# FORECASTING DIESEL CONSUMPTION IN SRI LANKA USING MULTIVARIATE TIME SERIES MODELS

Thusitha Tharindu Weerawardhana

(168849E)

Master of Science in Business Statistics

Department of Mathematics

University of Moratuwa
Sri Lanka

July 2020

# FORECASTING DIESEL CONSUMPTION IN SRI LANKA USING MULTIVARIATE TIME SERIES MODELS


Thusitha Tharindu Weerawardhana

(168849E)


Dissertation submitted in partial fulfillment of the requirements for the
Master of Science in Business Statistics


Department of Mathematics

University of Moratuwa
Sri Lanka


July 2020

## DECLARATION OF THE CANDIDATE

"I declare that this is my own work and this thesis/dissertation2 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 thesis/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 of Candidate Date

(T. T. Weerawardhana)

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

Senior Prof. T. S. G. Peiris

Senior Professor in Applied Statistics,

Department of Mathematics

Faculty of Engineering

University of Moratuwa

## ABSTRACT

In the modern trends of industrialization and development, energy has become one of the most important aspects of every economy in the world. Among the energy sources that are available today, diesel has a considerable consumption for various activities such as production of goods and transportation in many countries including Sri Lanka. Using monthly data of ten explanatory variables (January 1998 to July 2018) vector error correction model of order 2: VECM (2) was developed to model monthly consumption of diesel in Sri Lanka. The diesel consumption has been increasing due to various activities. The most significant influential variables are Exchange rates of USD to LKR, Merchandize Imports, Number of Tourists Arrivals, National Consumer Price Index, and Electricity Power Generated. The errors of the model were found to be white noise. The percentage errors of the fitted data using the VECM (2) model for both trained and validated set vary from -8.4 to +8.5%. Further it was found that Exchange rates of USD to LKR, Merchandize Imports, and Number of Tourists Arrivals show significant long run positive association with diesel consumption while National Consumer Price Index and Electricity Power Generated indicate significant long run negative relationship with diesel consumption in Sri Lanka. This model is suitable only short term prediction and it is recommended to develop the model so that it can be used for long run prediction. Nevertheless, the model provides the analyst with the ability to make decisions using various predicted intervals with different membership values by controlling the explanatory variables.



**Keywords:** Vector Error Correction Model, Multivariate Time Series, Correlation, Consumption of Diesel,

# ACKNOWLEDGEMENT

The author wishes to extend his sincere gratitude to the following people, for their generous support guidance, views and comments, which paved the way to the successful completion of this project:

Senior Prof. T.S.G. Peiris, Senior Professor in Applied Statistics and former Head of Department of Mathematics, Faculty of Engineering, University of Moratuwa and the Course Coordinator of M. Sc. / Post Graduate Diploma in Business Statistics, for being a visionary throughout the lifetime of this research project. His valuable insights and continuous comments helped me to complete this project successfully.

To the academic and non-academic staff at the Department of Mathematics for the all wishes and good will.

To my family and loved ones whom have offered their fullest support and strength of mind throughout my academic life.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| Abbreviation | Description |
| --- | --- |
| AR | Autoregressive Model |
| ARCH | Autoregressive Conditional Heteroskedasticity |
| ARMA | Autoregressive–moving-average Model |
| ARIMA | Autoregressive–Integrated Moving-Average Model |
| IID | Independent and Identically Distributed |
| OLS | Ordinary Least Square |
| MT | Metric Ton |
| VAR | Vector Autoregressive Model |
| VECM | Vector Error Correction Model |

# CHAPTER 1

# INTORDUCTION

## 1.1 Background

In the modern trends of industrialization and development, energy has become one of the most important aspects of every economy in the world. This energy requirement suggests that, the world economies are densely dependent on energy. Alam (2006) claims that, “energy is the indispensable force driving all economic activities”. In other words, when the energy consumption increases, there will more economic development activities commence in the country. Therefore, a country with greater economy emerges. Energy is an essential component of economic development of a country. Its demand is connected to economic factors such as prices of energy, income and population, urbanization, and level of technological development of the economy. Furthermore, Prentice & Poppitt (1996) insist that for developed nations, energy facilitates the productions and services and also sustain the economy of the country.

Energy is not only associated with the economical development, but also has a relationship with the social development of the country. Energy is an essential element and has a decisive role in our daily life, agriculture, industry and social services. At present, energy is required to fulfill basic human requirements such as clean water, sanitation and healthcare. Furthermore, the society of a country consumes the energy in order to have productive lighting, heating, cooking, and mechanical power, transport, and telecommunication services. The energy sector therefore is one vital sector for a country’s socio-economic development, production, and better standard of living.

### 1.2 World Energy Consumption

Some of the main energy sources available today are coal, petroleum, natural gas, renewable energy, nuclear, and fossil fuels. According to Wikipedia (2019), fossil fuel is the largest consumed energy source in the world while the nuclear energy is the lowest consumed energy. The pie chart of present energy consumption is shown in the Figure 1.1.

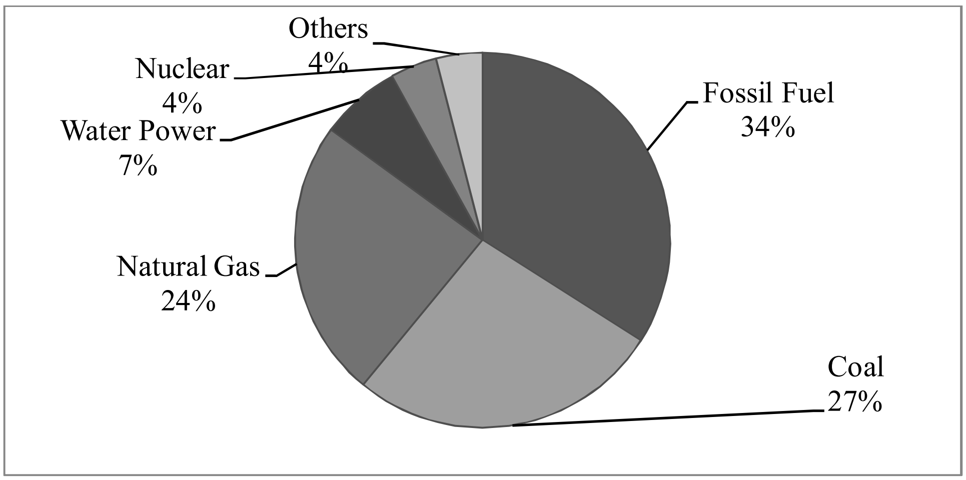


Figure 1. 1 The Distribution of Energy Consumption in the World.
*Source*: Wikipedia (2019)

According to the Figure 1.1, it is clear that 34% of the consumption of energy is acquired by the fossil fuels and hence it is the largest consumed energy source in the world. The second largest energy source consumed at present is coal, which shows 27%. Furthermore, the consumption of natural gas is third largest consumed source, which has the percentage of 24%. The fourth largest energy source consumed at present is water, which shows 7% of the total energy. The lowest consumed energy sources are nuclear power and other renewable energy, each of which has only 4% of usage.

## 1.3 Energy Consumption in Sri Lanka

The consumption of the energy in Sri Lanka, which is described in Figure 1.2, is somewhat similar to the word which described in Figure 1.1. The highest consumed energy source is fossil fuel while the lowest energy consumed is renewable energy.

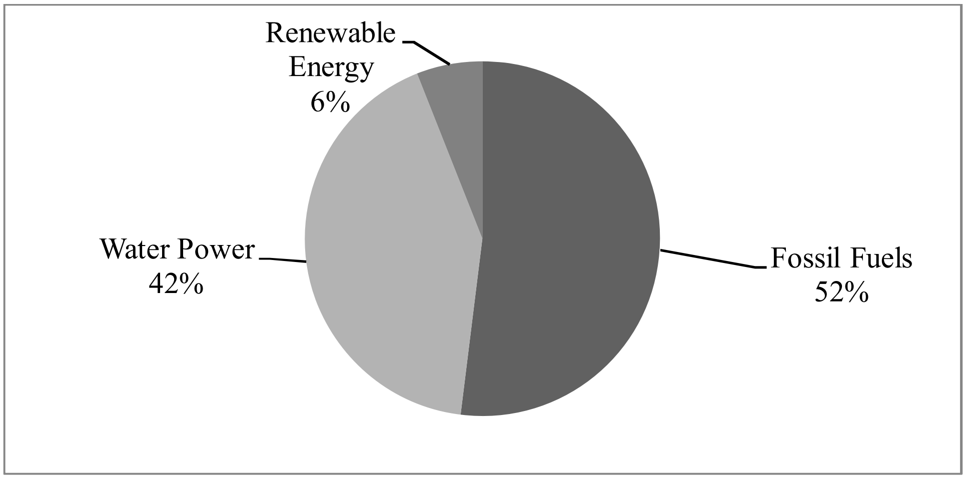


Figure 1. 2 Energy Consumption in Sri Lanka
*Source*: Energy consumption in Sri Lanka (2019)

According to the Figure 1.2, it is clear that 52% of the consumption of energy is obtained by the fossil fuels and hence similar to the Figure 1.1 it is the largest consumed energy source in Sri Lanka. However, compared to the Figure 1.1, the second largest energy source consumed in the country is not coal but water power, which shows 42%. Furthermore, the least consumed energy is renewable energy which has only 6% of the total consumption.

Among the energy sources that are available today, diesel has a considerable consumption in the world as heavy vehicles such as buses and trains and machines used in factories require diesel to operate. The demand for diesel and other energy sources has increased with the expansion of the word economy (World Energy Consumption, 2019). It is also stated that the demand of the diesel will increase in the future as more industrial activities of developed countries are likely to be increased dramatically in next few years (Clemente, 2016). According to Energy consumption in Sri Lanka (2014), Sri Lanka has also increased the consumption of diesel along with other petroleum products such as petrol, kerosene and lubricants. Furthermore, it is stated that the diesel is the most frequently consumed petroleum fuel in Sri Lanka (Central Bank of Sri Lanka, 2019). The average consumption of petroleum products in Sri Lanka from 1998 to 2018 (Figure 1.3) can be ranked as diesel (70%) > petrol (23%) > kerosene (7%).

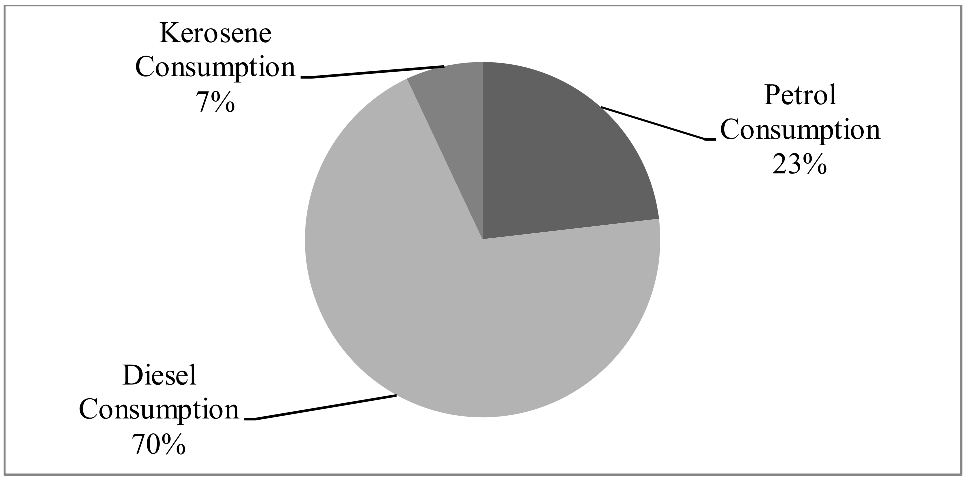


Figure 1. 3 Comparison of Consumption of Petroleum Products
*Source*: Central Bank of Sri Lanka (2019)

## 1.4 Problems in Fossil Fuel

One of the major problems in fossil fuel is occasional scarcity in the supply. Decrease in supply of fossil fuel leads to increase in the prices of it, and also it is a considerable impact to the countries that imports fossil fuels because of failing to fulfill the demand for the fuel (Mork, 1994). The failure to meet the demand for fossil fuel has a negative impact not only on the consumption of petroleum but also on the investments in the country, which leads to lower the growth rate in the economy of the country. When the scarcity in the supply of petroleum fuel exists over a longer period, the consumers of the fossil fuel in the production sector may have to change the production structure and tend to adopt non-fossil fuel energy sources. Therefore, this may lead to distortions for their production and consequent loses from their investments. Furthermore, the investors have to reallocate their labors in the firms and capital that invested across sectors in response to shortage in supply of fossil fuel. This can affect the increment in unemployment in the long term (Loungani, 1986). Therefore, the countries that import fossil fuel such as Sri Lanka remain vulnerable to sudden variations in supply and price of fossil fuel, particularly because Sri Lanka is more non-export-oriented country.

## 1.5 Present Situation in Sri Lanka

Since economic diversification is still low in Sri Lankan economy, sudden distortions in supply of fossil fuel have the enormous potential to impact the economy of the country negatively. For these reasons consumers of petroleum fuel in the country and the government must be assisted with information on the consumption pattern of the petroleum products. Therefore the government and the consumers in Sri Lanka will be able to make an incisive decision on the consumption, supply and the prices of fossil fuel by forecasting the future consumption of the fuel. Hence the unexpected problems related to scarcity of petroleum products can be eluded.

## 1.6 Importance of Forecasting

There are many factors that have an association between the consumption of diesel. Thus it is important to assess the relationship between the consumption of diesel in Sri Lanka and the factors related to the economy of the country as diesel is the major consumed petroleum fuel in the country. The forecasting system could be used to inform the government the future consumption pattern of diesel in the country in order to prevent shortage of diesel. Furthermore, the relationship between the consumption of diesel and the factors affecting it is useful for the government and policy makers to manage and control these factors cohesively in order to maintain the persistent supply of diesel island-wide. Therefore, the negative impacts, which are caused by the shortage of supply of diesel, can be eluded. Thus the negative impacts to the economy and to the investors can be ceased. Moreover, forecasting will be helpful for the smooth operation in the industrial and economic sectors of the country.

The results of the study can be used for various aspects in economic such as business decisions, hedging decisions, and volatility.

### 1.6.1 Business decision

A forecast of diesel consumption movements can be useful when making a range of business decisions, particularly in sectors such as transportation, construction, agriculture, electricity, and military sectors of the country (Use of Diesel, 2019). Most of the goods are transported by heavy vehicles such as trucks and trains with diesel engines, and most construction, farming, and military vehicles and equipment also have diesel engines because of high performance, efficiency, and safety features. The policy makers of these sectors expect the efficient and useful results and services from these sectors. On the other hand the investors of expects profits. Therefore, in order to make these sectors more efficient and maximize the utilization of the investments, the policy makers will want to assess the movements of consumption both of long-term and short-term consumption of diesel. Treasurers will also need to assess the accompanying requirements for funding and short-term investments.

### 1.6.2 Hedging decisions

In addition, forecasts can be helpful when setting or assessing a hedging strategy. Hedging strategies depend on external factors as well as the internal requirements of a business. In particular, if the fuel corporation fixes the unit price of the diesel and the market price of diesel in the world continues to increase, the corporation is exposed to the risk of having losses and eventually become a debtor to the Treasury. Moreover, it is important to investigate the relationship between the consumption of diesel and the external factors in order to decide the prices accurately. Therefore, the unit price of diesel has to be decided based on the present and future consumption of diesel considering the factors affecting the consumption. Thus it is important to model the present consumption of diesel and forecast its future for better hedging decisions.

### 1.6.3 Volatility

The other purpose for forecasting is to identify the volatility in the consumption of diesel. Consumption of diesel can have a significant effect on foreign currency earnings of the county. This may not be a problem in the long term, if the effect is understood by investors. However, it will affect the hedging strategy of the country, particularly over the shorter term, such that it may mean a higher percentage of short-term exposures are hedged.

## 1.7 Objectives of the Study

On the view of the above explanation, the objectives of the study are to;

- Identify the factors that are useful for predicting the future consumption of diesel in Sri Lanka
- Develop a model to predict annual consumption of diesel
- Validate the model

## 1.8 Outline of the Dissertation

The dissertation will be organized with nine chapters. Literature reviews of the study are discussed in the Chapter 2. Material and methodologies encompassing in the study with corresponding theoretical backgrounds are described in Chapter 3. Temporal variability of the consumption of diesel and the explanatory variables are described in Chapter 4 along with explanatory data analysis. The development of VEC model with all 10 explanatory variables is discussed in Chapter 5. Another VEC model using 5 selected explanatory variables is described in Chapter 6. In

Chapter 7, the 2 models are compared to identify the best fitted model. Conclusions and recommendations based on the inferences derived from the study are highlighted in Chapter 8.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Introduction

This chapter takes into consideration to review some forecasting techniques for petroleum and electricity carried out by various researches. These techniques are limited to VEC Model, ARIMA model technique, Seasonal ARIMA technique, Grey-Markov model, Bayesian linear regression theory, Markov Chain Monte Carlo method and artificial neural network.

## 2.2 Use of VEC Model

Warr & Ayres (2010) analyzed the relationship between Gross Domestic Product (GDP) and two different variables of energy inputs. The two variables are exergy, (energy that is available to be used) and useful work in US. The time period was from 1946 to 2000. After confirming that the variables are cointegrated they used a vector error correction model (VECM) to test for both short-term and long-term relationships. They discovered that there exists the evidence of granger causality from exergy to GDP, which is an increase in exergy impact GDP to increase both a short-term and long-term. On the other hand, the variable, useful-work has no short-run impact on GDP however it implied a long-term relationship with GDP. They further claimed that an increase of exergy alone is sufficient to impact the GDP in the short-run, while GDP shows positive relationship to increased exergy over a period of several years and useful work influence GDP by re-adjusting to the long-run equilibrium relationship.

Narayan et al., (2009) studied oil production in states of Australia using the annual data for the period from 1985 to 2006 and used VEC model with two explanatory variables namely, oil price and income of the states of the country. They found that the existence of a long run the impact of oil price on oil products is not statistically significant. Further, the income of the states of the country showed a significance negative relationship on oil production. They state that the growth in the economy of the country influences an increase in the production of oil which causes the upward trend in oil price, and eventually increment in the inflation as well.

### 2.3 Use of ARIMA Models

Liu (1991) studied the relationships between petrol prices, crude oil prices, and the stock of petrol by using monthly data for the period from January 1973 to December 1987 in Unites States of America (USA). Box-Jenkins ARIMA and transfer function models were used in this study. He found that the petrol price of USA is mainly influenced by the price of crude oil. The stock of petrol has little or no influence on the price of petrol during the first half of the period, and showed slight influence during the second half of the period. It has been discovered that the relationship between the prices of petrol and crude oil varies with the time, shifting from a long lag to a short lag. The models were estimated using with and without outlier adjustment. An iterative method for the joint estimation of model parameters and outlier effects was used for model estimation with outlier adjustment.

