.5 Model Selection

After an adequate model has been found, forecasts can be made. Prediction intervals based on the forecasts can also be constructed. In time series analysis, sometimes more than one model can fit the data equally well. Under those circumstances, system knowledge can help to choose the more relevant model. For this purpose, numerical criterion such as the Akaike information criterion (AIC) and the Schwartz's Bayesian criterion (SBC) are used to select the best model. The best model is the one which gives the lowest AIC and SBC values. The coefficient of determination (R^2) is also taken into account for the section of best model. The best model gives the largest R^2 value.

3.3.5.1 Akaike Information Criterion

Akaike's information criterion (AIC) is often used for model selection, especially in the time series. This criterion includes a penalty for over parameterization similar to the adjusted R^2 in regression analysis, but is more generally applicable. This criterion is equal to minus 2 times the log likelihood function plus 2 times the number of free parameters in the model. That is for sample size n, the expression of AIC is given below:

$\text{AIC}(k) = n\ln\left(\hat{\sigma}^2\right) + 2k$, where k is the number of parameters in the model and $\hat{\sigma}^2$ is the sample variance of the residuals.

3.3.5.2 Schwartz's Bayesian Criterion

Schwartz's Bayesian Criterion (SBC) is another mostly used for model selection in time series analysis. SBC is a measure of goodness of fit. This measure is often used for order selection of models. This criterion is equal to the number of free parameters times the natural log of the number of residuals minus 2 times the natural log of the likelihood function. For sample size n, the expression of SBC is given as:

$\mathrm{SBC}(k) = n\ln\left(\hat{\sigma}^2\right) + k\ln(n)$, where k is the number of parameters in the model and $\hat{\sigma}^2$ is the sample variance of the residuals.

3.3.5.3 Coefficient of determination

The coefficient of determination, R^2, is the proportion of variance of a dependent variable explained by the model. R^2 is used as a measure of fit and it is used to evaluate the goodness of fitted model.

In this study, *Eviews* is used to get all the output of ARIMA model development and forecasting.

3.3.6 Accuracy of the model

It is important to evaluate performance of fitted model on the basis of the fit of the forecasting. This goodness of fit approach often uses the residuals and does not really reflect the capability of the forecasting technique to successfully predict future observations. The user generally of the forecasts is very concerned about the accuracy of future forecasts, not model goodness of fit, so it is important to evaluate this aspect of any recommended technique. Measure of forecast accuracy should always be evaluated as part of a model validation effort.

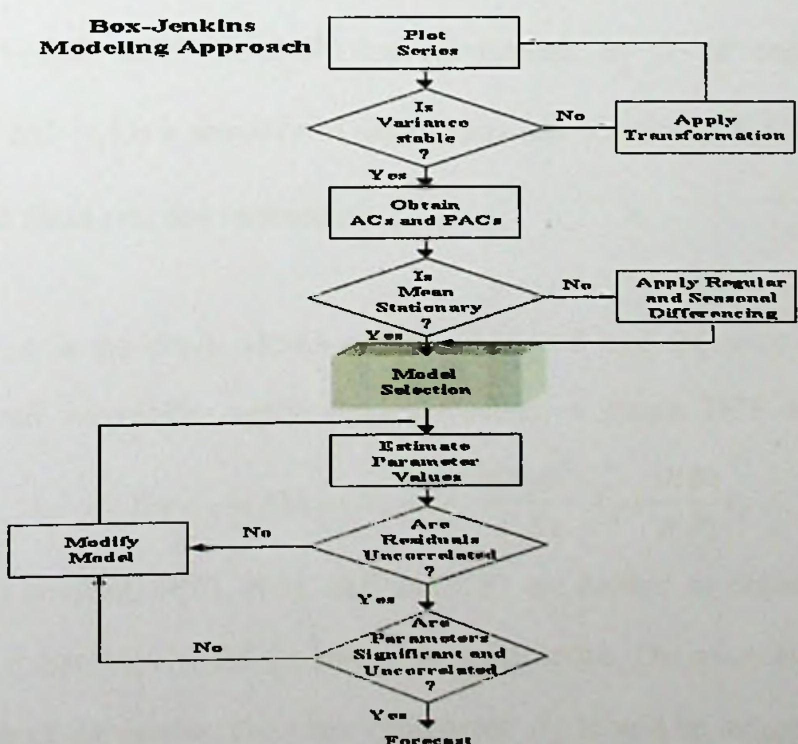
3.3.6.1 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) is the average of the sum of the absolute values of the percentage errors. It is generally used for evaluation of the forecast against the validation sample. To compare the average forecast accuracy of different models, MAPE (which is less sensitive to outlier distortion than the other techniques) statistics is used. It is defined as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100$$

Generally if MAPE is less than 10% then the fitted model is acceptable. However less than 15% also can be considered and this decision is purely subjective.

3.3.7 Flow chart of ARIMA modelling Approach



Adapted: Angel & Carles (2007)

3.4 Model development and forecasting using Dynamic Transfer Function method

Dynamic transfer function (DTF) model is a statistical model describing the relationship between an output variable Y_t and one or more input variables X_t. It has many applications in business and economics, especially in forecasting turning points. These models can have pulse, step, or continuous inputs. They have decay rates that are ordered according to the number of decay rate parameters in the model.

3.4.1 Dynamic Transfer Function – Noise Model

In practice, the output Y_t is not a deterministic function of X_t. It is often disturbed by some noise or has its own dynamic structure. It is denoted that the noise component as N_t. The noise may be serially correlated, and it is assumed that N_t follows an ARMA(p, q) model as $\phi(B)N_t = \theta(B)e_t$, where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$ are polynomials in B of degree p and q respectively, and $\{e_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance σ_e^2.

It is noted that in the above ARMA model, $E(N_t) = 0$ and the usual conditions of stationarity and invertibility apply. Putting together, a simple DTF model can be obtained as $$Y_t = c + v(B)X_t + N_t = c + \frac{\omega(B)B^b}{\delta(B)} X_t + \frac{\theta(B)}{\phi(B)} e_t ,$$

where c is a constant, $\theta(B)$, $\phi(B)$, $\omega(B)$ and $\delta(B)$ are defined as before with degree q, p, s, and r respectively, and $\{e_t\}$ are white noise series. The parameter b is called the decay rate of the system. The noise component N_t should be independent of X_t; otherwise, the model is not identifiable.

Further it is noted that when $b > 0$ the DTF model is useful in predicting the turning points of Y_t given those of X_t.

When there are two input variables X_{1t} and X_{2t}, the DTF model becomes

$Y_t = c + \frac{\omega_1(B)B^{b1}}{\delta_1(B)} X_{1t} + \frac{\omega_2(B)B^{b2}}{\delta_2(B)} X_{2t} + \frac{\theta(B)}{\phi(B)} e_t$, where ω_i and δ_i are similarly defined as above.

Similarly it can be extended for a finite number of input variables as well.

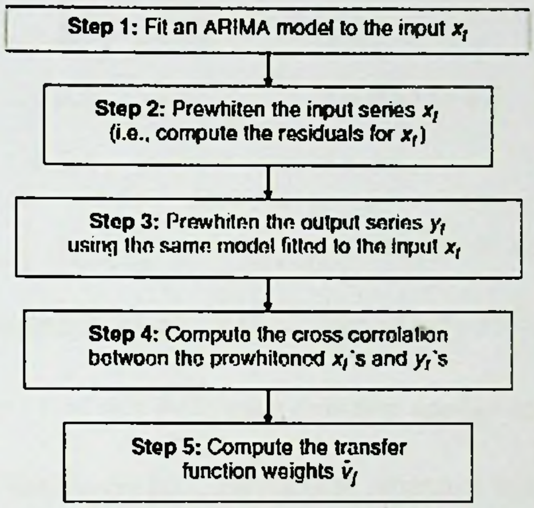
3.4.2 Dynamic Transfer Function model for Univariate Time Series Process

A general form of a DTF model can be expressed as

$Y_t = c + \sum_{i=1}^{m} \frac{\omega_i(B)B^{bi}}{\delta_i(B)} X_{it} + \frac{\theta(B)}{\phi(B)} e_t$ where $\omega_i(B) = \omega_0 + \omega_1 B + \omega_2 B^2 + ... + \omega_i B^i$, $\delta_i(B) = \delta_0 + \delta_1 B + \delta_2 B^2 + ... + \delta_i B^i$, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$ are polynomials in B of degree q, p, s, and r respectively , and $\{e_t\}$ are white noise series. The parameter bi is called the decay rate with the i[th] variable. The order of the DTF is said to be (r, s, b) and the added noise model is of order (p, q).

Since the DTF model is a straight forward extension of the ARMA model, for $\omega_i(B) = 0$, the model is equivalent to ***univariate time series process*** (Dominique *et al.* (2002)). Thus the DTF for univariate time series can be simply written as $Y_t = \frac{\theta(B)}{\phi(B)} e_t$.

3.4.3 Flow chart of computing dynamic transfer function parameters



Adapted from: Søren and Murat (2011)

3.4.4 Estimating Parameters of DTF model

The parameters of DTF are estimated as follows:

Since DTF is expressed as $Y_t = c + v(B)X_t + N_t = c + \frac{\omega(B)B^b}{\delta(B)}X_t + \frac{\theta(B)}{\varphi(B)}e_t$

The coefficients of the polynomials $\theta(B)$ and $\phi(B)$ are generally obtained by fitting ARMA model to the input N_t. On the other hand, dependent variable Y_t can be written as: $Y_t = v(B)X_t = \frac{\omega(B)B^b}{\delta(B)}X_t$. Hence $v(B) = \frac{\omega(B)B^b}{\delta(B)}$.

The infinite number of parameters in $v(B)$ can be computed from a finite number of parameters in $\omega(B)$ and $\delta(B)$ as follows (From George *et al.* (2008)):

$$\delta(B)v(B) = \omega(B)B^b$$

$$(1-\delta_1 B-\delta_2 B^2-...-\delta_r B^r)(v_0+v_1 B+v_2 B+...)=(\omega_0-\omega_1 B-\omega_2 B^2-...-\omega_S B^S)$$

This means by equating coefficients of B, it can be written as

$$v_j = \begin{cases} 0 & j < b \\ \delta_1 v_{j-1} + \delta_2 v_{j-2} + \ldots + \delta_r v_{j-r} + \omega_0 & j = b \\ \delta_1 v_{j-1} + \delta_2 v_{j-2} + \ldots + \delta_r v_{j-r} - \omega_{j-b} & j = b+1, b+2, \ldots, b+s \\ \delta_1 v_{j-1} + \delta_2 v_{j-2} + \ldots + \delta_r v_{j-r} & j > b+s \end{cases}$$

The weights $v_{b+s}, v_{b+s-1}, v_{b+s-2}, \ldots, v_{b+s-r+1}$ supply r starting values for the homogeneous difference equation $\delta(B)v_j = 0$ for $j > b+s$.

The solution $v_j = f(\delta, \omega, j)$ of this difference equation applies to all the values v_j for which $j \geq b+s-r+1$. Thus, in general, the impulse response weights v_j consist of b zero values $v_0, v_1, v_2, \ldots, v_{b-1}$.

A further $s-r+1$ values $v_b, v_{b+1}, v_{b+2}, \ldots, v_{b+s-1}$ following no fixed pattern (no such values occur if $s < r$). Values v_j with $j \geq b+s-r+1$ following the pattern dictated by the rth order difference equation, which has r starting values $v_{b+s}, v_{b+s-1}, v_{b+s-2}, \ldots, v_{b+s-r+1}$. Starting values v_j for $j < b$ will, of course, be zero.

In this study, *SAS* programme is used to get the output for the estimation of the parameters of DTF models.

3.5 Model development and forecasting using State Space modeling method

State Space (SS) models are models of jointly stationary multivariate time series processes that have dynamic interactions and that are formed from two basic equations. The state transition equation consists of a state vector of auxiliary variables as a function of a transition matrix and an input matrix, whereas the measurement equation consists of a state vector canonically extracted from observable variables. These vector models are estimated with a recursive protocol and can be used for multivariate forecasting.

In many applications, the driving forces behind the evolution of economic variables are not observable or measurable. When explanatory variables are not observable, standard vector autoregressive (VAR) models cannot be applied to study the evolution of the endogenous variables. However, it is easy to extend the VAR framework to analyze scenarios with unobservable explanatory variables by using state space models.

State space models allow to model an observed (multiple) time series, Y_t, as being explained by a vector of (possibly unobserved) state variables, S_t, which are driven by a stochastic process. A basic linear state space model takes the following form (Robert (2007)):

$$Y_t = HS_t + v_t \ , \ v_t \sim N(0, \Sigma_V) \qquad \textit{measurement equation}$$

$$S_{t+1} = BS_t + Gw_t \ , \ w_t \sim N(0, \Sigma_W) \qquad \textit{transition equation}$$

where H , B and G are system matrices (known matrices) with order 2×2, v_t and w_t are uncorrelated sequences of white noises.

The first equation (measurement equation) describes the relation between the observed time series, Y_t, and the (possibly unobserved) state S_t. In general, it is assumed that the series Y_t are measured with error, which is reflected in the measurement error v_t. The standard approach is to model v_t as a Gaussian error term, $v_t \sim N(0, \Sigma_V)$.

The second equation (transition equation) describes the evolution of the state variables as being driven by the stochastic process of innovations w_t. Typically one assumes normal innovations, such that $w_t \sim N(0, \Sigma_W)$.

It is noted that, an ARMA model can expressed as a state space form in "infinite" many ways (Akaike's approach, Aoki's method etc...). Conversely, for a given state space model, there is an ARMA representation. Moreover, a seasonal ARIMA model also can be put into state space form (Durbin & Koopman ,(2012)).

3.5.1 Estimation of State Space model

In practical applications, the system matrices H and B together with the variances Σ_V and Σ_W are unknown and have to be estimated. Obviously, whenever the explanatory variables are not observable, least squares estimation is not a way to go. The Kalman filter allows constructing the likelihood function associated with a state space model.

3.5.2 **Kalman Filtering** (by Tsay (2005))

Kalman filter is a set of recursive equation that allows us to update the information in a state space model. It basically decomposes an observation into conditional mean and predictive residual sequentially.

Kalman filter can be classified as follows:

- *Initialization-* Make inference on S_t and B
- *Prediction-* Draw inference about S_{t+h} with $h > 0$, given B
- *Smoothing-* Make inference about S_t given the data

3.5.3 Prediction and updating steps in Kalman filter method

From the state space model, it can be written as $S_{t+1|t} = BS_{t|t}$ and $Y_{t+1|t} = HS_{t+1|t}$

1. $P_{t+1|t} = BP_{t|t}B' + GQG'$

2. $V_{t+1|t} = HP_{t+1|t}H' + R$

3. $C_{t+1|t} = HP_{t+1|t}$

where $P_{t+j|t}$ be the conditional covariance matrix of $S_{t+j|t}$, $Q = \text{covariance}(w_t)$, $V_{t+1|t}$ is the conditional variance of Y_{t+1} , $R = \text{covariance}(v_t)$, $C_{t+1|t}$ denotes the conditional covariance between Y_{t+1} and S_{t+1}.

The above three equations are needed for prediction.

For the updating, the following two equations are needed:

$$S_{t+1|t+1} = S_{t+1|t} + P_{t+1|t}H'\left[HP_{t+1|t}H' + R\right]^{-1}\left(Y_{t+1} - Y_{t+1|t}\right)$$

$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}H'\left[HP_{t+1|t}H' + R\right]^{-1} HP_{t+1|t}$$

3.5.4 Forecasting with State Space models

Having obtained the maximum likelihood estimates, it can be used the state space model to forecast the observables. In particular, the final state predictor implied by the maximum likelihood estimates together with the measurement and transition equation can be used to forecast the observations.

In this study, *SAS* programme is used to get the output for the estimation of the parameters of State Space models using Kalman filter.

3.5.5 Advantages of a State-Space model

- It easily handles structural breaks, shifts, time-varying parameters of some static model. It makes those parameters dynamic states of a state-space model and model will automatically adjust to any shifts in parameters
- It handles missing data very naturally
- It allows changing on-a-fly parameters of a state-space model itself (covariance of noises and transition/observation matrices)
- It allows using data from different sources simultaneously in the same model to estimate one underlying quantity
- It allows to constructing a model from several interpretable unobservable dynamic components and estimates them
- Any ARIMA model can be represented in a state-space form, but only simple state-space models can be represented exactly in ARIMA form.

3.6 Synopsis

In this Chapter, almost all the tests, relevant to preliminary analysis, prior to time series model development are explained. All the techniques for each and every modeling method are separately described with their fundamental statistical theories and statistical methodologies. Also this provides the advantages and disadvantages by employing the Holt-Winter's Seasonal Exponential Smoothing, Seasonal ARIMA modeling, Dynamic Transfer Function modeling and State Space modeling methods.

4. PRELIMINARY ANALYSIS

4.1 Preliminary Analysis of Tourist Arrivals in Overall Frame

In this section, the discussion is based on the overall data which cover the entire time frame of this study from January 1967 to December 2015. Thus, here it is named as '*overall frame*' data.

Table 4.1: Descriptive statistics of yearly tourist arrivals in overall frame

Year	Total	Mean	Year	Total	Mean
1967	23,666	1972	1992	383,669	31972
1968	28,272	2356	1993	392,250	32688
1969	40,204	3350	1994	407,511	33959
1970	46,247	3854	1995	403,101	33592
1971	39,654	3305	1996	302,265	25189
1972	56,047	4671	1997	366,165	30514
1973	77,888	6491	1998	381,063	31755
1974	85,011	7084	1999	436,440	36370
1975	103,204	8600	2000	400,414	33368
1976	118,971	9914	2001	336,794	28066
1977	153,665	12805	2002	393,171	32764
1978	192,592	16049	2003	500,642	41720
1979	250,164	20847	2004	566,202	47184
1980	316,780	26398	2005	549,308	45776
1981	370,742	30895	2006	559,603	46634
1982	407,230	33936	2007	494,008	41167
1983	337,530	28128	2008	438,475	36540
1984	317,734	26478	2009	447,890	37324
1985	257,456	21455	2010	654,476	54540
1986	230,106	19176	2011	855,975	71331
1987	182,620	15218	2012	1,005,605	83800
1988	182,662	15222	2013	1,274,593	106216
1989	184,732	15394	2014	1,527,153	127263
1990	297,888	24824	2015	1,798,380	149865
1991	317,703	26475	**Cum. Total**	**19,493,921**	

Table 4.1 summaries the total number of arrivals in each year with mean of yearly arrivals of tourists in the overall frame. Based on these statistics, it can be seen that, nearly 19.5 million international tourists had come to Sri Lanka till December 2015. The yearly arrivals, on the whole, increase with year though there are few flops in particular periods. However, at the later stage there is a dramatic increase in number.

When the average tourist arrivals per month in every year is concerned, approximately from 2,000 tourists in the year 1967 to 150,000 tourists in the year 2015 had visited monthly to Sri Lanka in the last 49 years. It is a vast increase in the tourism industry.

Figure 1 shows the yearly total tourist arrivals from the year 1967 to 2015 in Sri Lanka.

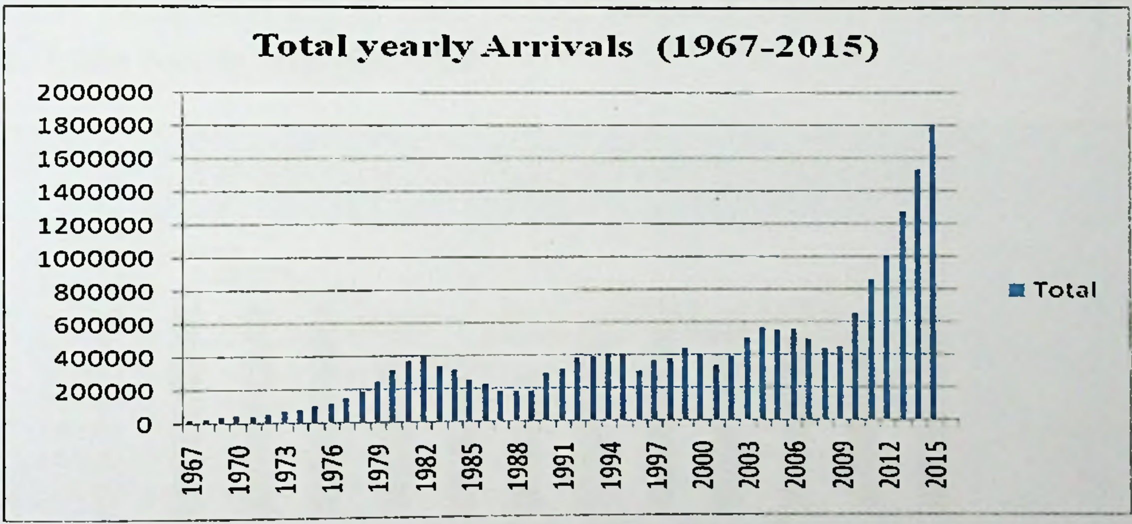


Figure 4.1: Plot of yearly tourist arrivals in overall frame

From Figure 4.1, it is clearly observed that, up to first 16 years (from 1967 to 1983) there is an upward trend, there after ups and downs in total arrivals till 43rd year (that is from 1983 to 2009). Meantime, from years 43 to 49 (from 2009 to 2015), there is a remarkable upward increase in the total number of international tourist arrivals to the island.

Therefore in this study, it was decided to consider mainly two windows such as Window I and Window II. Window I consist of the time period from 1967 January to 2009 May and consider as *before the internal conflict*. Window II consist of the period from 2009 June to 2015 December as *after the internal conflict*.

In addition to this classification of time frame, it is decided to further split the window I as Phase I, *before the begging of the conflict*, and Phase II as *during the conflict*. Accordingly, Phase I covers the period from 1967 January to 1983 July. Meanwhile Phase II covers the period from 1983 August to 2009 May. Hence it clear that, Phase I and Phase II are sub windows in Window I.

In the rest of the chapters in the dissertation, the discussions will be based on these four time frames such as Window I, Phase I, Phase II and Window II.

Figure 4.2 shows the plot of monthly average number of international tourists visited to Sri Lanka from the years 1967 to 2015.

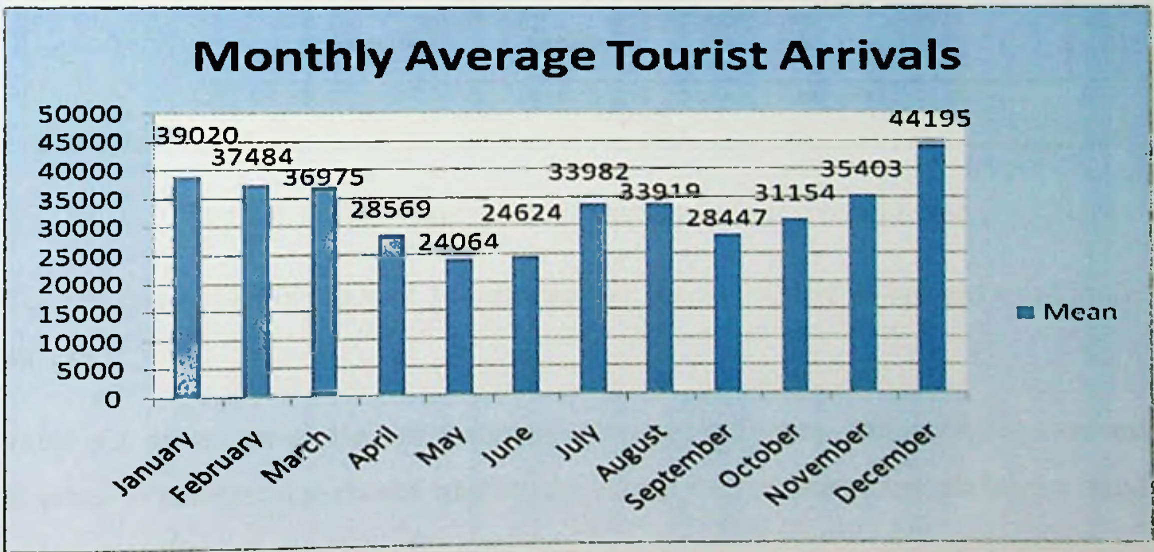


Figure 4.2: Plot of monthly average tourist arrivals in overall frame

It can be clearly observed from Figure 4.2 that, on the average January to May numbers of average arrivals decrease, and slight increase can be seen in June, July and August. Though there is a drop From August to September, the numbers of tourist arrivals increase thereafter. Ultimately the peak is reached in every December of the year. Thus the plot indicates that there are seasonal patterns in the tourist arrivals to Sri Lanka.

Based on the average monthly arrivals, it can be said that December as peak month while January, February and March as mini peak months. July, August and November are at moderated level in tourist arrivals. On the other hand, the lowest numbers of arrivals are recorded in the months of May and June in every year.

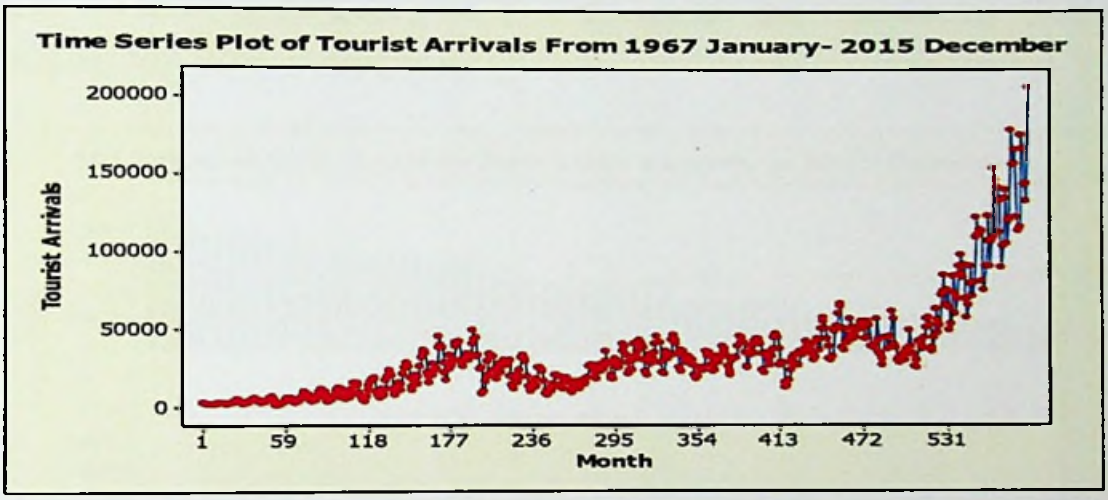


Figure 4.3: Time series plot of tourist arrivals in overall frame

The time series plot in Figure 4.3 illustrates that, there is clearly an upward trend in the arrivals.

Table 4.2 below shows the test results of Augmented Dickey- Fuller (ADF) test and Kruskal- Wallis test that checks whether the overall data of tourist arrivals have a trend and seasonality respectively.

Table 4.2: Test results of trend and seasonality of monthly tourist arrivals

Overall Frame	**ADF**	**Kruskal- Wallis**
Test Statistics	3.82	512.24
p- value	1.00	0.00

The p- value of Augmented Dickey- Fuller (ADF) test, in Table 4.2, very strongly support the null hypothesis. That means, it can be concluded with 95% confidence that there is a trend in the entire time frame data.

From Figure 4.3, it is hard to find a seasonal pattern visually. However, p-value of Kruskal- Wallis test in Table 4.2 confirms the existence of seasonality in the overall series with 95% confidence. Addition to this from Figure 4.4, ACF graph shows that there is a seasonal pattern (high spikes appearing in 12^{th}, 24^{th} and 36^{th} lags etc…). It also indicates the existence of the seasonality in the original series in the overall frame.

Figure 4.4 below is the graph of ACF of the monthly data from January 1967 to December 2015 in the overall frame.

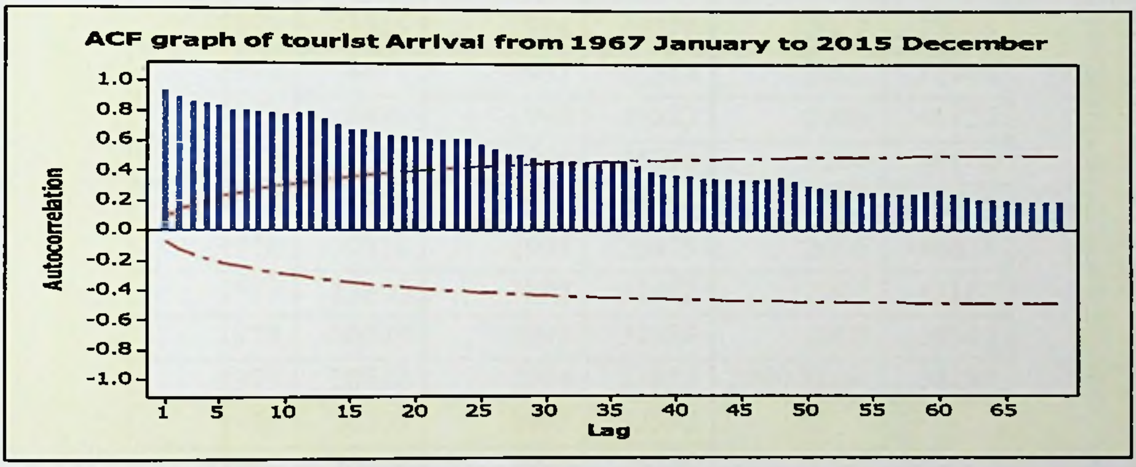


Figure 4.4: ACF graph of the series of tourist arrivals in overall frame

Further it is noted that from ACF graph in Figure 4.4, the series is non-stationary as ACF is not decaying exponentially with lags. To fit time series models, original series has to be converted to a stationary series.