Ayeni et al., (2001) explored the use of the Autoregressive Integrated Moving Average (ARIMA) theory in forecasting and estimating the demand of Liquefied Petroleum Gas (LPG), premix fuel and petrol in Ghana. They used monthly data from January 1999 to December 2010 and analyzed and forecasted the demand of the 3 variables for one year period. Auto Correlation Function and Partial Auto Correlation Function of the first differenced data were considered to generate possible models. The minimum value of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to select the best fitted model. The best fit models for the demand of petrol, LPG and premix fuel were suggested as ARIMA(1,1,3), ARIMA(2,1,3) and SARIMA(3,1,0)(2,0,0) respectively.

## 2.4 Use of Other Models

Chai et al., (2012) analyzed petrol and diesel consumption in the transportation sector of China based on the Bayesian linear regression theory and Markov Chain Monte Carlo method (MCMC). They established a demand-forecast model of petrol and diesel consumption with four explanatory variables. The four variables are urbanization level, turnover of passenger in aggregate (TPA), turnover of freight in aggregate (TFA), and civilian vehicle number (CVN). Furthermore, they forecasted the future consumption of both petrol and diesel for the period from 2010 to 2015 using data from 1985 to 2009. The results of the study showed that urbanization is the sensitive factor with strongest effect on petrol and diesel consumption in transportation sector. The forecasted results from other explanatory variables namely, GDP, TPA, and TFA showed less effect than the urbanization level, although the relationships between these 3 variables and the consumption of petrol and diesel are significant.

Liu and Lin. (1991) studied the consumption of natural gas in households of Taiwan. In this study, the authors investigated the relationships among several time series variables (temperature of service areas, price of natural gas) and developed models for forecasting the future consumption of natural gas. They found that the temperature of service areas and the price of natural gas showed strong impact in forecasting the consumption of natural gas in household. Furthermore, since the government price control policy, it was found that the price variable employed in modeling and forecasting of natural gas consumption needs to be used sensibly.

A study by carried out by Kumar and Jain (2010) applied three methods, namely, Grey-Markov model, Grey-Model with rolling mechanism, and singular spectrum analysis (SSA) to forecast the consumption of energy (crude-petroleum, coal, electricity and natural gas) in India. Grey-Markov model has been used to forecast consumption of crude-petroleum while Grey-Model with rolling mechanism used to forecast consumption of coal and electricity. The model SSA was used to model consumption of natural gas. The models for each time series were selected by comparing the minimum of values of mean absolute percentage errors (MAPE) of each time series. The results obtained from the study were compared with those of Planning Commission of India's projection. The comparison clearly indicated that these time series models forecast the energy consumption.

Pao (2009) proposed two new hybrid nonlinear models that combine a linear model with an artificial neural network (ANN) to develop a model to describe the observed variation of the consumption of electricity and petroleum, taking into account heteroskedasticity in the data. Both of the hybrid models can decrease prediction errors in forecasting. They claimed that the new hybrid models dominate the forecasts from conventional linear models and this is due to the flexibility of the hybrid models to describe the complex and Heteroskedasticity nonlinear relationships that are not easily captured by linear models.

## 2.5 Summary of Chapter 2

Many authors have used ARIMA model and Seasonal ARIMA model to describe the temporal variability of the consumptions of petroleum products in different countries. The VEC model, Bayesian linear regression models and Markov Chain Monte Carlo models have also been used by authors to evaluate the relationship between the factors affecting the consumption. Some studies indicate that gasoline price is mainly influenced by the price of crude oil, and GDP has a positive relationship with energy consumption. Furthermore, techniques such as Grey-Markov model, Grey-Model with rolling mechanism, singular spectrum analysis (SSA) and artificial neural networks (ANN) with linear models have been also be used to model the consumption of petroleum products. The results gathered from the literature review are useful to carry out this study.

# CHAPTER 3

# MATERIAL AND METHODS

In this chapter the secondary data collected for the analysis and the theoretical framework used in this study are outlined, including theory for the specific statistical procedures.

## 3.1 Secondary Data

The following monthly data were obtained from the website of the Central Bank of Sri Lanka (CBSL) for the period from January 1998 to July 2018.

1. Diesel Consumption
2. National Consumer Price Index (NCPI) (Base year =2010)
3. Wage Rate Index of Government Employees (December 1978=100)
4. Distance Operated in Public road transport (kilometers)
5. Distance Operated in Rail transport (kilometers)
6. Number of Vehicle registrations
7. Merchandise Imports (USD Million)
8. Merchandise Exports (USD Million)
9. Total Tourist Arrivals
10. Average Exchange Rates
11. Electricity Power generation (Giga Watt hours)

The model is developed using the data from the period from January 1998 to July 2017, where the sample size is 235. Since it is important to validate the models after developing, the model is validated using the data from August 2017 to July 2018, where the sample size is 12

Table 3.1 shows the breakdown of the number of data used for the modeling and validating the models.

Table 3. 1 Breakdown of the Data Collected

| Description | Period | Sample Size |
|---|---|---|
| Data for the model | January 1998 – July 2017 | 235 |
| Validation | August 2017 – July 2018 | 12 |

Generally it is a better to use at least one third of the original data set for validation. However when VAR/VECM models are used with more variables and the lag number is greater than 2 the above criteria cannot be met. Thus in this case only 12 points was used for validation

## 3.2 Description of Ten Explanatory Variables

### 3.2.1 Consumption of Diesel (DC)

Consumption of Diesel (in Metric Ton) is defined as the total monthly consumption of diesel in Sri Lanka.

### 3.2.2 National Consumer Price Index (NCPI)

The National Consumer Price Index (NCPI) is a measure that evaluates the weighted average of prices of a pool of consumer goods and services in Sri Lanka (National Consumer Price Index for Sri Lanka, 2020). NCPI is calculated by taking price changes for each item in the predetermined pool of goods and averaging them. The NCPI is one of the most frequently used statistics for identifying periods of inflation or deflation. The base year that is considered for the study is 2010.

### 3.2.3 Wage Rate of the Employees (WR)

The wage rate of the employees is defined as the percentage of the present monthly salary of the employees relative to the salary in December 1978 as given below.

$$\text{Wages Rate} = \frac{\text{Present Salary}}{\text{Salary in December 1978}} \times 100$$

### 3.2.4 Distance Operated in Public Road Transport (PT)

This variable describes the total distance operated (in Km) in public transport such as busses and Lorries per month.

### 3.2.5 Distance Operated in Rail Transport (RT)

Similar to the variable described in 3.2.4, this variable describes the total distance operated in train transport per month in Km.

### 3.2.6 Number of Vehicle Registered (VR)

The monthly total number of vehicle which are registered in Sri Lanka

### 3.2.7 Merchandise Imports (MI)

The variable merchandize imports are defined as the amount paid for importing goods and services per month in US $.

### 3.2.8 Merchandise Exports (ME)

This variable is defined as the amount earned for exporting goods and services per month in US $.

### 3.2.9 Total Tourist Arrivals (TA)

The variable total tourist arrival is defined as the number of tourists arrived in a month.

### 3.2.10 Monthly Average Exchange Rate (EX)

Monthly average exchange rate is the average value of one US Dollar in Sri Lanka Rupees.

### 3.2.11 Electricity Power Generation (EP)

This is the total electricity generated per month in Giga Watt hours (GWh).

## 3.3 Abbreviations of the Variables

Since the log transformation ($\log_e$) is applied for the analysis of the data in the Chapter 4, the following abbreviations were used (Table 3.2).

Table 3. 2 Abbreviations of the Variables

| **Abbreviation** | **Description of the Variable Name** |
|---|---|
| LNDC | $\text{Log}_e$ of Diesel Consumption |
| LNNCPI | $\text{Log}_e$ of National Consumer Price Index |
| LNWR | $\text{Log}_e$ of Wages Rate of Government Employees |
| LNPT | $\text{Log}_e$ of Distance Operated in Public Road Transport |
| LNRT | $\text{Log}_e$ of Distance Operated in Railroad |
| LNVR | $\text{Log}_e$ of Number of Vehicle Registered |
| LNMI | $\text{Log}_e$ of Merchandize Imports |
| LNME | $\text{Log}_e$ of Merchandize Exports |
| LNTA | $\text{Log}_e$ of Number of Tourists Arrived |
| LNEX | $\text{Log}_e$ of Exchange Rate of USD to LKR |
| LNEP | $\text{Log}_e$ of Electricity Power Generated |

## 3.4 Time Series

A time series is a set of observations measured successively in time. The time series of n observations can be notified as $\{X_t, t= 1, 2, 3\dots n\}$ where $X_t$ is the observation at time t.

### 3.4.1 Stationary Time Series

A time series $\{X_t\}$ is stationary if the joint probability density function of the random variables $\{X_{t1}, X_{t2},\ldots,X_{tn}\}$ is equal to the joint probability density function of the random variables $\{X_{t1+m}, X_{t2+m},\ldots,X_{tn+m}\}$ for arbitrary points $t_1, t_2,\ldots, t_n$

### 3.4.2 Strictly Stationary Time Series

A time series $\{X_t\}$ is said to be strictly stationary if

$F_X(x_{t1+r},\ldots,x_{tn+r}) = F_x(x_{t1},\ldots,x_{tn})$ for all r, $t_1,\ldots,t_n$

Where $F_X(x_{t1+r},\ldots,x_{tn+r})$ is the cumulative probability density function of joint probability density function of $\{X_t\}$ at time $t_{1+r},\ldots,t_{n+r}$

### 3.4.3 Weekly Stationary Time Series

A time series $\{X_t\}$ is said to be weakly stationary if

I. $E(X_t) = \mu$ for all t
II. $V(X_t) = \sigma^2$ for all t
III. $Cov(X_t, X_s) = Cov(X_{t+h}, X_{s+h})$ for all t, s, h

### 3.4.4 Augmented Dickey-Fuller test

If a time series appears non-stationary one may verify the existence of a unit root in an AR (p) series by performing an augmented Dickey-Fuller (ADF) test. The null hypothesis

$H_0$: $\beta = 1$ is tested against the alternative $H_a$: $\beta \leq 1$ using the regression

$$X_t = c_t + \beta X_{t-1} + \sum_{i=1}^{p-1} \Delta X_{t-i} + e_t$$

Where $c_t$ is a deterministic function of the time index t and $\Delta X_j = X_j - X_{j-1}$ is the differenced series of $X_t$. Thus, the ADF-test is the t-ratio of $\hat{\beta} - 1$ expressed as

$$\text{ADF Test}=\frac{\hat{\beta}-1}{std(\hat{\beta})}$$

Where $\hat{\beta}$ is the least-squares estimate of $\beta$. The interpretation of the ADF test is if the null hypothesis is rejected, then the time series is stationary.

### 3.4.5 Autocorrelation

The autocorrelation is defined as the linear dependence between $X_t$ and the past values of $X_{t-i}$ .

In other words, the autocorrelation of time series $X_t$ is given by as;

$$Corr(X_t, X_{t-1}),\ k=1,2,\ldots$$

Where k is lag

### 3.4.6 Autocorrelation Function (ACF)

The plot of autocorrelation vs. lag is known as autocorrelation function

### 3.4.7 Partial Autocorrelation Functions (PACF)

The partial autocorrelation function (PACF) is a function of ACF and is the amount of correlation between a variable and a lag of itself that is not explained by correlations at all lower-order-lags. Considering the AR models:

$X_t=\emptyset_{0,1}+\emptyset_{1,1}X_{t-1}+e_{1t},$

$X_t=\emptyset_{0,2}+\emptyset_{1,2}X_{t-1}+\emptyset_{2,2}X_{t-2}+e_{2t},$

$X_t=\emptyset_{0,3}+\emptyset_{1,3}X_{t-1}+\emptyset_{2,3}X_{t-2}+\emptyset_{3,3}X_{t-3}+e_{3t},$

$X_t=\emptyset_{0,4}+\emptyset_{1,4}X_{t-1}+\emptyset_{2,4}X_{t-2}+\emptyset_{3,4}X_{t-3}+\emptyset_{4,4}X_{t-4}+e_{4t},$

.

.

.

Where $\varphi_{0,j}$ , $\varphi_{i,j}$ and $e_{it}$ are, respectively, the constant term, the coefficient of $X_{t-i}$ and the error term of an AR(j) model. Since the equations are in the form of a multiple linear regression we may estimate the coefficients using the ordinary least-square method. The estimates $\widehat{\varphi}_{1,1}$, $\widehat{\varphi}_{2,2}$ and $\widehat{\varphi}_{k,k}$ of respective equation are called the lag-1, lag-2 and lag-k sample PACF of $X_t$. Thus, the complete sample PACF describes the time series' serial correlation with its previous values of a specific lag controlling for the values of the time series at all shorter lags.

### 3.4.8 Use of ACF and PACF

The ACF and PACF of stationary series can identify the possible ARIMA models by comparing both sample ACF and sample PACF with the corresponding theoretical ACF and theoretical PACF. Therefore in order to identify parsimonious models for the observed time series, it is required to make the series stationary and then sample ACF and PACF of the stationary series must be compared with corresponding theoretical ACF and PACF

### 3.4.9 Ljung-Box Portmanteau test (Q-Statistic)

Ljung-Box Portmanteau tests any group of autocorrelations of a time series are different from zero. The null hypothesis is;

$H_0 : \rho_1 = ... = \rho_m = 0$

And alternative hypothesis is;

$H_a : \rho_i \neq 0$ for some $i \in \{1, ..., m\}$

$H_0$ implies the data is random while $H_a$ represents the data is not random.

The test statistic is;

$$Q = n(n+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$$

Where n denotes the sample size, $\hat{\rho}_k^2$ the sample autocorrelation at lag k and h is the number of lag tested. Q is asymptotically follows a $\chi^2(h)$. The null hypothesis is rejected if $Q > \chi^2_{\alpha,h}$, where $\chi^2_{\alpha}$ denotes the $100(1-\alpha)^{th}$ percentile of a chi-squared distribution with h degrees of freedom.

### 3.5 Vector Auto Regressive Model (VAR)

Ordinary models usually consider a unidirectional relationship where the variable of interest is influenced by the predictor variables, but not the opposite way. However, in many macroeconomic models the reversed is often also true - all the variables have an effect on each other. When studying a set of macroeconomic time series vector autoregressive (VAR) models are frequently used. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables. With vector autoregressive models it is possible to approximate the actual process by arbitrarily choosing lagged variables. Thereby, one can form economic variables into a time series model without an explicit theoretical idea of the dynamic relations. The basic model for a set of K time series variables of order p, a VAR (p) model, has the form

$$y_t = \mathbf{A}_1 y_{t-1} + \ldots + \mathbf{A}_p y_{t-p} + u_t$$

Where, the $\mathbf{A}_i$'s are (K×K) coefficient matrices and $\mathbf{u_t}$ is a vector of assumed zero-mean independent white noise processes. The covariance matrix of the error terms, E $(\mathbf{u_t u_t}^T) = \mathbf{\Sigma_u}$, then assumes to be time-invariant and positive definite. The error terms $u_{i,t}$ may be contemporaneously correlated, but are uncorrelated with any past or future disturbances and thus allowing for estimation following the ordinary least square (OLS) method. By introducing the notation $\mathbf{Y} = [y_1, \ldots, y_T]$, $\mathbf{A} = [A_1 : \ldots : A_p]$ $\mathbf{U} = [u_1, \ldots, u_T]$ and $\mathbf{Z} = [Z_0, \ldots, Z_{T-1}]$, where

$$Z = \begin{bmatrix} u_{t-1} \\ \cdot \\ u_{t-p} \end{bmatrix}$$

The model can be expressed as

$$\mathbf{Y = AZ + U}$$

and the OLS estimator of A is

$$\widehat{\mathbf{A}} = \widehat{A}_1 \ldots\ldots \widehat{A}_p = \mathbf{YZ'(ZZ')^{-1}}$$

The covariance matrix $\mathbf{\Sigma_u}$ may be estimated in the usual way. By denoting the OLS residuals as $\mathbf{\hat{u}=y_t\text{-}\hat{A}\ Z_{t\text{-}1}}$ the matrix

$$\mathbf{\hat{\Sigma}_u = \frac{1}{T\text{-}K_p}\sum_{t=1}^{T} \hat{u}_t\hat{u}_t'}$$

Where T is the number of observations and $\mathbf{\hat{\Sigma}_u}$ is an estimator which is consistent and asymptotically normally distributed independent of $\mathbf{\hat{A}}$. Furthermore, the process is defined as stable if the determinant of the autoregressive operator has no root in or on the complex unit circle. Otherwise, some or all of the time series variables are integrated.

### 3.5.1 Johansen cointegration test

Johansen cointegration test uses two test statistics do determine the number of cointegration vectors. The first, the maximum Eigen value statistic, tests the null hypothesis of;

$H_0 : r \leq j - 1$

Cointegrating relations against the alternative of;

$H_a : r = j$ cointegrating relations for $j \in \{1, 2, ..., n\}$.

It is computed as:

$$LR_{max}(j-1, j) = -T \cdot \log(1-\lambda_j) = \lambda_{max}(j-1)$$

Where, T is the sample size.

Thus, the null hypothesis of no cointegrating relationship against the alternative of one cointegrating relationship is tested by $LR_{max}$ (0, 1) = $-T \cdot \log(1 - \lambda_1)$ where $\lambda_1$ is the largest Eigen value.

The second test statistic, the trace statistic, tests the null hypothesis

$H_0 : r \leq j - 1$

Against the alternative

$H_a : r \geq j$ for $j \in \{1, 2, ..., n\}$,

And is computed as;

$$LR_{trace}(j\text{-}1,n) = -T\left[\sum_{i=j}^{n} \log(1\text{-}\lambda_i)\right] = \lambda_{trace}(j\text{-}1)$$

Both tests reject the null hypothesis for large values of the test statistic. Thus, if *cv* stands for the critical value of the test and $\lambda$ $(j - 1)$ the statistic, the form of the test is:

$$\text{Reject } H_0 \text{ if } \lambda(j\text{-}1) > cv$$

The critical values for the two tests are different in general (except when $j = n$) and come from non-standard null distributions that are dependent on the sample size T and the number of co-integrating vectors being tested for.

### 3.5.2 Phillips–Perron test

The Phillips–Perron test is used in time series analysis to test the null hypothesis that a time series is integrated of order 1. It builds on the Dickey–Fuller test of the null hypothesis $\rho = 1$ in $\Delta y_t = (\rho\text{-}1)y_{t-1} + \varepsilon_t$, where is the first difference operator.

### 3.5.3 Test for Normality

Testing for normality is a standard tool to conduct a diagnostic check to identify a model before it can be used for forecasting. The test designed to determine the normality residual of data. The purpose of this test is to determine whether the residuals from the data are normally distributed or not. In order to testing for normality, we can use the Jarque-Bera (JB) Test of Normality. This test used the measure of skewness and kurtosis. In its application to decide whether the null hypothesis is rejected or not, the value of JB with the value of chi-square with 2 degrees of freedom is used. The calculation of JB is as follows:

$$JB=\frac{n-m}{6}\left(s^2+\frac{(k-3)^2}{4}\right)$$

Where:

n: Number of Sample

s: Expected Skewness

k: Expected Excess Kurtosis

m: Number of Independent variable

### 3.6 Vector Error Correction Model (VECM)

When the data used are stationary at the same level of differencing and there is a cointegration, then the model VAR, which described in Section 3.4.3, will be combined with Error Correction Model to become Vector Error Correction Model (VECM). VAR model can be applied if all the variables are stationary. However, when the one or more time series variables show co-integration relationship VECM is used. VECM is VAR which has been designed for use with non-stationary data having co-integration relationship.

VECM is one of the time series modeling which can directly estimate the level to which a variable can be brought back to equilibrium condition after a shock on other variables. VECM is very useful by which to estimate the short term effect for both variables and the long run effect of the time series data.

The VECM (p) with the co-integration rank $r \leq k$ is as given below.

$$\Delta y_t = c + \Pi\, Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta\, Y_{t-1} + \varepsilon_t$$

Where:

$\Delta$: Operator differencing, where $\Delta y_t = y_t - y_{t-1}$

$y_{t-1}$: Vector variable endogenous with 1st lag

$\varepsilon_t$: Vector Residual

c: Vector Intercept

$\Pi$: Matrix coefficient of co-integration $\Pi = \alpha\beta'$;

α= Vector adjustment, matrix with order $k \times r$ and

$\beta'$=vector co-integration (long-run parameter) matrix $k \times r$

$\Gamma_i$: Matrix with order $k \times k$ of coefficient endogenous of the $i^{th}$ variable

### 3.6.1 The Optimum Lag Length

In order to determine the optimum lag length of the VEC Model, the minimum of values the following criteria can be used.

I. Final Prediction Error (FPE)

$$\mathrm{FPE}=\frac{\mathrm{T+m}}{\mathrm{T\text{-}m}}\times\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}(u_t{}^p)^2$$

II. Akaike Information Criterion (AIC)

$$\mathrm{AIC}=\ln\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}(u_t{}^p)^2+\mathrm{m}\frac{2}{\mathrm{T}}$$

III. Bayesian Criterion of Gideon Schwarz (SC)

$$\mathrm{SC}=\ln\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}(u_t{}^p)^2+\mathrm{m}\frac{\ln\mathrm{T}}{\mathrm{T}}$$

IV. Hannan-Quinn Criterion (HQ)

$$\mathrm{HQ}=\ln\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}(u_t{}^p)^2+\mathrm{m}\frac{2\ln(\ln\mathrm{T})}{\mathrm{T}}$$

Where $u_t{}^p$denotes the residuals estimation from the model VAR(p), $m$ is the number of dependent variables. $T$is the number of observations and $p$ is the length of VAR.

# CHAPTER 4

# TEMPORAL VARIABILITY OF TIME SERIES

The most important step in building econometric models is to get an understanding of the characteristics of the time series variables. This chapter will provide the temporal variability and explanatory data analysis of all the variables.

## 4.1 Temporal Variability of the Monthly Consumption of Diesel (DC)

### 4.1.1 Monthly Consumption of Diesel in Sri Lanka

The Figure 4.1 describes the fluctuation of the monthly diesel consumption in Sri Lanka (in Metric Ton - MT) from January 1998 to July 2017.

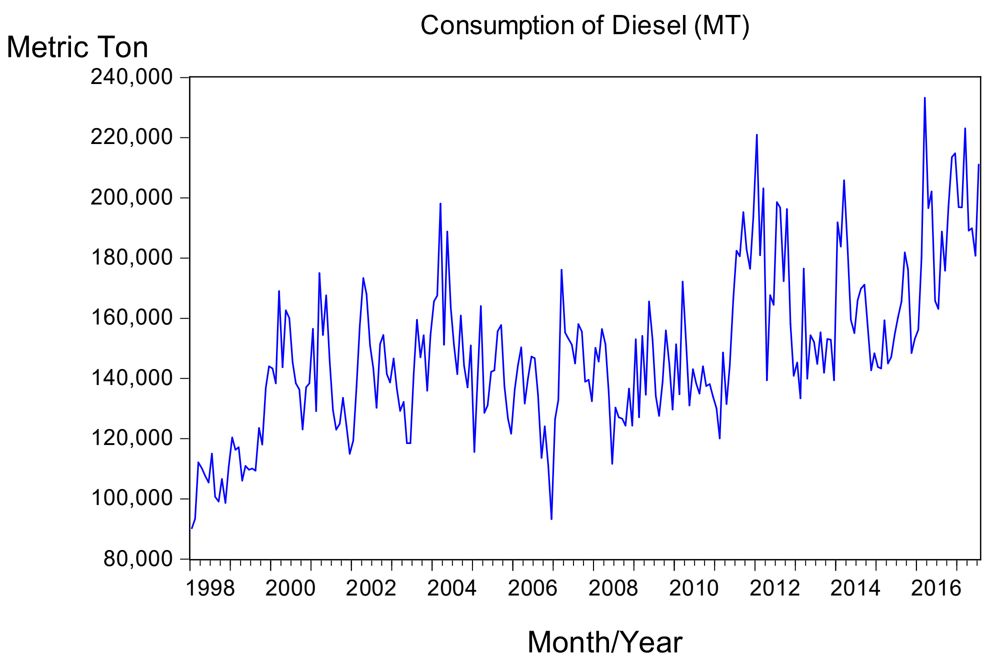


Figure 4.1 Temporal Variability of the Monthly Diesel Consumption in Sri Lanka

According to the Figure 4.1, the consumption of diesel in Sri Lanka fluctuated considerably and appears to exhibit non-stationary series with upward trend starting from 90,000 MT in the January 1998 and records approximately 210,000 MT in July 2017. Furthermore, by observing the figure, it seems that there are no outliers, seasonal patterns or cyclic movements in the plot.

### 4.1.2 Distribution of Monthly Consumption of Diesel

The distribution of the monthly consumption of diesel is shown in the Figure 4.2.

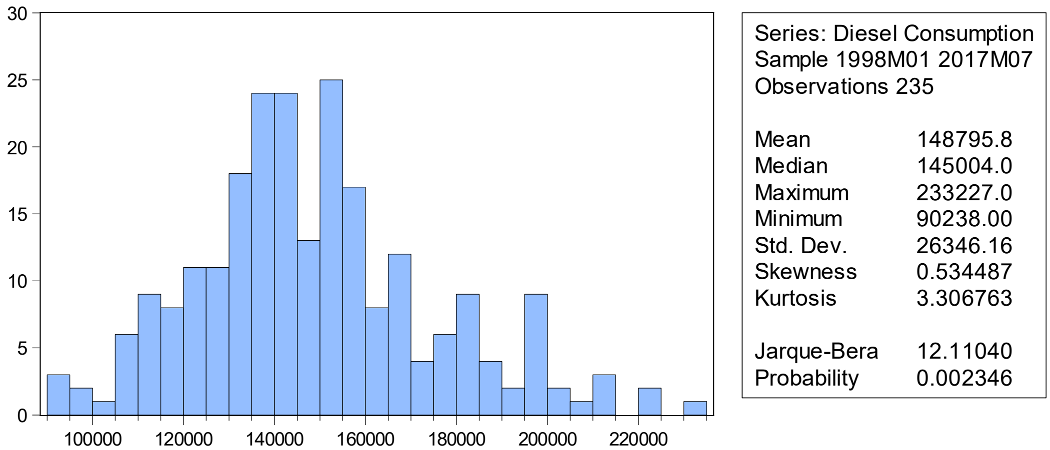


Figure 4.2 Distribution of Diesel Consumption

According to the Figure 4.2 the monthly consumption of diesel varies between 90,238 MT (January 1998) and 233,227 MT (March 2016) with the mean 148,795.80 MT during the period. Examining the histogram of the Figure 4.2, it is clear that there is one peak in the distribution with right tail implying that the distribution could not be normal. This is further proved statistically by the Jarque-Bera Test, which shows the p-value of 0.002 for the test statistic of 12.11 confirming that the distribution of the consumption of diesel is significantly deviate from normal distribution at 5% level of significance.

## 4.2 The Total Annual Diesel Consumption in Sri Lanka

The annual total DC from 1998 to 2017 in Sri Lanka is shown in the Table 4.1.

Table 4.1 Annual Total Diesel Consumption (DC)

| Year | Annual Total DC (MT) in Million | Year | Annual Total DC (MT) in Million |
|---|---|---|---|
| 1998 | 1.25 | 2008 | 1.62 |
| 1999 | 1.42 | 2009 | 1.72 |
| 2000 | 1.74 | 2010 | 1.71 |
| 2001 | 1.68 | 2011 | 1.95 |
| 2002 | 1.77 | 2012 | 2.14 |
| 2003 | 1.67 | 2013 | 1.79 |
| 2004 | 1.92 | 2014 | 2.03 |
| 2005 | 1.66 | 2015 | 1.88 |
| 2006 | 1.57 | 2016 | 2.29 |
| 2007 | 1.76 | 2017 | 2.30 |

According to the Table 4.1, it is clear that the total DC in 1998 is 1.25 Million MT and it increased up to 1.74 Million MT in 2000. Although, there was a slight decrement in DC in 2001, where the total DC was 1.68 Million MT, the consumption increased up to 1.77 Million MT in the next year (2002). However, the consumption dropped again up to 1.67 Million MT in 2003, which is approximately equal to the amount in 2001. In next year, (2004) the consumption increased rapidly by 250,000 MT and recorded 1.92 Million MT and decreased rapidly in 2005 and 2006. Thus the consumption in 2006 is 1.57 Million MT. Then the consumption showed a slight upward trend in the period 2006 – 2010 and recorded 1.71 Million MT in 2010. However, the consumption has increased rapidly up to 2.14 Million MT in 2011 to 2012 period and dropped suddenly by 350,000 MT in 2013. Then the consumption managed to increase slightly by 240,000 approximately in the next year before dropping again in 2015, where the consumption is 1.88 Million MT. However, the consumption increased rapidly up to 2.3 Million MT for the period from 2016 to 2017.

### 4.3 Temporal Variability of the 10 Explanatory Variables

The following Figure 4.3 to 4.12 describes the temporal variability of the National Consumer Price Index (NCPI), Generation of Electricity (EP), Merchandize Exports (ME), Merchandize Imports (MI), Monthly Average Exchange Rate of US Dollar to LKR (EX), Distance Operated in Public Transport (PT), Distance Operated in Rail Transport (RT), Number of Tourists arrived (TA), Number of Vehicle Registered (VR), and Wage Rate Index of Government Employees (WR). The time period is the same as that of DC, which is from January 1998 to July 2017.

#### 4.3.1 Temporal Variability of NCPI

Time plot of Figure 4.3 describes the temporal variability of the National Consumer Price Index (NCPI). The base year is 2010.

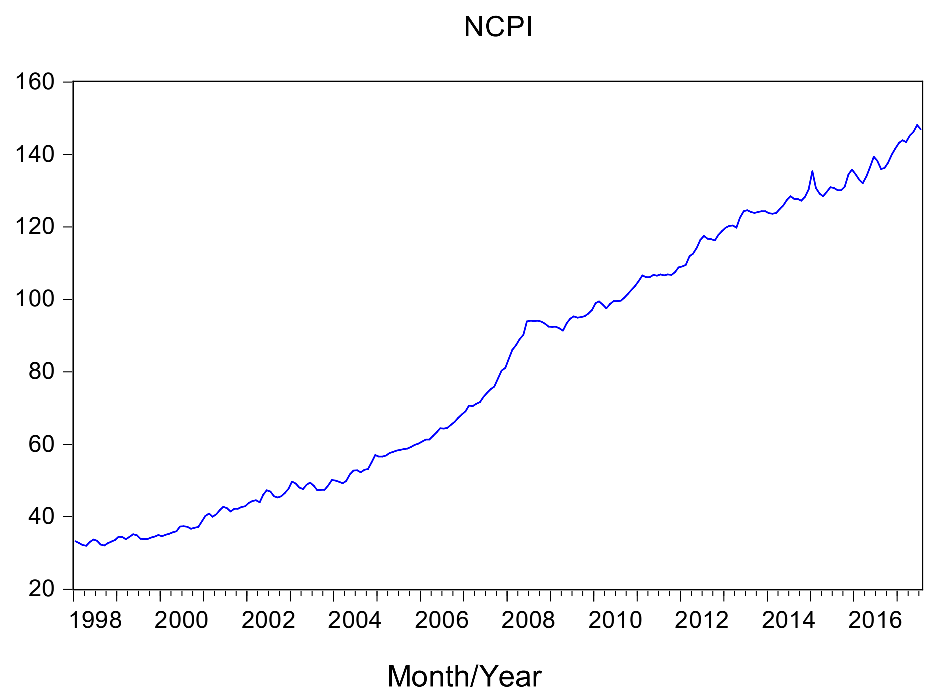


Figure 4.3 Temporal Variability of monthly NCPI

According to the Figure 4.3, the time series generally shows an upward trend during the period without seasonal pattern or outliers. The NCPI has started from under 40 in January 1998 and it has grown gradually and recorded approximately 145 in July 2017. Therefore, it can be easily concluded that the series is not stationary.

The basic statistics of NCPI is shown in the Table 4.2

Table 4. 2 Basic Statistics of NCPI

| **Statistic** | **Value** |
|---|---|
| Mean | 81.53 |
| Maximum | 148.09 |
| Minimum | 31.94 |
| Standard Error of the mean | 2.39 |
| Coefficient of variation | 45.11 |
| Test Statistic of Jarque-Bera | 21.41 (p = 0.0) |

Results in Table 4.2 indicate that the annual NCPI varies between 31.94 (April, 1998) and 148.09 (June, 2017) with overall average of 81.53 and coefficient of variation 45.11. Based on the results of Jarque-Bera test statistic (21.41, p=0.0), it can be concluded that NCPI is significantly deviate from the normal distribution.

### 4.3.2 Temporal Variability of EP

The time plot of the Figure 4.4 shows the growth of Electricity Power Generation (EP) in Sri Lanka in GWh.

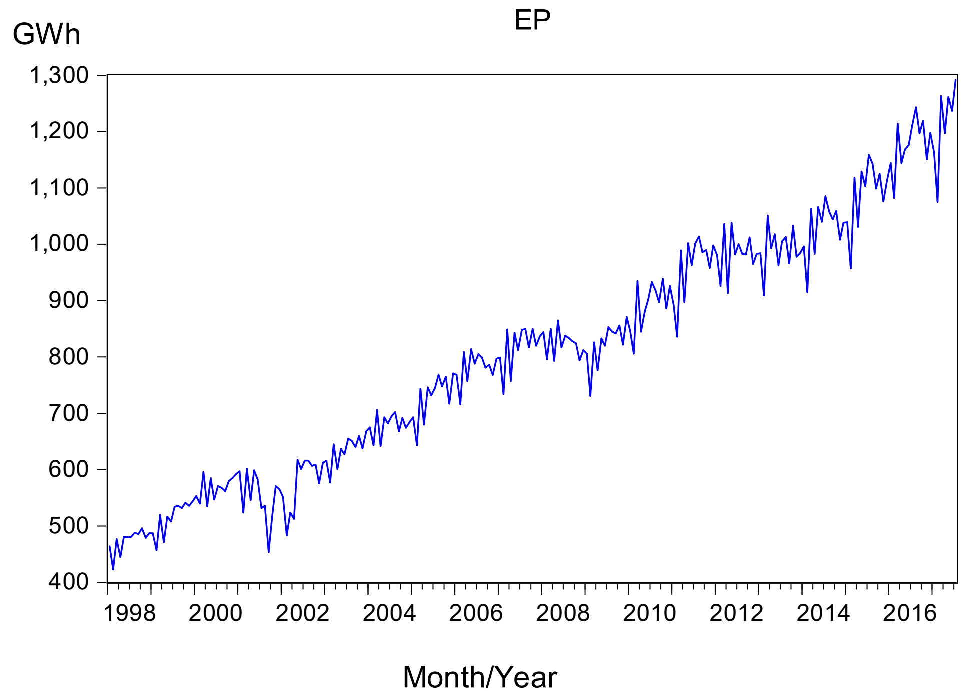


Figure 4.4 Temporal Variability of EP

According to the Figure 4.4, it is evident that EP shows an upward trend for the period from January 1998 (460 GWh) to July 2017 (1,292 GWh). Furthermore, it suggests that the data has no outliers or seasonal pattern. Therefore, it is obvious that the series is not stationary.

The basic statistics of EP is shown in the Table 4.3

Table 4. 3 Basic Statistics of EP

| **Statistic** | **Value** |
|---|---|
| Mean | 807.49 |
| Maximum | 1292 |
| Minimum | 423 |
| Standard Error of the mean | 14.11 |
| Coefficient of variation | 26.79 |
| Test Statistic of Jarque-Bera | 10.22 (p = 0.01) |

According to the Table 4.3 above, the average electricity power generation (EP) is 807.49 GWh for the period from January 1998 to July 2017. The data has spread between 423.0 GWh (February, 1998) and 1292.0 GWh (July, 2017) with the standard error 14.11 while the coefficient of variation is 26.79. Thus the data shows a considerable higher spread around the mean. Jarque-Bera test statistic of the time series is 10.22 with p-value 0.01. Thus it can be concluded with 95% of confidence that EP is significantly deviate from the normality.

### 4.3.3 Temporal Variability of EX

The time plot of the Figure 4.5 shows the growth of exchange rates (EX) US Dollar to Sri Lanka Rupees.

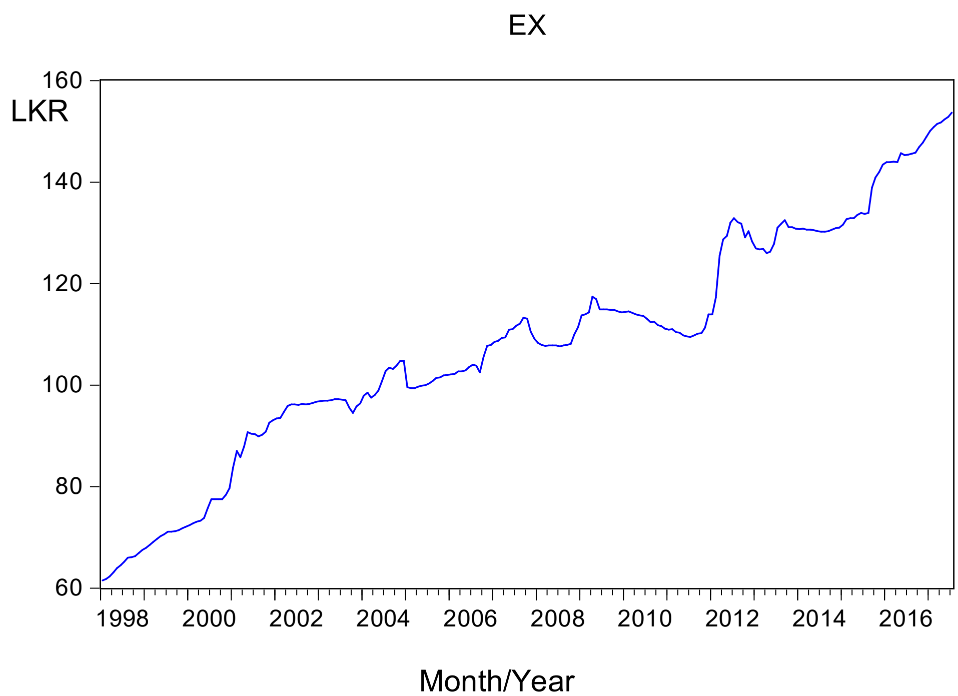


Figure 4.5 Temporal Variability of EX

According to the Figure 4.5, the plot shows that the EX has an upward trend with no outliers or seasonal pattern during the period January 1998 to July 2017. Since the plot shows the increasing trend it is clear that the series is not stationary.