4.2 Preliminary Analysis of Tourist Arrivals in Window I

As was defined Section 4.1, window I covers the time frame from January 1967 to May 2009. In this study, the window I is considered as the period before the internal conflict. In this time period, the number of international tourists who had come to Sri Lanka is 12,117,578 in total. It is about 62.16% of the total number of tourist arrivals (19,493,921 till 2015 December) in the entire period of overall frame. That is nearly in forty two and half years, 62.16% of total arrivals had happened while the balance 37.84% of arrivals just happened only in last six and half years.

Table 4.3 below provides the average number of international tourists who had visited to Sri Lanka from January 1967 to May 2009 in window I.

Table 4.3: Average tourist arrivals in window I

Year	Mean	Year	Mean	Year	Mean
1967	1972	1982	33936	1997	30514
1968	2356	1983	28128	1998	31755
1969	3350	1984	26478	1999	36370
1970	3854	1985	21455	2000	33368
1971	3305	1986	19176	2001	28066
1972	4671	1987	15218	2002	32764
1973	6491	1988	15222	2003	41720
1974	7084	1989	15394	2004	47184
1975	8600	1990	24824	2005	45776
1976	9914	1991	26475	2006	46634
1977	12805	1992	31972	2007	41167
1978	16049	1993	32688	2008	36540
1979	20847	1994	33959	2009 May	31288
1980	26398	1995	33592		
1981	30895	1996	25189		

As per the average arrivals in Table 4.3, it can be noted that there are ups and downs in arrivals in this period. Also the peak hits in the year 2004 while the lowest is at the beginning of the year 1967. It is also noted that from the year 2004 to 2009 May (end of the conflict) the average arrivals decrease slowly. The arrivals drop in 2005 might be due to the Tsunami in December 2004.

It is clear from Figure 4.5 that here also almost similar pattern exists with the overall time frame pattern. Thus in window I also, it can be said that December as peak month while January, February and March as mini peak months. July, August and November are at moderated level in tourist arrivals. On the other hand, the lowest numbers of arrivals are recorded in the months May and June on the average.

Figure 4.5 shows the plot of monthly average number of international tourists visited to Sri Lanka from January 1967 to May 2009 in Sri Lanka.

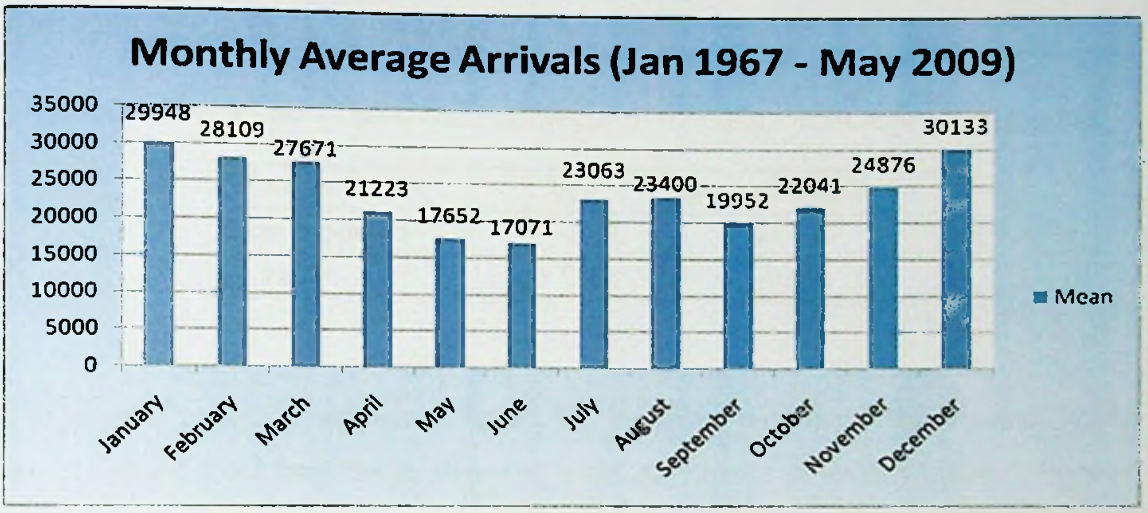


Figure 4.5: Plot of monthly average tourist arrivals in window I

From Figure 4.6, it is clearly observed that around first 200 months (from 1967 to middle of 1983) there is an upward trend and suddenly after a drop, high variations can be seen thereafter in monthly arrivals till the end of 510 months (May 2009). Moreover, from the p-value of the ADF test in Table 4.4, it can be concluded with 95% confidence that there is a trend in monthly arrivals in window I.

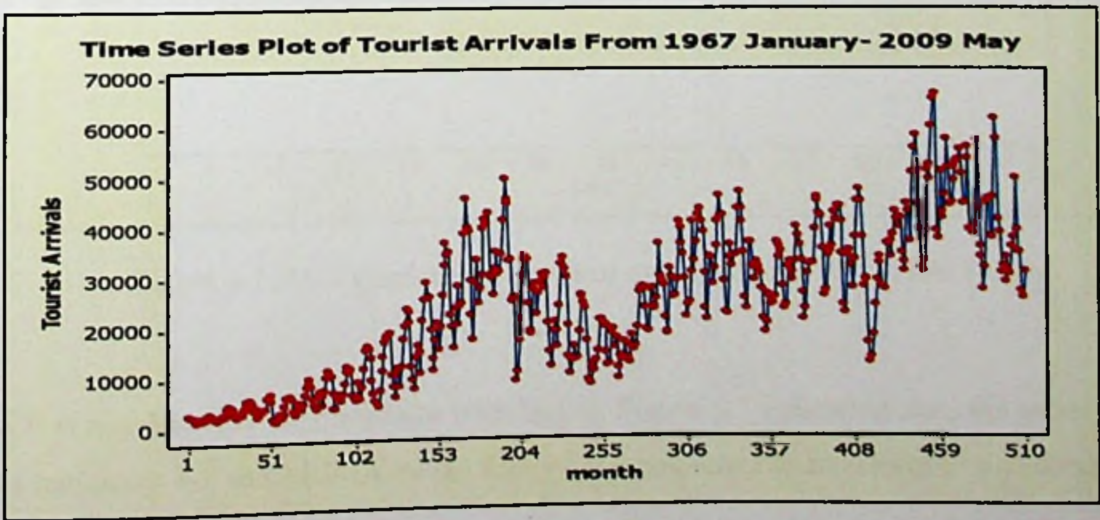


Figure 4.6: Time series plot of tourist arrivals in window I

Table 4.4 below shows that the test results of Augmented Dickey- Fuller (ADF) test and Kruskal - Wallis test to check whether the overall data of tourist arrivals have a tend and seasonality, respectively, in the widow I.

Table 4.4: Test results of trend and seasonality of monthly tourist arrivals in window I

Window I	ADF	Kruskal- Wallis
Test Statistics	-1.73	417.50
p- value	0.42	0.00

The seasonal pattern is not clear in Figure 4.6. However, p- value of the Kruskal- Wallis test in Table 4.4 confirms the existence of seasonality in the series of window I. Further, it can be clearly visualized in the graph of the ACF in Figure 4.7 as cyclic pattern appears in every 12^{th} lag.

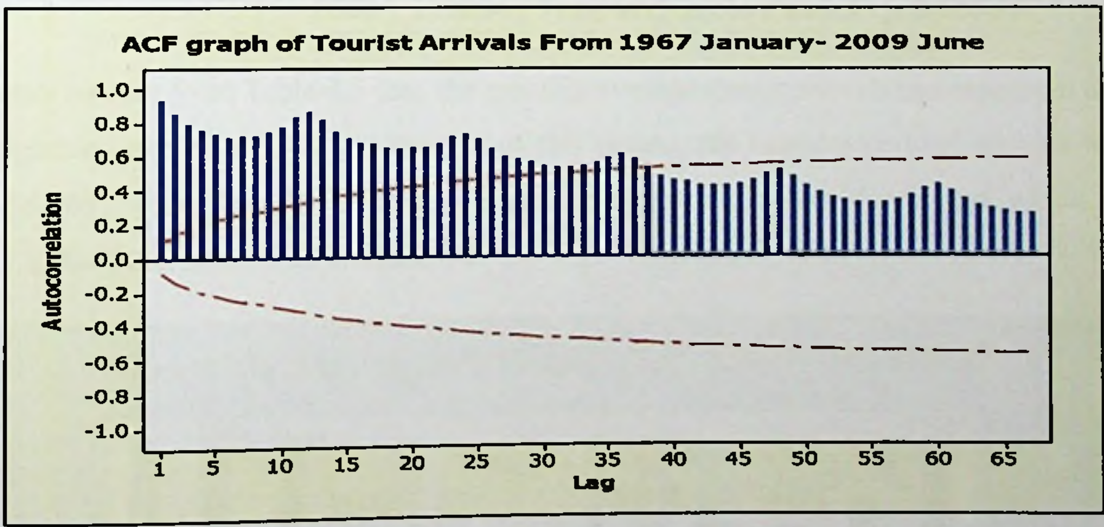


Figure 4.7: ACF graph of the series of tourist arrivals in window I

ACF is not decaying exponentially with lags in Figure 4.7 indicating that, the series is not stationary. For the ARIMA model fittings, it is necessary to transform to a stationary series by taking differences.

4.3 Preliminary Analysis of Tourist Arrivals in Phase I

Phase I covers the time frame as the beginning of internal conflict from January 1967 to July 1983. The Table 4.5 represents the average and total number of international tourist arrivals in phase I.

Table 4.5: Average and total tourist arrivals in phase I

Year	Mean	Total	Year	Mean	Total
1967	1972	23666	1976	9914	118971
1968	2356	28272	1977	12805	153665
1969	3350	40204	1978	16049	192592
1970	3854	46247	1979	20847	250164
1971	3305	39654	1980	26398	316780
1972	4671	56047	1981	30895	370742
1973	6491	77888	1982	33936	407230
1974	7084	85011	July 1983	36103	252720
1975	8600	103204	**Total arrivals in Phase 1**		**2,563,057**

It can be seen from Table 4.5 that, the monthly average tourist arrivals increase from the beginning except in 1971. At the end of this period, the cumulative total arrivals are 2,563,057 which is 13.15% total tourist arrivals in the overall frame and which is 21.15% of tourist arrivals in window I.

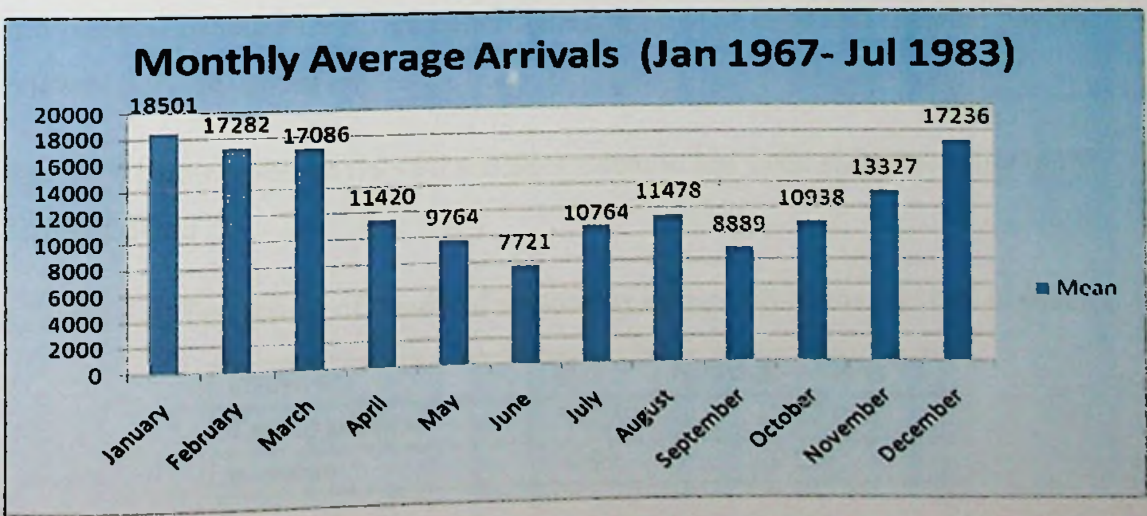


Figure 4.8: Plot of monthly average tourist arrivals in phase I

From Figure 4.8, it can be seen that the same that pattern of arrival as in the overall frame and window I. Here too, it is clear that, December as peak month while January, February and March as mini peak months. July, August and November are at moderated level in tourist arrivals. On the other hand, the lowest numbers of arrivals are recorded in the months of May and June on the average in phase I.

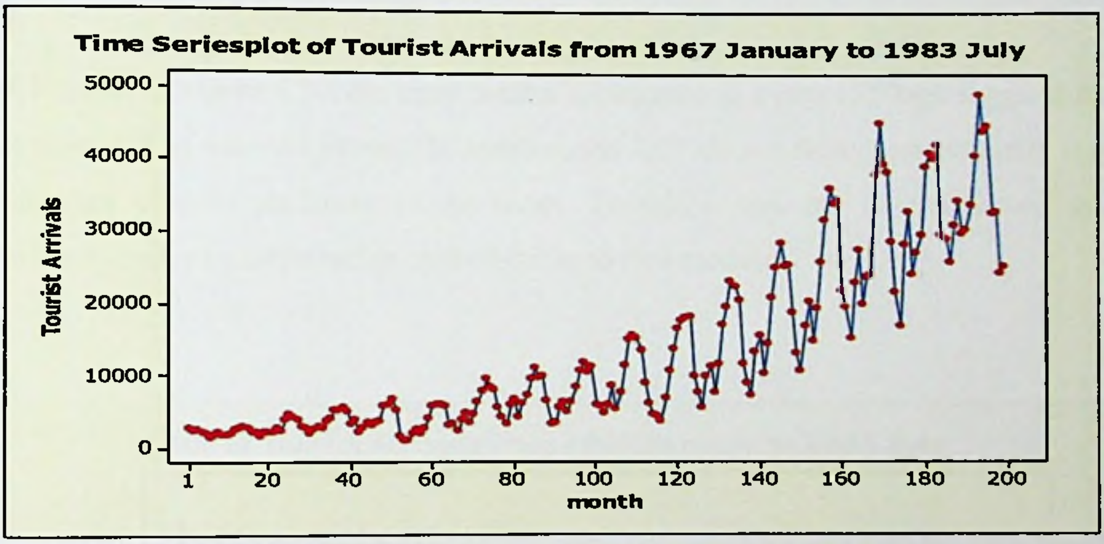


Figure 4.9: Time series plot of tourist arrivals in phase I

From the time series plot in Figure 4.9, it can be clearly seen a positive trend. Also there is a seasonal pattern as well. To confirm the existence of trend and seasonal pattern, the relevant tests are carried out.

Table 4.6 below gives the relevant statistics values to test trend and seasonality in the series of phase I.

Table 4.6: Test results of trend and seasonality of monthly tourist arrivals in phase I

Phase I	ADF	Kruskal- Wallis
Test Statistics	1.29	172.71
p- value	0.99	0.00

The p-value of ADF test in Table 4.6 is 0.99 and thus it can be concluded with 95% confidence that, there is a trend in the series. Meantime, the 0 p-value of Kruskal- Wallis test confirms the existence of the seasonal pattern in the series of phase I. Therefore, existence of trend and seasonality confirms that the series is not stationary. The ACF graph in Figure 4.10 further confirms the non-stationary of the series.

ACF graph in Figure 4.10, the same pattern appearance in every 12th lags suggests that the existence of seasonal pattern. In addition, the ACF do not decay exponentially is an indication of non- stationary of the series. Therefore, this can be transferred to a stationary series by differencing methods prior to fit a model.

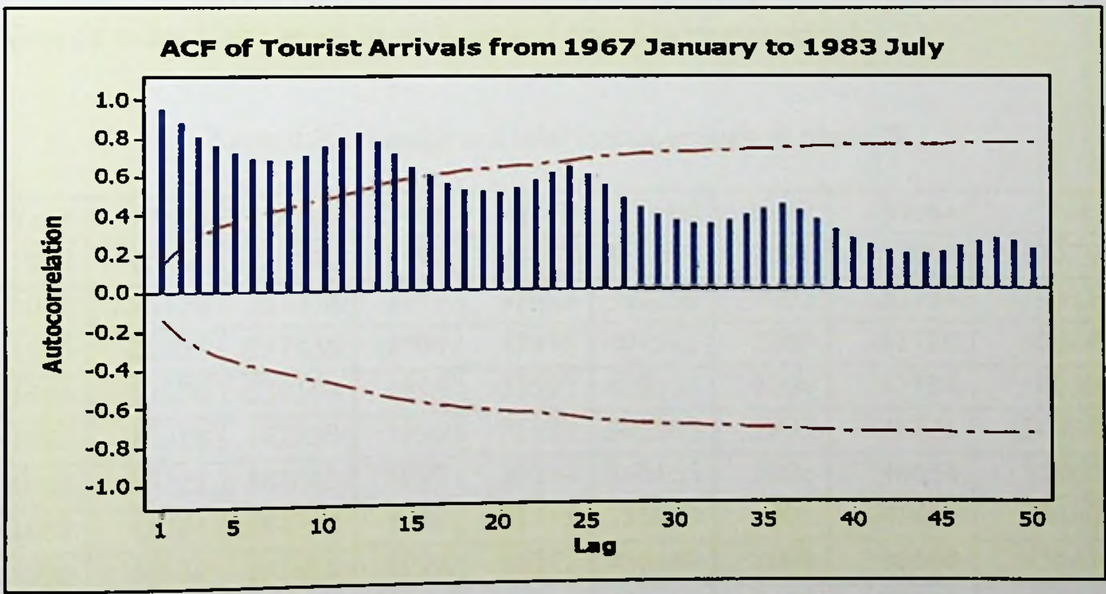


Figure 4.10: ACF graph of the series of tourist arrivals in phase I

4.4 Preliminary Analysis of Tourist Arrivals in Phase II

As it is defined in the section 4.1, the phase II covers the time period from August 1983 to May 2009. This is the period that the internal conflict took place in Sri Lanka. During this conflict period, 9,554,521 tourists had come to Sri Lanka which is 49.01% of total tourist arrivals in overall frame and which is 78.85% of tourist arrivals in window I. Hence approximately 49% of total tourist had visited Sri Lanka during the conflict period which took place for nearly 3 decades.

From Table 4.7, when average of arrivals in every month is concern, the lower values appear in the years 1987, 1988 and 1989. On the other hand, the higher average number of tourists had come to Sri Lanka in the years 2004, 2005 and 2006. After that till the end of the conflict, arrivals have decreased. But in between the lower values and higher values there are high variations with ups and downs in tourist arrivals.

Table 4.7: Average and total tourist arrivals in phase II

Year	Mean	Total	Year	Mean	Total	Year	Mean	Total
1983	16962	84810	1992	31972	383669	2001	28066	336794
1984	26478	317734	1993	32688	392250	2002	32764	393171
1985	21455	257456	1994	33959	407511	2003	41720	500642
1986	19176	230106	1995	33592	403101	2004	47184	566202
1987	15218	182620	1996	25189	302265	2005	45776	549308
1988	15222	182662	1997	30514	366165	2006	46634	559603
1989	15394	184732	1998	31755	381063	2007	41167	494008
1990	24824	297888	1999	36370	436440	2008	36540	438475
1991	26475	317703	2000	33368	400414	2009	31288	187729
				Total tourist arrivals in phase II				**9,554,521**

Without any doubt, it can be seen, from Figure 4.11 as well, the similar pattern exists with overall frame, window I and phase II. In other words, December usually is as peak month while January, February and March as mini peak months. July, August and

November are at moderated level in tourist arrivals. On the other hand, the lowest numbers of arrivals are recorded in the months May and June in every year.

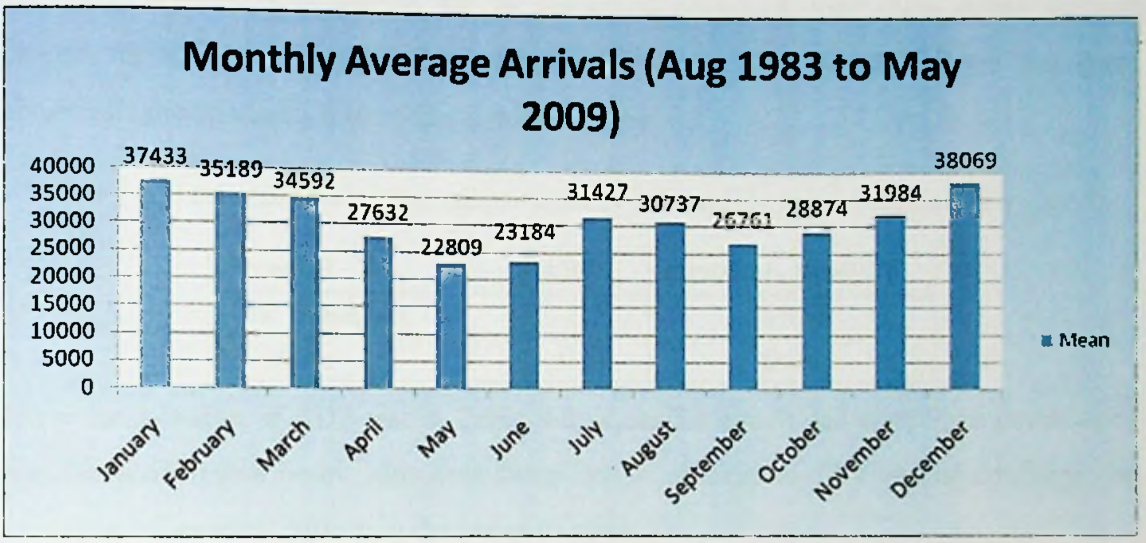


Figure 4.11: Plot of monthly average tourist arrivals in phase II

The figure below shows the plot of average monthly arrivals in phase II

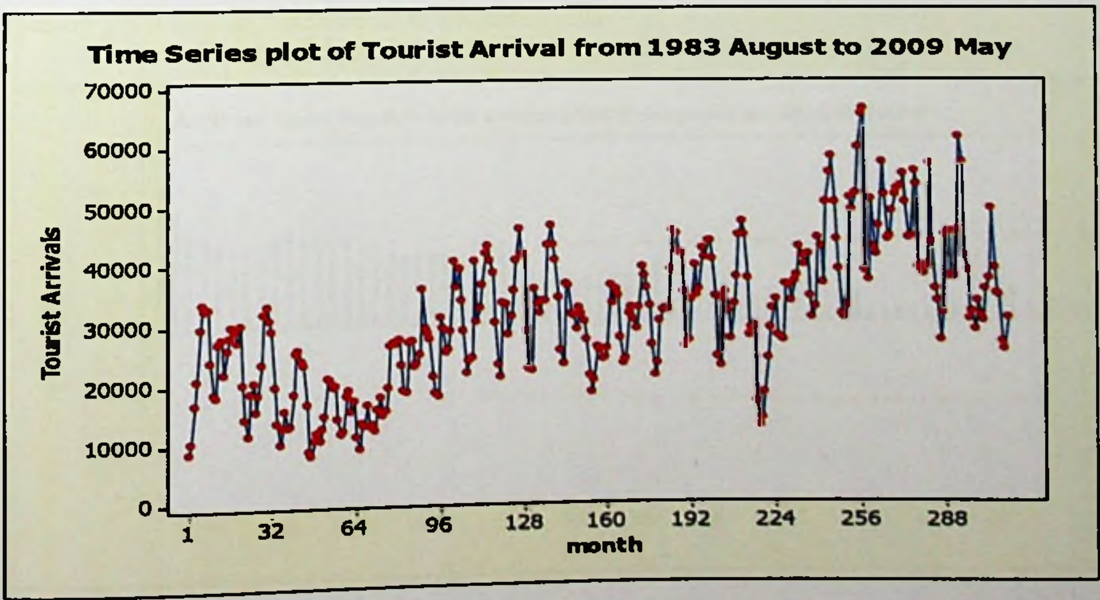


Figure 4.12: Time series plot of tourist arrivals in phase II

The time series plot in Figure 4.12 provides the information about existence of trend and seasonal pattern. In addition to that, in some months (or years) there are some drops in the tourist arrivals. It is hard to discuss the actual reasons behind these drops without proper scientific information. Nevertheless it can be said that based on the data, abnormal pattern occurs during this conflict period.

Table 4.8: Test results of trend and seasonality of monthly tourist arrivals in phase II

Phase II	ADF	Kruskal- Wallis
Test Statistics	-1.57	196.89
p- value	0.50	0.00

From the p- value of ADF test in Table 4.8, it can be concluded with 95% confidence that the series has a trend. Moreover the p- value of Kruskal- Wallis test confirms the existence of seasonal pattern in the series of phase II.

It is very clear from the ACF graph in Figure 4.13 that, the series has the seasonal pattern (in every 12th lag a similar pattern can be seen). Also it does not decay exponentially with lags. This implies that the series in phase II is non-stationary. Thus for the time series model development purposes it has to be transformed to a stationary series by performing methods of transformations.

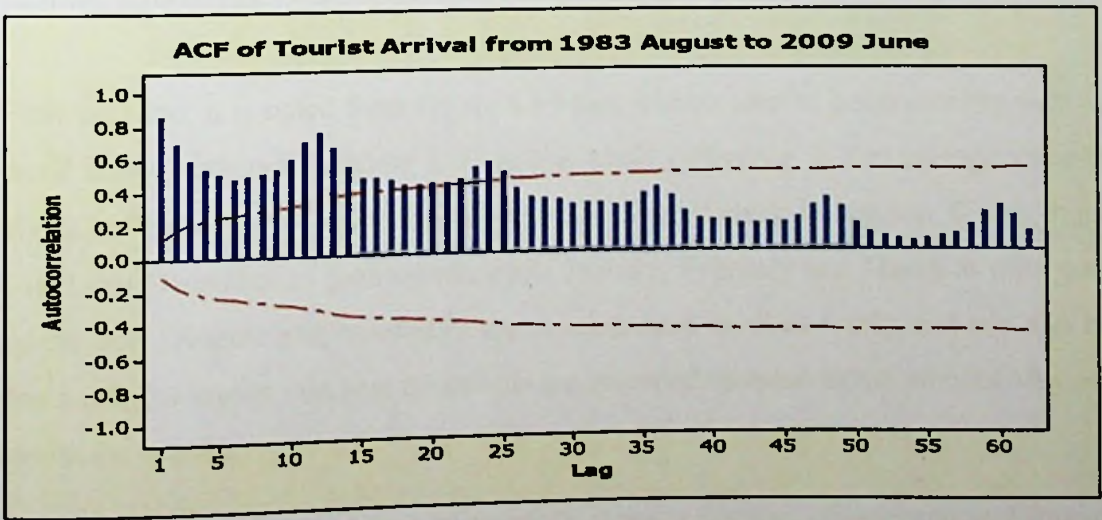


Figure 4.13: ACF graph of the series of tourist arrivals in phase II

4.5 Preliminary Analysis of Tourist Arrivals in Window II

Since the window II covers the time period from June 2009 to December 2015, here it is defined as the after the internal conflict. In window II, the total number of international tourists had visited the Island is 7,376,343 (see Table 4.9), which is about 37.84% of total tourist arrivals from the year 1967 to the year 2015. It implies that more than one third of the total tourists (nearly 19.5 million) had come to Sri Lanka after the conflict.

Table 4.9: Descriptive statistics of tourist arrivals in window II

Year	Mean	Total	Growth Rate	Year	Mean	Total	Growth Rate
2009	43360	260161	2.15%	**2013**	106216	1274593	26.75%
2010	54540	654476	46.12%	**2014**	127263	1527153	19.81%
2011	71331	855975	30.79%	**2015**	149865	1798380	17.76%
2012	83800	1005605	17.48%	**Total tourist arrivals**		**7,376,343**	

From the statistics appeared in Table 4.9, it can be clearly observed that there is dramatic increase in average monthly tourist arrivals from 43,360 to 149,865 in six and half years. Also only in the year 2015 nearly 1.8 million tourists had visited the island and which is the biggest hit in tourism history of Sri Lanka. Further it is noted that, every year there is a positive growth rate.

In this case too, it is noted from Figure 4.14 that, almost similar pattern exists with the overall frame and with window I. Here the small difference is the average monthly arrivals in February exceed the January arrivals. Nevertheless, in window II also, it can be said that December as peak month while January, February and March as mini peak months. July, August and November are at moderated level in tourist arrivals. On the other hand, the lowest numbers of arrivals are recorded as usual in the months May and June on the average.