The basic statistics of EX is shown in the Table 4.4

Table 4. 4 Basic Statistics of EX

| **Statistic** | **Value** |
|---|---|
| Mean | 107.88 |
| Maximum | 153.67 |
| Minimum | 61.5 |
| Standard Error of the mean | 1.47 |
| Coefficient of variation | 20.91 |
| Test Statistic of Jarque-Bera | 3.02 (p = 0.22) |

Table 4.4 describes that the average monthly exchange rate (EX) is 107.88, which means that the 1 US Dollar is equal to 107.88 Sri Lanka Rupees on average for the period from January 1998 to July 2017. The EX has spread between 61.50 (January, 1998) and 153.67 (July, 2017) with the standard error of 1.47. The coefficient of variation is 20.91. Thus it can be concluded that the variation of EX is high. Furthermore, the p-value of the Jarque-Bera test ($p>0.05$) confirms that the distribution of EX is not significantly deviate from normal distribution.

### 4.3.4 Temporal Variability of ME

The time plot of the Figure 4.6 shows the financial value of Merchandize Exports (ME) in US Dollars.

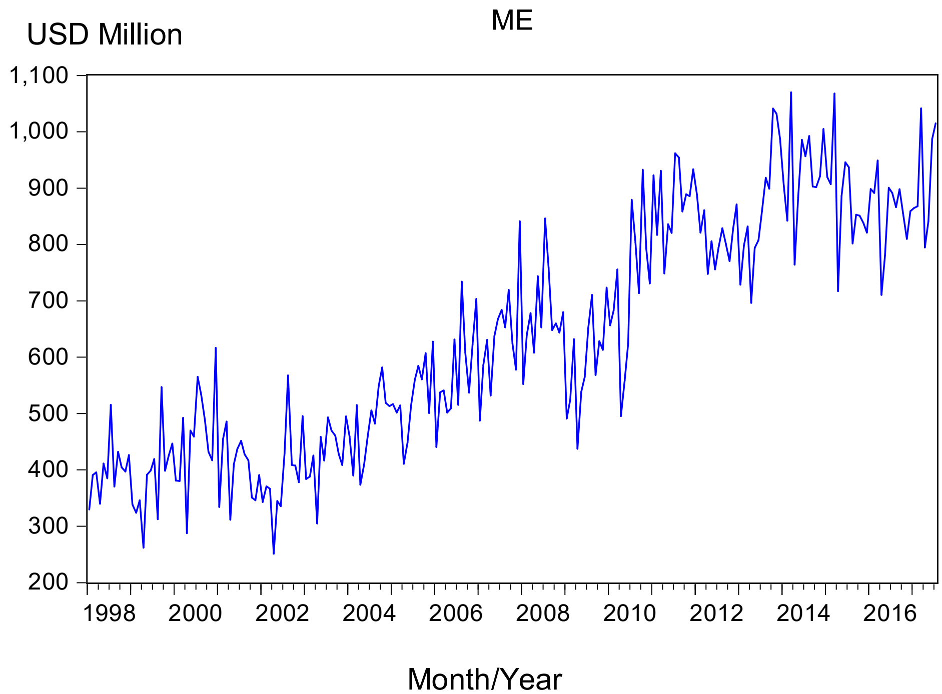


Figure 4.6 Temporal Variability of ME

According to the Figure 4.6, the plot shows that the ME has an upward trend during the period January 1998 to July 2017. Moreover, it time series plot shows that there exists no outliers or seasonal pattern in the data. The value of ME in January 1998 is almost US $ 350 Million and recorded above US $ 1,000 Million in July 2017. Therefore, it implies that the series is not stationary.

The basic statistics of ME is shown in the Table 4.5

Table 4. 5 Basic Statistics of ME

| **Statistic** | **Value** |
|---|---|
| Mean | 634.69 |
| Maximum | 1070.1 |
| Minimum | 251.3 |
| Standard Error of the mean | 13.67 |
| Coefficient of variation | 33.01 |
| Test Statistic of Jarque-Bera | 15.56 (p = 0.0) |

Table 4.5 illustrates the average merchandize exports (ME) for the period is US $ 634.69 Million and the standard error or the same is US $ 13.67 Million The range of the data is spread is US $ 251.3 Million (April, 2002) to 1070.1 Million (March, 2014). Since the coefficient of variation is 33.01, it can be concluded that the data dispersion of the data around the mean is high. Moreover, considering the Jarque-Bera test statistic of ME, it is evident that ME is significantly deviate from the normal distribution as the p-value of the statistic is less than 5%.

### 4.3.5 Temporal Variability of MI

The time plot of the Figure 4.7 shows the financial value of Merchandize Imports (MI) in US Dollar.

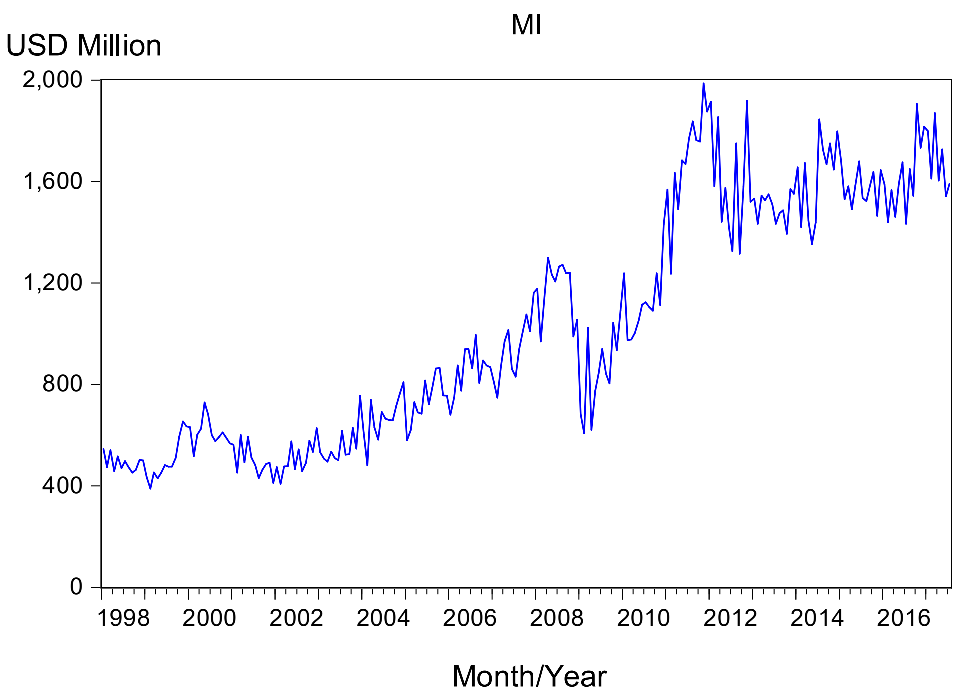


Figure 4.7 Temporal Variability of MI

According to the Figure 4.7, the plot shows that the MI shows an upward trend during the period January 1998 to July 2017 with no outliers and seasonal patterns in the period similar to the previous time plots. In January 1998, MI shows almost US $ 550 Million and recorded above US $ 1,600 Million in July 2017. Therefore, the plot shows that the series is not stationary.

The basic statistics of MI is shown in the Table 4.6

Table 4. 6 Basic Statistics of MI

| Statistic | Value |
|---|---|
| Mean | 1025.77 |
| Maximum | 1986.4 |
| Minimum | 389.1 |
| Standard Error of the mean | 30.65 |
| Coefficient of variation | 45.81 |
| Test Statistic of Jarque-Bera | 22.23 (p = 0.01) |

In view of the statistics of merchandize imports (MI) of all goods in the Table 4.6, it is evident that the average of the variable for the period US $ 1025.77 Million, while the standard error is 30.65. The range of the data is US $ 389.1 Million (February, 1999) to US $ 1986.4 Million (November, 2011). The coefficient of variation is 45.81, which confirms that the there is higher variation of the data around the mean. Based on the p-value of the Jarque–Bera test, it can be concluded with 95% confidence that the distribution of MI is significantly deviate from the normal distribution.

### 4.3.6 Temporal Variability of PT

The temporal variability of the distance operated in public transport (PT) in Km is shown in the Figure 4.8 and the basic statistics are shown in the Table 4.7.

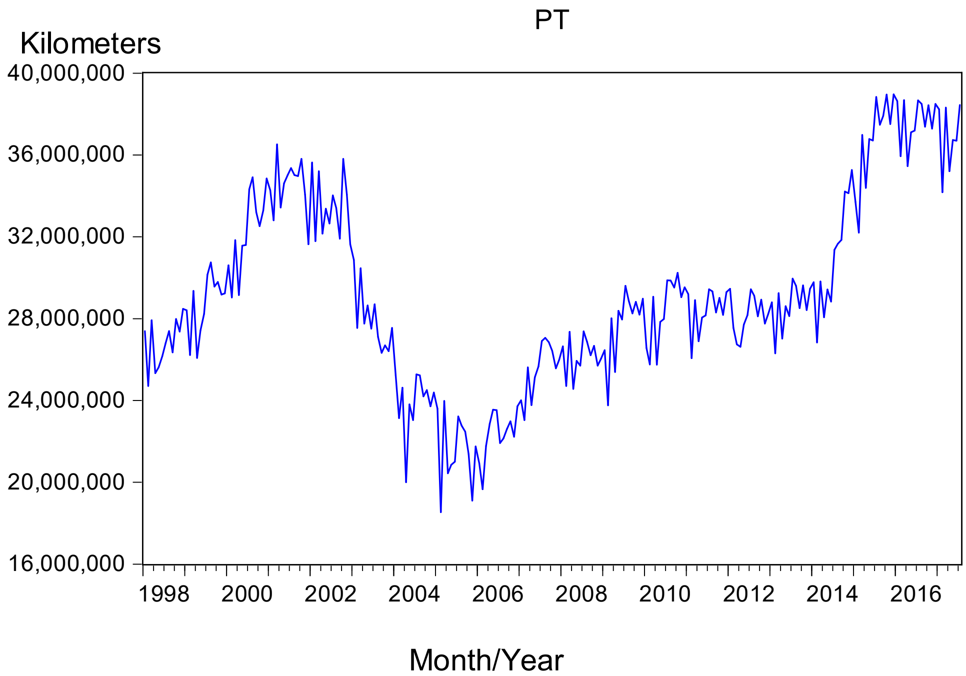


Figure 4.8 Temporal Variability of PT

According to the Figure 4.8, generally PT shows a slight upward trend for the period from January 1998 to January 2002. By observing the Figure 4.8 more closely, it is clear that the plot shows shifts in the trend during the period. PT increased gradually from almost 27.5 Million Km to more than 36 Million Km for the period from January 1998 to March 2001. PT shows a rapid downward trend until January 2006, where PT shows 18.5 Million Km approximately. The trend has become upward and continued until the September 2014, where PT shows approximately 28 Million Km. After October 2014 PT shows a rapid upward trend until June 2015 and record 39 Million Km approximately.

Table 4. 7 Basic Statistics of PT

| Statistic | Value |
|---|---|
| Mean | 29,169,197 |
| Maximum | 38,971,074 |
| Minimum | 18,533,139 |
| Standard Error of the mean | 306,322.4 |
| Coefficient of variation | 16.09 |
| Test Statistic of Jarque-Bera | 6.34 (p = 0.04) |

It can be seen that PT varies between 18,533,139 Km (February, 2005) and 38,971,074 Km (December, 2015) with mean of 29,169,197 Km and Coefficient of variation of 16.09. Considering the Jarque-Bera test statistic of PT, it is evident that the p-value of the test statistic is 0.04, which confirms that PT is significantly deviate from the normal distribution.

### 4.3.7 Temporal Variability of RT

The time plot of shows the distance operated in Railway Transport (RT) in Km. and basic statistics are shown in the Figure 4.9 and Table 4.8 respectively.

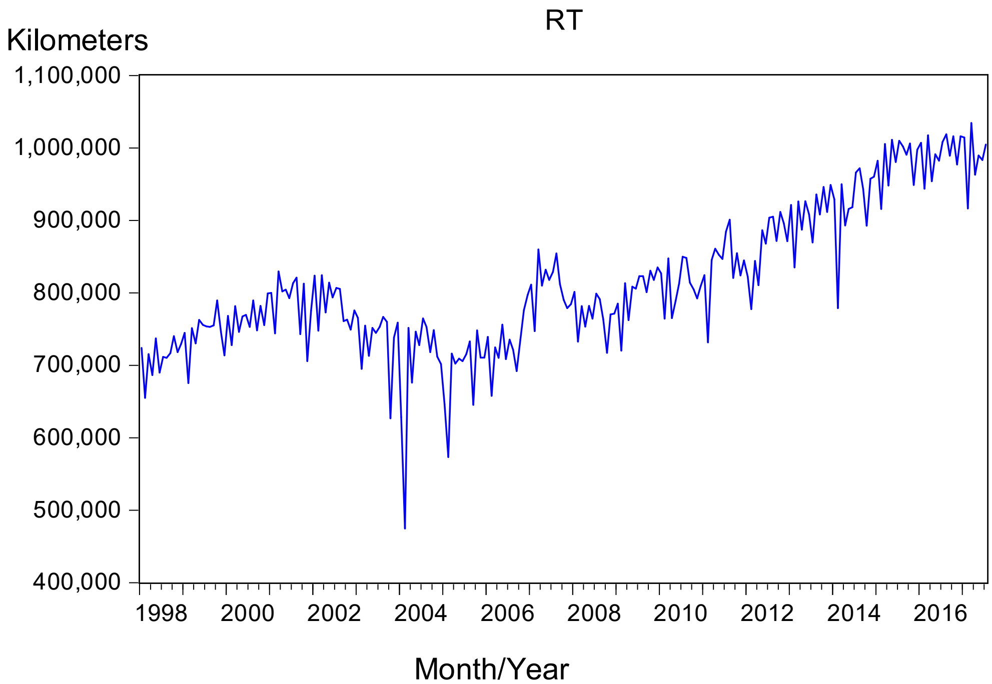


Figure 4.9 Temporal Variability of RT

According to the Figure 4.9, the plot shows that the RT has an upward trend during the period January 1998 to July 2017, although the series shows a sudden fall in February 2004. The value of RT in January 1998 is approximately 700,000 Km and recorded approximately 1 Million in July 2017. Therefore, it implies that the series is not stationary. Moreover, the Figure 4.9 implies that the data does not contain a seasonal pattern in the period.

Table 4. 8 Basic Statistics of RT

| **Statistic** | **Value** |
|---|---|
| Mean | 813,664 |
| Maximum | 1,034,442 |
| Minimum | 474,717 |
| Standard Error of the mean | 6,467.27 |
| Coefficient of variation | 12.18 |
| Test Statistic of Jarque-Bera | 5.17 (p = 0.08) |

Considering the statistics in the Table 4.8, it is evident that the average of the RT is 813,664 Km, while the standard error of the same is 6,467.27 Km. The range of the data spread is from 474,717 (February, 2004) to 1.03 Million Km (March, 2017) while the coefficient of variation is 12.18. Thus it is evident that the spread of the data around the mean is high. The Jarque-Bera test statistic is 5.17 with p-value of 0.08 confirming that the distribution of RT is normal with 95% confidence.

### 4.3.8 Temporal Variability of TA

The time plot of the Figure 4.10 shows the temporal variability of total number of Tourists Arrivals (TA)

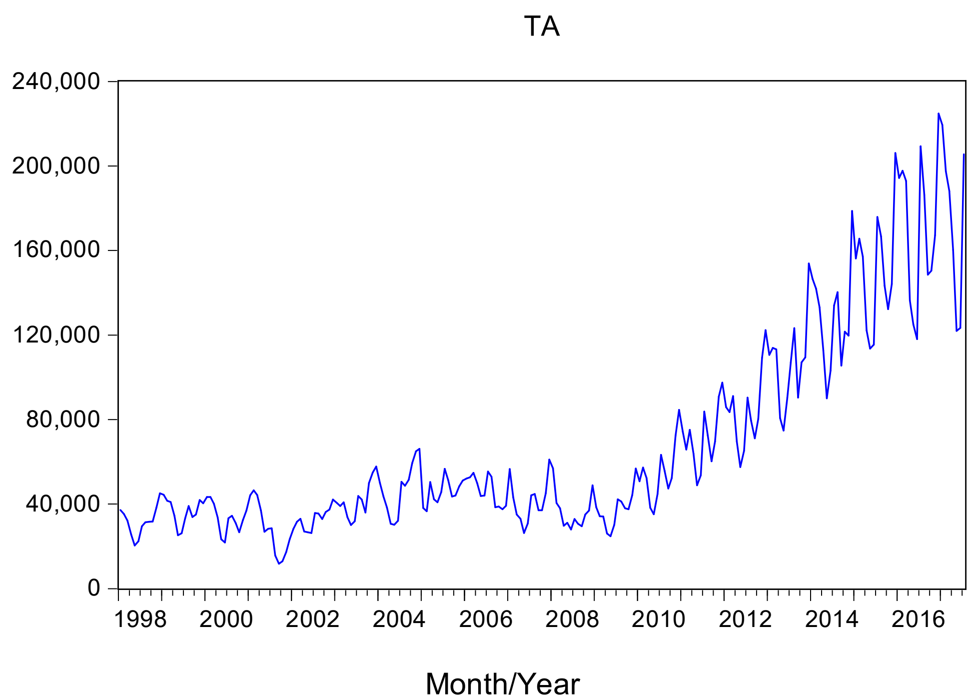


Figure 4.10 Temporal Variability of TA

According to the Figure 4.10, the plot shows that the TA has fluctuated around 40,000 tourists per month during the period January 1998 to July 2010. After August 2010, TA shows a rapid increment with a multiplicative seasonal pattern and recorded approximately 200,000 in July 2017. Therefore, it implies that the series TA is not stationary and no outlier exists in the series.

The basic statistics of TA is shown in the Table 4.9

Table 4. 9 Basic Statistics of TA

| **Statistic** | **Value** |
|---|---|
| Mean | 67,604 |
| Maximum | 224,791 |
| Minimum | 11,758 |
| Standard Error of the mean | 3,143 |
| Coefficient of variation | 71.27 |
| Test Statistic of Jarque-Bera | 94.51 ($p = 0.0$) |

Table 4.9 describes that the average monthly number of tourists arrived (TA) is 67,604, which means that more than 67,000 tourists arrive on average per month for the period from January 1998 to July 2017. The number of tourist arrival has spread between 11,758 and 224,791 with the standard error of 3,143. Thus the standard error is considerably high similar to the variables describes described above. Furthermore, the coefficient of variation is 71.27 implying that the dispersion of the data is very high. Jarque-Bera test statistic of TA is 94.51 with p-value 0.0. Thus it can be concluded with 95% of confidence that TA is significantly deviate from the normality.

### 4.3.9 Temporal Variability of Monthly VR

The time plot of the Figure 4.11 shows the number of vehicle registered (VR).