Figure 4.14 shows that, the plot of monthly average number of international tourists visited to Sri Lanka in window II.

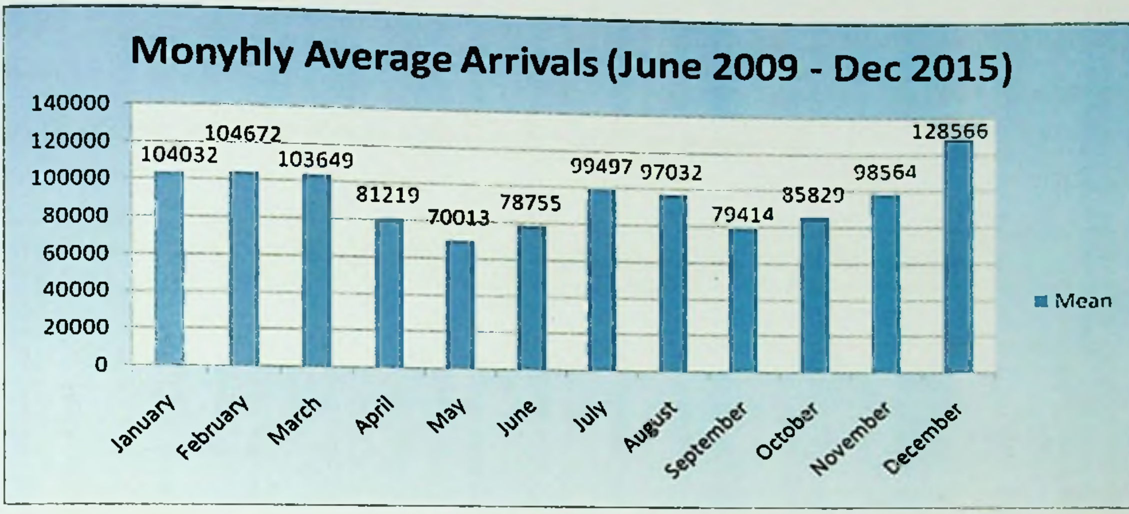


Figure 4.14: Plot of monthly average tourist arrivals in window II

The figure below shows the plot of average monthly arrivals in window II

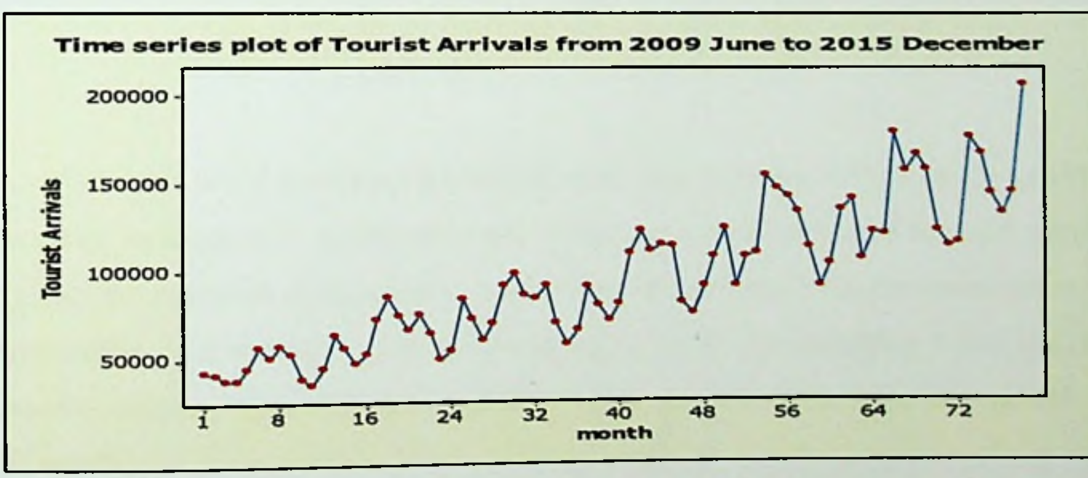


Figure 4.15: Time series plot of tourist arrivals in window II

From the time series plot appeared in Figure 4.15, it can be clearly observed a positive trend with a seasonal pattern. Further it is proved statistically by taking the p- values of ADF and Kruskal- Wallis tests in Table 4.10.

Table 4.10: Test results of trend and seasonality of monthly tourist arrivals in window II

Window II	ADF	Kruskal- Wallis
Test Statistics	-1.73	62.11
p- value	0.99	0.00

Since the p-value of ADF test is 0.99, it can be concluded with 95% confidence that there is a trend in the series. At the same time, the p- value of Kruskal- Wallis test confirms the existence of seasonality in the series of window II.

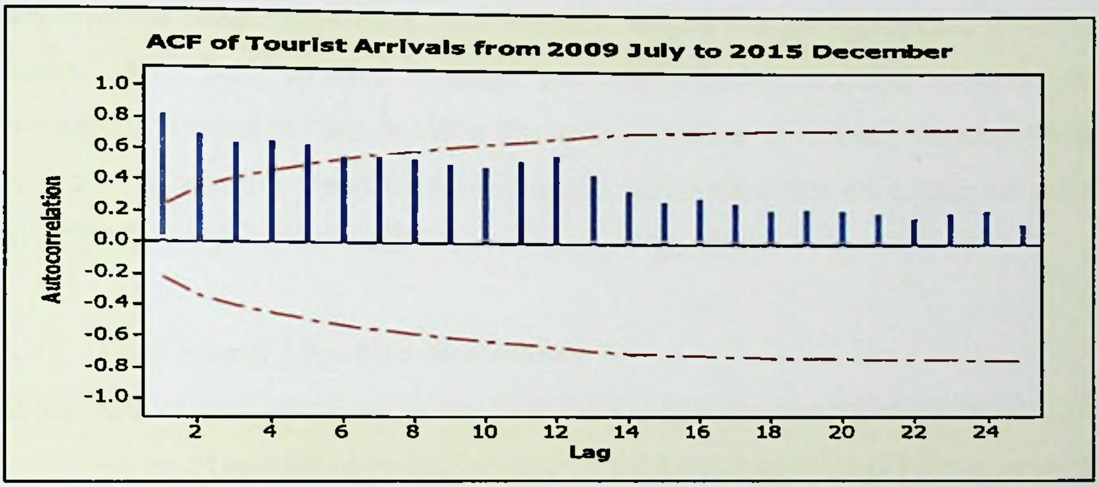


Figure 4.16: ACF graph of the series of tourist arrivals in window II

The ACF graph is not decaying exponentially with lags in Figure 4.16, it can be said that the series in window II is non-stationary at the same time there is a seasonal pattern suggests the existence of seasonality in the series. In window II too, the series has to be transformed as a stationary before developing a model by removing trend and the seasonal pattern. Generally this transformation can be done by the method of differencing.

4.6 Synopsis

In the section of preliminary analysis, all the relevant tests prior to time series model development are carried out separately for overall frame, Window I , Window II, Phase I and Phase II. It includes, stationary of the series checking, behavior of monthly average arrivals, checking of trend and seasonality in the series. This reveals that, on the average, almost same pattern in arrivals is observed in all the time spans.

5. MODEL DEVELOPMENT IN WINDOW I

5.1 Model Development using Exponential Smoothing method in window I

Since the original series in window I is already shown in Chapter 4 that, it has a trend and seasonal components. Thus Holt- Winter's method is more appropriate to fit the model. Also both additive seasonal and multiplicative seasonal models are considered to select the best model in this section. Further to estimate the smoothing constants, auto search algorithm as well as grid search algorithm are employed. Also *STATISTICA* software is used for the two search algorithms.

5.1.1 Grid Search Algorithm in window I

In the case of grid search algorithm, *STATISTICA* provides the estimates of best 10 combinations of smoothing constants based on the ascending order of sum of squares of errors. It is noted that, the values are chosen by setting the initial values as 0.01 and incremental step value by 0.1 and 0.05 for all the cases.

Accordingly, the best combination of estimated smoothing constants for additive seasonal model are: $\alpha = 0.91$, $\beta = 0.01$, and $\gamma = 0.01$. Also the best combination of estimated smoothing constants for multiplicative seasonal model are: $\alpha = 0.95$, $\beta = 0.25$, and $\gamma = 0.05$.

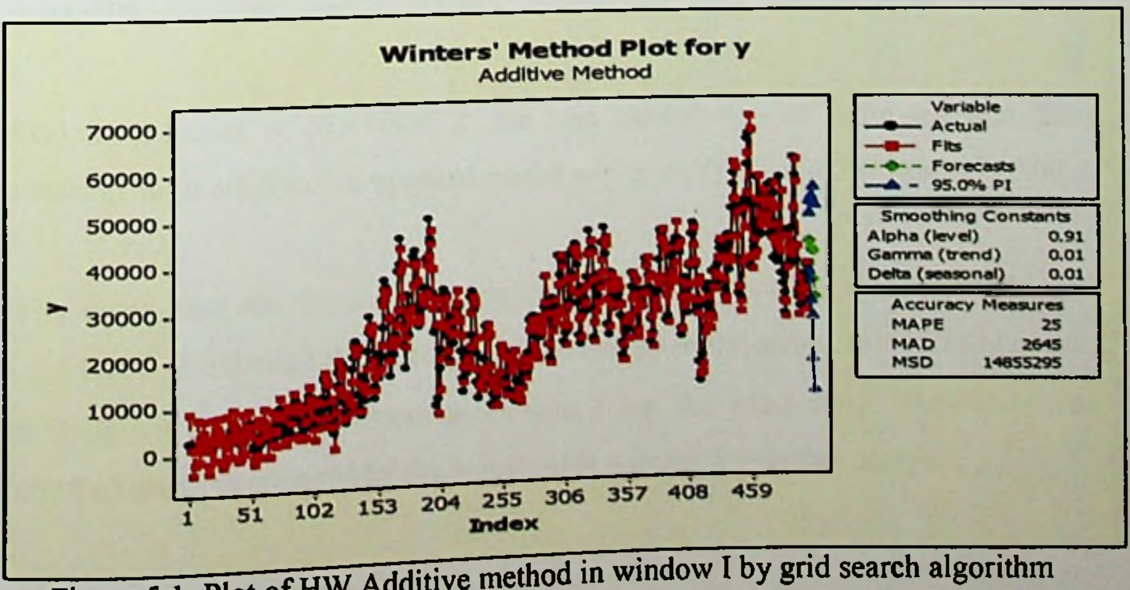


Figure 5.1: Plot of HW Additive method in window I by grid search algorithm

Figure 5.1 and Figure 5.2 represent the Holt – Winter's Multiplicative and Holt – Winter's Additive method plots, respectively, for the series in window I. In this case the smoothing constants are estimated by using grid search algorithm.

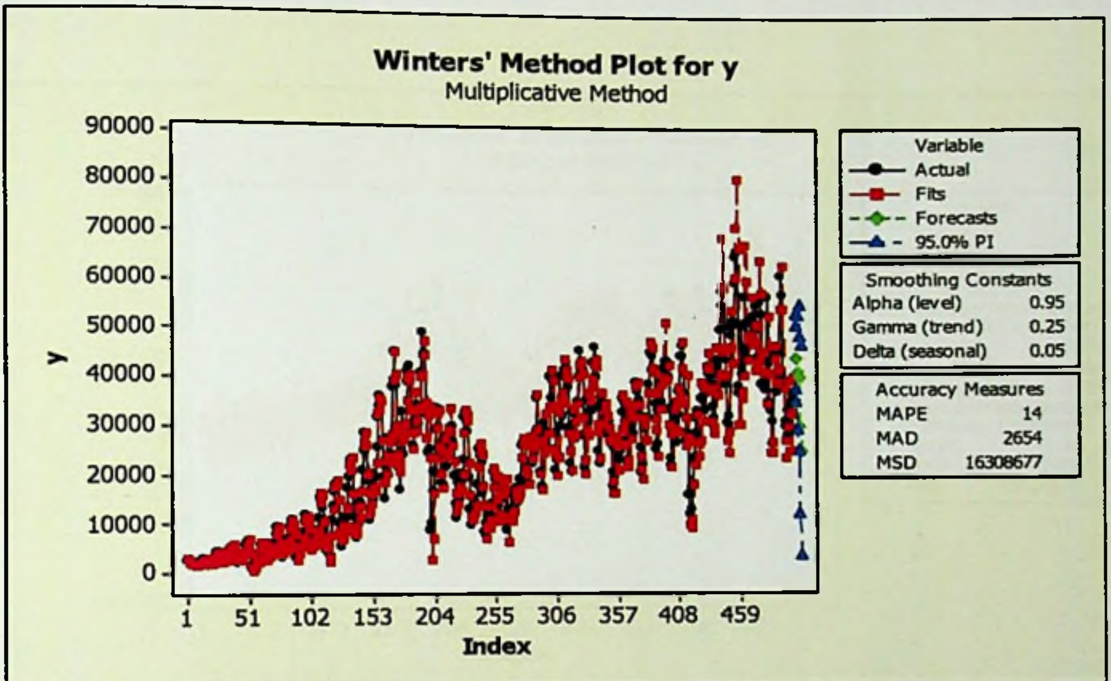


Figure 5.2: Plot of HW Multiplicative method in window I by grid search algorithm

5.1.2 Auto Search Algorithm in window I

In the case of auto search algorithm, from output of *STATISTICA* the estimated smoothing constants for additive seasonal model are: $\alpha = 0.095$, $\beta = 0.905$, and $\gamma = 0.036$.

Also from output of *STATISTICA*, the best combination of estimated smoothing constants for multiplicative seasonal model are: $\alpha = 1.00$, $\beta = 0.597$, and $\gamma = 0.00$.

It is noted that the third smoothing constant corresponding to seasonality is 0 ($\gamma = 0.00$). It indicates that there is no seasonality in the series. It is already proved in Section 4.2 that, the series in window I has the seasonality. Therefore, auto searched multiplicative model for window I is rejected for further analysis.

Figure 5.3 and Figure 5.4 represent the Holt – Winter's Multiplicative and Holt – Winter's Additive method plots, respectively, for the series in window I. Also it is noted in this case that, the smoothing constants are estimated by using auto search algorithm.

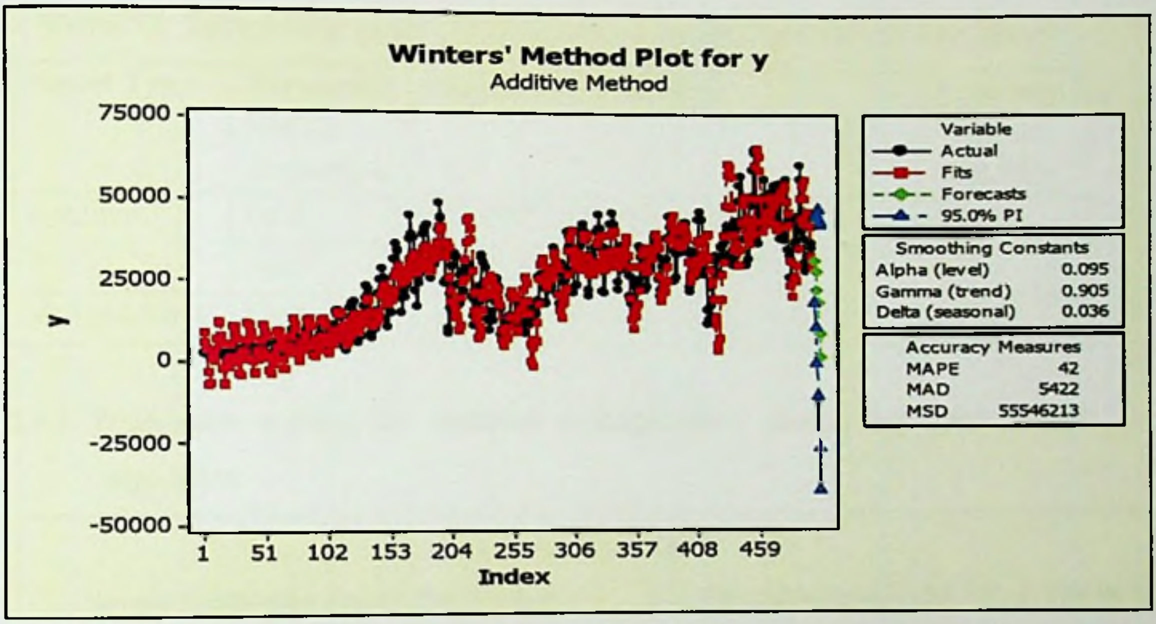


Figure 5.3: Plot of HW Additive method in window I by auto search algorithm

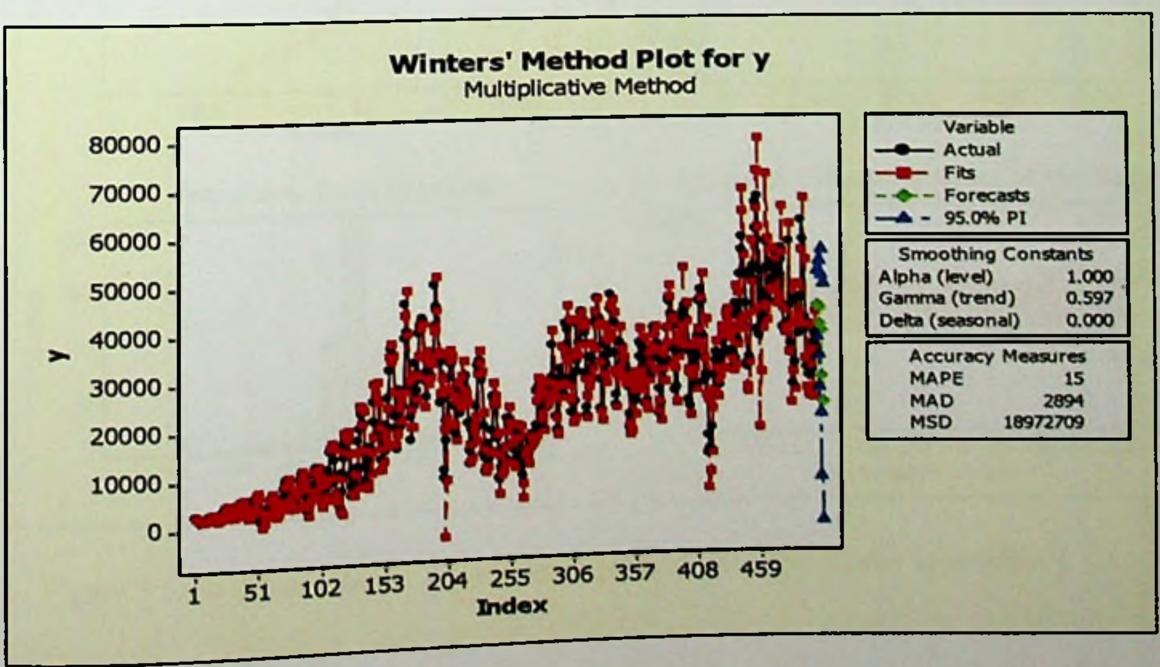


Figure 5.4: Plot of HW Multiplicative method in window I by auto search algorithm

Table 5.1 below summaries the three different models obtained from exponential smoothing method. However based on the MAPE value of each model, the grid searched multiplicative model (whose MAPE values is 14) is considered, at the moment, for further analysis such as for diagnostic testing and ex-ante forecasting.

Table 5.1: Summary of models by Exponential Smoothing method in window I

Model Type	Parameter Search Algorithm	Parameters Estimates	Accuracy Measure (MAPE)
Additive	Grid	$\alpha=0.91$, $\beta=0.01$, and $\gamma=0.01$	25
	Auto	$\alpha=0.095$, $\beta=0.905$, and $\gamma=0.036$	42
Multiplicative	Grid	$\alpha=0.95$, $\beta=0.25$, and $\gamma=0.05$	14

5.1.3 Diagnostic testing for selected multiplicative model by grid search algorithm

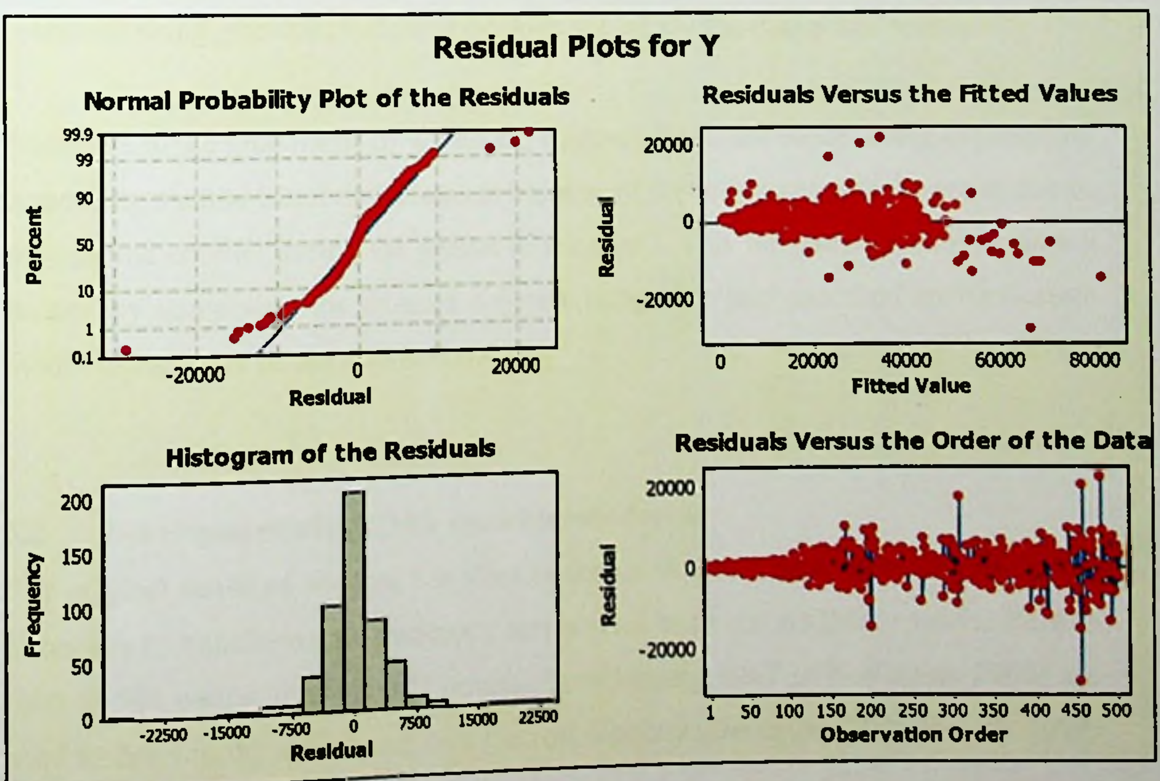


Figure 5.5: Residual plots of the grid searched multiplicative model in window I

According to the p- value (< 0.005) of Anderson- Darling test (obtained by plotting only the plot of probability separately), it can be concluded with 95% confidence that the residual of multiplicative model (under grid search algorithm) do not follow normal distribution. Also it can be seen in Figure 5.5 that, the normal probability plot of the residuals is not linear. This too suggests that the residuals do not follow normal distribution.

From, the plot of residuals versus fitted values in Figure 5.5, it can be seen that the residuals do not scatted randomly, thus the residuals are not independently distributed. In Figure 5.5 of the plot of residuals versus order of the observation suggests that the variance of residual reasonably constant as it looks like symmetric about 0 though there are few outliers. Also it does not follow any pattern.

Ultimately it can be concluded that the residuals of the multiplicative model (obtained using grid search algorithm) does not satisfy the diagnostic testing.

Therefore in the time frame of window I, finding significant model using exponential smoothing method is not that adequate because of the long swings in the series due to the internal conflict during the period in window I. This inappropriateness is shown further by computing the ex-ante forecast using this grid searched multiplicative model in Table D1 of the APPENDIX D.

5.2 Development of ARIMA model in window I

The original series of window I is already shown that as a non-stationary. Thus it is necessary to transform as a stationary series prior to fit the ARIMA models. Here in total of 509 points, the first 503 points (from January 1967 to November 2008) are used to develop the model and then the rest 6 points (December 2008 to May 2009) are reserved for the model validation.

5.2.1 Transformation of series in window I

Previously section 4.2 confirms that the series has a positive trend and hence it is non- stationary. To remove the trend, the regular difference technique is carried out. Accordingly, the new series is named as D1Y (the original series after 1 difference) and the new time series plot for D1Y is plotted in Figure 5.6.

It seems visually from the plot of time series of D1Y in Figure 5.6 that, there is no trend but it has to be proved statistically using the test called ADF test as well. The test statistic value of ADF test is -7.98 and its corresponding p- value is 0.00. Therefore it can be concluded with 95% confidence that, the D1Y series has no trend.

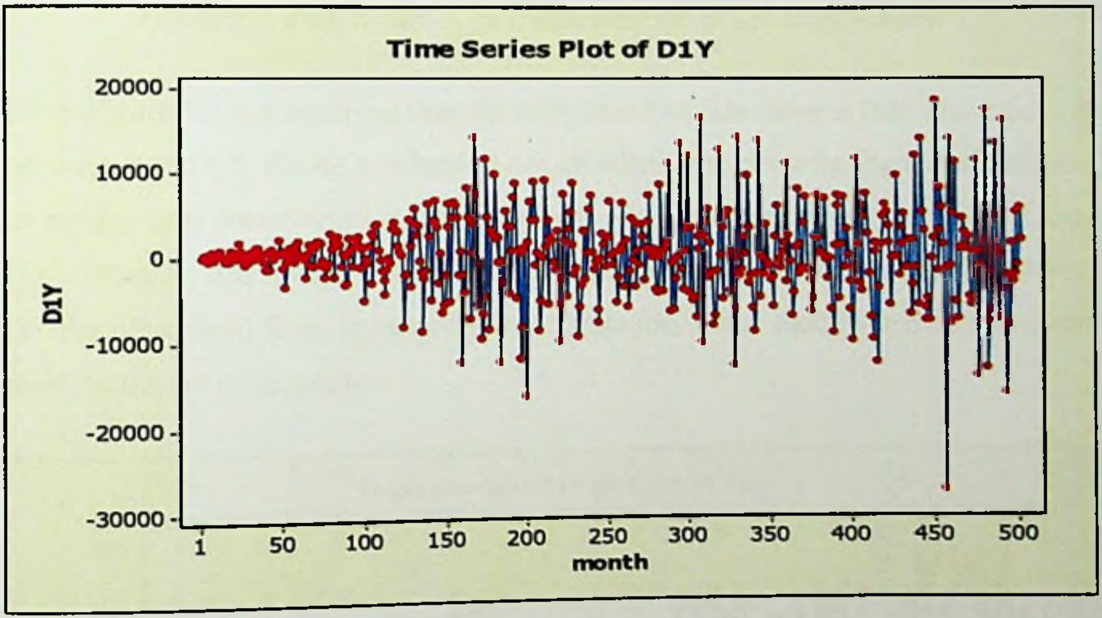


Figure 5.6: Time series plot of D1Y series in window I

Though, there is no information regarding the seasonal pattern from Figure 5.6, some high fluctuations can be seen in the middle. Before checking the ACF graph for existence of seasonal pattern, it is better to do the Box- Cox Transformation for the original series. Then the Box-Cox plot is plotted in Figure 5.7 below.

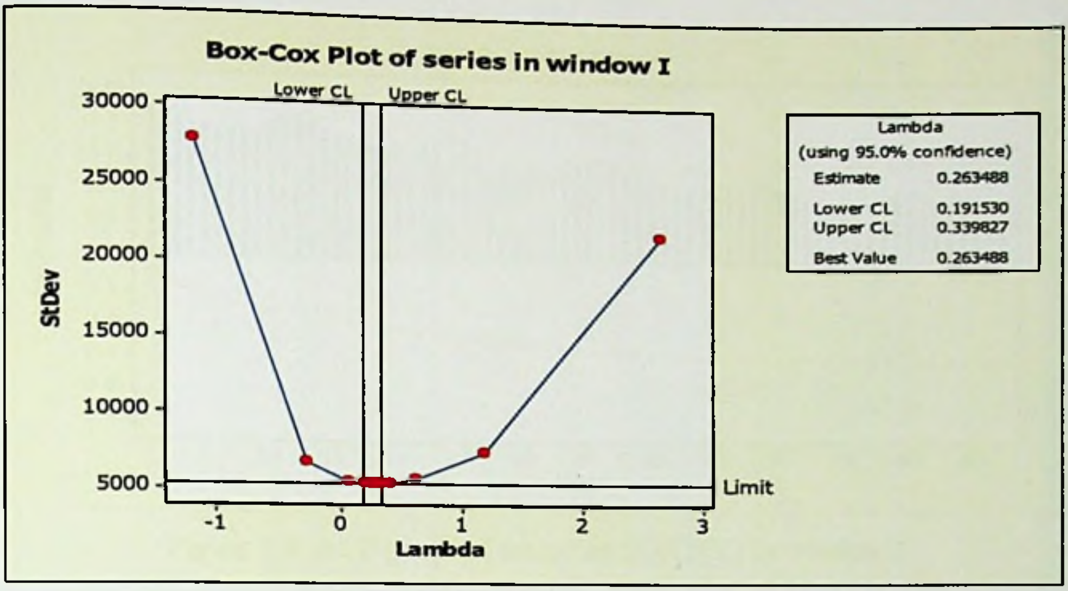


Figure 5.7: Plot of Box- Cox transformation of series in window I

From Figure 5.7 it is observed that, the estimated Lambda value is 0.26 and it falls in between 0 and 0.5. Hence it is hard to decide whether to consider log transformation or square root transformation. Nevertheless, both transformations are carried out. Unfortunately, it is unable to obtain any significant model for the log transformation. On the other hand from square root transformation, some models are obtained and their details are given below:

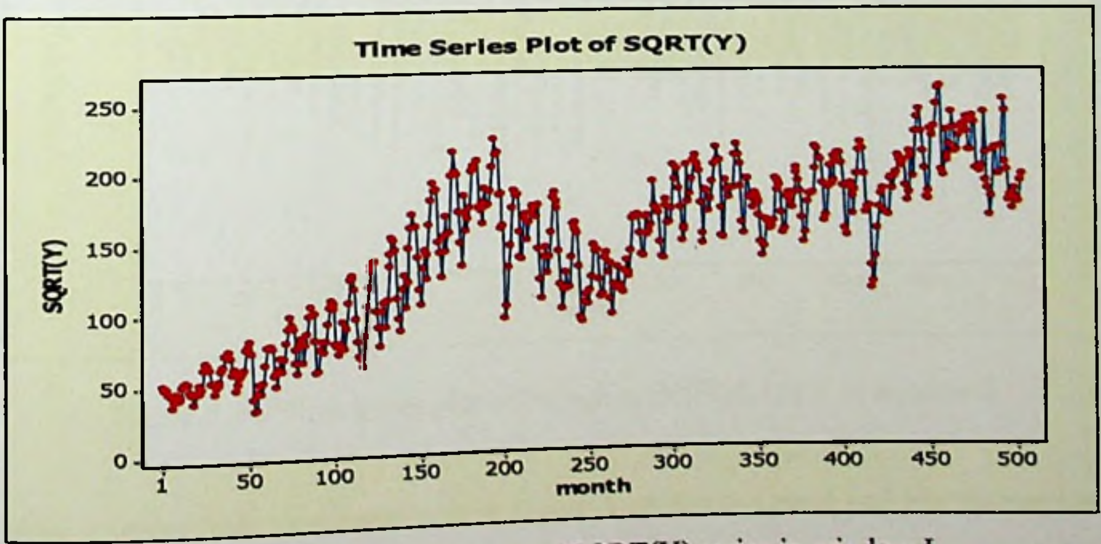


Figure 5.8: Time series plot of SQRT(Y) series in window I

It is noted from the time series plot of SQRT (Y) in Figure 5.8 that, the exactly the same pattern (with original series Y) of trend and seasons are exists here as well.

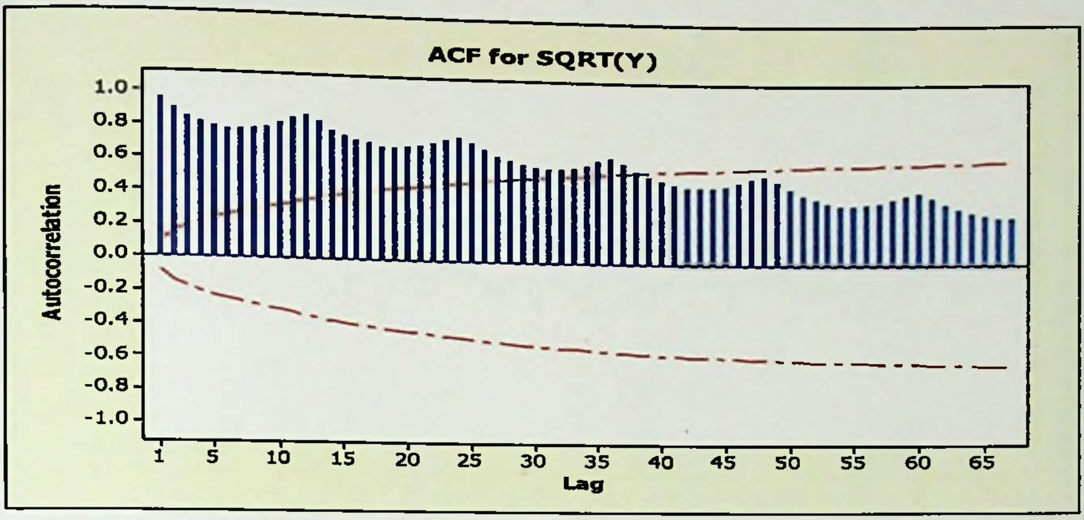


Figure 5.9: ACF graph of the series SQRT(Y) in window I

ACF graph in Figure 5.9 too gives the same style and pattern with ACF of the original series in Figure 4.7. Thus to remove the trend, the first difference for the series SQRT(Y) is taken.

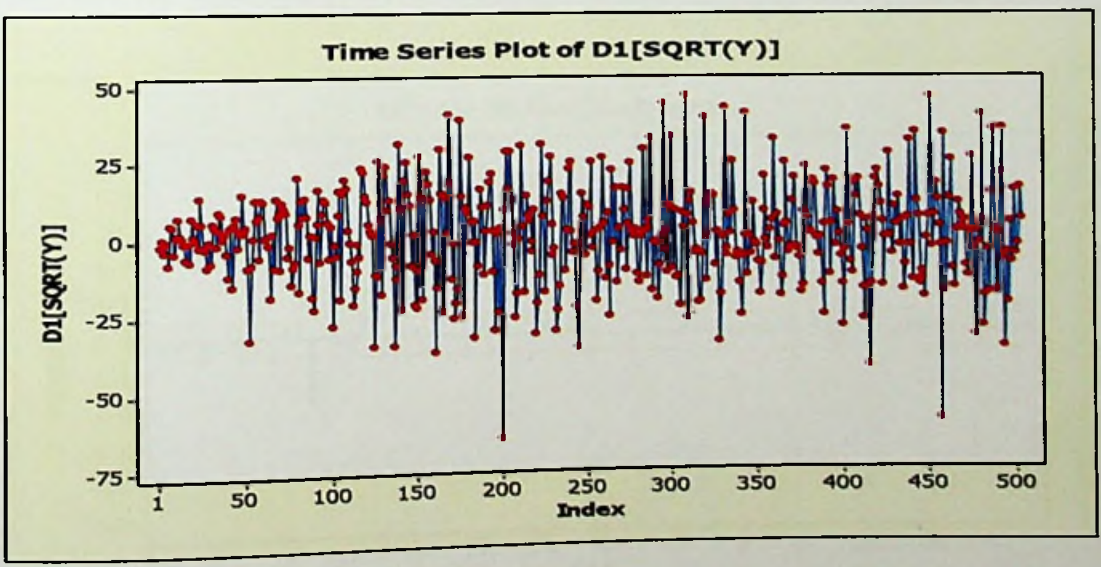


Figure 5.10: Time series plot of the series D1[SQRT(Y)] in window I

Now it seems from time series plot in Figure 5.10 that, no trend and less fluctuation compare with D1Y series in Figure 5.6. Further, the test statistic of ADF test (-10.46) with its p-value (0.00) strongly suggests that there is no trend in the series of D1[SQRT(Y)].

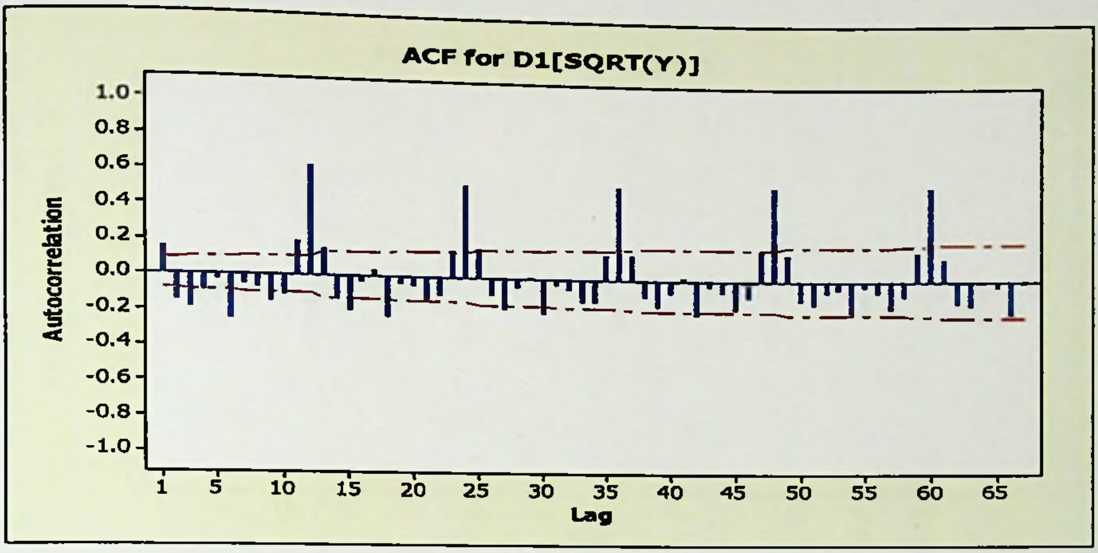


Figure 5.11: ACF graph of the series D1[SQRT(Y)] in window I

ACF graph in Figure 5.11 clearly shows that, the series D1[SQRT(Y)] has seasonal pattern with every 12th lag. Therefore, to remove the seasonal pattern, the 12th difference for the series D1[SQRT(Y)] is taken and its ACF graph is shown below.

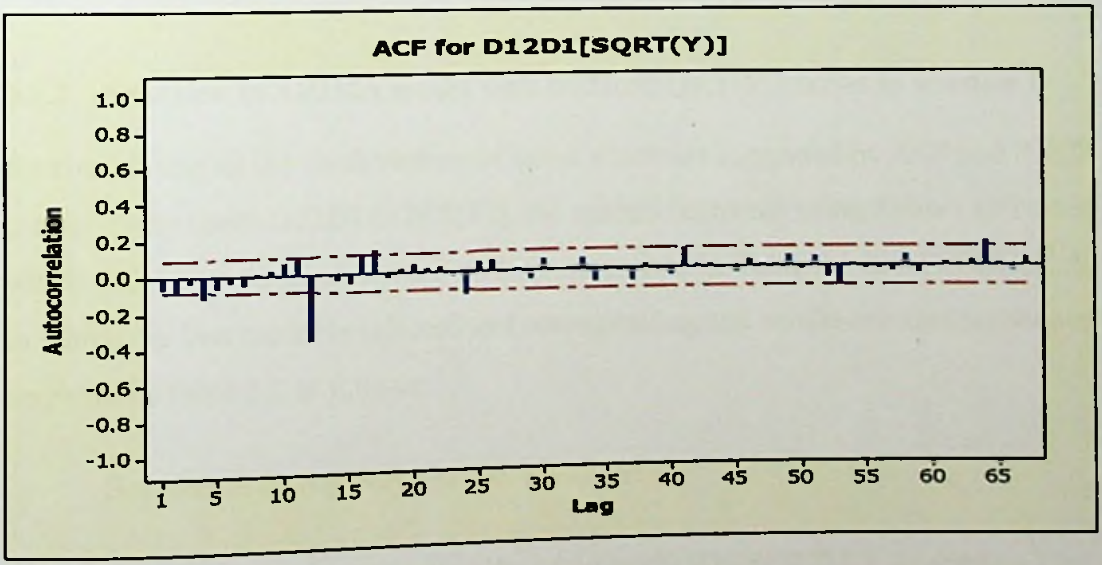


Figure 5.12: ACF graph of the series D12D1[SQRT(Y)] in window I

It seems that from the ACF graph in Figure 5.12 that, the seasonality is removed and the model may contain the Seasonal MA terms such as SMA(1), SMA(2). With this, PACF also has to be checked prior to confirm the stationary of series.

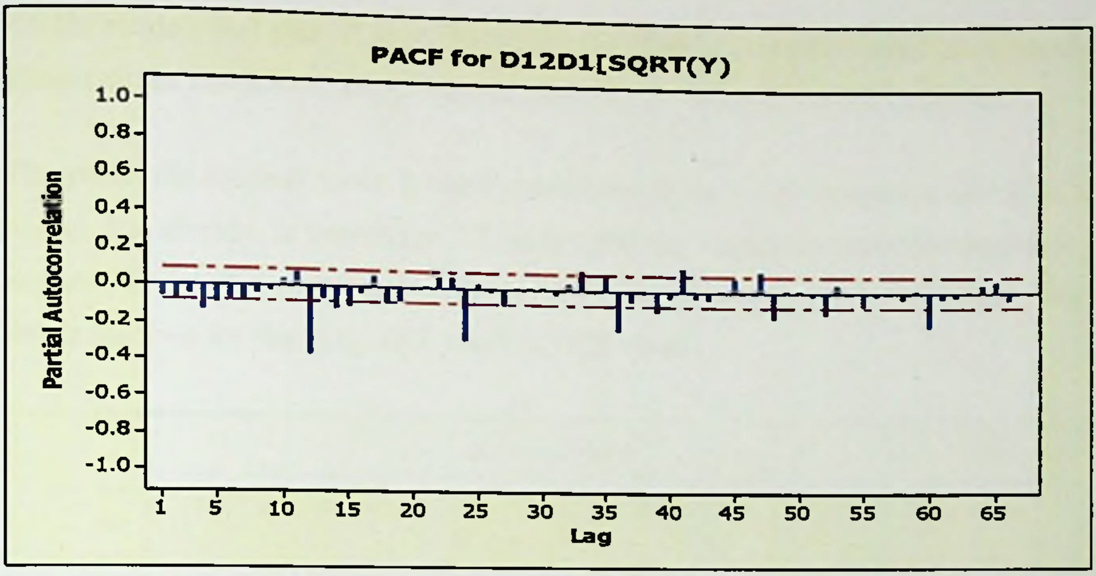


Figure 5.13: PACF graph of the series D12D1[SQRT(Y)] in window I

From PACF graph in Figure 5.13 suggests that, the model may have Seasonal AR terms such as SAR(1), SAR(2). Furthermore, from ACF graph in Figure 5.12 and PACF graph in Figure 5.13 confirm that the series D12D1[SQRT(Y)] is stationary and now the Seasonal ARIMA model can be developed.

5.2.2 Selection of ARIMA model with D12D1[SQRT(Y)] series in window I

By considering all the combinations of terms which are suggested by ACF and PACF graphs of the series D12D1[SQRT(Y)], the models (obtained using *Eviews* software) which satisfy the required conditions are summarised in Table A1 in APPENDIX A. In which the best model is selected and corresponding test results are also mentioned herewith in Table 5.2 as follows:

Best model is: **ARIMA (0, 1, 0) (1, 1, 1)$_{12}$**

Table 5.2: Test results of best model of series D12D1[SQRT(Y)] in window I

Skewness	Kurtosis	White's General Test (p-value)	Lagrange's Multiplier Test (p-value)	Durbin Watson Statistics	R^2 Value	AIC	SBC
-0.93	**10.27**	0.85	0.17	2.05	0.38	7.61	7.63

All the models in Table A1 in APPENDIX A including this best model have satisfy almost all the conditions, except that the Kurtosis of residuals are not closer to 3.

Therefore, the original series Y itself is considered for the development of ARIMA model. It is already, in the section 4.2, shown that the original series is non-stationary and time series plot of its trend removed series (D1Y) in Figure 5.6 is obtained. Now this is the time for checking ACF graph of D1Y series.

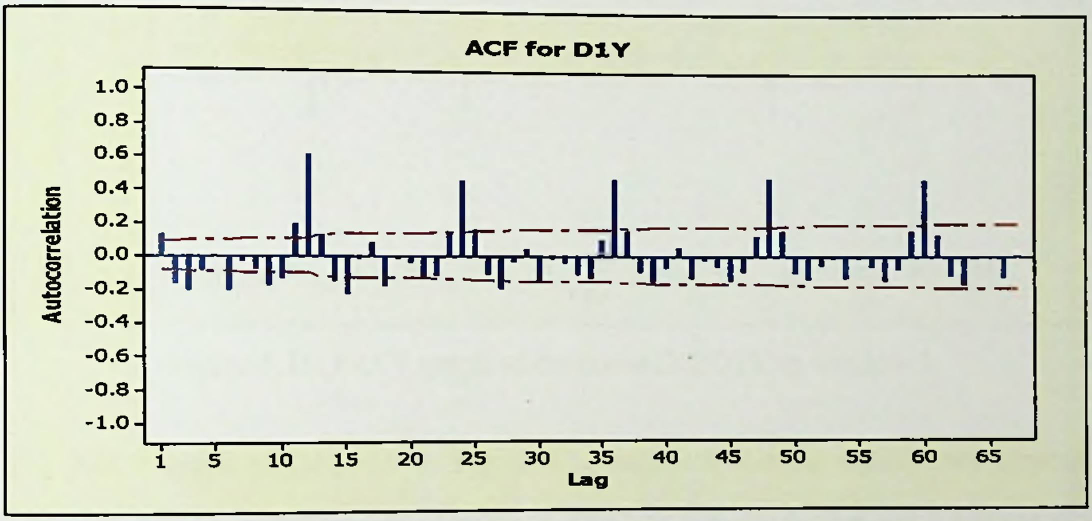


Figure 5.14: ACF graph of the series D1Y in window I

In the ACF graph of the series D1Y in Figure 5.14 also, it is noted the almost same pattern like in the ACF graph of the series D1[SQRT(Y)] in Figure 5.11. The 12th difference is also considered for further investigation.

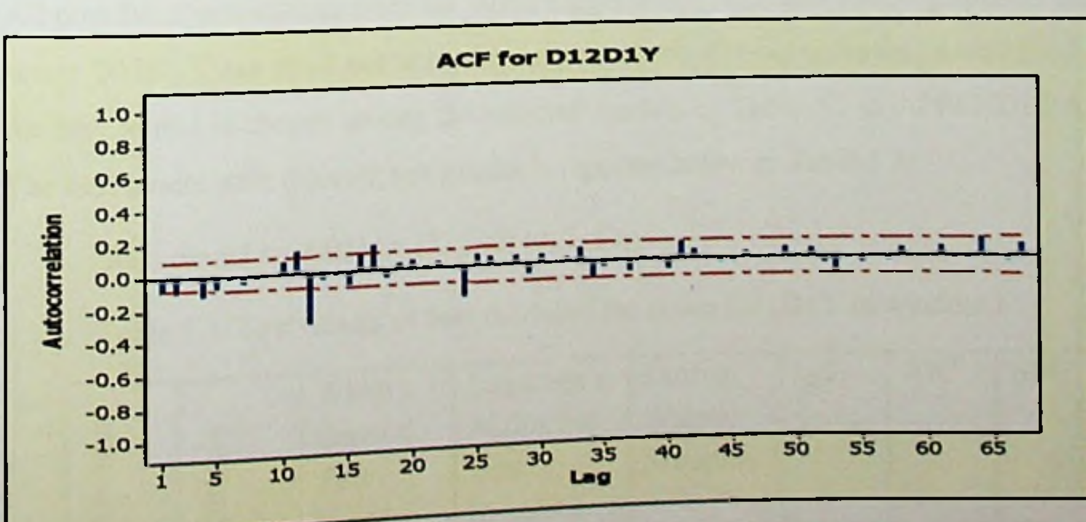


Figure 5.15: ACF graph of the series D12D1Y in window I

This ACF graph of D12D1Y in Figure 5.15 suggests that the models may contain Seasonal MA(1), Seasonal MA(2) as their terms. To confirm the stationarity, the PACF is also investigated.

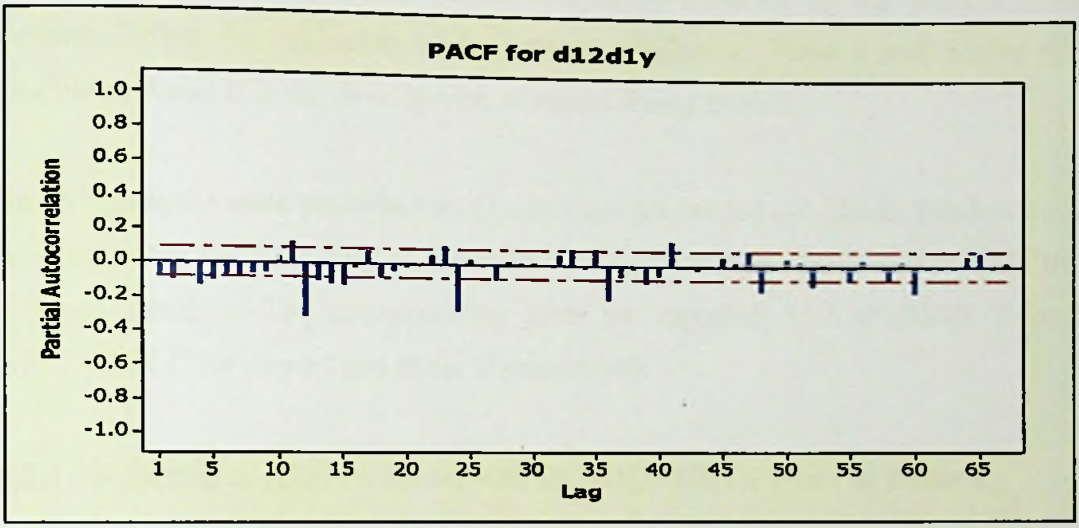


Figure 5.16: PACF graph of the series D12D1Y in window I

The PACF graph of D12D1Y in Figure 5.16 suggests that the model may contain Seasonal AR(1), Seasonal AR(2) terms. Further to this, both ACF (in Figure 5.15) and PACF (in Figure 5.16) confirm that the series is stationary. Thus further model development can be carried out with D12D1Y.

5.2.3 Selection of ARIMA model with D12D1Y series in window I

All possible combinations from the terms suggested by ACF and PACF graphs of the series D12D1Y are tried out with the support of the *Eviews* software. Accordingly the best model is chosen among the selected models in Table A2 in APPENDIX A. The best model with relevant test results is reported below in Table 5.3:

Best model is: **ARIMA (1, 1, 0) (0, 1, 1)$_{12}$**

Table 5.3: Test results of best model of the series D12D1Y in window I

Skewness	Kurtosis	White's General Test (p-value)	Lagrange's Multiplier Test (p-value)	Durbin Watson Statistics	R^2 Value	AIC	SBC
-0.64	9.46	0.14	0.07	2.09	0.30	19.37	19.36

The Kurtosis value in Table 5.3 of the model for the series D12D1Y is also not closer to 3. However this model, as well, meets all other required conditions. Since the normality condition is not satisfied by the models of the series D12D1[SRQT(Y)] and D12D1Y, it is decided to further analyze by breaking up the window I as follows: before the beginning of the internal conflict as Phase I and during the conflict as Phase II in the development of model fitting as well.

Since almost the same procedure and techniques are carried out like in window I, in this part, only the final step of the best model selection is shown below with the required statistics. The corresponding plots are reported in APPENDIX B and APPENDIX C for phase I and phase II respectively.

5.2.4 Selection of ARIMA model with D12D1[LOG(Y)] series in phase I

Best model for phase I is: **ARIMA (2, 1, 0) (0, 1, 1)$_{12}$**

Table 5.4: Test results of best model of the series D12D1[LOG(Y)] in phase I

Skewness	Kurtosis	White's General Test (p-value)	Lagrange's Multiplier Test (p-value)	Durbin Watson Statistics	R^2 Value	AIC	SBC
-0.59	**8.56**	0.20	0.10	2.05	0.50	7.68	7.74

The Kurtosis value of residuals in Table 5.4 is somewhat less than those for the other two series D12D1[SRQT(Y)] and D12D1Y in window I. But it is also not closer to 3 and thus indicates that, still the residuals do not follow normal distribution.

5.2.5 Selection of ARIMA model with D12D1[SQRT(Y)] series in phase II

Best model for phase I is: **ARIMA (0, 1, 0) (0, 1, 1)$_{12}$**

From Table 5.5, it can be observed that, in this case too, the Kurtosis value (10.86) is more than that of the value obtained in phase I (8.56). Therefore it can be claimed that this abnormal condition in window I occurs due to phase II.

Table 5.5: Test results of best model of the series D12D1[SQRT(Y)] in phase II

Skewness	Kurtosis	White's General Test (p-value)	Lagrange's Multiplier Test (p-value)	Durbin Watson Statistics	R^2 Value	AIC	SBC
-0.57	**10.86**	0.98	0.14	2.07	0.35	19.58	19.59

In other words, the high fluctuations in the series in phase II affect the entire series in window I. These high fluctuations (due to outliers) have to be removed to get a better Kurtosis value for the best model in window I. Until that, it is assumed **ARIMA (1, 1, 0) (0, 1, 1)$_{12}$** as best model for the series D12D1Y so far constructed for window I.

5.3 Model development using Dynamic Transfer Function method in window I

As it is described in the methodology chapter, the Dynamic Transfer Function (DTF) for a univariate time series can be written as $Y_t = \frac{\theta(B)}{\phi(B)} e_t$ where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$ are polynomials in B (backshift operator) of degree p and q respectively, and $\{e_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance σ_e^2.

Though to fit DTF model, non-stationary series can be used, stationary series is much better to obtain more significant model. Meanwhile it is already shown from the section 5.2.2 that, D12D1Y as the stationary series for window I. In this section too the same series is employed to fit DTF model.

5.3.1 Parameter Estimation of DTF model in window I

In the case of estimating the parameters in *SAS* programme, several models are tried out by assigning different values for p and q, finally it is end up with the same model Seasonal ARIMA (1, 1, 0)(0, 1, 1)$_{12}$ which obtained in the previous section 5.2.3. This may be due to the reason that, in the univariate time series process of DTF, the DTF model is equivalent to ARIMA model.

According to the *SAS* output, the parameter $\theta_1 = 1.91091$ and its p- value is less than 0.0005 indicates that the AR(1) parameter is significant. Similarly the parameter $\phi_1 = 0.0101$ and its p- value is also less than 0.0005 indicates that MA(12) parameter is significant. Therefore it can be concluded with 95% confidence that the parameters are significant. Hence

Autoregressive (AR) Factor: $1-1.91091*B$

Moving Average (MA) Factor 1: $1-0.0101*B^{12}$

Here it is noted that only one model is significant and therefore there is no need of comparing information criterions. However for information, those criterions are listed as the goodness of fitted model: AIC=9994.356, SBC=10090.87, $R^2 = 75.87\%$

5.3.2 Diagnostic checking for the residuals of fitted DTF model in window I

The p-value (<0.005) of Anderson Darling test in Figure 5.17 suggests that the residuals do not follow normal distribution. Further plot of residuals versus predicted values in Figure 5.18 is not random indicates that the residuals are not independently distributed. Thus it can be concluded that the residuals of fitted model do not satisfy the diagnostic testing.