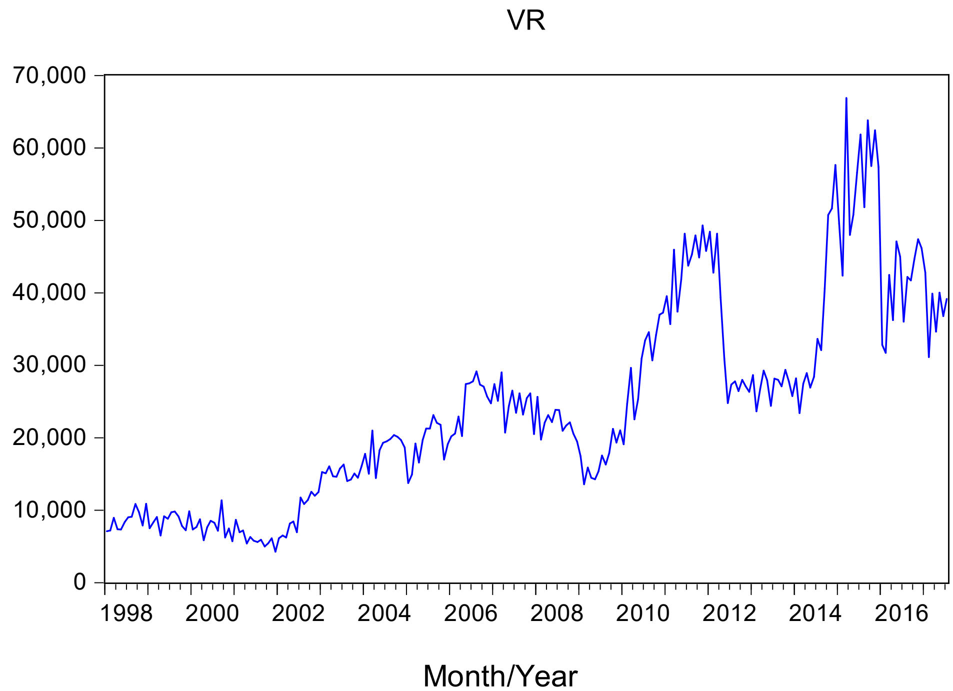


Figure 4.11 Temporal Variability of VR

According to the Figure 4.11, the plot shows generally that the VR has an upward trend during the period January 1998 to July 2017. The value of VR in January 1998 is under 10,000 and recorded above 40,000 in July 2017. Furthermore, it is clear that there are 2 downward shifts, which are recorded in June 2012 and January 2016. The Figure 4.11 further shows that there is 1 upward shift in January 2015. Therefore, it implies that the series is not stationary without any outliers.

The basic statistics of VR is shown in the Table 4.10

Table 4. 10 Basic Statistics of VR

| **Statistic** | **Value** |
|---|---|
| Mean | 24,052 |
| Maximum | 66,889 |
| Minimum | 4,258 |
| Standard Error of the mean | 923.87 |
| Coefficient of variation | 58.88 |
| Test Statistic of Jarque-Bera | 21.9 (p = 0.0) |

According to the Table 4.10, the average number of vehicle registered (VR) for the period from January 1998 to July 2017 is 24,052 and the standard error of the mean is 923.87. VR varies between 4,258 (December, 2001) and 66,889 (March, 2015). The coefficient of variation (58.88) is considerably high. Based on the results of Jarque-Bera test it can be concluded that with 95% confidence that the distribution of VR is also significantly deviate from normal distribution.

### 4.3.10 Temporal Variability of WR

The time plot and basic statistic of the wages rate of government employees (WR) are shown in the Figure 4.12 and Table 4.11 respectively.

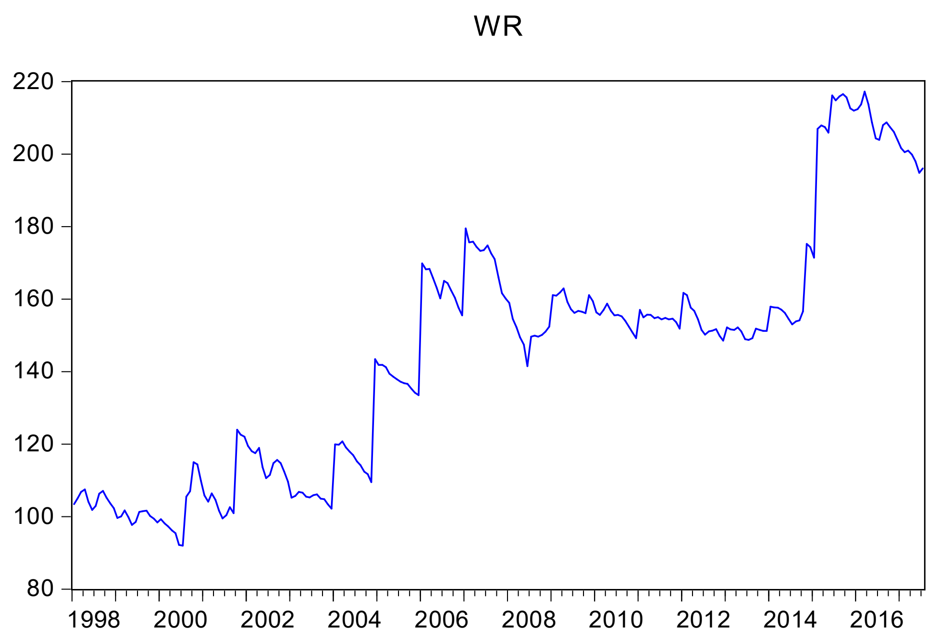


Figure 4.12 Temporal Variability of WR

According to the Figure 4.12, the plot shows that the WR has an upward trend during the period January 1998 to July 2017 with an upward shift in December 2014. The value of WR in January 1998 is above 100 and recorded above 200 in July 2017. Therefore, it can be hypothesized that the series is not stationary.

Table 4. 11 Basic Statistics of WR

| **Statistic** | **Value** |
|---|---|
| Mean | 145.11 |
| Maximum | 217.2 |
| Minimum | 92 |
| Standard Error of the mean | 2.21 |
| Coefficient of variation | 23.31 |
| Test Statistic of Jarque-Bera | 8.54 ($p = 0.01$) |

According to the Table 4.11, it is evident that the average of the variable is 145.1 implying that if monthly wages of a government employee Rs. 100 in 1978, the wage of the employees is Rs. 145.1 on average per month during the period 1998 to 2017. Further, the standard error of the same is 2.21. The range of the data spread is from 92 to 217. 2. The coefficient of variation is 23.31. Thus it is clear that the spread of the data around the mean is high. The Jarque-Bera test statistic of WR is 8.54 with the p-value of 0.01, which confirms that time series is significantly deviate from the normal distribution at 5% level of significance.

## 4.4 Stabilize the Variations of the Observed Time Series

In the section 4.3 the temporal variability of the consumption of diesel with the 10 explanatory variables was described and found that the variability of all the variables is increasing with the time. Thus it is obvious that the variance is not homogeneous with the time. In order to minimize the heteroskedasticity of the variances, all the time series were transformed in to its logarithm ($Log_e$). The plot of log transformed time series, consumption of diesel (LNDC) is shown in Figure 4.13

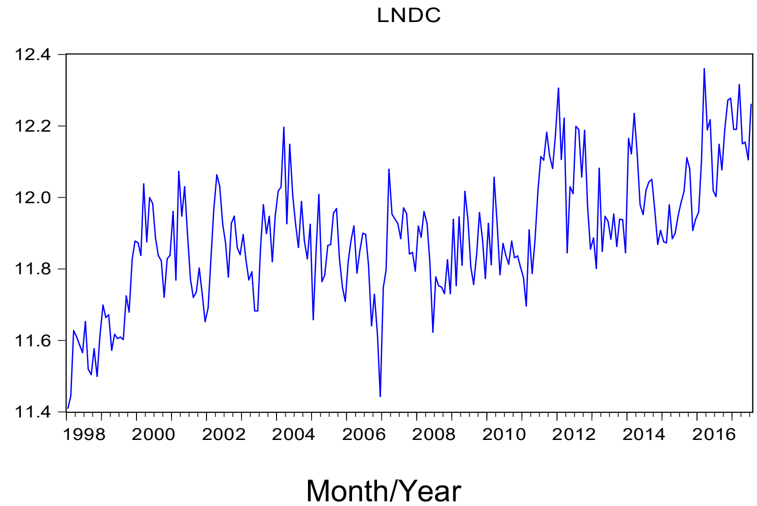


Figure 4. 13 Time Plot of LNDC

According to the Figure 4.13, there is an upward trend in LNDC time series and it is similar to its original time series described in the Figure 4.1. The values of LNDC in January 1998 and July 2017 are 11.4 and 12.2 respectively. Thus it seems that LNDC is not stationary.

The plots of log transformed time series, of the 10 explanatory variables are shown in Figure 4.14.

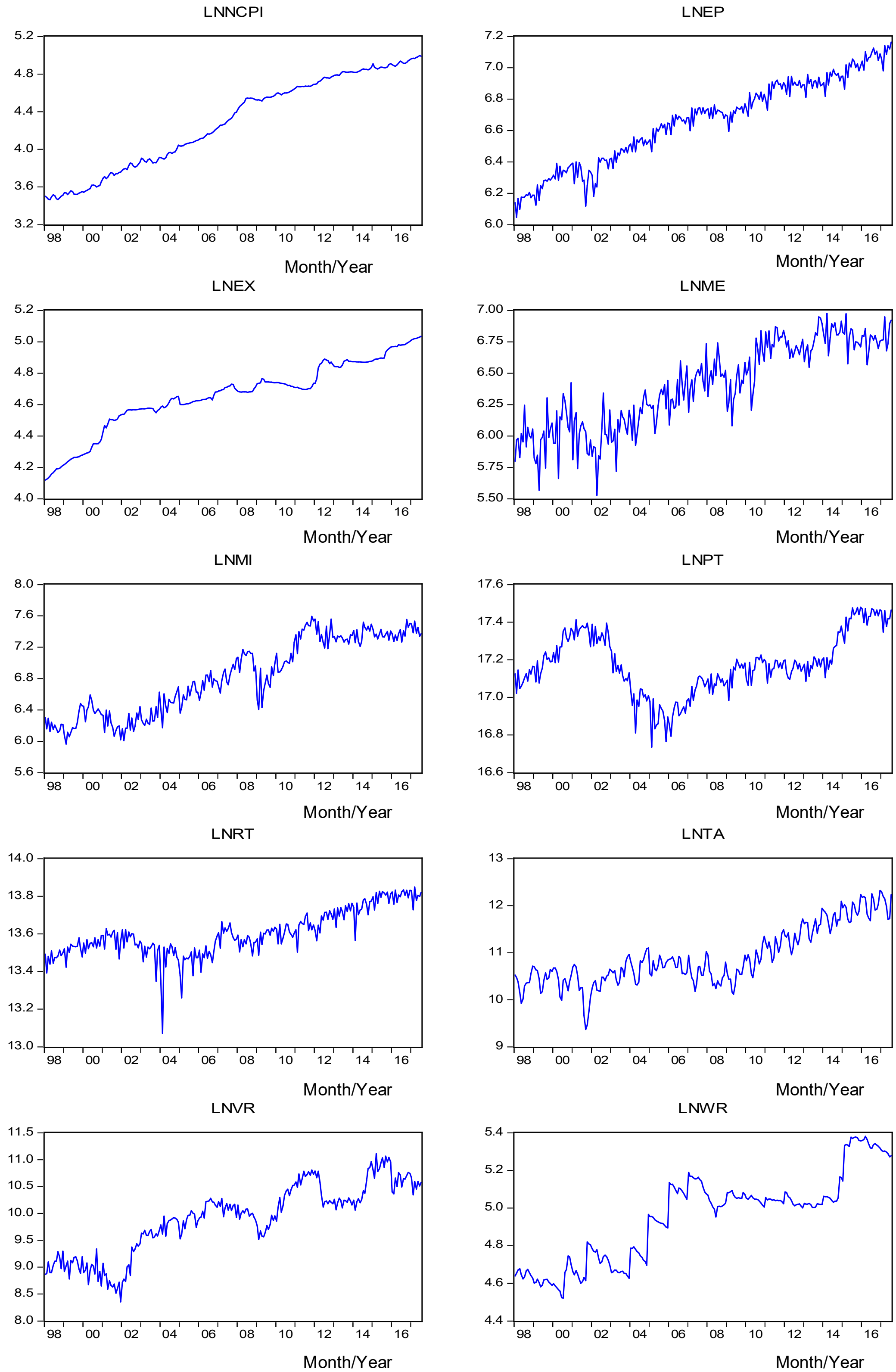


Figure 4. 14 Time Plots of 10 Explanatory Variables (After Log Transformation)

According to the Figure 4.14, all the time plots show that there is upward trend in all the time series and the trends are similar to the corresponding original time series described in the Figure 4.3 to 4.12. Thus it is clear that all the explanatory variables are time variant.

## 4.5 Distribution of Each Log Time Series

The graphical representations of the distributions of log-transformed time series are shown in the Figure 4.15 to 4.25