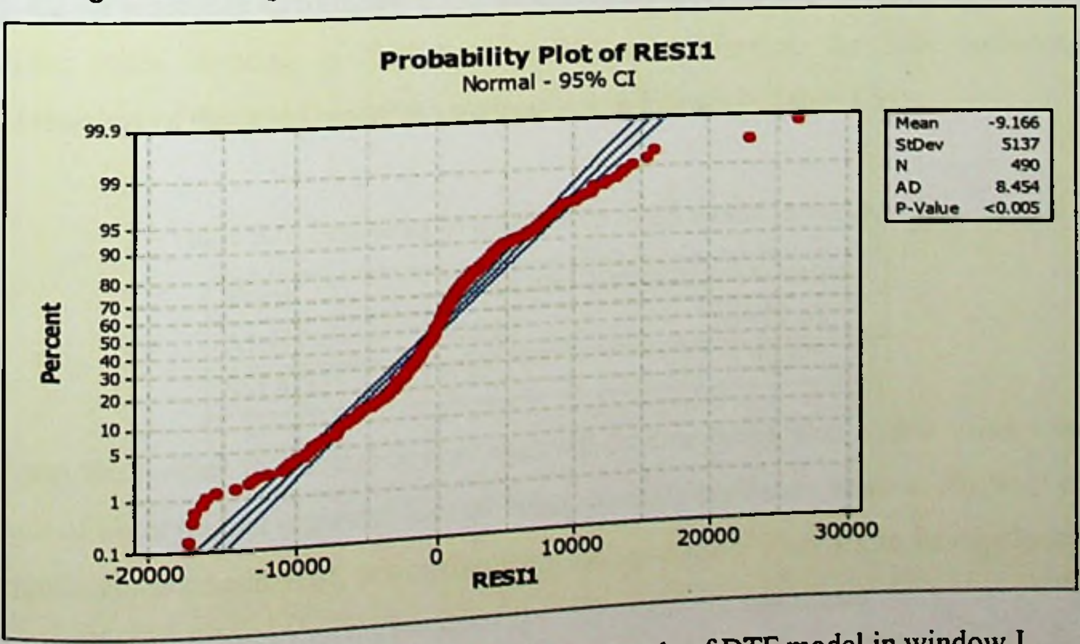


Figure 5.17: Normal probability plot of residuals of DTF model in window I

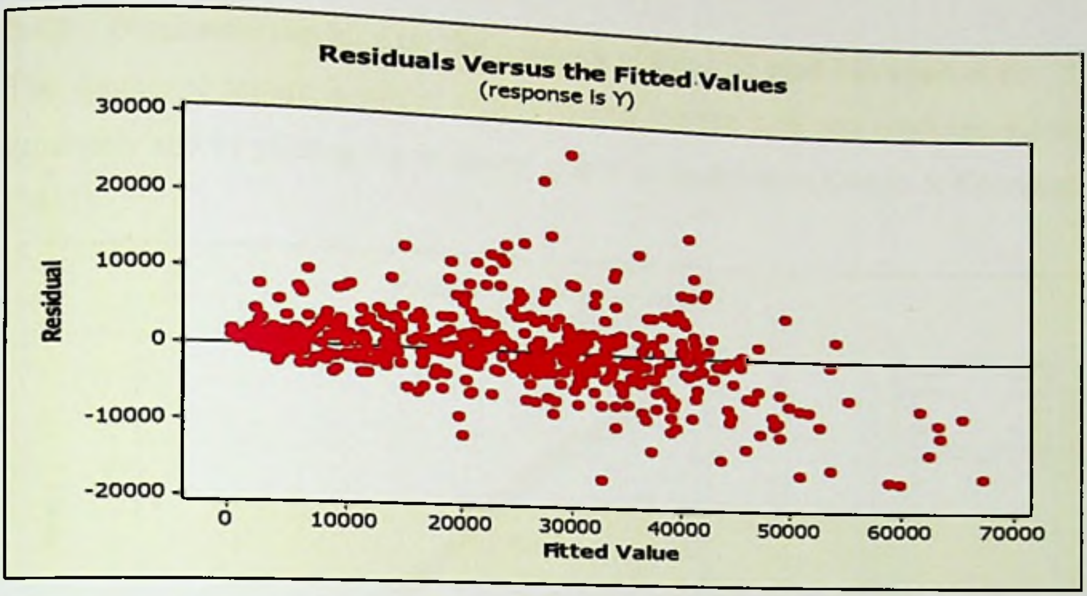


Figure 5.18: Plot of Residuals versus Predicted values of DTF model in window I

5.4 Model development using State Space method in window I

Stationary condition is not necessary to fit a model using State Space (SS) method. Nevertheless, the stationary series obtained in the section 5.2.2, D12D1Y is employed in this section as well. The parameters of SS models are estimated by using Kalman filter method.

5.4.1 Parameter Estimation of SS model by Kalman filter in window I

After some iteration in *SAS* from preliminary estimates, the final estimated parameters of the fitted model are summarised as follows in Table 5.6:

Table 5.6: Estimates of parameters of SS model in window I

Term	Parameter Estimates	t- value
AR(1)	-0.023	-17.62
SMA(1)	0.82	22.92

From the t-value in Table 5.6, it is observed by comparing with t-table value 1.96 (no. of observations is greater than 60 hence normal distribution value at 5% level of significant is considered), both parameters are significant. Thus it can be concluded with 95% confidence that the parameters of SS model are significant.

5.4.2 Diagnostic checking for the residuals of fitted SS model in window I

The diagnostic testing is carried out by storing the residuals and predicted values separately and by plotting the following graphs as described in Durbin & Koopman, (2012).

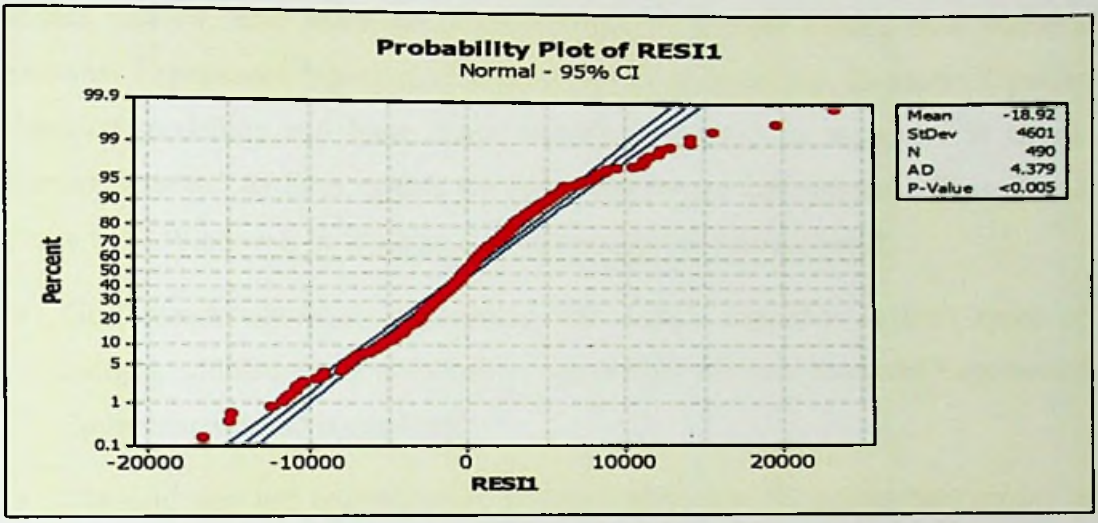


Figure 5.19: Normal probability plot of residuals of SS model in window I

The p-value (<0.0005) of the Anderson Darling test in the probability plot of residuals of the fitted model in Figure 5.19, it can be concluded with 95% confidence that the residuals are not normally distributed.

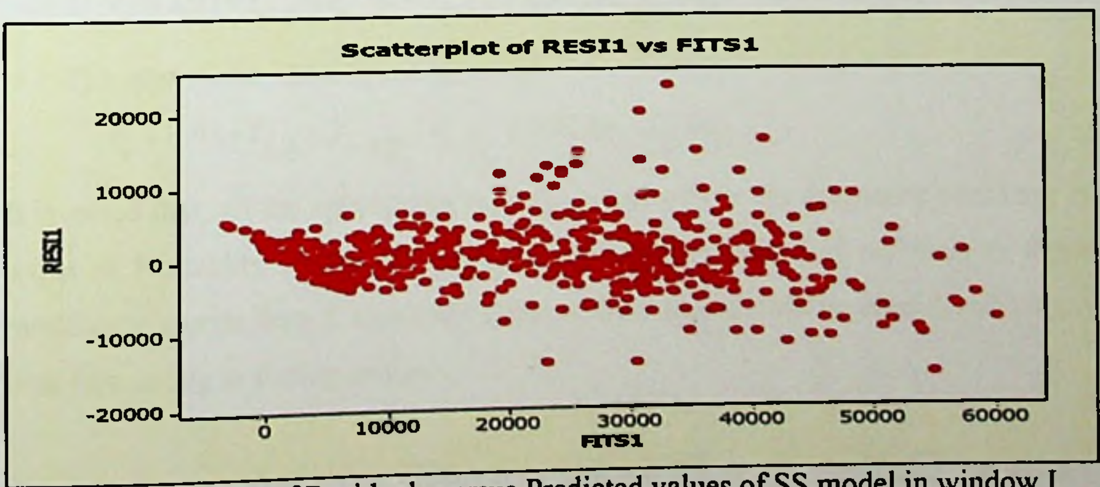


Figure 5.20: Plot of Residuals versus Predicted values of SS model in window I

Further, from the plot of residuals versus predicted values of the fitted model in Figure 5.20, it can be seen that the residuals are not randomly scatted. Thus it can be concluded that the residuals do not follow normal distribution and are not independently distributed, it can be concluded that the fitted model by SS method does not satisfy the diagnostic testing.

5.5 Synopsis

In this Chapter, time series model development in window I using Holt-Winter's Seasonal Exponential Smoothing, Seasonal ARIMA modelling, Dynamic Transfer Function modelling and State Space modelling methods is explained in detail. Further Seasonal ARIMA models are developed for the sub windows, Phase I and Phase II, of Window I. In Window I, the followings can be concluded:

- Grid search algorithm outperforms auto search algorithm in both types of models, additive and multiplicative, when Holt-Winter's Seasonal Exponential Smoothing method is employed.
- The grid searched multiplicative model is selected as the appropriate model in Holt-Winter's Seasonal Exponential Smoothing method.
- The appropriate Seasonal ARIMA model is

  $$Y_t = 1.11 * Y_{t-1} + Y_{t-12} - Y_{t-13} - 0.88 * e_{t-1} + e_t$$

- The appropriate Dynamic Transfer Function model is

  $$Y_t = 2.91 * Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.01 * e_{t-12} + e_t$$

- The appropriate State Space model is

  $$Y_t = 0.98 * Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.82 * e_{t-12} + e_t$$

It is noted that, all the appropriate models do not satisfy the diagnostic checking in terms of Normality testing. Particularly, the Kurtosis values of residuals of those models are greater than 3. However, none of these models will be considered for ex-post forecasting in further analysis.

Based on Ex-ante forecast in Table D1 in APPENDIX D, the significant models are only from the methods of Seasonal ARIMA modelling and Dynamic Transfer Function modelling. In which Seasonal ARIMA modelling outperform other methods.

6. MODEL DEVELOPMENT IN WINDOW II

6.1 Model Development using Exponential Smoothing method in window II

The series in window II is also shown already that it has a trend and seasonal components. Thus the more appropriate technique Holt- Winter's method is employed to fit the models. In window II also both types of models additive seasonal and multiplicative seasonal are considered to select the best model. For the estimation of smoothing constants, auto search algorithm and grid search algorithm are applied. Also *STATISTICA* software is used for the two search algorithms.

6.1.1 Grid Search Algorithm in window II

The output of *STATISTICA* provides the estimates of best 10 combinations of smoothing constants (for both types of models) based on the ascending order of sum of squares of errors. In addition, the values are chosen by setting the initial values as 0.01 and incremental step value by 0.1 and 0.05 for all the cases.

Accordingly, the best combination of estimated smoothing constants for additive seasonal model are: $\alpha = 0.15$, $\beta = 0.05$, and $\gamma = 0.10$. Also the best combination of estimated smoothing constants for multiplicative seasonal model are: $\alpha = 0.25$, $\beta = 0.05$, and $\gamma = 0.05$.

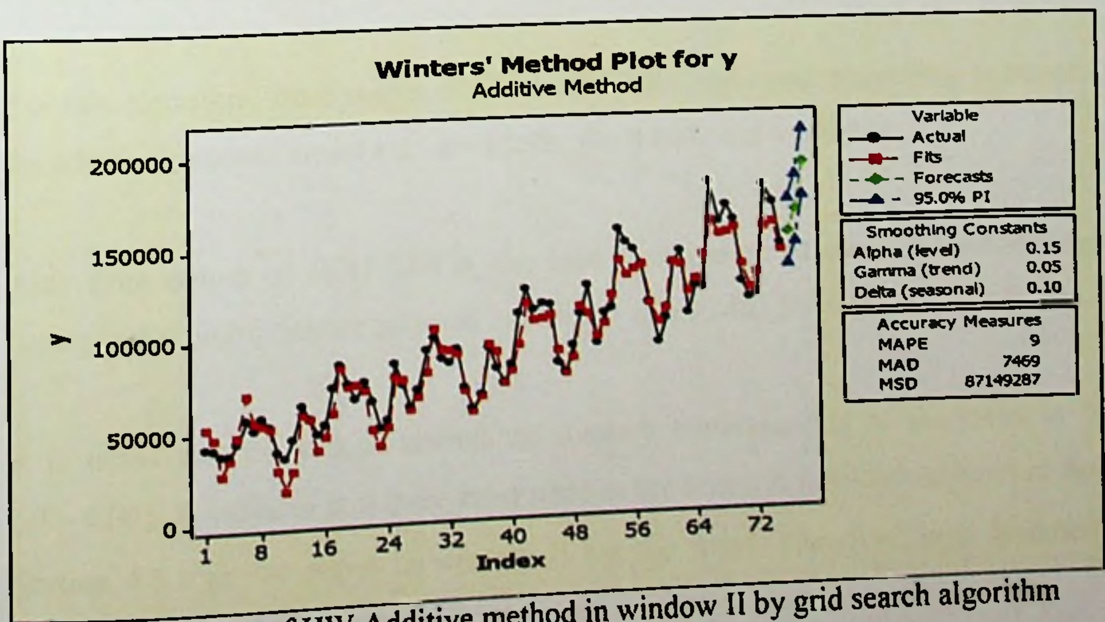


Figure 6.1: Plot of HW Additive method in window II by grid search algorithm

Figure 6.1 and Figure 6.2 represent the Holt – Winter's Additive method and Holt – Winter's Multiplicative method plots, respectively, for the series in window II. Further the smoothing constants are estimated by using grid search algorithm

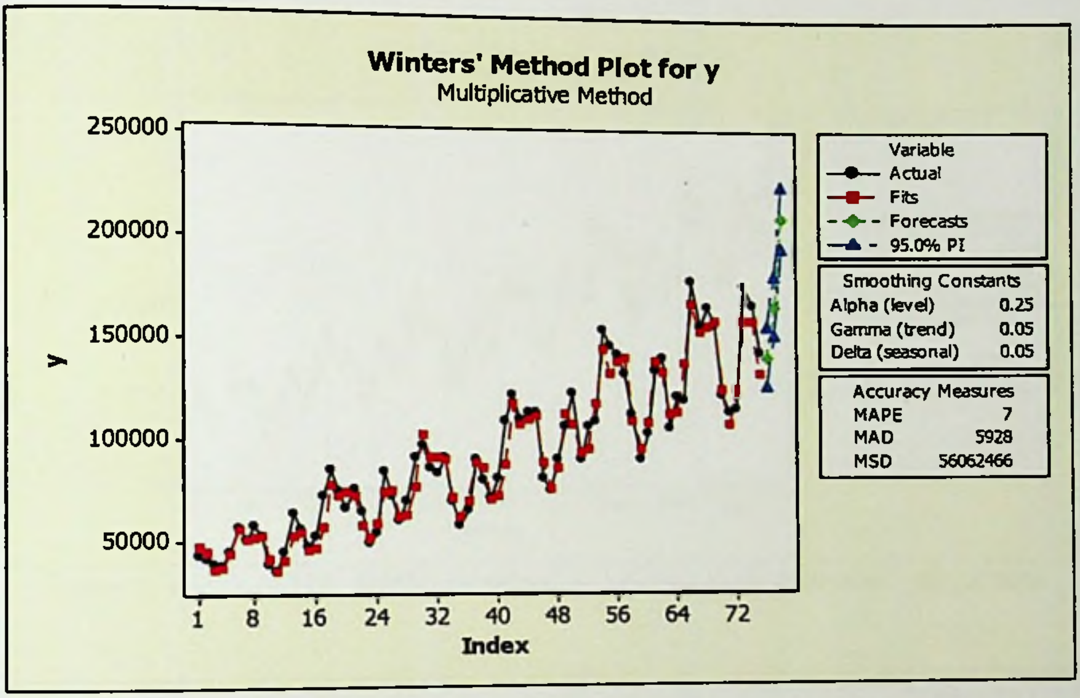


Figure 6.2: Plot of HW Multiplicative method in window II by grid search algorithm

6.1.2 Auto Search Algorithm in window II

For this algorithm, from output of *STATISTICA* the estimated smoothing constants for additive seasonal model are: $\alpha = 0.292$, $\beta = 0.060$, and $\gamma = 0.058$

Also from output of *STATISTICA*, the best combination of estimated smoothing constants for multiplicative seasonal model are: $\alpha = 0.246$, $\beta = 0.00$, and $\gamma = 0.75$.

It is noted that the second smoothing constant corresponding to the trend is 0 ($\beta = 0.00$). It indicates that there is no trend in the series. It is already proved in the Section 4.5 that, the series in window II has the trend. Therefore, auto searched multiplicative model for window II is rejected for further analysis.

Figure 6.3 and Figure 6.4 represent the Holt – Winter's Additive and Holt – Winter's Multiplicative method plots, respectively, for the series in window II. Further the smoothing constants are estimated by using auto search algorithm.

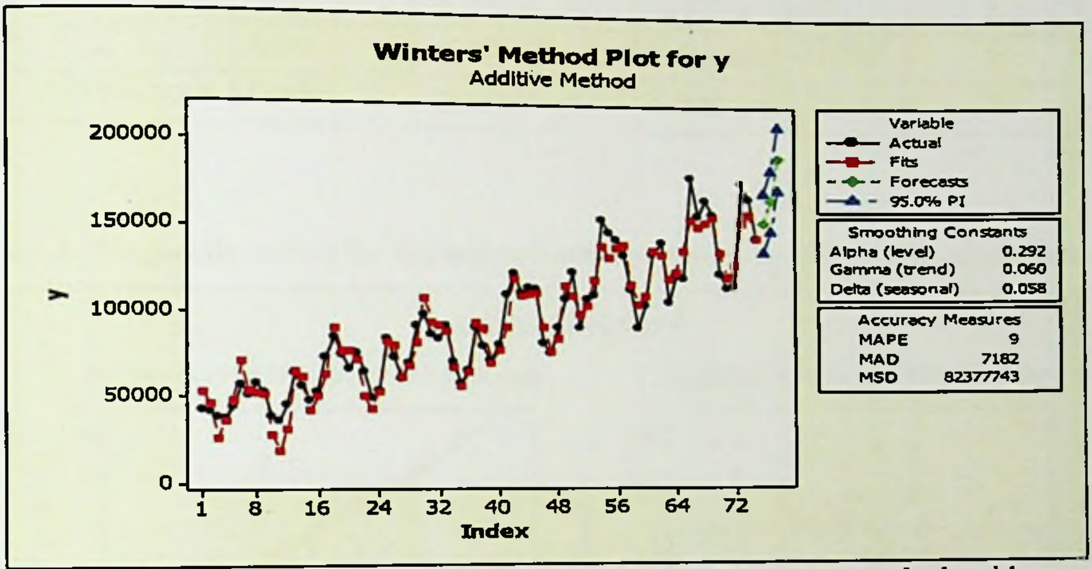


Figure 6.3: Plot of HW Additive method in window II by auto search algorithm

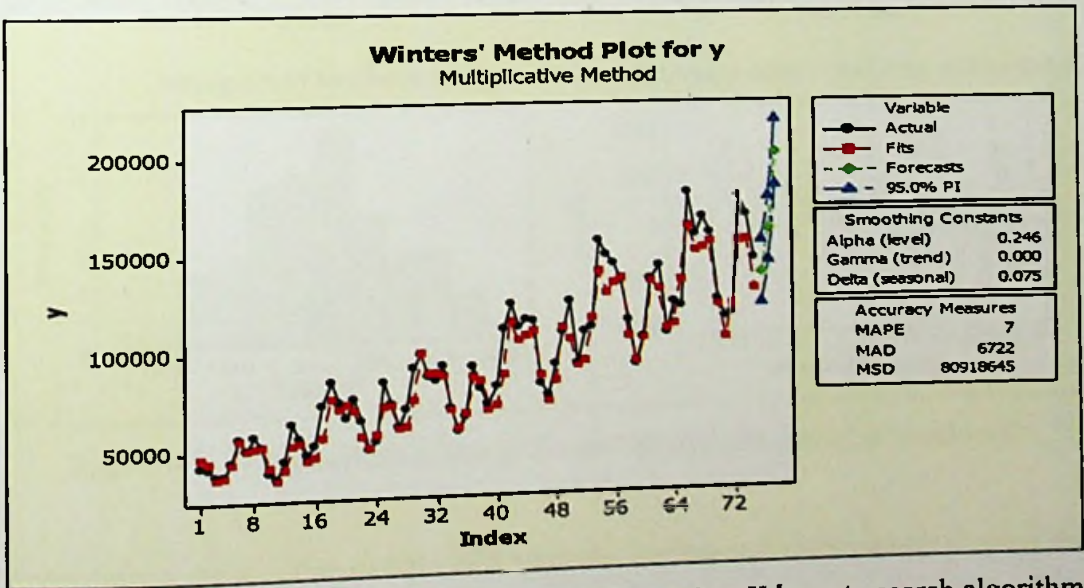


Figure 6.4: Plot of HW Multiplicative method in window II by auto search algorithm

Table 6.1 below summaries the three different models obtained from exponential smoothing technique. Also based on the MAPE value of each model (all are less than 10), all three models are considered for further analysis such as for diagnostic testing and ex-ante forecasting.

Table 6.1: Summary of models by Exponential Smoothing method in window II

Model Type	Parameter Search Algorithm	Parameters Estimates	Accuracy Measure (MAPE)
Additive	Grid	$\alpha=0.15$, $\beta=0.05$, and $\gamma=0.10$	9
	Auto	$\alpha=0.292$, $\beta=0.060$, and $\gamma=0.058$	9
Multiplicative	Grid	$\alpha=0.25$, $\beta=0.05$, and $\gamma=0.05$	7

6.1.3 Diagnostic testing for the selected additive model by grid search algorithm

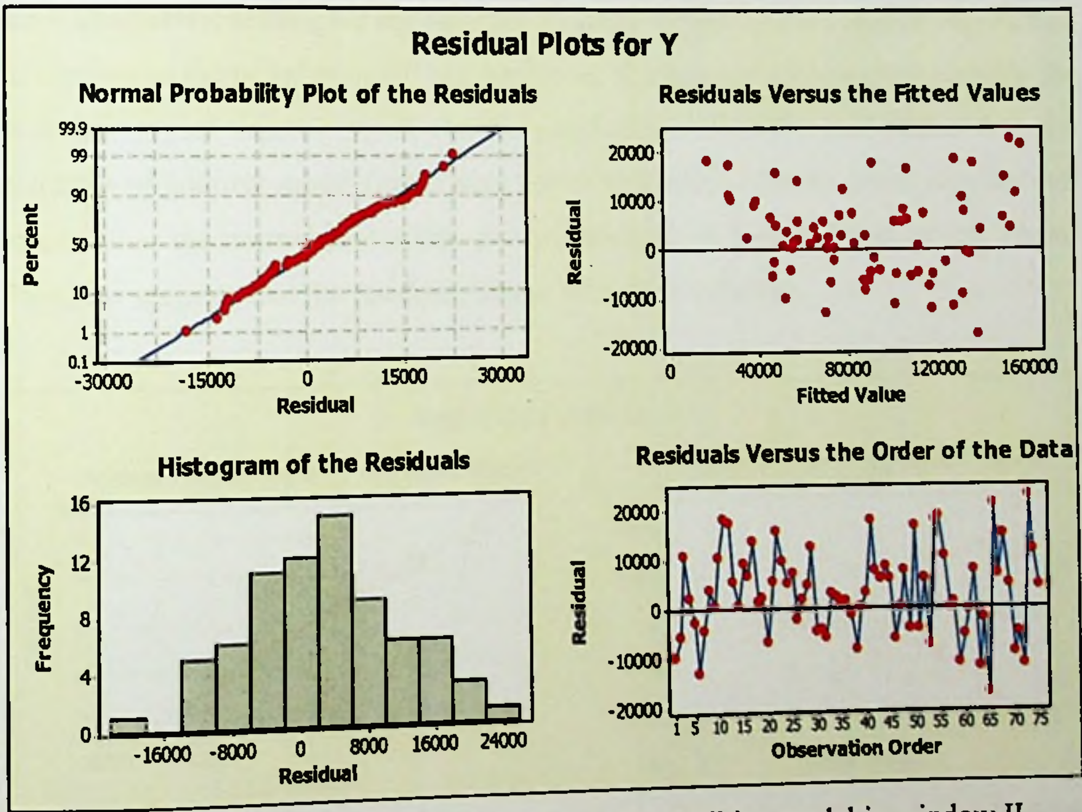


Figure 6.5: Residual plots of the grid searched additive model in window II

According to the p- value (0.831) of Anderson- Darling test (obtained by plotting the probability plot separately), it can be concluded with 95% confidence that the residuals of additive model (under grid search algorithm) follow normal distribution. Also it can be seen in Figure 6.5 that, the normal probability plot of residuals is almost linear. This too suggests that the residuals follow normal distribution.

From the plot of residuals versus fitted values in Figure 6.5, it can be seen that the residuals are almost scatted randomly, thus it can be said that the residuals are independently distributed.

In Figure 6.5, the plot of residuals versus order of the observations suggests that the variance of residuals is reasonably constant as it looks like symmetric about 0. Therefore this model satisfies the diagnostic test for residuals. Hence HW additive grid searched model can be considered further for ex-ante forecast.

6.1.4 Diagnostic testing for the selected additive model by auto search algorithm

According to the p- value (0.401) of Anderson- Darling test (obtained by plotting the probability plot separately), it can be concluded with 95% confidence that the residuals of additive model (under auto search algorithm) follow normal distribution. In addition, the normal probability plot of residuals in Figure 6.6 is almost linear. This also suggests that the residuals follow normal distribution.

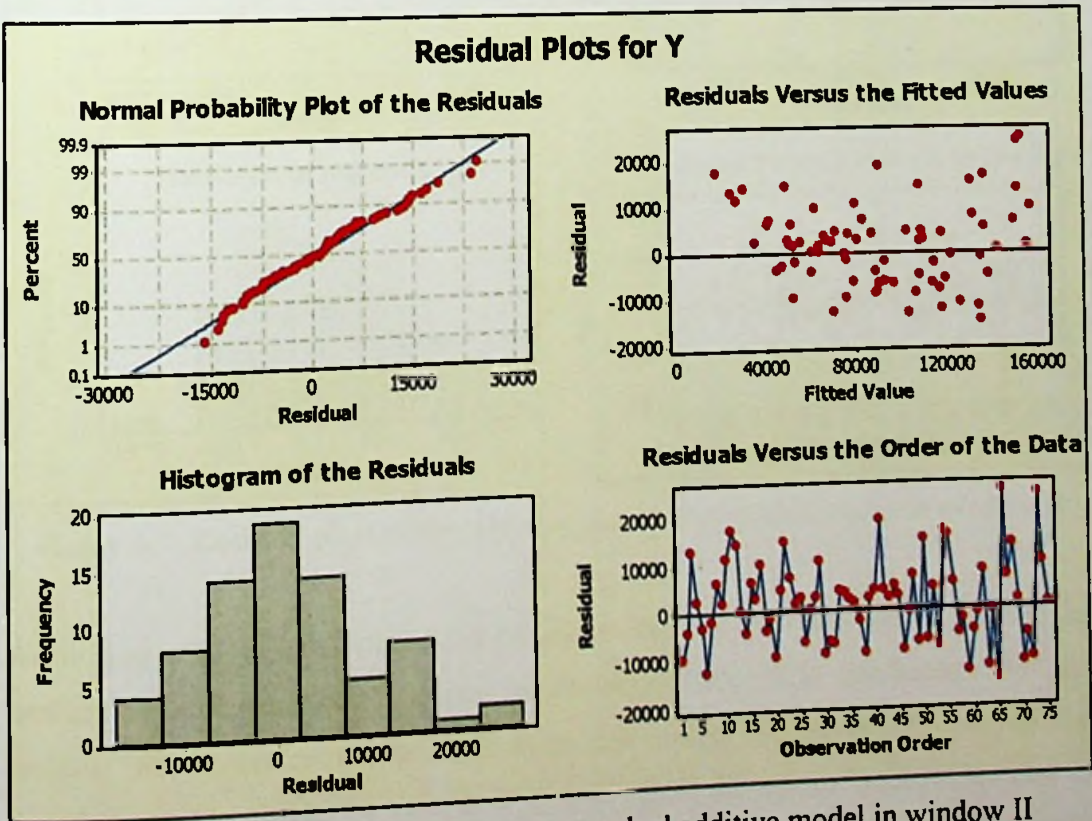


Figure 6.6: Residual plots of the auto searched additive model in window II

From the plot of residuals versus fitted values in Figure 6.6, it can be seen that the residuals are almost scatted randomly, thus it can be said that the residuals are independently distributed. In Figure 6.6 of the plot of residuals versus order of the observations suggests that the variance of residual reasonably constant as it looks like symmetric about 0. Also it does not follow any systematic pattern.

Therefore this model satisfies the diagnostic test for residuals. Hence HW additive auto searched model can be considered further for ex-ante forecast.

6.1.5 Diagnostic checking for the selected multiplicative model by grid search algorithm

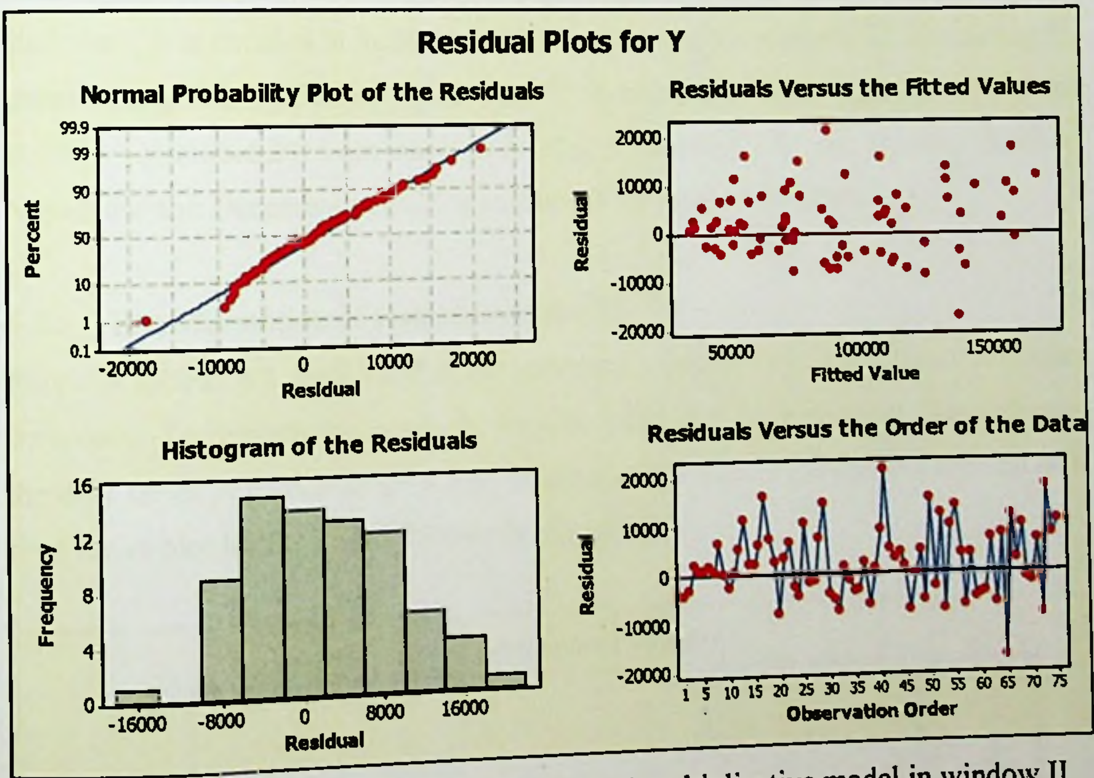


Figure 6.7: Residual plots of the grid searched multiplicative model in window II

According to the p- value (0.518) of Anderson- Darling test (obtained by plotting the probability plot separately), it can be concluded with 95% confidence that the residual of multiplicative model (under grid search algorithm) follow normal distribution. Also it can be seen in Figure 6.7 that, the normal probability plot of residuals is almost linear. This too suggests that the residuals follow normal distribution.

From the plot of residuals versus fitted values in Figure 6.7, it can be seen that the residuals are almost scatted randomly, thus it can be said that the residuals are independently distributed. In Figure 6.7, the plot of residuals versus order of the observations suggests that the variance of residual reasonably constant as it looks like symmetric about 0. Also it does not follow any systematic pattern.

Therefore this model satisfies the diagnostic test for residuals. Hence HW multiplicative grid searched model can be considered further for ex-ante forecast.

6.2 Development of ARIMA model in window II

Since from the previous section 4.5, the original series for window II is non-stationary, it is decided to build an ARIMA model for the purposes of forecasting by transforming into stationary series. Here it is noted that 76 data points from June 2009 to September 2015 are taken to develop the models. The rest 3 points October, November and December of 2015 are reserved for model validation.

6.2.1 Transformation of series in window II

Previous section 4.5 confirms that the series has a positive trend and hence it is non-stationary. To remove the trend, the regular difference is carried out. Accordingly, the new series is named as D1Y (the original series after 1 difference) and the new time series plot for D1Y is as follows in Figure 6.8.

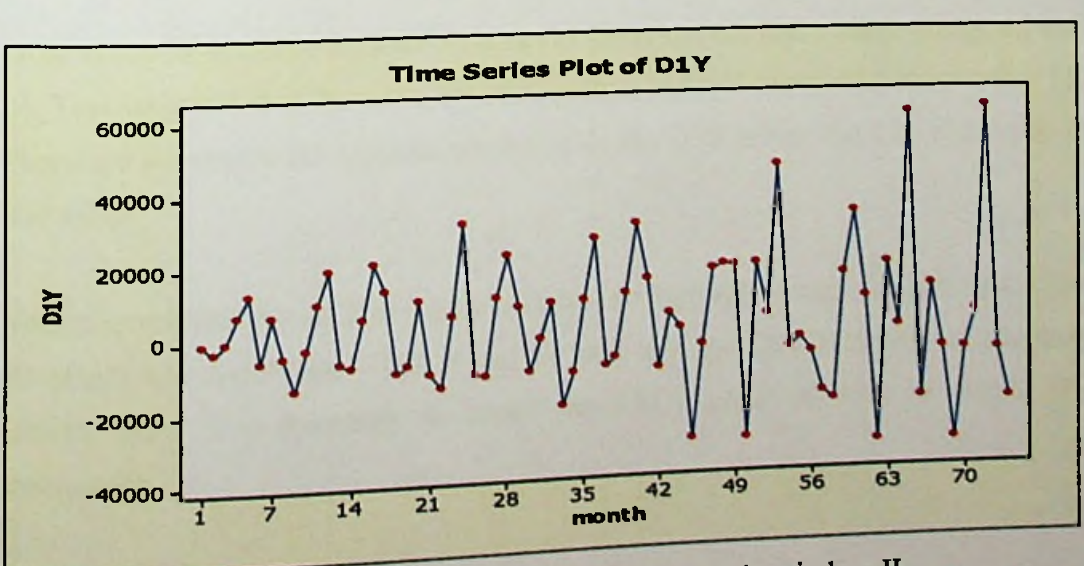


Figure 6.8: Time series plot of D1Y series in window II

From the plot of time series of D1Y in Figure 6.8, it seems visually that there is no trend but it is has to be proved using the statistical test called ADF test. The test statistic value and corresponding p- value of ADF test are -11.75 and 0.00 respectively.

Therefore it can be concluded with 95% confidence that, the D1Y series has no trend. However, no clear information can be obtained regarding the seasonal pattern from Figure 6.8. Thus the ACF graph is checked for this purpose. The ACF graph of series D1Y is given below in the Figure 6.9:

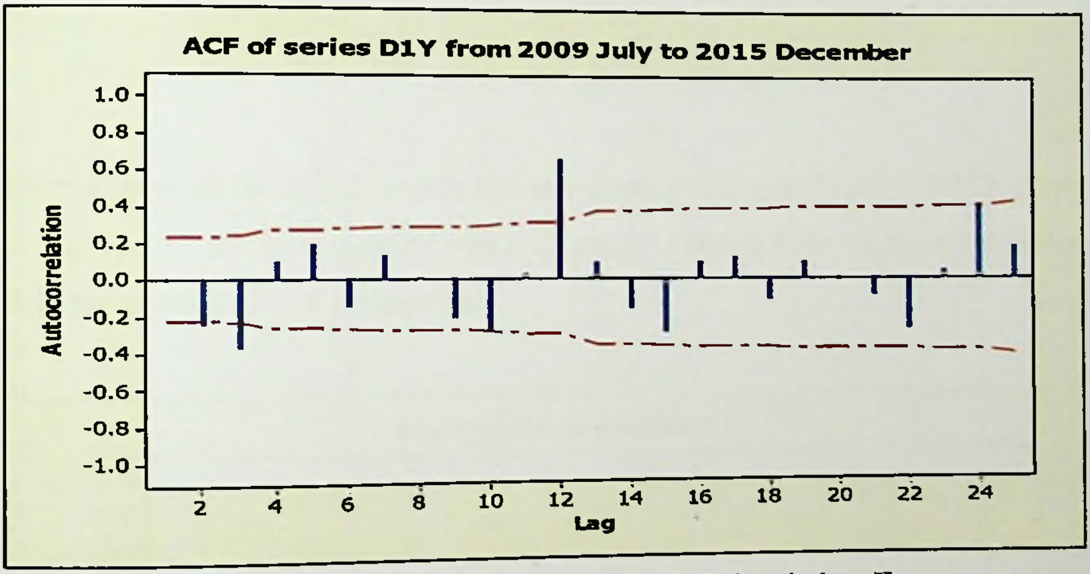


Figure 6.9: ACF graph of the series D1Y in window II

From the ACF graph in Figure 6.9, it can be observed the high spikes at lags 12 and 24. That indicates that there is a seasonal pattern in D1Y series with seasonality 12. Therefore to remove the seasonal pattern from the D1Y series, the 12th difference is also taken.

Now it seems that the ACF in Figure 6.10 has no high spikes after the first spike and all others are significant. Thus it can be said that the D12D1Y has no seasonal pattern. Now it is necessary to check the PACF graph as well to decide the stationarity.

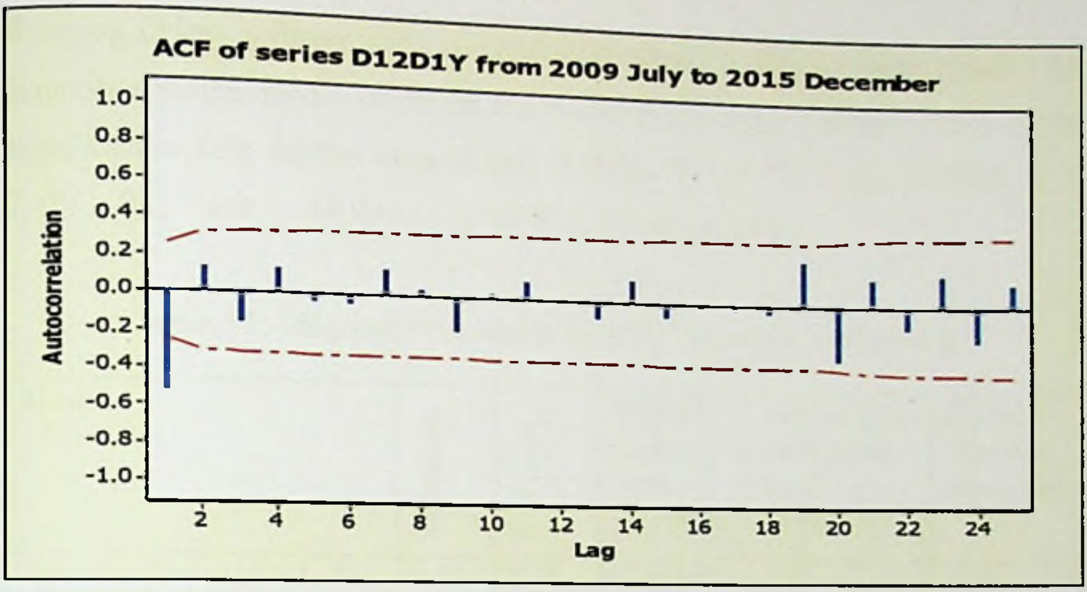


Figure 6.10: ACF graph of D12D1Y series in window II

Since almost all the spikes (except few at beginning) are significant of PACF graph in Figure 6.11 and by comparing with ACF graph in Figure 6.10, it can be confirmed that, the series D12D1Y is stationary.

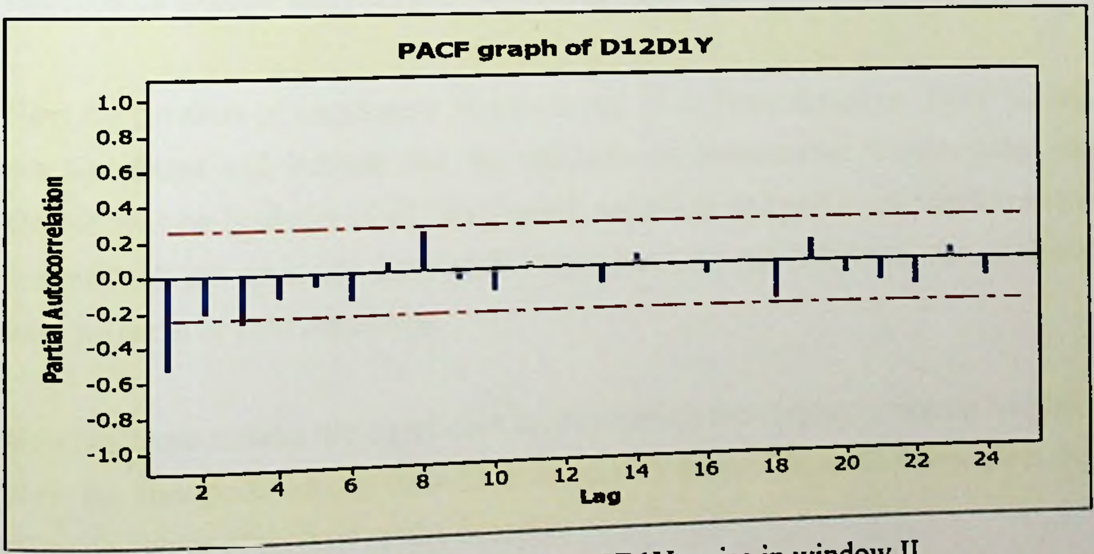


Figure 6.11: PACF graph of D12D1Y series in window II

Further form ACF and PACF graphs in Figure 6.10 and Figure 6.11 respectively, it can be suggested that the seasonal ARIMA model may have the terms MA (1), AR(1) with their seasonal AR and seasonal MA terms. Moreover, almost all the combination of these terms are tried out to check the significance of parameters.

By using *Eviews* software, the parameter estimation and other required statistics for diagnostic testing, model validation and model selection are obtained. From all the significant models, the best three models: ARIMA (0, 1, 1) $(0, 1, 0)_{12}$, ARIMA (1, 1, 0) $(0, 1, 0)_{12}$ and ARIMA (1, 1, 1) $(0, 1, 0)_{12}$ are selected.

Table 6.2: Diagnostic test results of ARIMA models in window II

Model	Skewness	Kurtosis	White's General Test (p-value)	Lagrange's Multiplier Test (p-value)	Durbin Watson Statistics
ARIMA (0, 1, 1) $(0, 1, 0)_{12}$	0.40	3.38	0.99	0.62	1.93
ARIMA (1, 1, 0) $(0, 1, 0)_{12}$	0.21	2.83	0.12	0.64	2.22
ARIMA (1, 1, 1) $(0, 1, 0)_{12}$	0.23	3.36	0.30	0.52	2.04

Since Skewness and Kurtosis values in all three models in Table 6.2 are closer to 0 and 3 respectively, it can be assumed that the residuals in all three models follow the normal distribution. In addition, the p-values of White's General test are not significance indicate that the variance of residuals are constants in all three models.

Next the p-values of Lagrange's Multiplier test of all three models in Table 6.2 are not significant and indicate that the residuals are independent. Furthermore, the Durbin Watson Statistics of all three models are closer to 2 and it also confirms that the residuals are randomly distributed. Hence it can be concluded that, the residuals have no serial or auto correlation.

Now all three models are significant as they satisfy the diagnostic testing as well. Now the best model among these three has to be selected. Table 6.3 below gives the criterion values of the best model selection.

Table 6.3: Information criterions of ARIMA models in window II

Model	R^2 Value	AIC	SBC
ARIMA (0, 1, 1) $(0, 1, 0)_{12}$	0.40	21.06	21.09
ARIMA (1, 1, 0) $(0, 1, 0)_{12}$	0.28	21.25	21.28
ARIMA (1, 1, 1) $(0, 1, 0)_{12}$	0.38	21.19	21.29

From the statistics values appeared in Table 6.3, coefficient of determination R^2 value is high in first model as well as it has the lowest values of AIC and SBC. Therefore the first model ARIMA (0, 1, 1) $(0, 1, 0)_{12}$, can be selected as the best model for the window II.

6.3 Model development using Dynamic Transfer Function method in window II

Since Dynamic Transfer Function (DTF) for a univariate time series can be written as $Y_t = \dfrac{\theta(B)}{\phi(B)} e_t$ where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$ are polynomials in B (backshift operator) of degree p and q respectively, and $\{e_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance σ_e^2 it can be considered as equivalent to a ARIMA model.

To this case, even non-stationary series can be used to fit DTF model. Also it is noted that, as per mentioned in the methodology chapter, any Seasonal ARIMA model can be converted to DTF model. It is already shown from the previous section 6.2.1 that, D12D1Y as the stationary series for window II. Therefore the same series is employed to fit DTF model as well.

6.3.1 Parameter Estimation of DTF model in window II

To estimate the parameters of DTF in *SAS* programme, several models are tried out by assigning the values for p and q. Ultimately, may be due to equivalence property, the same model ARIMA (1, 1, 0) $(0, 1, 1)_{12}$ which obtained in the previous section 6.2.1 is obtained here as well with different parameter estimates.

According to the *SAS* output, the parameter $\phi_1 = 0.81731$ and its p- value is less than 0.0001 which indicates that the MA (1) parameter is significant. Therefore it can be concluded with 95% confidence that the parameter of the model is significant. Hence Moving Average (MA) Factor 1: $1+0.81731*B$

The goodness of fitted model are: AIC=1308.214 and SBC=1312.468

6.3.2 Diagnostic checking for the residuals of fitted DTF model in window II

The p-value (0.820) of the Anderson Darling test in Figure 6.12 suggests that the residuals follow normal distribution. Thus it can be concluded with 95% confidence that the residuals are normally distributed.

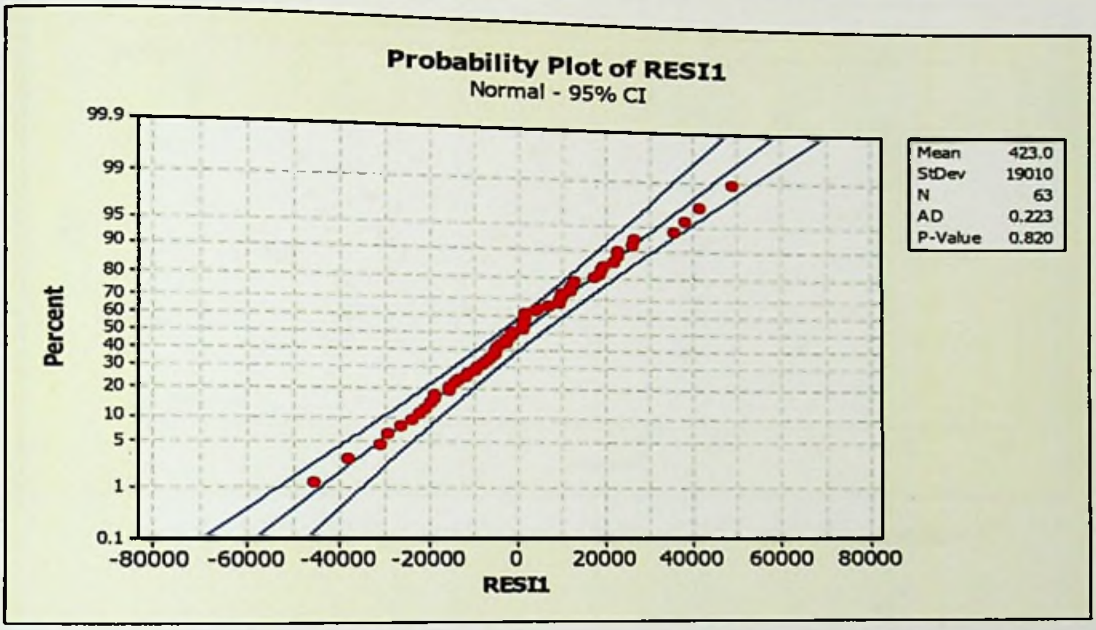


Figure 6.12: Normal probability plot of residuals of DTF model in window II

From the plot of residuals versus predicted values in Figure 6.13, it can be seen that the residuals are distributed randomly. Thus it can be said that the residuals are independently distributed.

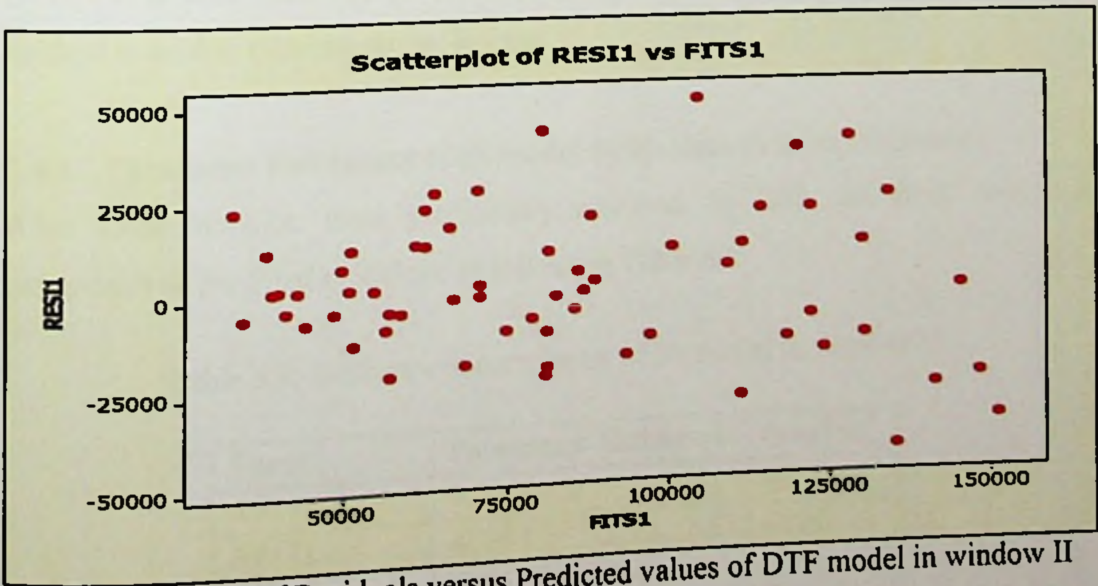


Figure 6.13: Plot of Residuals versus Predicted values of DTF model in window II

In addition, the plot of residuals versus observations order in Figure 6.14 shows that it does not follow any systematic pattern and symmetric about 0. Thus it can be claimed that the variance of the residuals is constant throughout.

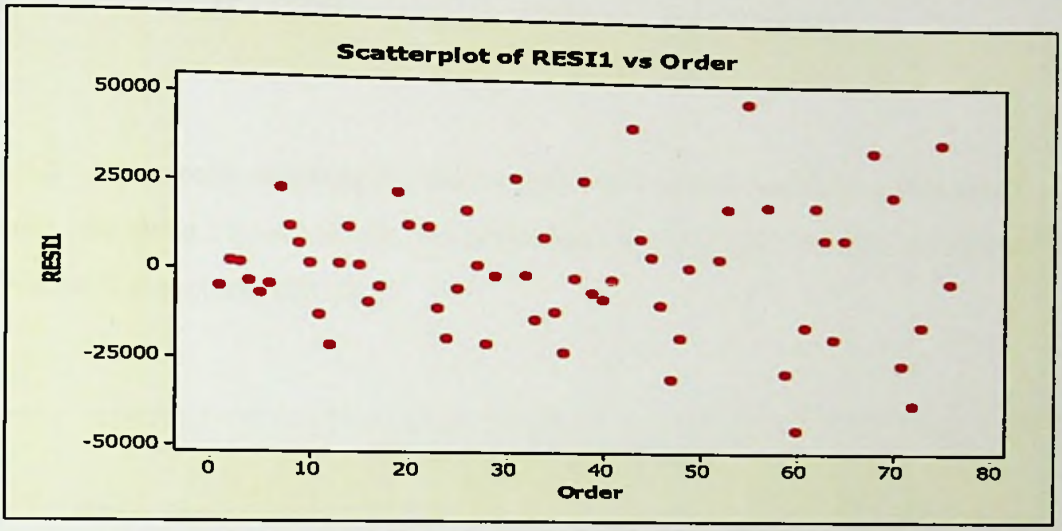


Figure 6.14: Plot of Residuals versus Order of DTF model in window II

Therefore it can be concluded that the residuals of the fitted DTF model satisfy the diagnostic testing.

6.4 Model development using State Space method in window II

The same stationary series obtained after transformation in the section 6.2.1, D12D1Y is used here too to fit State Space (SS) model. Further, Kalman filter method is used to estimate the parameters.

6.4.1 Parameter Estimation of SS model by Kalman filter in window II

After some iteration from preliminary estimates in *SAS*, the final estimated parameters of the fitted model are as follows in Table 6.4:

Table 6.4: Estimates of parameters of SS model in window II

Term	Parameter Estimates	t- value
AR(1)	0.05	2.27
AR(2)	-0.06	-3.24
MA(1)	-0.75	-5.95

From the t-value in Table 6.4, it is observed by comparing with t-calculated value 1.96 (number of observations is more than 60 hence normal distribution value at 5% level of significant is considered), all three parameters are significant. Thus it can be concluded with 95% confidence that the parameters are significant.

6.4.2 Diagnostic checking for the residuals of fitted SS model in window II

From the three Figures below, the diagnostic checking are tested as mentioned in Durbin & Koopman, (2012).

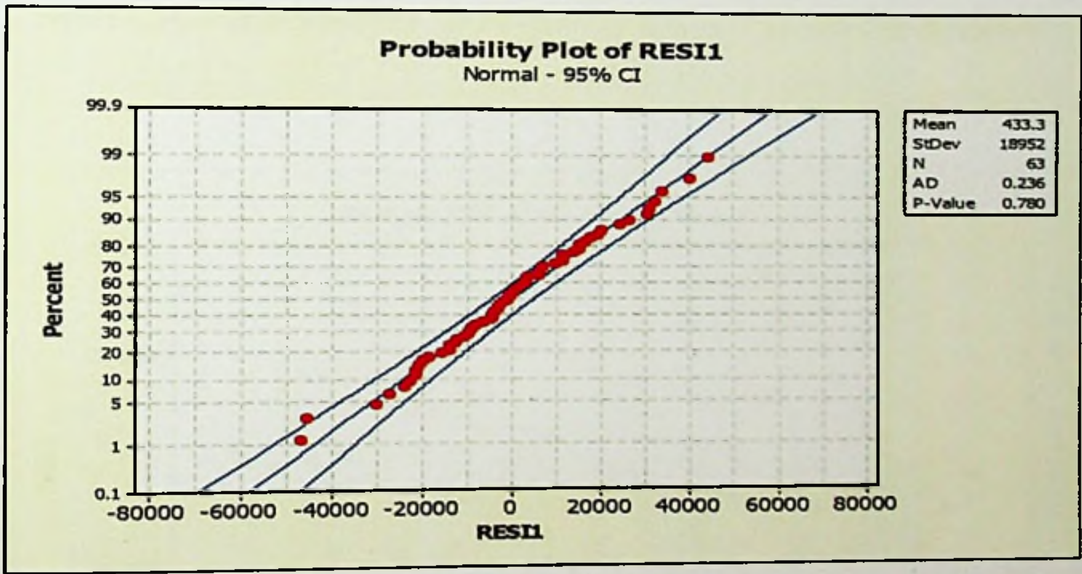


Figure 6.15: Normal probability plot of residuals of SS model in window II

The p-value (0.780) of Anderson Darling test in Figure 6.15 suggests that the residuals follow a normal distribution. Further it can be seen from the plots of residuals versus predicted in Figure 6.16, and residual versus order of observations in Figure 6.17, that the residuals are randomly scatted and its variance is constant respectively.

Therefore it can be concluded that, the fitted model for the window II obtained through State Space method satisfies the diagnostic test.