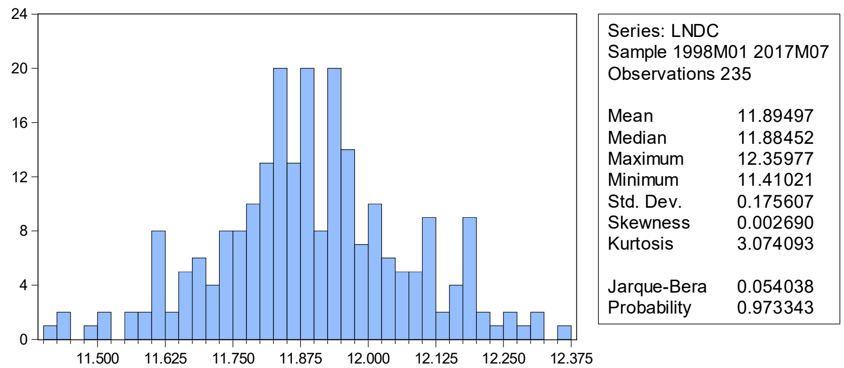


Figure 4. 15 Distribution of LNDC

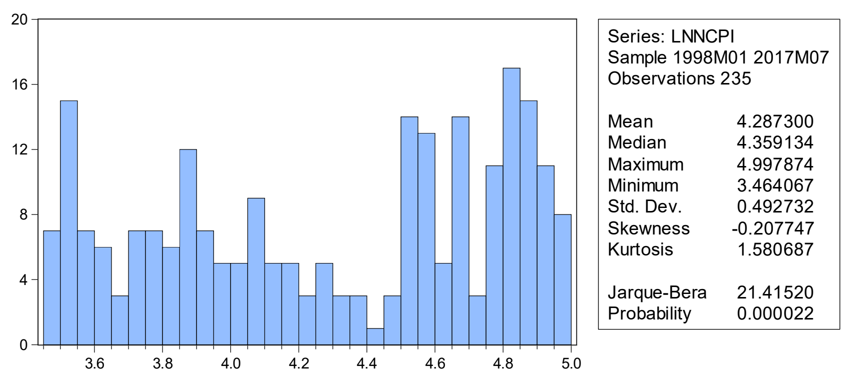


Figure 4. 16 Distribution of LNNCPI

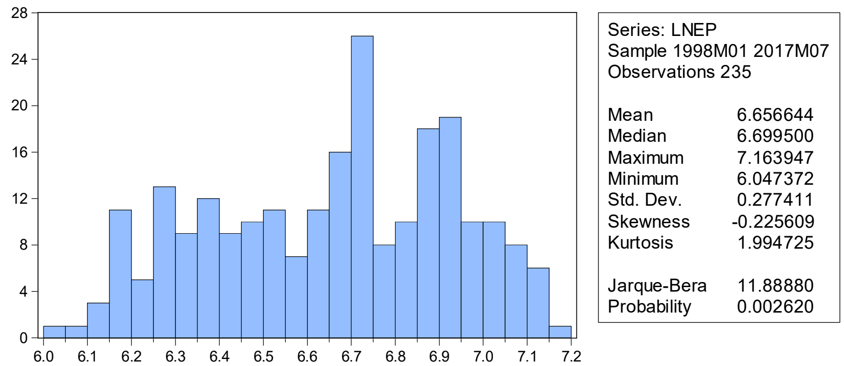


Figure 4. 17 Distribution of LNEP

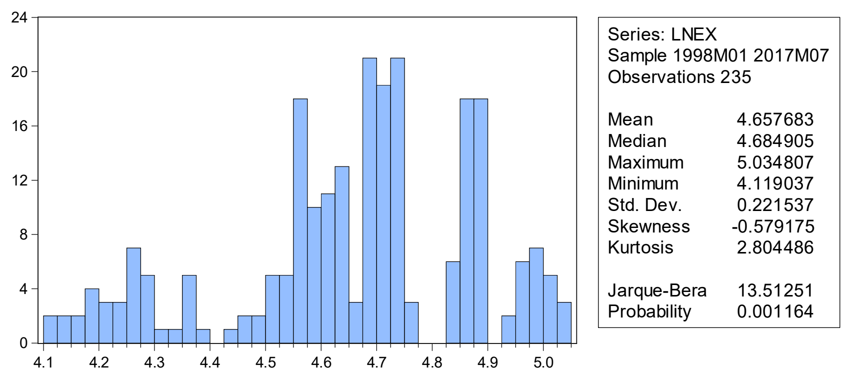


Figure 4. 18 Distribution of LNEX

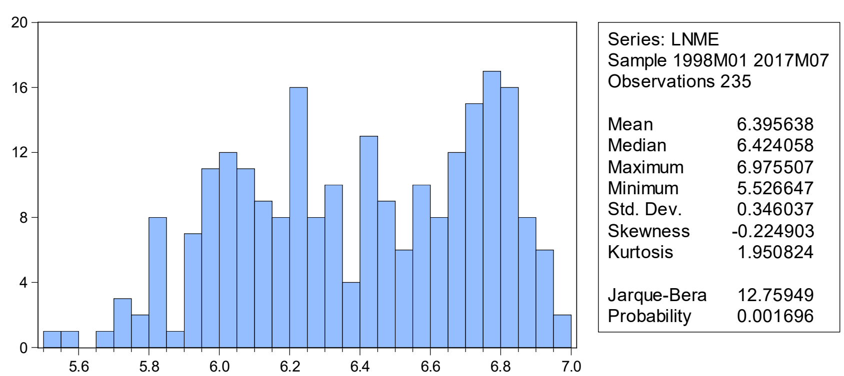


Figure 4. 19 Distribution of LNME

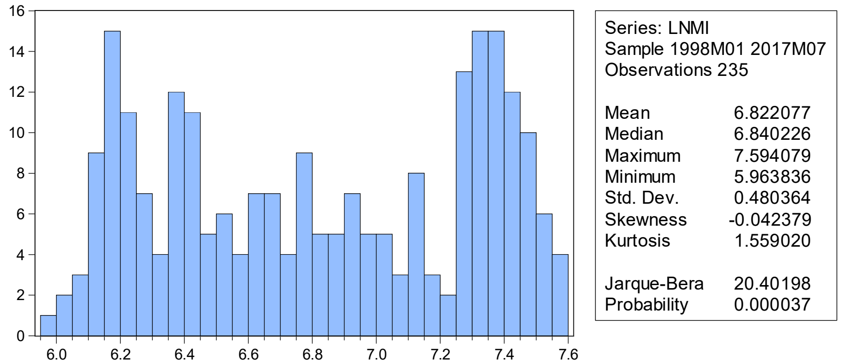


Figure 4. 20 Distribution of LNMI

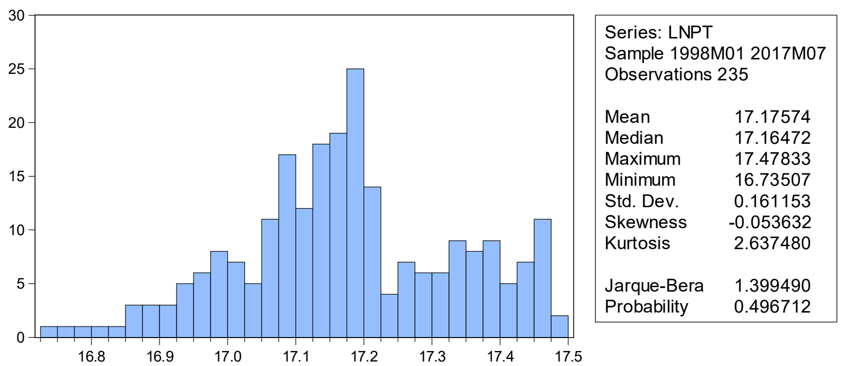


Figure 4. 21 Distribution of LNPT

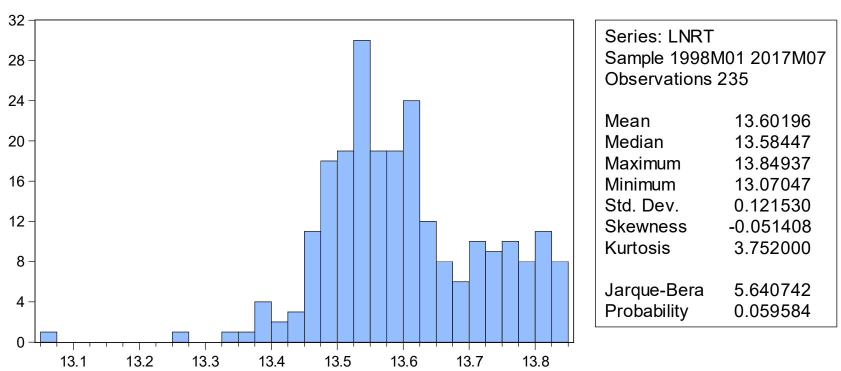


Figure 4. 22 Distribution of LNRT

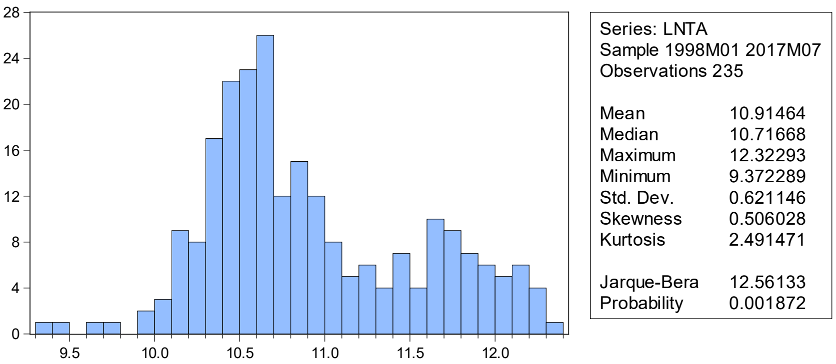


Figure 4. 23 Distribution of LNTA

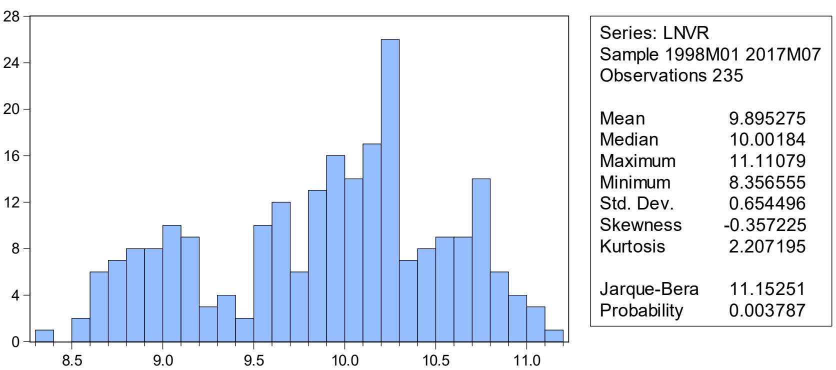


Figure 4. 24 Distribution of LNVR

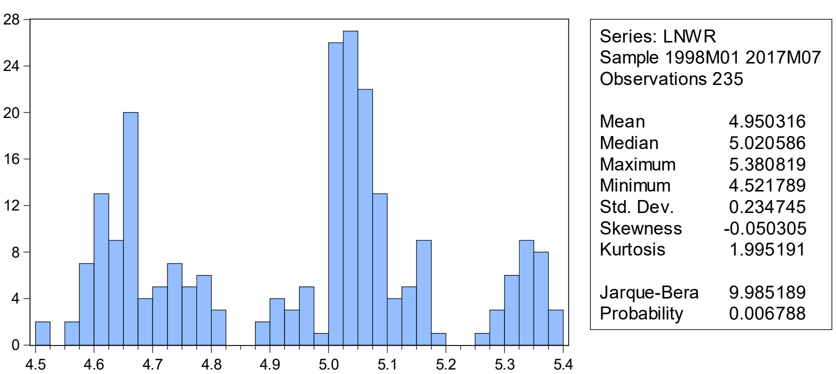


Figure 4. 25 Distribution of LNWR

According to Figure 4.15, it is evident that the mean of LNDC is 11.89 and the data varied between 11.41 (minimum) and 12.35 (maximum) for the period from January 1998 to July 2017. Since the p-value (0.973) of the Jarque-Bera test (test statistic = 0.054) is greater than 0.05, it can be concluded with 95% confidence that the distribution of LNDC is not significantly deviate from normal distribution.

The Figures 4.15 – 4.25 clearly indicate that the log transformation was able to minimize the heteroskedasticity of all the variables and the difference between maximum value and the corresponding minimum value is much smaller than the corresponding range for the raw data. In other words log transformation was useful to stabilize the variance of the series. This generally helps for substantial forecasting improvement. However, it should be pointed out that LNPT and LNRT is not significantly deviate from normal distribution while the rest of log transformed the variables deviate significantly from normal distribution.

## 4.6 Correlation between the Observed Series

The correlations of the observed time series are described in the Table 4.12 and corresponding p-values of the correlations are shown in the Table 4.13.

Table 4. 12 Correlations the Log-Transformed Time Series

| | LN DC | LN NCPI | LN EP | LN EX | LN ME | LN MI | LN PT | LN RT | LN TA | LN VR | LN WR |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LNDC | 1.0 | 0.6 | 0.6 | 0.7 | 0.5 | 0.6 | 0.4 | 0.5 | 0.6 | 0.5 | 0.5 |
| LNNCPI | 0.6 | 1.0 | 1.0 | 0.9 | 0.9 | 0.9 | 0.2 | 0.7 | 0.8 | 0.9 | 0.9 |
| LNEP | 0.6 | 1.0 | 1.0 | 0.9 | 0.9 | 0.9 | 0.3 | 0.8 | 0.8 | 0.9 | 0.9 |
| LNEX | 0.7 | 0.9 | 0.9 | 1.0 | 0.8 | 0.8 | 0.3 | 0.7 | 0.8 | 0.8 | 0.9 |
| LNME | 0.5 | 0.9 | 0.9 | 0.8 | 1.0 | 0.9 | 0.2 | 0.7 | 0.8 | 0.9 | 0.8 |
| LNMI | 0.6 | 0.9 | 0.9 | 0.8 | 0.9 | 1.0 | 0.2 | 0.7 | 0.8 | 0.9 | 0.8 |
| LNPT | 0.4 | 0.2 | 0.3 | 0.3 | 0.2 | 0.2 | 1.0 | 0.7 | 0.4 | 0.1 | 0.2 |
| LNRT | 0.5 | 0.7 | 0.8 | 0.7 | 0.7 | 0.7 | 0.7 | 1.0 | 0.7 | 0.6 | 0.7 |
| LNTA | 0.6 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.4 | 0.7 | 1.0 | 0.8 | 0.7 |
| LNVR | 0.5 | 0.9 | 0.9 | 0.8 | 0.9 | 0.9 | 0.1 | 0.6 | 0.8 | 1.0 | 0.9 |
| LNWR | 0.5 | 0.9 | 0.9 | 0.9 | 0.8 | 0.8 | 0.2 | 0.7 | 0.7 | 0.9 | 1.0 |

Table 4. 13 The p-values of the Corresponding Correlations in Table 4.12

| | LN DC | LN NCPI | LN EP | LN EX | LN ME | LN MI | LN PT | LN RT | LN TA | LN VR | LN WR |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LNDC | - | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LNNCPI | 0.0 | - | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LNEP | 0.0 | 0.0 | - | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LNEX | 0.0 | 0.0 | 0.0 | - | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LNME | 0.0 | 0.0 | 0.0 | 0.0 | - | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LNMI | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LNPT | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 0.0 | 0.0 | 0.1 | 0.0 |
| LNRT | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 0.0 | 0.0 | 0.0 |
| LNTA | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 0.0 | 0.0 |
| LNVR | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | - | 0.0 |
| LNWR | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - |

According to the Table 4.12 it is evident that all the correlation coefficients are greater than 0.5 except few pairs. However, based on the corresponding p-values, it can be concluded that the correlation between any pair is significantly higher than zero. All the correlations are positively correlated. Furthermore, according to the Table 4.12, it is clear that LNNCPI, LNEP, LNEX, LNMI, and LNTA show strong correlation with LNDC than the other variables, as the corresponding correlation coefficients are greater than or equal 0.6 with p-value of 0.0.

Moreover, all the p-values shown in the Table 4.13 are zero except the p-value of the correlation between LNPT and LNVR, which is 0.1. Therefore, it can be concluded with 95% confidence that correlation between LNPT and LNVR is not significantly different from zero.

## 4.7 Summary of Chapter 4

In this chapter, the temporal variability of the DC and all 10 explanatory variables were studied. It was found that none of the series are stationary. Furthermore all the time series significantly deviate from normal distribution except EX and RT. Moreover, it is found that the variances of all time series are considerably high and hence log transformation was used to stabilize the variance of the all the variables. After the log transformation, the correlations between each pair of variables are significantly greater than zero except correlation between LNPT and LNVR., Furthermore, it was found that LNNCPI, LNEP, LNEX, LNMI, and LNTA has significantly strong correlations with LNDC.

# CHAPTER 5

# DEVELOPMENT OF VECTOR ERROR CORRECTION MODEL USING ALL EXPLANATORY VARIABLES

This chapter will contain extensive analysis of the consumption of diesel in Sri Lanka. The initial part of the analysis in this chapter is to discover whether the observed series are stationary since the stationary time series play a major part in later sections of the study. More extensive analysis of LNDC will be performed with VEC model using all the 10 log transformed ($Log_e$) explanatory time series variables and the predicting accuracy of the model will be performed.

## 5.1 Stationary of the Original Log Series

The results of Dicky Fuller test for each of the 11 series after log transformation are shown in Table 5.1-5.11.

Table 5. 1 Test for Stationarity of LNDC (Level)

Null Hypothesis: LNDC has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -5.402372 | 0.0001 |
| Test critical values: | 1% level | -3.997930 | |
| | 5% level | -3.429229 | |
| | 10% level | -3.138092 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 2 Test for Stationarity of LNEP (Level)

Null Hypothesis: LNEP has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 12 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -2.927949 | 0.1557 |
| Test critical values: | 1% level | -3.999930 | |
| | 5% level | -3.430196 | |
| | 10% level | -3.138663 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 3 Test for Stationarity of LNEX (Level)

Null Hypothesis: LNEX has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -2.888185 | 0.1683 |
| Test critical values: | 1% level | -3.997930 | |
| | 5% level | -3.429229 | |
| | 10% level | -3.138092 | |

Table 5. 4 Test for Stationarity of LNME (Level)

Null Hypothesis: LNME has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 13 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -2.763791 | 0.2125 |
| Test critical values: | 1% level | -4.000122 | |
| | 5% level | -3.430289 | |
| | 10% level | -3.138717 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 5 Test for Stationarity of LNMI (Level)

Null Hypothesis: LNMI has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -3.099745 | 0.1088 |
| Test critical values: | 1% level | -3.998104 | |
| | 5% level | -3.429313 | |
| | 10% level | -3.138142 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 6 Test for Stationarity of LNCPI (Level)

Null Hypothesis: LNNCPI has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 3 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -0.287805 | 0.9906 |
| Test critical values: | 1% level | -3.998280 | |
| | 5% level | -3.429398 | |
| | 10% level | -3.138192 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 7 Test for Stationarity of LNPT (Level)

Null Hypothesis: LNPT has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 14 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -2.319754 | 0.4211 |
| Test critical values: | 1% level | -4.000316 | |
| | 5% level | -3.430383 | |
| | 10% level | -3.138772 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 8 Test for Stationarity of LNRT (Level)

Null Hypothesis: LNRT has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 12 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -1.418795 | 0.8531 |
| Test critical values: | 1% level | -3.999930 | |
| | 5% level | -3.430196 | |
| | 10% level | -3.138663 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 9 Test for Stationarity of LNTA (Level)

Null Hypothesis: LNTA has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 12 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -1.966374 | 0.6160 |
| Test critical values: | 1% level | -3.999930 | |
| | 5% level | -3.430196 | |
| | 10% level | -3.138663 | |

Table 5. 10 Test for Stationarity of LNVR (Level)

Null Hypothesis: LNVR has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -2.617615 | 0.2729 |
| Test critical values: | 1% level | -3.997930 | |
| | 5% level | -3.429229 | |
| | 10% level | -3.138092 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 11 Test for Stationarity of LNWR (Level)

Null Hypothesis: LNWR has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -2.880492 | 0.1708 |
| Test critical values: | 1% level | -3.997758 | |
| | 5% level | -3.429146 | |
| | 10% level | -3.138043 | |

*MacKinnon (1996) one-sided p-values.

Based on the results of Augmented Dickey Fuller tests (-5.40, p-value = 0.0) in the Table 5.1, it can be concluded that LNDC is stationary time series at 5% level of significant. However the results of Table 5.2 – 5.11 indicate that the all the 10 explanatory time series are not stationary as the corresponding p-values are greater than 5%.

## 5.2 Stationary of the First Difference Series

The results of Dicky Fuller test for each of the 11 first difference log series are shown in Table 5.12-5.22.

Table 5. 12 Test for Stationarity of LNDC (First Difference)

Null Hypothesis: D(LNDC) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 3 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -11.24097 | 0.0000 |
| Test critical values: | 1% level | -3.998457 | |
| | 5% level | -3.429484 | |
| | 10% level | -3.138243 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 13 Test for Stationarity of LNEP (First Difference)

Null Hypothesis: D(LNEP) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 11 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -5.004831 | 0.0003 |
| Test critical values: | 1% level | -3.999930 | |
| | 5% level | -3.430196 | |
| | 10% level | -3.138663 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 14 Test for Stationarity of LNEX (First Difference)

Null Hypothesis: D(LNEX) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -10.37929 | 0.0000 |
| Test critical values: | 1% level | -3.997930 | |
| | 5% level | -3.429229 | |
| | 10% level | -3.138092 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 15 Test for Stationarity of LNME (First Difference)

Null Hypothesis: D(LNME) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 12 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -4.913798 | 0.0004 |
| Test critical values: | 1% level | -4.000122 | |
| | 5% level | -3.430289 | |
| | 10% level | -3.138717 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 16 Test for Stationarity of LNMI (First Difference)

Null Hypothesis: D(LNMI) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -16.84219 | 0.0000 |
| Test critical values: | 1% level | -3.998104 | |
| | 5% level | -3.429313 | |
| | 10% level | -3.138142 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 17 Test for Stationarity of LNCPI (First Difference)

Null Hypothesis: D(LNNCPI) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 3 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -7.909739 | 0.0000 |
| Test critical values: | 1% level | -3.998457 | |
| | 5% level | -3.429484 | |
| | 10% level | -3.138243 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 18 Test for Stationarity of LNPT (First Difference)

Null Hypothesis: D(LNPT) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 12 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -7.895003 | 0.0000 |
| Test critical values: | 1% level | -4.000122 | |
| | 5% level | -3.430289 | |
| | 10% level | -3.138717 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 19 Test for Stationarity of LNRT (First Difference)

Null Hypothesis: D(LNRT) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 11 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -6.150922 | 0.0000 |
| Test critical values: | 1% level | -3.999930 | |
| | 5% level | -3.430196 | |
| | 10% level | -3.138663 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 20 Test for Stationarity of LNTA (First Difference)

Null Hypothesis: D(LNTA) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 11 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -4.865197 | 0.0005 |
| Test critical values: | 1% level | -3.999930 | |
| | 5% level | -3.430196 | |
| | 10% level | -3.138663 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 21 Test for Stationarity of LNVR (First Difference)

Null Hypothesis: D(LNVR) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -25.15600 | 0.0000 |
| Test critical values: | 1% level | -3.997930 | |
| | 5% level | -3.429229 | |
| | 10% level | -3.138092 | |

*MacKinnon (1996) one-sided p-values.

Table 5. 22 Test for Stationarity of LNWR (First Difference)

Null Hypothesis: D(LNWR) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -15.81448 | 0.0000 |
| Test critical values: | 1% level | -3.997930 | |
| | 5% level | -3.429229 | |
| | 10% level | -3.138092 | |

*MacKinnon (1996) one-sided p-values.

Based on the results of the Augmented Dickey Fuller tests shown from Table 5.12 to Table 5.22, it can be confirmed that first difference of all the series are significant at 5% level and hence the first difference of the log-transformed series are stationary. Thus it can be concluded all series are I (1), which means that all the series are integrated at one.

## 5.3 Existence of Cointegration

Since the results of the Augmented Dickey Fuller tests indicate that all series are stationary in their first difference, Jahansen co-integration test was carried out to detect the existence of a linear combination of the stationary time series variables. The results of the Johansen test for co-integration using the maximum eigenvalue statistic and the trace statistic test are shown in Table 5.23 and Table 5.24 respectively.