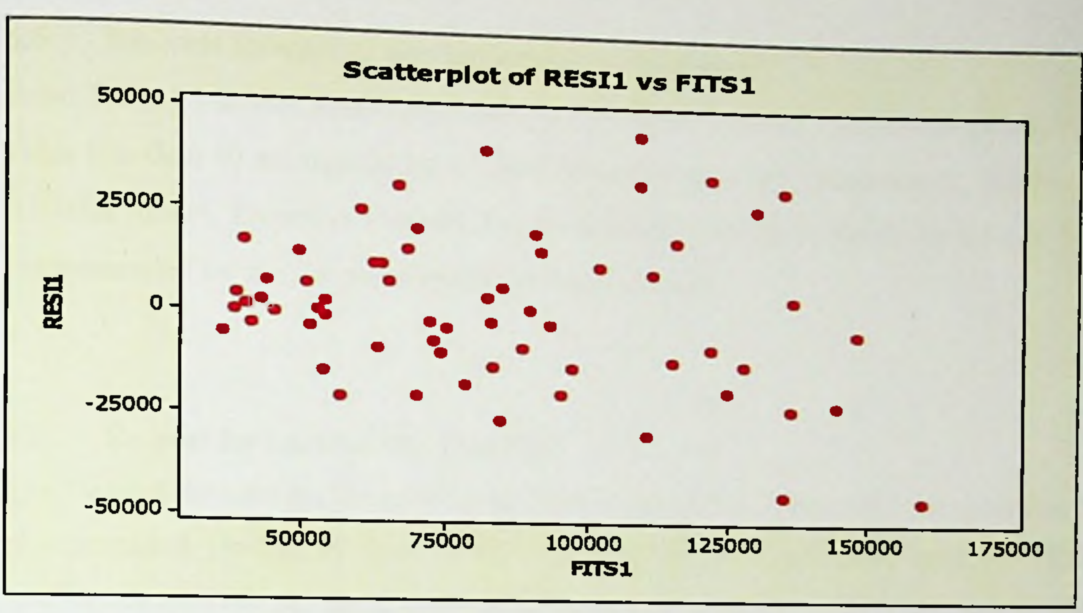


Figure 6.16: Plot of Residuals versus Predicted values of SS model in window II

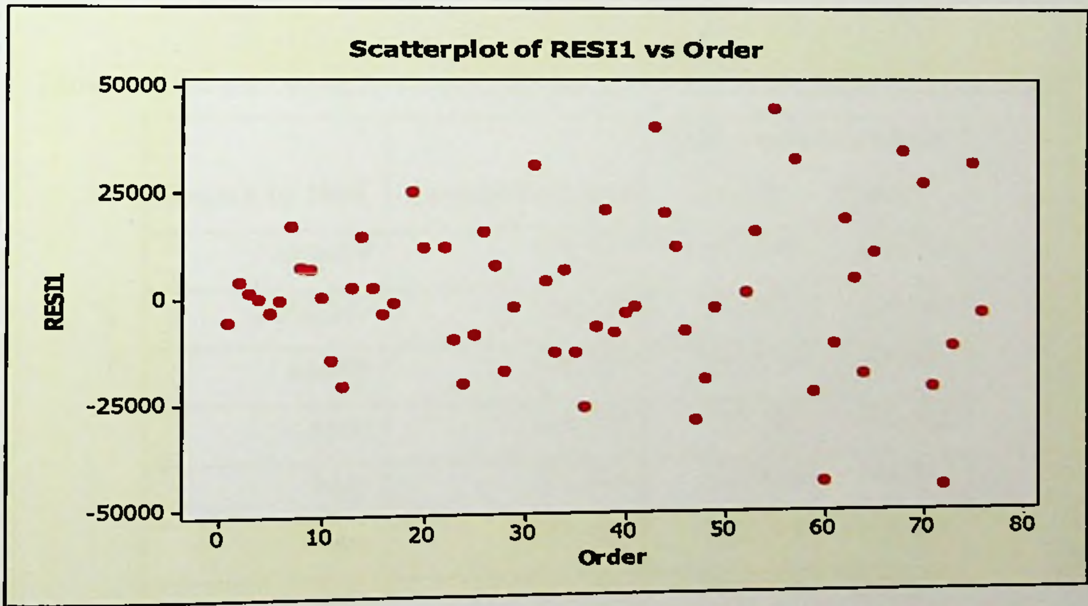


Figure 6.17: Plot of Residuals versus Order of SS model in window II

Ultimately, it can be concluded that three models from Exponential smoothing method, one model from Seasonal ARIMA method, one model from Dynamic Transfer Function method and one model from State Space method satisfy all the required conditions in diagnostic checking. Therefore, all these six models can be considered for Ex-ante forecast.

6.5 Ex- ante forecast of selected models in window II

From Table E1 in APPENDIX E, it can be concluded that, the models whose MAPE value less than 10 are significant in terms of model accuracy. Accordingly, Seasonal ARIMA model, Dynamic Transfer Function model and State Space model can be recommended for the Ex-post forecast of future arrivals.

6.6 Ex-post forecast for the year 2016

The Ex-post forecast for the entire year 2016 is calculated, separately, using all three recommended models in window II. The monthly wise estimates with its 95% confidence intervals are summarised as follows:

Table 6.5 below represents the monthly Ex-post forecast for the year 2016 using the seasonal ARIMA model.

Table 6.5: Ex-post monthly forecast for the year 2016 by Seasonal ARIMA model

Month in 2016	Ex-post Forecast	95% Confidence level	
		Lower	Upper
January	179,927	161,646	196,653
February	189,355	170,821	206,068
March	180,789	162,212	197,697
April	146,014	127,259	162,982
May	137,363	118,454	154,411
June	139,362	120,275	156,466
July	199,758	180,496	216,919
August	190,628	171,187	207,840
September	167,456	147,836	184,718
October	156,427	136,628	173,738
November	168,359	148,383	185,719
December	230,391	210,237	247,798
Total Arrivals	**2,085,829**	**1,855,431**	**2,291,007**

From Table 6.5, it can be concluded that, by the Seasonal ARIMA model, in the year of 2016 approximately 2.085 million international tourist arrivals can be expected to Sri Lanka with (1.855 million, 2.291 million) as 95% confidence interval. It is approximately from 3.17% to 27.39% increase in growth with the year 2015.

Table 6.6 below represents the monthly Ex-post forecast for the year 2016 using the Dynamic Transfer Function model.

Table 6.6: Ex-post monthly forecast for the year 2016 by DTF model

Month in 2016	**Ex-post Forecast**	**95% Confidence level**	
		Lower	**Upper**
January	179,003	161,534	196,471
February	188,298	170,726	205,870
March	179,808	162,133	197,483
April	144,974	127,196	162,751
May	136,286	118,407	154,165
June	138,224	120,243	156,204
July	198,561	180,480	216,642
August	189,367	171,186	207,548
September	166,131	147,850	184,411
October	155,037	136,657	173,416
November	166,904	148,426	185,382
December	228,871	210,295	247,446
Total Arrivals	**2,071,464**	**1,855,133**	**2,287,789**

From Table 6.6, it can be concluded that, by the Dynamic Transfer Function model, in the year of 2016 approximately 2.071 million international tourist arrivals can be expected to Sri Lanka with (1.855 million, 2.287 million) as 95% confidence interval. It is approximately from 3.16% to 27.21% increase in growth with the year 2015.

Table 6.7 represents the monthly Ex-post forecast for the year 2016 using the State Space model.

Table 6.7: Ex-post monthly forecast for the year 2016 by SS model

Month in 2016	Ex-post Forecast	95% Confidence level	
		Lower	Upper
January	178,235	160,679	195,792
February	187,673	170,016	205,330
March	178,349	160,316	196,382
April	143,488	125,405	161,571
May	134,669	116,525	152,813
June	136,595	118,416	154,775
July	196,911	178,697	215,126
August	187,714	169,467	205,961
September	164,475	146,196	182,754
October	153,380	135,070	171,690
November	165,247	146,905	183,588
December	227,214	208,841	245,586
Total Arrivals	**2,053,950**	**1,836,533**	**2,271,368**

From Table 6.7, it can be concluded that, by the State Space model, in the year of 2016 nearly 2.054 million international tourist arrivals can be expected to Sri Lanka with (1.836 million, 2.271 million) as 95% confidence interval. It is approximately from 2.12% to 26.30% increase in growth with the year 2015.

6.7 Synopsis

Time series model development in window II using Holt-Winter's Seasonal Exponential Smoothing, Seasonal ARIMA modelling, Dynamic Transfer Function modelling and State Space modelling methods is discussed in this Chapter in detail.

The followings can be concluded in window II:

- Grid search algorithm and auto search algorithm perform same in both types of models, additive and multiplicative, when Holt-Winter's Seasonal Exponential Smoothing method is employed.

- The grid searched multiplicative, grid searched additive and auto searched additive models are selected as the appropriate models in Holt-Winter's Seasonal Exponential Smoothing method.

- The appropriate Seasonal ARIMA model is

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.83 * e_{t-1} + e_t$$

- The appropriate Dynamic Transfer Function model is

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.82 * e_{t-1} + e_t$$

- The appropriate State Space model is

$$Y_t = 0.05 * Y_{t-1} - 0.06 * Y_{t-2} - 0.75 * e_{t-1} + e_t$$

- It is noted that, all the appropriate models selected in all four method are satisfy the diagnostic checking. Therefore these models are considered for Ex-ante forecasting in further analysis as well as for Ex-post forecast of future arrivals.

- Based on Ex-ante forecast, the significant models are only from the methods of Seasonal ARIMA modelling, Dynamic Transfer Function modelling and State Space modelling. In which Seasonal ARIMA modelling outperform all other methods.

- It is concluded that from Ex-post forecast, the expected international tourists are approximately 2.085 million, 2.071 million and 2.054 million by using Seasonal ARIMA modelling, Dynamic Transfer Function modelling and State Space modelling methods respectively.

7. FACTORS INFLUENCED BY TOURISM IN SRI LANKA

In this chapter, the discussion is on some factors which are influenced by tourism in Sri Lanka. Three factors are mainly taken into account for deep discussion, which are: accommodation, direct employment and foreign exchange earnings by tourism. At the latter part of the chapter, some highlights relevant to Sri Lankan tourism are reported briefly. Further this discussion concerns only from the year 2009 to 2014 where the data are available in window II. Also the data for the year 2015 is excluded as they are not available at the moment.

7.1 Accommodation facilities for tourists

It is clear that the purpose of forecasting is mainly for planning. In that concept, it is necessary to check whether Sri Lanka is capable to cater the international tourists upon their visits in future. If the accommodation capacity is not sufficient, then the country should take the necessary actions to make available the lodging facilities by implementing infrastructure developments.

Table 7.1 below shows the total capacity of rooms available for tourists from the year 2009 to 2014. It contains the room availability (registered at SLTDA) in tourist hotels as well as in supplementary establishments such as Boutique Hotels & Villas, Home Stay Units, Bed & Breakfast (BB) Units, Guest Houses and Heritage Homes.

Table 7.1: Accommodation capacity in Sri Lanka from 2009 to 2014

Year	2009	2010	2011	2012	2013	2014
No. of Rooms in Tourist Hotels	14461	14714	14653	15510	16655	18510
No. of Rooms in Supplementary Establishments	5946	5895	6141	8207	8513	9916
Total Capacity	20407	20609	20794	23717	25168	28426
Growth Rate	1.07%	0.99%	0.90%	14.06%	6.12%	12.94%

From Table 7.1, it is observed that, 3258 new rooms are added to total capacity of accommodation in the year 2014 and which is of 12.94% increase with the year 2013. Also in the year 2013 newly added rooms to the industry is 1451 and it is only 6.12% increase with the year 2012. It seems an increasing trend in both categories of rooms. Though it does not increase in the increasing rate, a positive increase rate can be seen. From the year 2009 to 2010, the rate decreases whereas in 2012 the peak can be observed and then decreases in 2013 and increases in 2014.

However, the registered accommodations are not adequate for tourists. The tourists those who use to stay at unregistered places which sometimes give negative image to the world. Therefore, it is one of the main reasons that to construct more tourists' rooms island wide. As per the estimation done in the "Tourism Development Strategy 2011-2016'', for the year 2016 will be 2.5 million tourists, then the estimated hotel rooms for them were 45,000. Therefore in the year 2015 and 2016, 16574 rooms have to be constructed. With the rate of increase of international tourist in window II (refer Table 4.9), the rate of increase of volume of accommodation should be increase.

7.2 Direct Employment by tourism

It is know that, the tourism creates direct as well as indirect employments in the industries. Promoting tourism definitely is a solution to the unemployment in a country. This section is mainly concerned on direct employment in Sri Lanka due to the tourism.

Table 7.2 below gives only the direct employment created due to tourism in Sri Lanka. The data for the indirect employment and data for the year 2015 are excluded as they are not available at the moment.

Table 7.2: Direct Employment by Sri Lankan tourism

Year	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
Direct Employment	52,071	55,023	57,788	67,882	112,550	129,790

From Table 7.2, it can be clearly seen that the number of direct employment increases every year from 2009 to 2014. The trend of growth rate of direct employment is very clearly shown in Figure 7.1 below:

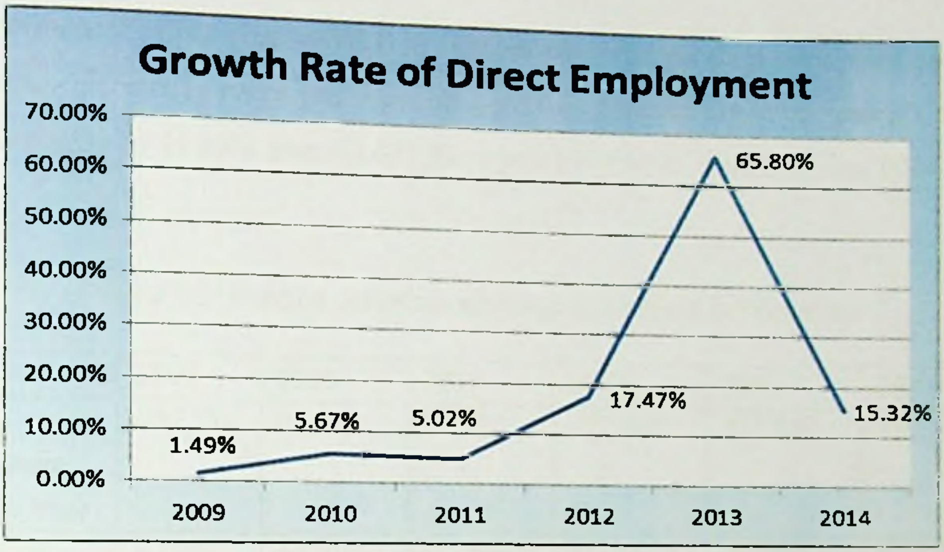


Figure 7.1: Growth Rate of Direct Employment in Sri Lanka

From Figure 7.1, it is noted that in the year 2013, annual growth rate 65.80% is the top in recent years in Sri Lanka. It is assumed that the rate of indirect employment also follows comparatively the same with the direct employment in Sri Lanka.

7.3 Foreign Exchange (FE) Earning through tourism

Foreign exchange earnings are one of the most important incomes to Sri Lanka. According to the annual report of SLTDA, in the year 2014 tourism has increased its rank up to third level, as the largest source of Foreign Exchange Earner of the national economy in 2014, from fourth level in the year 2013.

In Sri Lanka the total receipt by foreign exchange earnings and FE receipt per tourist per day from the year 2009 to 2014 are summarized in Table 7.3.

From Table 7.3 it is observed that, the foreign exchange earnings have dramatically increased over the last six years. Particularly, these receipts increased by 43.57% from Rs. 221,147.1 million in 2013 to Rs. 317,501.7 million in 2014. But in the year

2013, growth rate is 67% which is an increase from Rs 132,427 million to Rs 221,147.1 million in the year 2012 and which is the highest rate in the recent past in Sri Lanka.

Furthermore it can be observed that the foreign exchange earnings per tourist per day gradually increase in every year from 2009 to 2014. Specifically in the year 2014 rate of increase is by 19.83% from Rs 475.35 in year 2013 to Rs 569.60 in year 2014.

Table 7.3: Foreign exchange earnings by tourism in Sri Lanka

Year	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
FE Earnings (in Rs Mn)	40,133	65,018	91,926	132,427	221,147	317,501
Growth Rate	8.19%	62.01%	41.39%	44.06%	67.00%	43.57%
FE Receipt per tourist per day (in Rs)	196.48	272.17	294.23	360.79	475.35	569.60

The below Figure 7.2 provides the trend of foreign exchange earnings per tourist per day from the year 2009 to 2014 in Sri Lanka.

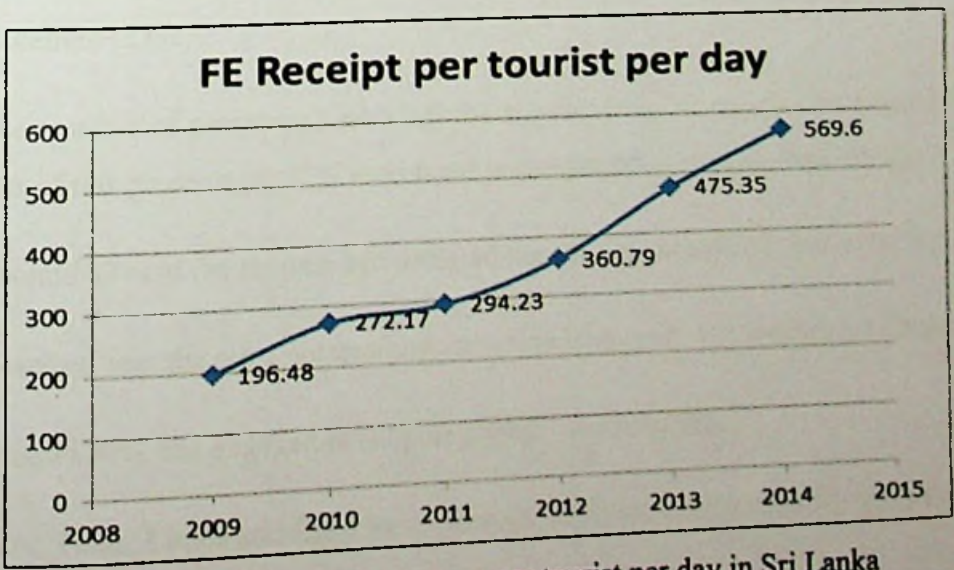


Figure 7.2: Plot of FE earnings per tourist per day in Sri Lanka

As far as the above three factors are concerned, tourism industry in Sri Lanka after the internal conflict is dramatically growing. Based on the data available, the factors described above which are influenced by tourism also show positive growth from the year 2009 to 2014. Further from FE receipts, it can be claimed that its contribution to the national income is increasing every year.

Moreover the records in the annual report of SLTDA reveal that the purpose of most of the international tourists visit is to spend their leisure time in Sri Lanka. This is mainly because of the natural beauty of Sri Lanka.

7.4 Some highlights relevant to Sri Lankan tourism

In this section, some highlights which relevant to tourism in Sri Lanka are reported in points form. These information are extracted from survey findings by the SLTDA. These points are taken to be considered by the relevant officials when the implementation plans for infra structure taking place.

- Nearly 73% of the tourist had used hotels for accommodation whereas 16% stayed in guest houses and rest houses.
- The overall rating of accommodation facilities used was either Good (55%) or Excellent (27%).
- The number of passengers who left the airport while on transit was around 86%. Out of this percentage, 92% used hotel accommodation during their transit.
- Around 87% of the tourists had declared their main purpose of visit to be holiday.
- Beaches' was the most outstanding attraction of tourists for visiting Sri Lanka.
- Around 80% had engaged in shopping whilst in Sri Lanka.
- Cars, Taxis, Limos and Cabs were the popular mode of transport for many tourists (40%).

- A majority of the tourists were either delighted (65%) or satisfied (33%) with their overall visit to Sri Lanka.

- A vast majority of the tourists (73%) had said that they had an intention of visiting Sri Lanka again.

- More than half of the tourists (61%) have plans of visiting Sri Lanka in future.

- Comments given by tourists suggested the need of better roads and transport services for tourists.

- Some tourists were unhappy about the variation of foods and standard of the hotels.

7.5 Synopsis

The behaviour of some factors, which are influenced by tourism in Sri Lanka, is discussed in this Chapter with relevant figures. The factors such as capacity of accommodation, direct employments and foreign exchange by tourism gradually increase in every year. It indicates that, the tourism industry in Sri Lanka growing. At the same time improving this industry definitely develop the country in terms of economy.

Though tourism industry contributes more to the national income and country development, it has some negative impact also such as social problems and environmental problems. Thus, it is necessary to be aware of the negative impact when tourism development plans are taken place in Sri Lanka.

8. CONCLUSIONS

The conclusions to this study are based on the results and discussions in the previous chapters 4, 5 and 6. Also this combines the results from the window I, window II and overall frame from January 1967 to December 2015.

8.1 Proportion of arrivals in separate time frames

The entire time frame of this study is divided as *before the conflict* (January 1967- July 1983), *during the conflict* (August 1983- May 2009) and *after the conflict* (June 2009 to December 2015), the conflict period covers the most of the time frame of this study. Based on the tourist arrivals to Sri Lanka, it can be concluded that 13.15% of total arrivals have been happened before the conflict in 199 months whereas 49.01% of total tourists had come to Sri Lanka during the conflict period in 310 months. Nevertheless, after the conflict 37.84% of total international tourists had visited the island only in 79 months with dramatic increase in numbers.

8.2 The pattern of monthly arrivals

Though the monthly arrivals are taken into separate windows, it can be seen that the pattern is the same in both windows as well as in overall frame. Hence it can be concluded that December as the peak month while January, February and March as mini peak months. July, August and November are at moderate level in tourist arrivals. On the other hand, the lowest numbers of arrivals have been recorded in the months of May and June in every year.

8.3 Model for the period of before the conflict

Table D1 in APPENDIX D summarises the significant models, with their MAPE values, fitted in window I (before the conflict). These models are obtained mainly by employing four methods of time series model fittings. For the purpose of ex-ante forecast the six months period from December 2008 to May 2009 are taken.

According to the model accuracy MAPE value based on the ex-ante forecast, the models come from Seasonal ARIMA and Dynamic Transfer Function can be considered as the best models in the window I, because MAPE values of only those two models are less than 10%. The corresponding models are as follows:

Seasonal ARIMA model for window I

$$Y_t = 1.11 * Y_{t-1} + Y_{t-12} - Y_{t-13} - 0.88 * e_{t-1} + e_t$$

Dynamic Transfer Function model for window I

$$Y_t = 2.91 * Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.01 * e_{t-12} + e_t$$

8.4 Model for the period of after the conflict

Table E1 in APPENDIX E gives the significant models fitted in window II (after the conflict) by employing four methods of time series model fittings. Also it provides the MAPE values for the three months ex-ante forecast values from October 2015 to December 2015.

According to the model accuracy MAPE value, based on the ex-ante forecast, the models come from Seasonal ARIMA, Dynamic Transfer Function and State Space methods can be considered as the best models in the window II, since only their MAPE values are less than 10%. These three models can be used for ex-post forecast of the year 2016. The corresponding fitted models are as follows:

Seasonal ARIMA model for window II

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.83 * e_{t-1} + e_t$$

Dynamic Transfer Function model for window II

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.82 * e_{t-1} + e_t$$

State Space model for window II

$$Y_t = 0.05 * Y_{t-1} - 0.06 * Y_{t-2} - 0.75 * e_{t-1} + e_t$$

8.5 Conclusions on influenced factors by tourism

Some points can be made as conclusions for the factors which are affected by tourism in Sri Lanka.

The room capacity in tourist hotels increases every year. But the registered accommodation facilities are not sufficient at the moment to cater the future arrivals to Sri Lanka. Direct employment by the tourism is also increased. This indicates the unemployment can be reduced in future by certain level if the same trend of arrival goes up. The total foreign exchange (FE) earnings and FE earnings per tourist per day also comparatively has increases dramatically with year. Therefore tourism contributes in large scale to the national revenue in Sri Lanka.

8.6 Prediction for the year 2016

From the Ex-post forecast for the year 2016 by selected best three models in window II, the followings can be concluded:

- From Seasonal ARIMA model, approximately 2.085 million international tourist arrivals can be expected to Sri Lanka with (1.855 million, 2.291 million) as 95% confidence interval. It is approximately from 3.17% to 27.39% increase in growth with the year 2015.

- From Dynamic Transfer model, approximately 2.071 million international tourist arrivals can be expected to Sri Lanka with (1.855 million, 2.287 million) as 95% confidence interval. It is approximately from 3.16% to 27.21% increase in growth with the year 2015.

- From State Space model, nearly 2.054 million international tourist arrivals can be expected to Sri Lanka with (1.836 million, 2.271 million) as 95% confidence interval. It is approximately from 2.12% to 26.30% increase in growth with the year 2015.

Table 8.1: Estimates and actual arrivals till May 2016

Month	January	February	March	April	May	Total
Actual	194,280	197,697	192,841	136,367	125,044	846,229
SARIMA	179,927	189,355	180,789	146,014	137,363	833448
DTF	179,003	188,298	179,808	144,974	136,286	828369
SS	178,235	187,673	178,349	143,488	134,669	822414

It is noted that the actual monthly arrivals in Table 8.1 are with in the 95% confidence intervals of all three models (Refer Tables 6.5, 6.6 and 6.7). Further it can be observed from Table 8.1 that, the total arrivals till May 2016 are very closer to the total of the monthly wise estimates by all three models. Moreover, the total of estimated monthly arrivals of Seasonal ARIMA model is closest among all three models. It also indicates that, Seasonal ARIMA model outperform other models in terms of forecasting accuracy.

8.7 Synopsis

This Chapter summaries the proportion of arrivals in separate time frames, the pattern of arrivals in those time frames and the fitted models in both Windows.

Moreover it provides the Ex-post forecast of recommended models. Hence, it shows that, till May 2016 the estimated tourist arrivals by seasonal ARIMA model is 833,448, by DTF model is 828.369 and by SS model is 822,414. But the actual tourist arrivals from the period January to May in 2016 are 846,229. Thus, the seasonal ARIMA model has the least difference with actual arrivals. Hence, the seasonal ARIMA method outperforms other three methods.

9. RECOMMENDATIONS

Based on the study, here we provide some recommendations which can be made to improve the tourism industry in Sri Lanka.

9.1 Suggestions to improve the tourism industry

Since it can be observed that the tourist arrivals have been dramatically increased in recent past, particularly after the internal conflict in Sri Lanka, it is recommended for more attention on this industry is needed in the country.

The seasonal patterns are very clearly observed over this study. It is recommended that to promote some activities to attract the tourists in off periods as well. Establishing new locations and creating new programmes such as organizing international conferences, sports and games, especially in the months of April, May and June, may be a suggestion to increase the arrivals during the off seasons for tourists who are coming to Sri Lanka.

9.2 Recommendations by the study

It is known that, forecasting is important factor for the planning. Therefore to forecast the tourist arrivals, it is recommended to consider the data after the conflict as there is no disturbance, like internal conflict, to tourism in Sri Lanka other than natural disasters. Hence for this purpose, the models fitted using Seasonal ARIMA, Dynamic Transfer Function and State Space method are highly recommended.

It is estimated that approximately 2.085 million tourist are expected in the year 2016. But it seems that the registered accommodations are not adequate to cater them. Also unregistered accommodations sometimes may cause negative impact for tourists. Therefore it is highly recommended to make adequate the available registered accommodations with high quality of accommodations and services for the future tourists.

The national income has been increased due to the contribution from tourism as an industry and it generates employment opportunities to Sri Lankans workforce. Improving this industry systematically will definitely develop the economy in the country rapidly during the upcoming decades. Thus it is recommended and suggested to enhance the level of infrastructure facilities in the country which will be able to accommodate the future growth in this particular industry, and it is the responsibility of the relevant officials and decision makers in the country.

9.3 Recommendations for further studies

In the model development in the window I, the residuals of almost all the possible models do not satisfy the normality test (particularly the Kurtosis values are not closer to 3). Thus it is better to develop models by removing the extreme values of observations (by removing the outliers in observations).

Since tourism contributes to national revenue and its development, it is better to carry out a causal relation study of international tourist arrivals versus Gross Domestic Product (GDP) and then it would be useful to fit a Dynamic Transfer Function (DTF) model by considering GDP as the dependent variable and tourist arrivals as the independent variable.

9.4 Synopsis

This chapter provides some recommendations based on the study. Also it provides some suggestions without the significant results through this study to improve the industry in Sri Lanka. In addition, few recommendations by this study for further studies are made.

LIST OF REFERENCES

Akuno, A. O., Otieno, M. O., Mwangi, C. W. and Bichanga, L. A. (2015). Statistical models for forecasting tourists' arrival in Kenya. *Open Journal of Statistics*, *5*, 60-65.

Angel, A. J. and Carles S. (2007). Part II – Time series analysis C5 ARIMA (Box-Jenkins) models [ppt]. Retrieved from: http://www.docfoc.com/part-ii-time-series-analysis-c5-arima-box-jenkins-models-angel-a.

Arumugam, P. and Anithakumari, V. (2013). Seasonal time series and transfer function modelling for natural rubber forecasting in India. *International Journal of Computer Trends and Technology (IJCTT)* – *4*(5), 1366-1370.

Athanasopoulos, G. and Hyndman, R. J. (2006). Modelling and forecasting Australian domestic tourism. Preprint submitted to *Tourism Management*, Retrieved from: http://www.robjhyndman.com/papers/austourism.pdf.

Balaguer, J. and Cantavella, J. M. (2002). Tourism as a long-run economic growth factor: the Spanish case, *Applied Economics 34*, 877-884.

Bermudez, J. D., Segura, J. V. and Vercher, E. (2007). Holt- Winters forecasting: an alternative formulation applied to UK air passenger data. *Journal of Applied Statistics, 34*, 1075-1090.

Burger, C. J. S. C., Dohnal, M., Kathrada, M. and Law, R. (2001). A practitioners guide to time-series methods for tourism demand forecasting - a case study of Durban, South Africa. *Tourism Management, 22*, 403-409.

Chang, J. L., Hsueh, F. C. and Tian, S. L. (2011). Forecasting tourism demand using time series, artificial neural networks and multivariate adaptive regression splines: Evidence from Taiwan. *International Journal of Business Administration, 2*(2), 14-24.

Cho, V. (2001). Tourism forecasting and its relationship with leading economic indicators. *Journal of Hospitality and Tourism Research, 24*, 399-420.

Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management, 24*, 323-330.

Chu, F. (1998). Forecasting tourism demand in Asian-Pacific countries. *Annals of Tourism Research, 25*(3), 597- 615.

Crick, M. (1989). The hippy in Sri Lanka: A symbolic analysis of the imagery of school children in Kandy criticism. *Hearsay and Interpretation* 3, 37-54.

Crick, M. and Harrison, D. (1992). Life in the informal sector: street guide in Kandy, Sri Lanka. *Tourism and Less Developed Countries, D. Harison (ed), Belvern Press, London*, 135-147.

Dimitrios, G., Dimitris, P. and Daniel, S. (2012). Forecasting tourist arrivals in Greece and the impact of macroeconomic shocks from the countries of tourists' origin. *Annals of Tourism Research, 39* (2), 641 – 666.

Dominique, M. H., Leonard, J. P. and Randall, L. S. (2002). *Market response models- Econometric and Time Series Analysis* (2nd Edition). Dordrecht: Kluwer academic publishers, 286.

Douglas, C. M., Cheryl, L. J. and Murat, K. (2015). *Introduction to time series analysis and forecasting* (2nd Edition). Hoboken, New Jersey: John Wiley & Sons, Inc., 357.

Dritsakis, N. (2004). Tourism as a long-run economic growth factor: an empirical investigation for Greece using causality analysis. *Tourism Economics, 10*, 305-316.

Durbarry, R. (2004). Tourism and economic growth: the case of Mauritius. *Tourism Economics* 10, 389-401.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods* (2nd Edition). Oxford statistical science series, Oxford University press.

Gamage, A., Shaw, R. and Ihalanayake, R. (1997). Cost of political upheavals to tourism in Sri Lanka. *Asia Pacific Journal of Tourism Research 2*, 75-87.

Gamage, A. and King, B. (1999). Comparing migrant and non-migrant tourism impacts. *International Journal of Social Economics 26*, 312-324.

Gardner, J. E. S. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting, 4*, 1-28.

George, E. P. B., Gwilym M. J. and Gregory C. R. (2008). *Time series analysis-forecasting and control* (4th Edition). Hoboken, New Jersey: John Wiley & Sons, Inc. Publication, 449.

King, B. E. M. and Gamage, A. (1994). Maximizing the value of the ethnic connection: Expatriate travellers from Australia to Sri Lanka, *Journal of Travel Research 33*, 46-52.

Kurukulasooriya, N. and Lelwala, E. I. (2011). Short-term forecasting of international tourist arrivals to Sri Lanka: Comparative evidence using exponential smoothers. *The 6th International Research Conference on Management and Finance 2011*. Retrieved from: http://ssrn.com/abstract=2018372.

Kurukulasooriya, N. and Lelwala, E. (2014). Time series behaviour of burgeoning international tourist arrivals in Sri Lanka: The post- war experience. *Ruhuna Journal of Management and Finance, 1*(1), 1-14.

Law, R. (2000). Back-propagation learning in improving the accuracy of neural network based tourism demand forecasting. *Tourism Management, 21*, 331-340.

Lim, C. and McAleer, M. (2001). Forecasting tourism arrivals. *Annals of Tourism Research, 28*, 965–977.

Loganathan, N. and Yahaya, I. (2010). Forecasting international tourism demand in Malaysia using Box Jenkins SARIMA application. *South Asian Journal of Tourism and Heritage*, 3(2).

Ministry of Economic Development in Sri Lanka (2011). *Tourism Development Strategy 2011-2016.* Retrieved from: http://www.sltda.lk/sites/default/files/English.pdf

Prasert, C., Chukiat, C. and Ratchanee, M. (2008). Time series models for forecasting international visitor arrivals to Thailand. *International Conference on Applied Economics – ICOAE*, 159-163.

Robert, A. Y. and Monnie, M. (1999). *Introduction to time series analysis and forecasting with applications of SAS and SPSS*. Brooklyn, New York: Academic press, inc, 70.

Robert, K. (2007). *State space models and the Kalman filter*. Seminar paper prepared for 40461 Vektorautoregressive Methoden. Retrieved from: http://homepage.univie.ac.at/robert.kunst/statespace.pdf.

Saayman, A. and Saayman, M. (2010). Forecasting tourist arrivals in South Africa. *Acta Commercii 2010*, 281-293.

Song, H. and Lib, G. (2008). Tourism demand modelling and forecasting-A review of recent research. *Tourism Management, 23*, 203-220.

Søren, B. and Murat, K. (2011). *Time series analysis and forecasting by example.* Hoboken, New Jersey: John Wiley & Sons, Inc., 208.

Sri Lanka Tourism Development Authority - SLTDA (2014). *Annual Statistical Report 2014.* Retrieved from: http://www.sltda.lk/sites/default/files/Annual_Statistical_Report-2014.pdf

Tsay, R. S. (2005). Lecture 16: State Space Model and Kalman Filter, Bus 41910, Time Series Analysis [pdf]. Retrieved from: http://faculty.chicagobooth.edu/ruey.tsay/teaching/bs41910/lec16.pdf

Tse, R. Y. C. (2001). Estimating the impact of economic factors on tourism: evidence from Hong Kong. *Tourism Economics,* 7(3), 277-293.

Wanhill, S. R. C. (1994). Evaluating the worth of investment incentives for tourism development. *Journal of Travel Research (Fall)*, 33-39.

Wickremasinghe, G. B. and Ihalanayake, R. (2006). The causal relationship between tourism and economic growth in Sri Lanka: Some empirical evidence, *working paper series, School of Applied Economics.* Retrieved from: https://www.researchgate.net/publication/228587058_The_Causal_Relationship_between_Tourism_and_Economic_Growth_in_Sri_Lanka_Some_Empirical_Evidence

Witt, S. F., Newbould, G. D. and Watkins, A. J. (1992). Forecasting domestic tourism demand: application to Las Vegas arrivals data. *Journal of Travel Research, 31*, 36-41.

APPENDIX A

Selected Seasonal ARIMA models in window I

Table A1: Test results of selected models of series D12D1[SQRT(Y)] in window I

Model	R^2	DW	LM	Skewness	Kurtosis	AIC	SBC	White
ARIMA $(0, 1, 0)(3, 1, 0)_{12}$	28	2.08	0.13	-0.73	11.67	7.81	7.84	0.08
ARIMA $(0, 1, 0)(4, 1, 0)_{12}$	32	2.08	0.17	-0.87	11.97	7.78	7.82	0.66
ARIMA $(0, 1, 0)(5, 1, 0)_{12}$	36	2.06	0.38	-0.94	11.82	7.75	7.8	0.94
ARIMA $(0, 1, 0)(1, 1, 1)_{12}$	**38**	**2.05**	**0.17**	**-0.93**	**10.27**	**7.61**	**7.63**	**0.85**
ARIMA $(4, 1, 0)(1, 1, 1)_{12}$	39	2.08	0.06	-0.95	10.55	7.61	7.64	0.46
ARIMA $(6, 1, 0)(1, 1, 1)_{12}$	39	2.08	0.07	-0.88	10.60	7.61	7.63	0.51

Table A2: Test results of selected models of series D12D1Y in window I

Model	R^2	DW	LM	Skewness	Kurtosis	AIC	SBC	White
ARIMA $(1, 1, 0)(0, 1, 1)_{12}$	**30%**	**2.09**	**0.07**	**-0.64**	**9.46**	**19.37**	**19.36**	**0.14**
ARIMA $(4, 1, 0)(0, 1, 0)_{12}$	29%	2.19	0.32	0.28	9.90	19.68	19.68	0.38
ARIMA $(0, 1, 4)(0, 1, 0)_{12}$	26%	2.19	0.09	0.28	9.88	19.66	19.67	0.17

DW- Durbin Watson Statistic value
LM- Lagrange's Multiplier test p- value
White- White's General test p- value
AIC- Akaike Information Criterion
SBC- Schwartz's Bayesian Criterion

APPENDIX B

Relevant Figures for Seasonal ARIMA model development in Phase I

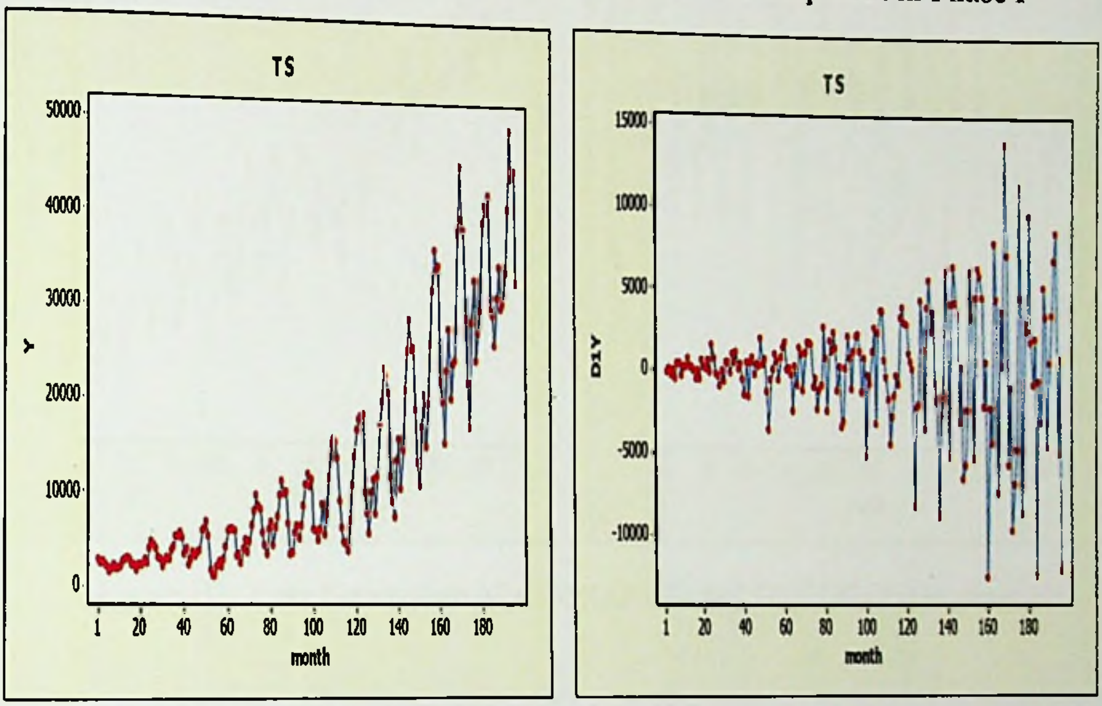


Figure B1: Time Series plots of series Y and D1Y in phase I

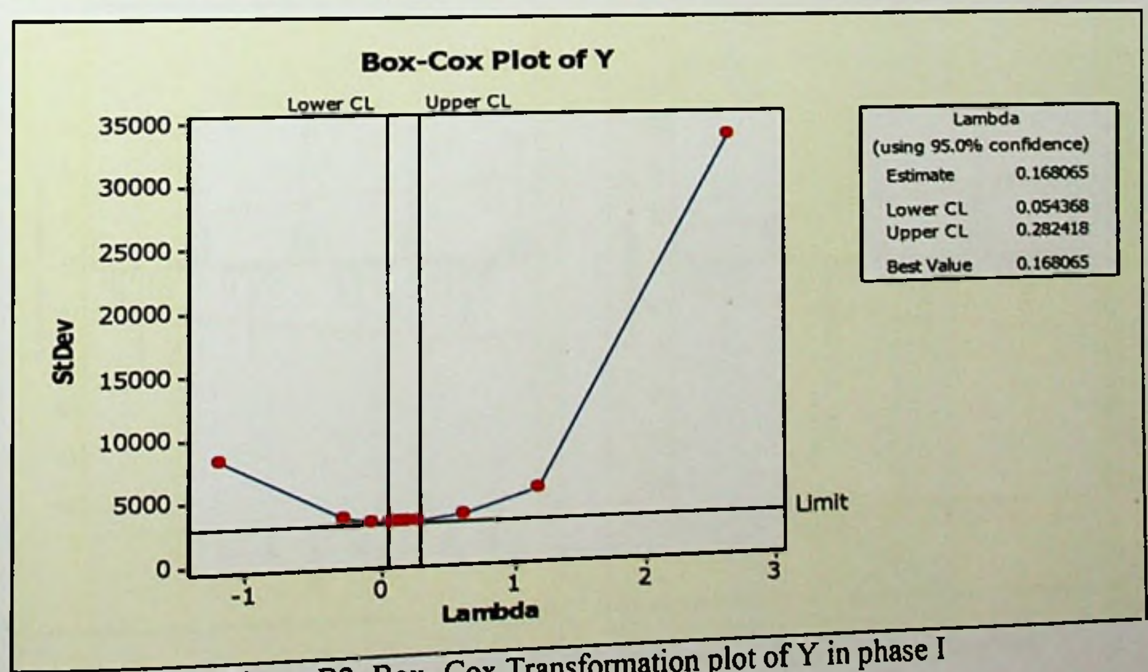
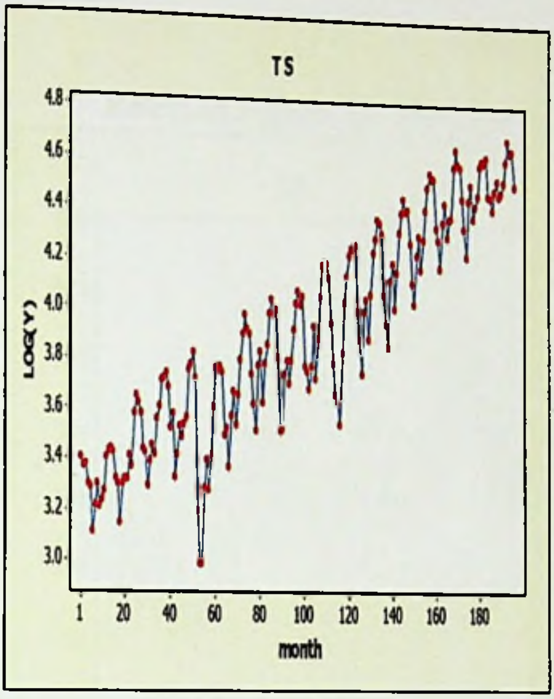


Figure B2: Box- Cox Transformation plot of Y in phase I



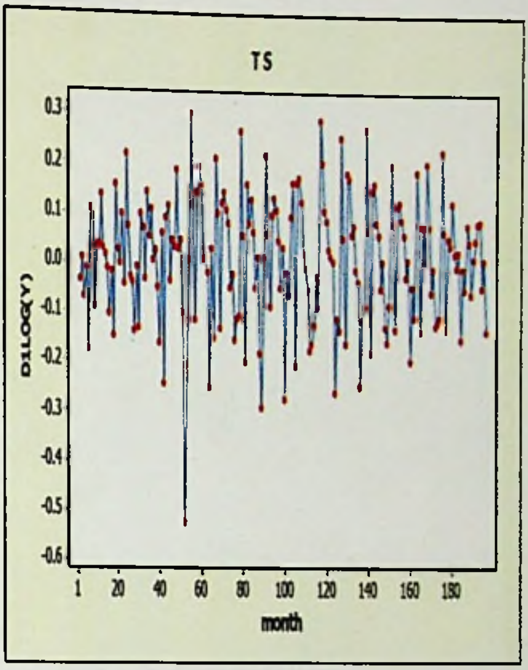
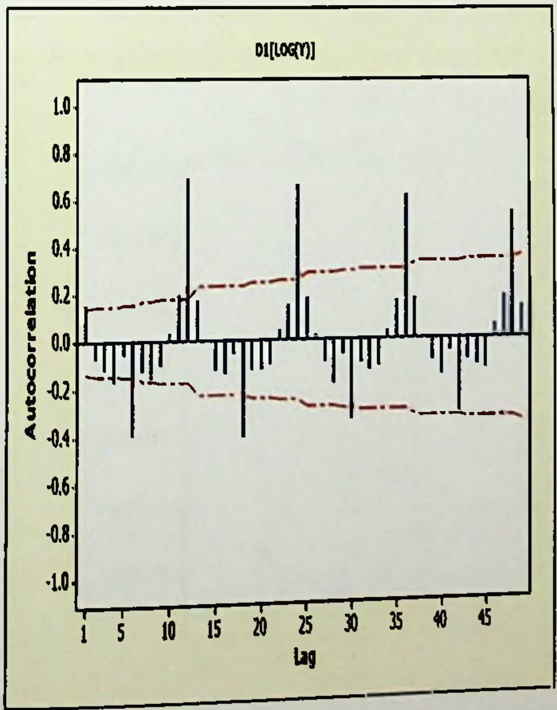


Figure B3: Time Series plots of series LOG(Y) and D1[LOG(Y)] in phase I



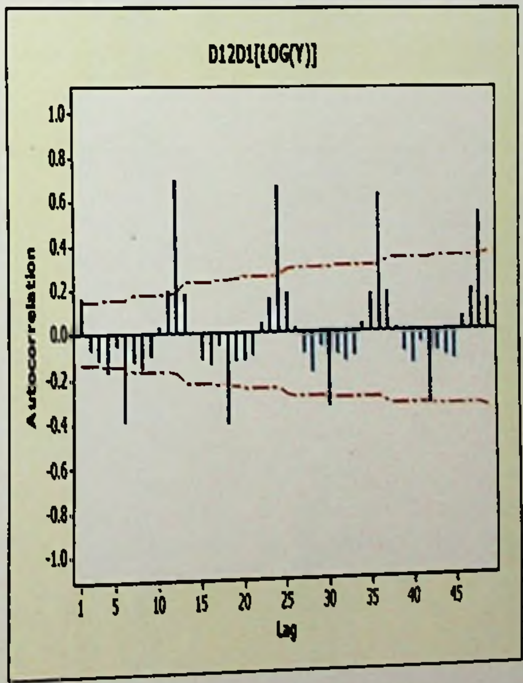


Figure B4: ACF plot of series D1[LOG(Y)] and D12D1[LOG(Y)] in phase I

APPENDIX C

Relevant Figures for Seasonal ARIMA model development in Phase II

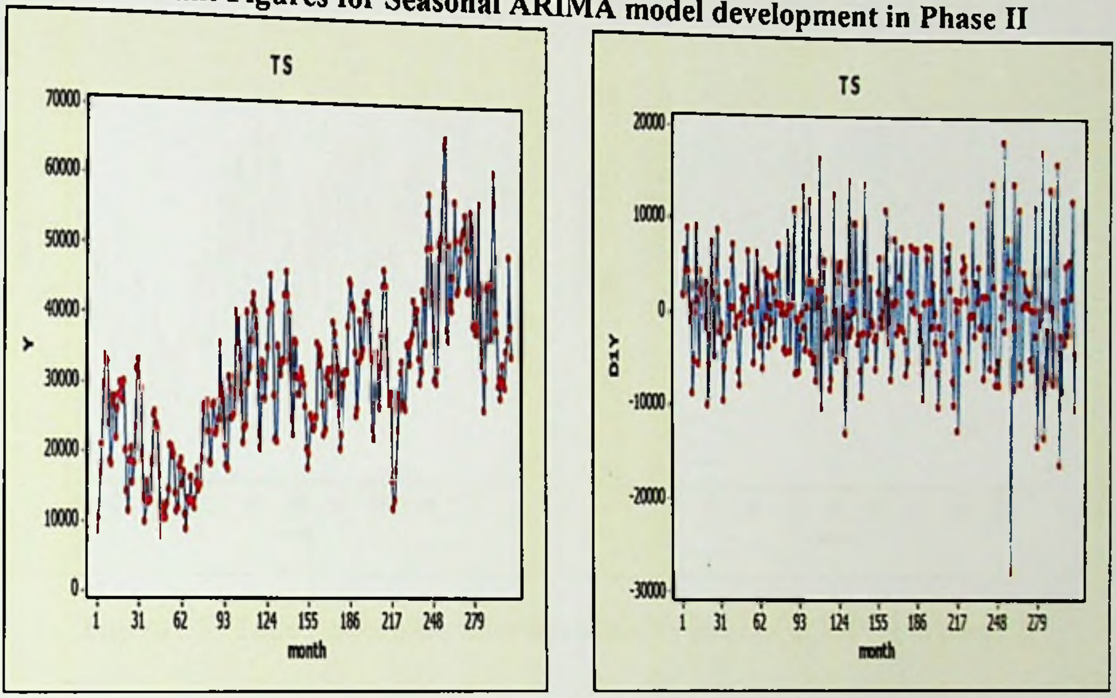


Figure C1: Time Series plots of series Y and D1Y in phase II

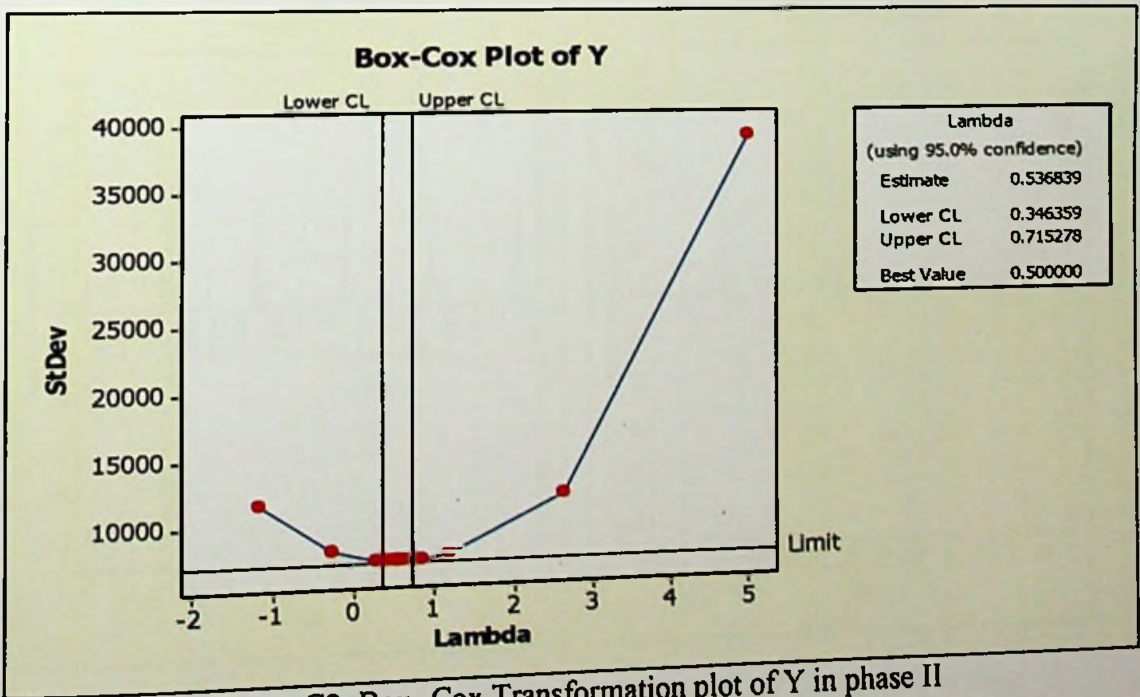
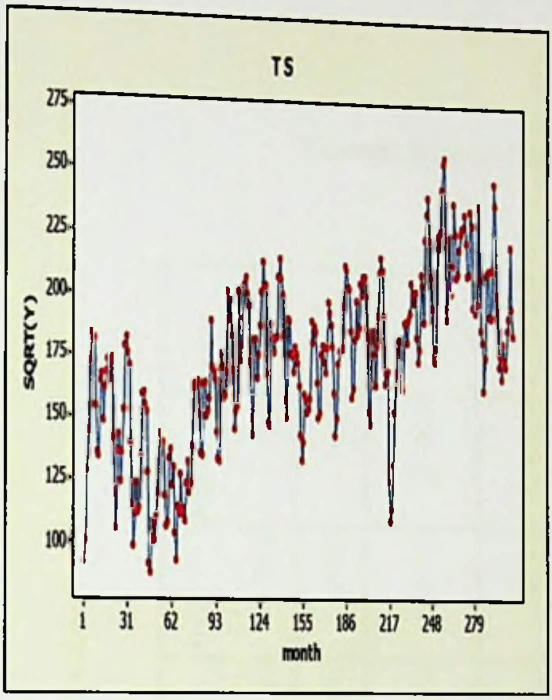


Figure C2: Box- Cox Transformation plot of Y in phase II



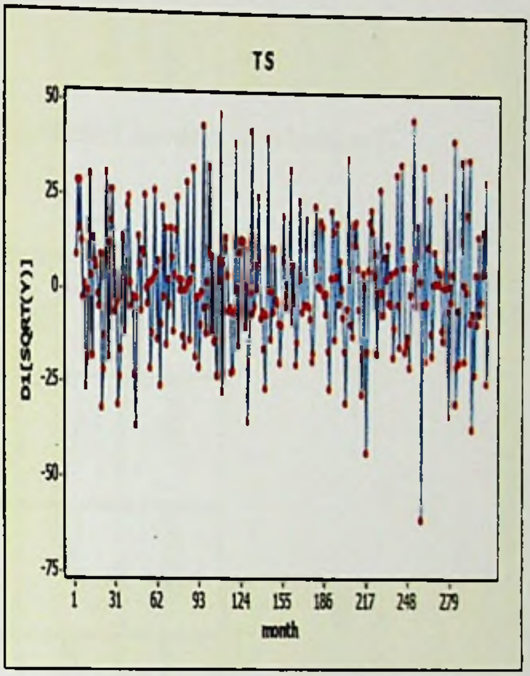
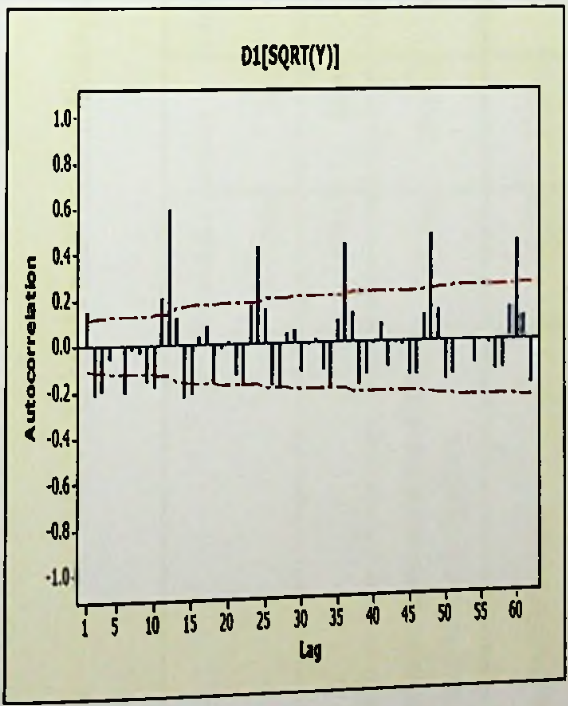


Figure C3: Time Series plots of series LOG(Y) and D1[LOG(Y)] in phase II



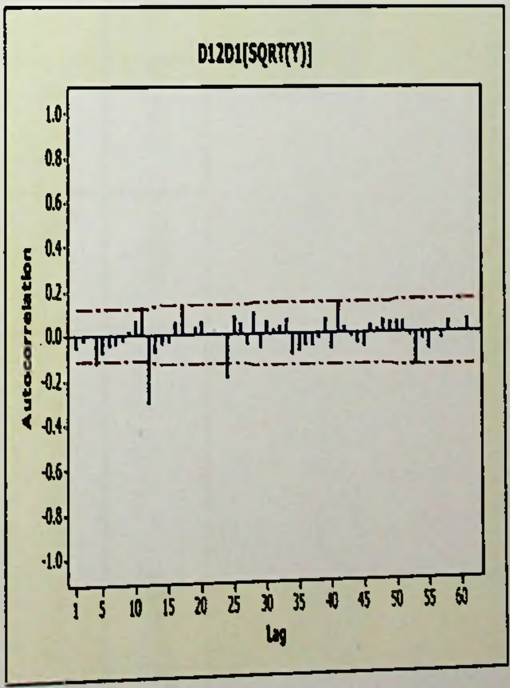


Figure C4: ACF plot of series D1[LOG(Y)] and D12D1[LOG(Y)] in phase II

APPENDIX D

Ex-ante forecast by the fitted models in window I

Table D1: Ex-ante forecastd values of fitted significant models in window I

Month	December 2008	January 2009	February 2009	March 2009	April 2009	May 2009	MAPE
Actual	48925	38468	34169	34065	26054	24739	
Multiplicative Exponential Smoothing model by grid search algorithm	44129.9	43982.5	40935.7	39867.04	29997.26	24740.6	12.69
Seasonal ARIMA model	43173	37804.7	38724.1	30861.3	28424.4	27625.4	9.5
Dynamic Transfer Function model	50746	34869.3	31849	24580.49	25044.52	23043.11	9.74
State Space model	46526.23	31729.34	30468.21	23143.72	25305.57	22729.17	12.72

APPENDIX E

Ex-ante forecast by the fitted models in window II

Table E1: Ex-ante forecasted values of fitted significant models in window II

Month	October 2015	November 2015	December 2015	MAPE
Actual	**132280**	**144147**	**206114**	
Additive Exponential Smoothing model by auto search algorithm	152044.1	165268.1	190673.9	12.36
Additive Exponential Smoothing model by grid search algorithm	149378.6	162280	188421.3	11.36
Multiplicative Exponential Smoothing model by grid search algorithm	154254.5	169275	197635	12.72
Seasonal ARIMA model	146594	144745	203690	**4.14**
Dynamic Transfer Function model	147095.55	145246.55	204191.55	**4.30**
State Space model	150732.6	150248.06	208930.56	**6.52**