Table 5. 23 Unrestricted Cointegration Rank Test (Maximum Eigen value)

| Unrestricted Cointegration Rank Test (Maximum Eigenvalue) | | | | |
|---|---|---|---|---|
| Hypothesized No. of CE(s) | Eigenvalue | Max-Eigen Statistic | 0.05 Critical Value | Prob.** |
| None * | 0.439172 | 141.6934 | 70.53513 | 0.0000 |
| At most 1 * | 0.337531 | 100.8863 | 64.50472 | 0.0000 |
| At most 2 * | 0.246176 | 69.23606 | 58.43354 | 0.0032 |
| At most 3 * | 0.207916 | 57.10667 | 52.36261 | 0.0152 |
| At most 4 | 0.151667 | 40.29819 | 46.23142 | 0.1881 |
| At most 5 | 0.109327 | 28.36548 | 40.07757 | 0.5359 |
| At most 6 | 0.070926 | 18.02379 | 33.87687 | 0.8764 |
| At most 7 | 0.063578 | 16.09387 | 27.58434 | 0.6578 |
| At most 8 | 0.038362 | 9.583687 | 21.13162 | 0.7827 |
| Max-eigenvalue test indicates 4 cointegrating eqn(s) at the 0.05 level | | | | |
| * denotes rejection of the hypothesis at the 0.05 level | | | | |
| **MacKinnon-Haug-Michelis (1999) p-values | | | | |

Table 5. 24 Unrestricted Cointegration Rank Test (Trace)

| Included observations: 233 after adjustments | | | | |
|---|---|---|---|---|
| Trend assumption: Linear deterministic trend | | | | |
| Series: LNDC LNNCPI LNEP LNEX LNME LNMI LNPT LNRT LNTA LNVR LNWR | | | | |
| Lags interval (in first differences): 1 to 1 | | | | |
| Unrestricted Cointegration Rank Test (Trace) | | | | |
| Hypothesized No. of CE(s) | Eigenvalue | Max-Eigen Statistic | 0.05 Critical Value | Prob.** |
| None * | 0.438148 | 479.3541 | 285.1425 | 0.0000 |
| At most 1 * | 0.348643 | 345.0257 | 239.2354 | 0.0000 |
| At most 2 * | 0.284394 | 245.1393 | 197.3709 | 0.0000 |
| At most 3 * | 0.204559 | 167.1714 | 159.5297 | 0.0179 |
| At most 4 | 0.149682 | 113.8473 | 125.6154 | 0.2080 |
| At most 5 | 0.114994 | 76.06756 | 95.75366 | 0.5040 |
| At most 6 | 0.070204 | 47.60422 | 69.81889 | 0.7376 |
| At most 7 | 0.063048 | 30.64408 | 47.85613 | 0.6855 |
| At most 8 | 0.037008 | 15.47044 | 29.79707 | 0.7488 |
| Max-eigenvalue test indicates 4 cointegrating eqn(s) at the 0.05 level | | | | |
| * denotes rejection of the hypothesis at the 0.05 level | | | | |
| **MacKinnon-Haug-Michelis (1999) p-values | | | | |

Results in Table 5.23 and Table 5.24 indicate that the both test results give the same results, confirming that there exists co-integration at 5% level of significance. Furthermore, the results in both Tables 5.23 and 5.24 indicate that both tests reject the three hypothesis: (i) there exists at most one, (ii) there exists at most two and (iii) there exist at most three, at the 5% significance level. Therefore, it can be concluded that there are 4 co-integrating linear combinations of the all 11 time series based on the two test statistics. Hence, it can be concluded that there is long run equilibrium relationship between LNDC and other 10 explanatory variables with 95%

confidence. That means all series move together for long term equilibrium. Thus, the VEC model can be developed for the first differenced series.

### 5.4 Optimal Lag Length

In order to use the VEC model, it is required to find the optimum lag length of the model using the results in the Table 5.25.

Table 5. 25 VAR Lag Order Selection Criteria

VAR Lag Order Selection Criteria
Endogenous variables: LNDC LNNCPI LNEP LNEX LNME LNMI LNPT LNRT LNTA LNVR LNWR
Exogenous variables: C
Sample: 1998M01 2017M07

| Lag | LogL | LR | FPE | AIC | SC | HQ |
|---|---|---|---|---|---|---|
| 0 | 1581.395 | NA | 3.25e-20 | -13.65561 | -13.49118 | -13.58928 |
| 1 | 3867.081 | 4332.866 | 2.18e-28 | -32.47896 | -30.50581* | -31.68303 |
| 2 | 4101.154 | 421.3309 | 8.19e-29 | -33.46221 | -29.68032 | -31.93667* |
| 3 | 4262.944 | 275.7461 | 5.84e-29 | -33.81690 | -28.22628 | -31.56176 |
| 4 | 4400.845 | 221.8417 | 5.21e-29 | -33.96387 | -26.56453 | -30.97912 |
| 5 | 4540.891 | 211.8954* | 4.66e-29* | -34.12949* | -24.92141 | -30.41513 |

* indicates lag order selected by the criterion

According to the Table 5.25, the listed criterion values are Likelihood Ratio (LR), Final Prediction Error (FPE), Akaike Information Criterion (AIC), Schwarz information criterion, (SC) and Hannan-Quinn information criterion (HQ). It is evident that the LR, FPE, and AIC criterion values show the minimum values at the lag 5 while SC and HQ shows that the optimum lag length is 1 and 2 respectively. Since 3 out of 5 criterion values imply that the optimum lag is 5, the lag 5 was selected as the optimum lag for long run equilibrium and the VEC model.

## 5.5 VEC Model

The following Table 5.26 and Table 5.27 describe the long run equilibrium of VEC Model – VECM(5) for the LNDC and the error correction terms of the model with optimum lag length as 5.

Table 5. 26 Vector Error Correction Model Estimates

Sample (adjusted): 1998M07 2017M07
Standard errors in ( ) & t-statistics in [ ]

| Cointegrating Eq: | CointEq1 |
|---|---|
| LNDC(-1) | 1.000000 |
| LNNCPI(-1) | 0.875758<br>(0.13590)<br>[ 6.44410] |
| LNEP(-1) | -0.481129<br>(0.40567)<br>[-1.18602] |
| LNME(-1) | -0.183398<br>(0.18840)<br>[-0.97343] |
| LNMI(-1) | -0.549691<br>(0.14964)<br>[-3.67339] |
| LNEX(-1) | -1.568711<br>(0.25862)<br>[-6.06570] |
| LNPT(-1) | -0.632189<br>(0.15426)<br>[-4.09819] |
| LNRT(-1) | 1.601092<br>(0.42186)<br>[ 3.79528] |
| LNTA(-1) | 0.029352<br>(0.04331)<br>[ 0.67776] |
| LNVR(-1) | 0.085120<br>(0.05065)<br>[ 1.68040] |
| LNWR(-1) | 0.195351<br>(0.12880)<br>[ 1.51673] |
| C | -13.26406 |

Table 5. 26 (Continued) Vector Error Correction Model Estimates

| Error Correction: | D(LNDC) |
|---|---|
| CointEq1 | -0.218828 |
| | (0.09855) |
| | [-2.22047] |
| D(LNDC(-1)) | -0.122576 |
| | (0.11129) |
| | [-1.10139] |
| D(LNDC(-2)) | 0.048950 |
| | (0.10631) |
| | [ 0.46043] |
| D(LNDC(-3)) | -0.117562 |
| | (0.09845) |
| | [-1.19411] |
| D(LNDC(-4)) | -0.169172 |
| | (0.09286) |
| | [-1.82176] |
| D(LNDC(-5)) | -0.028059 |
| | (0.08603) |
| | [-0.32615] |
| D(LNNCPI(-1)) | 0.169848 |
| | (0.66922) |
| | [ 0.25380] |
| D(LNNCPI(-2)) | -1.480083 |
| | (0.68725) |
| | [-2.15363] |
| D(LNNCPI(-3)) | 0.422221 |
| | (0.70974) |
| | [ 0.59489] |
| D(LNNCPI(-4)) | -0.603975 |
| | (0.67426) |
| | [-0.89577] |
| D(LNNCPI(-5)) | -0.811103 |
| | (0.64676) |
| | [-1.25410] |
| D(LNEP(-1)) | -0.198230 |
| | (0.23263) |
| | [-0.85214] |
| D(LNEP(-2)) | -0.265066 |
| | (0.23721) |
| | [-1.11743] |
| D(LNEP(-3)) | 0.228875 |
| | (0.24238) |
| | [ 0.94427] |

| Error Correction: | D(LNDC) |
|---|---|
| D(LNEP(-4)) | 0.297335 |
| | (0.25839) |
| | [ 1.15071] |
| D(LNEP(-5)) | 0.252009 |
| | (0.23095) |
| | [ 1.09116] |
| D(LNME(-1)) | 0.009949 |
| | (0.06747) |
| | [ 0.14745] |
| D(LNME(-2)) | -0.066091 |
| | (0.08430) |
| | [-0.78403] |
| D(LNME(-3)) | -0.122120 |
| | (0.08661) |
| | [-1.41004] |
| D(LNME(-4)) | -0.143104 |
| | (0.08042) |
| | [-1.77951] |
| D(LNME(-5)) | -0.089321 |
| | (0.06263) |
| | [-1.42627] |
| D(LNMI(-1)) | -0.237734 |
| | (0.08795) |
| | [-2.70315] |
| D(LNMI(-2)) | -0.018463 |
| | (0.09739) |
| | [-0.18959] |
| D(LNMI(-3)) | 0.076952 |
| | (0.10169) |
| | [ 0.75673] |
| D(LNMI(-4)) | -0.025094 |
| | (0.09909) |
| | [-0.25323] |
| D(LNMI(-5)) | 0.020271 |
| | (0.08184) |
| | [ 0.24771] |
| D(LNEX(-1)) | -1.094553 |
| | (0.75221) |
| | [-1.45512] |
| D(LNEX(-2)) | -0.668080 |
| | (0.83043) |
| | [-0.80450] |

| Error Correction: | D(LNDC) |
|---|---|
| D(LNEX(-3)) | 0.182672 |
| | (0.80560) |
| | [ 0.22675] |
| D(LNEX(-4)) | 0.717265 |
| | (0.80795) |
| | [ 0.88776] |
| D(LNEX(-5)) | -0.155576 |
| | (0.76478) |
| | [-0.20342] |
| D(LNPT(-1)) | -0.076294 |
| | (0.21144) |
| | [-0.36083] |
| D(LNPT(-2)) | -0.023829 |
| | (0.21639) |
| | [-0.11012] |
| D(LNPT(-3)) | -0.094762 |
| | (0.20966) |
| | [-0.45198] |
| D(LNPT(-4)) | -0.095593 |
| | (0.20164) |
| | [-0.47407] |
| D(LNPT(-5)) | -0.366493 |
| | (0.18831) |
| | [-1.94621] |
| D(LNRT(-1)) | 0.003720 |
| | (0.21382) |
| | [ 0.01740] |
| D(LNRT(-2)) | 0.401369 |
| | (0.22149) |
| | [ 1.81213] |
| D(LNRT(-3)) | 0.280194 |
| | (0.22838) |
| | [ 1.22687] |
| D(LNRT(-4)) | 0.338769 |
| | (0.19765) |
| | [ 1.71395] |
| D(LNRT(-5)) | 0.271583 |
| | (0.17905) |
| | [ 1.51682] |
| D(LNTA(-1)) | 0.061884 |
| | (0.04787) |
| | [ 1.29284] |

Table 5. 26 (Continued) Vector Error Correction Model Estimates

| Error Correction: | D(LNDC) |
|---|---|
| D(LNTA(-2)) | -0.008658 |
| | (0.04630) |
| | [-0.18699] |
| D(LNTA(-3)) | 0.126164 |
| | (0.04730) |
| | [ 2.66754] |
| D(LNTA(-4)) | 0.028253 |
| | (0.04817) |
| | [ 0.58656] |
| D(LNTA(-5)) | 0.024699 |
| | (0.04810) |
| | [ 0.51351] |
| D(LNVR(-1)) | -0.137538 |
| | (0.05647) |
| | [-2.43556] |

| Error Correction: | D(LNDC) |
|---|---|
| D(LNVR(-2)) | -0.097479 |
| | (0.06134) |
| | [-1.58904] |
| D(LNVR(-3)) | -0.034231 |
| | (0.06186) |
| | [-0.55334] |
| D(LNVR(-4)) | -0.018771 |
| | (0.05892) |
| | [-0.31860] |
| D(LNVR(-5)) | -0.129686 |
| | (0.05521) |
| | [-2.34905] |
| D(LNWR(-1)) | -0.135525 |
| | (0.20165) |
| | [-0.67210] |

| Error Correction: | D(LNDC) |
|---|---|
| D(LNWR(-2)) | -0.096523 |
| | (0.22885) |
| | [-0.42177] |
| D(LNWR(-3)) | 0.073616 |
| | (0.23100) |
| | [ 0.31868] |
| D(LNWR(-4)) | 0.265989 |
| | (0.23661) |
| | [ 1.12415] |
| D(LNWR(-5)) | 0.093093 |
| | (0.22996) |
| | [ 0.40482] |
| C | 0.023740 |
| | (0.01183) |
| | [ 2.00666] |
| R-squared | 0.524258 |
| Adj. R-squared | 0.369366 |
| S.E. equation | 0.093735 |
| F-statistic | 3.384657 |

Table 5. 27 Description of the Coefficients of the above VECM(5) Model

| | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | -0.218828 | 0.098550 | -2.220474 | 0.0277 |
| C(2) | -0.122576 | 0.111292 | -1.101390 | 0.2723 |
| C(3) | 0.048950 | 0.106313 | 0.460431 | 0.6458 |
| C(4) | -0.117562 | 0.098452 | -1.194110 | 0.2341 |
| C(5) | -0.169172 | 0.092862 | -1.821761 | 0.0702 |
| C(6) | -0.028059 | 0.086029 | -0.326155 | 0.7447 |
| C(7) | 0.009949 | 0.067469 | 0.147455 | 0.8829 |
| C(8) | -0.066091 | 0.084296 | -0.784032 | 0.4341 |
| C(9) | -0.122120 | 0.086607 | -1.410041 | 0.1603 |
| C(10) | -0.143104 | 0.080418 | -1.779508 | 0.0769 |
| C(11) | -0.089321 | 0.062625 | -1.426275 | 0.1556 |
| C(12) | -0.237734 | 0.087947 | -2.703147 | 0.0076 |
| C(13) | -0.018463 | 0.097386 | -0.189589 | 0.8499 |
| C(14) | 0.076952 | 0.101690 | 0.756733 | 0.4502 |
| C(15) | -0.025094 | 0.099095 | -0.253230 | 0.8004 |
| C(16) | 0.020271 | 0.081835 | 0.247708 | 0.8047 |
| C(17) | -1.094553 | 0.752207 | -1.455122 | 0.1475 |
| C(18) | -0.668080 | 0.830428 | -0.804500 | 0.4222 |
| C(19) | 0.182672 | 0.805601 | 0.226753 | 0.8209 |
| C(20) | 0.717265 | 0.807953 | 0.887756 | 0.3759 |
| C(21) | -0.155576 | 0.764783 | -0.203424 | 0.8390 |
| C(22) | -0.076294 | 0.211438 | -0.360834 | 0.7187 |
| C(23) | -0.023829 | 0.216392 | -0.110120 | 0.9124 |
| C(24) | -0.094762 | 0.209661 | -0.451977 | 0.6519 |
| C(25) | -0.095593 | 0.201645 | -0.474066 | 0.6361 |
| C(26) | -0.366493 | 0.188312 | -1.946206 | 0.0533 |
| C(27) | 0.003720 | 0.213815 | 0.017396 | 0.9861 |
| C(28) | 0.401369 | 0.221491 | 1.812125 | 0.0717 |
| C(29) | 0.280194 | 0.228381 | 1.226871 | 0.2215 |
| C(30) | 0.338769 | 0.197654 | 1.713953 | 0.0883 |
| C(31) | 0.271583 | 0.179047 | 1.516824 | 0.1311 |
| C(32) | 0.061884 | 0.047867 | 1.292844 | 0.1978 |
| C(33) | -0.008658 | 0.046300 | -0.186994 | 0.8519 |
| C(34) | 0.126164 | 0.047296 | 2.667545 | 0.0084 |
| C(35) | 0.028253 | 0.048167 | 0.586562 | 0.5583 |
| C(36) | 0.024699 | 0.048097 | 0.513512 | 0.6083 |
| C(37) | -0.137538 | 0.056471 | -2.435558 | 0.0159 |
| C(38) | -0.097479 | 0.061345 | -1.589038 | 0.1139 |
| C(39) | -0.034231 | 0.061862 | -0.553342 | 0.5807 |
| C(40) | -0.018771 | 0.058917 | -0.318596 | 0.7504 |
| C(41) | -0.129686 | 0.055208 | -2.349051 | 0.0200 |
| C(42) | -0.135525 | 0.201645 | -0.672098 | 0.5024 |
| C(43) | -0.096523 | 0.228851 | -0.421773 | 0.6737 |
| C(44) | 0.073616 | 0.231003 | 0.318679 | 0.7504 |
| C(45) | 0.265989 | 0.236613 | 1.124154 | 0.2625 |
| C(46) | 0.093093 | 0.229962 | 0.404819 | 0.6861 |
| C(47) | 0.169848 | 0.669224 | 0.253798 | 0.8000 |
| C(48) | -1.480083 | 0.687251 | -2.153627 | 0.0327 |
| C(49) | 0.422221 | 0.709743 | 0.594892 | 0.5527 |
| C(50) | -0.603975 | 0.674255 | -0.895766 | 0.3716 |
| C(51) | -0.811103 | 0.646761 | -1.254100 | 0.2115 |
| C(52) | -0.198230 | 0.232626 | -0.852139 | 0.3953 |
| C(53) | -0.265066 | 0.237211 | -1.117427 | 0.2654 |
| C(54) | 0.228875 | 0.242383 | 0.944270 | 0.3464 |
| C(55) | 0.297335 | 0.258392 | 1.150711 | 0.2514 |
| C(56) | 0.252009 | 0.230955 | 1.091162 | 0.2767 |
| C(57) | 0.023740 | 0.011831 | 2.006665 | 0.0464 |

According to the Table 5.26 and Table 5.27 the coefficient of the cointegration equation of the VECM (5) model for the LNDC is -0.218828 and p-value is 0.0277, which is less than 0.05. Thus the coefficient is significant at 5% level and hence the model is significant. Therefore, it implies that 21.88% deviations from long run equilibrium are corrected by the VECM (5) model in each month by making changes in LNDC. Moreover, all the coefficients of the co-integration equation are significant at 5% level.

According to the Table 5.26 and 5.27, it is clear that the coefficients of LNEP, LNME, LNMI, LNEX, and LNPT are positive while the coefficients of LNNCPI, LNRT, LNTA, LNVR, and LNWR are negative. Since the model is significant at 5% level of significance, it can be concluded that the long run association between the LNDC and all 10 variables significantly influence. Furthermore, it can be concluded that LNEP, LNME, LNMI, LNEX, and LNPT have long term positive relationship with LNDC while LNNCPI, LNRT, LNTA, LNVR, and LNWR have long term negative relationship with LNDC.

## 5.6 Diagnostics Tests for Residuals of the VECM(5) Model

### 5.6.1 Randomness

The following Figure 5.1 represents the correlogram of the residuals of the VEC model described above.

**Correlogram of Residuals**

Date: 11/08/19 Time: 20:35
Sample: 1998M01 2017M07
Included observations: 229
Q-statistic probabilities adjusted for 56 dynamic regressors

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob* |
|---|---|---|---|---|---|---|
| | | 1 | -0.002 | -0.002 | 0.0006 | 0.980 |
| | | 2 | -0.034 | -0.034 | 0.2685 | 0.874 |
| | | 3 | -0.014 | -0.014 | 0.3111 | 0.958 |
| | | 4 | 0.021 | 0.020 | 0.4178 | 0.981 |
| | | 5 | 0.085 | 0.085 | 2.1323 | 0.831 |
| | | 6 | 0.063 | 0.065 | 3.0655 | 0.801 |
| | | 7 | -0.044 | -0.038 | 3.5322 | 0.832 |
| | | 8 | 0.025 | 0.030 | 3.6782 | 0.885 |
| | | 9 | -0.116 | -0.123 | 6.9229 | 0.645 |
| | | 10 | 0.068 | 0.060 | 8.0401 | 0.625 |
| | | 11 | -0.083 | -0.104 | 9.7259 | 0.555 |
| | | 12 | -0.019 | -0.013 | 9.8151 | 0.632 |
| | | 13 | -0.095 | -0.100 | 12.021 | 0.526 |
| | | 14 | 0.140 | 0.160 | 16.846 | 0.264 |
| | | 15 | -0.053 | -0.062 | 17.544 | 0.287 |
| | | 16 | -0.112 | -0.102 | 20.665 | 0.192 |
| | | 17 | -0.067 | -0.045 | 21.800 | 0.193 |
| | | 18 | 0.057 | 0.045 | 22.624 | 0.205 |
| | | 19 | -0.081 | -0.090 | 24.258 | 0.186 |
| | | 20 | 0.068 | 0.042 | 25.413 | 0.186 |
| | | 21 | -0.004 | 0.041 | 25.417 | 0.230 |
| | | 22 | 0.020 | 0.009 | 25.524 | 0.273 |
| | | 23 | -0.013 | 0.026 | 25.566 | 0.322 |
| | | 24 | 0.006 | -0.034 | 25.574 | 0.375 |
| | | 25 | 0.038 | 0.055 | 25.958 | 0.410 |
| | | 26 | 0.136 | 0.109 | 30.769 | 0.237 |
| | | 27 | -0.024 | 0.006 | 30.918 | 0.275 |
| | | 28 | -0.039 | -0.119 | 31.309 | 0.304 |
| | | 29 | -0.105 | -0.084 | 34.253 | 0.230 |
| | | 30 | -0.020 | -0.029 | 34.358 | 0.267 |
| | | 31 | -0.005 | -0.021 | 34.364 | 0.310 |

Figure 5. 1 Correlogram of the Residual with (Q Statistic) of VEC Model

According to the Figure 5.1, the autocorrelation of the residuals is not significant up to lag 31 in Q-statistic. Thus it can be concluded that the there is no significant autocorrelations among the residuals in the model as none of the lags shows lower p-value than 5%. Thus errors are random.

### 5.6.2 ARCH Effect

Figure 5.2 describe the correlogram of squared residuals of the VEC model above.

**Correlogram of Residuals Squared**

Date: 11/08/19 Time: 20:48
Sample: 1998M01 2017M07
Included observations: 229

| | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|
| 1 | 0.101 | 0.101 | 2.3550 | 0.125 |
| 2 | 0.103 | 0.094 | 4.8357 | 0.089 |
| 3 | 0.191 | 0.176 | 13.385 | 0.004 |
| 4 | -0.012 | -0.055 | 13.420 | 0.009 |
| 5 | -0.011 | -0.041 | 13.449 | 0.020 |
| 6 | -0.068 | -0.096 | 14.536 | 0.024 |
| 7 | -0.037 | -0.007 | 14.856 | 0.038 |
| 8 | 0.005 | 0.035 | 14.863 | 0.062 |
| 9 | -0.005 | 0.028 | 14.870 | 0.095 |
| 10 | 0.015 | 0.016 | 14.927 | 0.135 |
| 11 | 0.048 | 0.033 | 15.498 | 0.161 |
| 12 | 0.007 | -0.015 | 15.509 | 0.215 |
| 13 | 0.032 | 0.016 | 15.760 | 0.262 |
| 14 | -0.028 | -0.048 | 15.957 | 0.316 |
| 15 | 0.010 | 0.019 | 15.982 | 0.383 |
| 16 | -0.018 | -0.019 | 16.061 | 0.449 |
| 17 | -0.032 | -0.007 | 16.316 | 0.502 |
| 18 | 0.064 | 0.071 | 17.350 | 0.499 |
| 19 | -0.079 | -0.083 | 18.922 | 0.462 |
| 20 | -0.008 | -0.006 | 18.940 | 0.526 |
| 21 | -0.011 | -0.030 | 18.972 | 0.587 |
| 22 | -0.041 | -0.008 | 19.411 | 0.620 |
| 23 | 0.117 | 0.134 | 22.944 | 0.464 |
| 24 | 0.054 | 0.058 | 23.705 | 0.479 |
| 25 | 0.052 | 0.030 | 24.406 | 0.496 |
| 26 | 0.014 | -0.069 | 24.458 | 0.550 |
| 27 | -0.128 | -0.171 | 28.742 | 0.374 |
| 28 | 0.029 | 0.042 | 28.961 | 0.414 |
| 29 | 0.029 | 0.093 | 29.191 | 0.455 |
| 30 | -0.068 | 0.014 | 30.405 | 0.445 |
| 31 | 0.061 | 0.054 | 31.391 | 0.447 |

Figure 5. 2 Correlogram of Squared Residuals in VEC model

According to Figure 5.2, it is clear that all the squared residuals lie between the 95% confidence interval limits in the correlogram. . It can be concluded that residual has no ARCH effect.

### 5.6.3 Normality

Figure 5.3 illustrates the test for normality of the residuals.

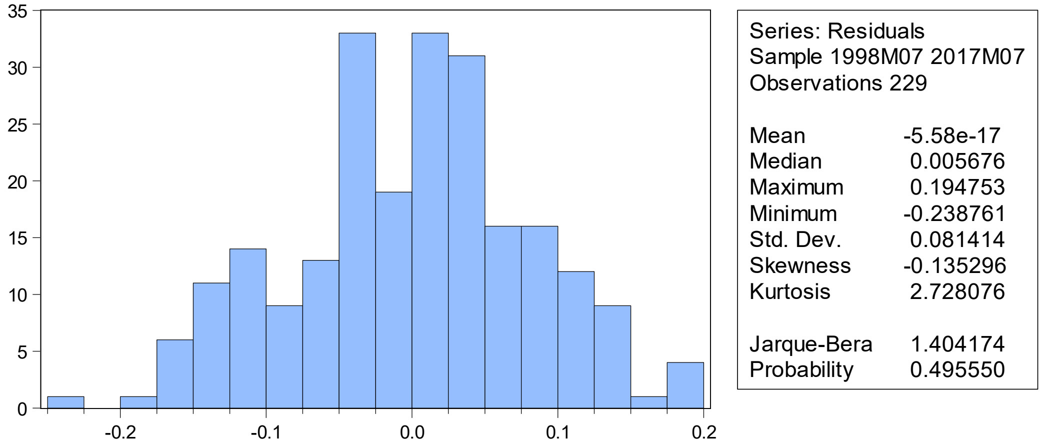


Figure 5. 3 Normality Test for the Residuals

According to Jarque-Bera test statistic in the Figure 5.3, it is clear that the errors are not significantly deviated from normally distributed at 5% level of significant since the test statistics is 1.4 with p-value is greater than 5%.

### 5.6.4 Serial Correlation

Test for serial correlation of the residual of the model is described in the Table 5.28 below.

Table 5. 28 Test for Serial Correlation

Breusch-Godfrey Serial Correlation LM Test:

| | | | |
|---|---|---|---|
| F-statistic | 0.730427 | Prob. F(2,170) | 0.4832 |
| Obs*R-squared | 1.951091 | Prob. Chi-Square(2) | 0.3770 |

According to the Table 5.28, it is evident that the F-Statistic is 0.730 with the corresponding p-value of 0.4832 confirming that the there is no significant serial correlation in the residuals of the model as the p-value is greater than 5%.

### 5.6.5 Heteroskedasticity

Test for Heteroskedasticity of the model is described in the table 5.29 below.

Table 5. 29 Test for Heteroskedasticity

| | | | |
|---|---|---|---|
| F-statistic | 2.370138 | Prob. F(1,226) | 0.1251 |
| Obs*R-squared | 2.366297 | Prob. Chi-Square(1) | 0.1240 |

According to the Table 5.29, it is evident that the F-Statistic is 2.37 with the p-value of 0.1251 confirming that the there is no significant heteroskedasticity in the residuals of the model. Thus the residuals are of constant variance.

### 5.6.6 Errors of the Fitted Model

The VECM(5) model is fitted to the training data set and calculated the root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), which are displayed in the Figure 5.4.

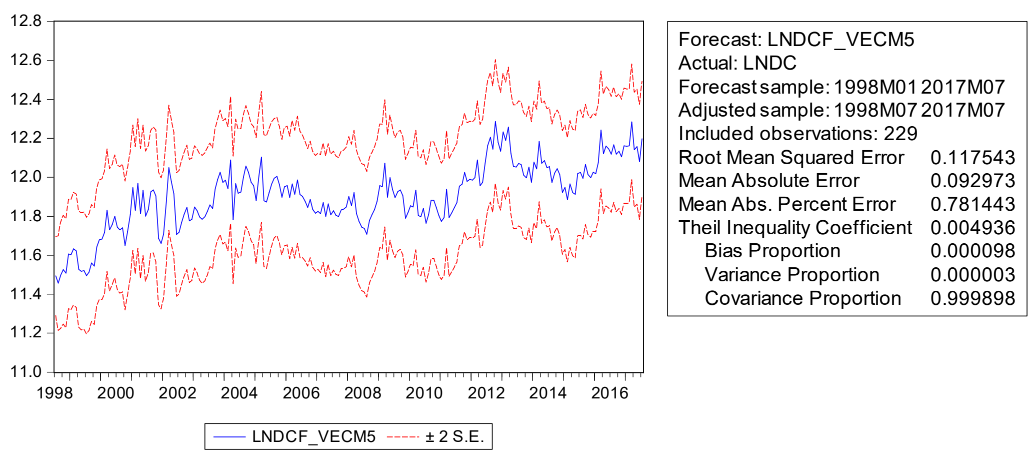


Figure 5. 4 Error of VECM(5) model

According to the Figure 5.4, RMSE of the fitted model to the training data set is 0.117 while MAE and MAPE is 0.09 and 0.78 respectively. Thus it can be concluded that the predicting errors of the training data are fairly low.

## 5.7 VECM(5) Model

As all the requirements were satisfied by the errors of the model, the, the equation of the VECM(5) model can be written as:

$$D(LNDC) = -0.218[CE]+[e] \qquad \text{(Model 1)}$$

$$CE = LNDC_{t-1}-0.18LNME_{t-1}-0.55LNMI_{t-1}-1.57LNEX_{t-1}-0.63LNPT_{t-1}+1.60LNRT_{t-1} +0.03LNTA_{t-1}+0.08LNVR_{t-1}+0.19LNWR_{t-1}+0.88LNNCPI_{t-1}-0.48LNEP_{t-1}-13.26$$

$$\begin{aligned}
e = & -0.122d(LNDC_{t-1})+0.048d(LNDC_{t-2})-0.117d(LNDC_{t-3})-0.169d(LNDC_{t-4}) - \\
& -0.028d(LNDC_{t-5})+ 0.009d(LNME_{t-1})-0.066d(LNME_{t-2})-0.122d(LNME_{t-3})- \\
& -0.143d(LNME_{t-4})-0.089d(LNME_{t-5})-0.237d(LNMI_{t-1})-0.018d(LNMI_{t-2})+ \\
& +0.076d(LNMI_{t-3})-0.025d(LNMI_{t-4})+0.02d(LNMI_{t-5})-1.094d(LNEX_{t-1})- \\
& -0.668d(LNEX_{t-2})+0.182d(LNEX_{t-3})+0.717d(LNEX_{t-4})-0.155d(LNEX_{t-5})- \\
& -0.076d(LNPT_{t-1})-0.023d(LNPT_{t-2})-0.094d(LNPT_{t-3})-0.095d(LNPT_{t-4})- \\
& -0.366d(LNPT_{t-5})+0.003d(LNRT_{t-1})+0.401d(LNRT_{t-2})+0.280d(LNRT_{t-3})+ \\
& 0.338d(LNRT_{t-4})+0.271d(LNRT_{t-5})+0.061d(LNTA_{t-1})-0.008d(LNTA_{t-2})+ \\
& 0.126d(LNTA_{t-3})+0.028d(LNTA_{t-4})+0.024d(LNTA_{t-5})-0.137d(LNVR_{t-1})- \\
& -0.097d(LNVR_{t-2})-0.034d(LNVR_{t-3})-0.018d(LNVR_{t-4})-0.129d(LNVR_{t-5})- \\
& -0.135d(LNWR_{t-1})-0.096d(LNWR_{t-2})+0.073d(LNWR_{t-3})+0.265d(LNWR_{t-4})+ \\
& 0.093d(LNWR_{t-5})+169d(LNNCPI_{t-1})-1.48d(LNNCPI_{t-2})+0.422d(LNNCPI_{t-3}) - \\
& -0.603d(LNNCPI_{t-4})-0.811d(LNNCPI_{t-5})-0.198d(LNEX_{t-1})-0265d(LNEX_{t-2})+ \\
& +0.228d(LNEX_{t-3})+0.297d(LNEX_{t-4})+0.252d(LNEX_{t-5})+0.023
\end{aligned}$$

Where, $d(X_t)$ : First difference of time series $X_t$

$X_{t-i}$ : $i^{th}$ lag of time series $X_t$

*Other abbreviations of the variables in the VEC model mentioned above are described in the section 3.3 of chapter 3.*

## 5.8 Validation of the VEC Model

Above Model (1) is validated using monthly data specified in the Table 3.1(Section 3.1). The results of the forecasted diesel consumption and corresponding percentages of the errors are summarized in the Table 5.30.

Table 5. 30 Validation of the VECM(5) Model

| **Time** | **Actual Diesel Consumption (Metric Ton)** | **Predicted Diesel Consumption (Metric Ton)** | **Percentage of Error (%)** |
|---|---|---|---|
| 2017-August | 194,671 | 213,502.6 | -9.67 |
| 2017-September | 183,343 | 191,629.9 | -4.51 |
| 2017-October | 172,930 | 197,644.2 | -14.29 |
| 2017-November | 178,297 | 179,088.6 | -0.44 |
| 2017-December | 183,808 | 174,207.6 | 5.22 |
| 2018-January | 178,589 | 202,703.2 | -13.50 |
| 2018-February | 183,788 | 158,347.9 | 13.84 |
| 2018-March | 204,830 | 197,504.8 | 3.57 |
| 2018-April | 191,143 | 192,693.9 | -0.81 |
| 2018-May | 190,501 | 192,028.0 | -0.80 |
| 2018-Jun | 185,649 | 175,454.2 | 5.49 |
| 2018-July | 190,676 | 179,147.8 | 6.05 |

According to the Table 5.30, it is clear that the percentage of error varied between -14.29% and 13.84%. This range can be considered as reasonable. Thus it can be considered that the model forecasted diesel consumption do not deviate considerably with the true diesel consumption as all the percentage of errors are less than 15%. Thus it can be concluded that the errors are fairly low.

### 5.9 Summary of Chapter 5

There exists long term equilibrium between the LNDC and the 10 explanatory variables according to the VECM(5) model, which was developed to model LNDC. The errors of the fitted model found random and constant variance. The model (1) was validated for an independent data set. The percentage error of the fitted model for trained set and validation set varies from -14.29 to +13.84%. The LNEP, LNME, LNMI, LNEX, and LNPT have long term positive relationship with LNDC while LNNCPI, LNRT, LNTA, LNVR, and LNWR have long term negative relationship with LNDC.

# CHAPTER 6

# ALTERNATIVE VECTOR ERROR CORRECTION MODEL: SCREENING VARAIBLES

In Chapter 5, VECM (5) was developed using all ten explanatory variables without considering the internal association among those explanatory variables. In this Chapter few variables are selected to develop a VECM model. $Log_e$ transformation series are used.

## 6.1 Selecting Variables for the VEC Model

Though 10 explanatory variables were considered, it is not flexible to use all the variables in the model. Therefore, based on the magnitude and the significant level of the correlation coefficient, some variables were discarded to the VEC model.

According to the Table 4.12 and 4.13 (Section 4.6) it is clear that LNNCPI, LNEP, LNEX, LNMI, and LNTA show strong significant correlation with LNDC than the other explanatory variables, as the corresponding correlation coefficients are greater than or equal 0.6 with p-value of 0.0. Therefore only these five explanatory variables are selected for the development of VEC model using log transformation series.

## 6.2 Stationary of Each Time Series

The results of Augmented Dickey-Fuller test (Section 5.1 & 5.2) confirmed the first differences of six series are I(1). The six time series are: LNNCPI, LNEP, LNEX, LNMI, LNTA and LNDC.

### 6.3 Existence of Cointegration

The results of the Johansen test for co-integration using the maximum Eigen value statistic can be found in Table 6.1 whereas the test with trace statistic can be found in Table 6.2 for LNDC and the 5 selected time series.

Table 6. 1 Unrestricted Cointegration Rank Test (Maximum Eigen value)

| Unrestricted Cointegration Rank Test (Maximum Eigenvalue) | | | | |
|---|---|---|---|---|
| Hypothesized No. of CE(s) | Eigenvalue | Max-Eigen Statistic | 0.05 Critical Value | Prob.** |
| None * | 0.201504 | 52.43087 | 40.07757 | 0.0013 |
| At most 1 * | 0.144322 | 36.31565 | 33.87687 | 0.0251 |
| At most 2 * | 0.130818 | 32.66722 | 27.58434 | 0.0102 |
| At most 3 | 0.060794 | 14.61396 | 21.13162 | 0.3169 |
| At most 4 | 0.028258 | 6.678823 | 14.26460 | 0.5278 |
| At most 5 | 0.007450 | 1.742331 | 3.841466 | 0.1868 |
| Max-eigenvalue test indicates 3 cointegratingeqn(s) at the 0.05 level | | | | |
| * denotes rejection of the hypothesis at the 0.05 level | | | | |
| **MacKinnon-Haug-Michelis (1999) p-values | | | | |

Table 6. 2 Unrestricted Cointegration Rank Test (Trace)

| | | | | |
|---|---|---|---|---|
| Sample (adjusted): 1998M03 2017M07 | | | | |
| Included observations: 233 after adjustments | | | | |
| Trend assumption: Linear deterministic trend | | | | |
| Series: LNDC LNNCPI LNEP LNEX LNMI LNTA | | | | |
| Lags interval (in first differences): 1 to 1 | | | | |
| Unrestricted Cointegration Rank Test (Trace) | | | | |
| Hypothesized No. of CE(s) | Eigenvalue | Max-Eigen Statistic | 0.05 Critical Value | Prob.** |
| None * | 0.201504 | 144.4489 | 95.75366 | 0.0000 |
| At most 1 * | 0.144322 | 92.01799 | 69.81889 | 0.0003 |
| At most 2 * | 0.130818 | 55.70234 | 47.85613 | 0.0077 |
| At most 3 | 0.060794 | 23.03512 | 29.79707 | 0.2443 |
| At most 4 | 0.028258 | 8.421153 | 15.49471 | 0.4215 |
| At most 5 | 0.007450 | 1.742331 | 3.841466 | 0.1868 |
| Trace test indicates 3 cointegratingeqn(s) at the 0.05 level | | | | |
| * denotes rejection of the hypothesis at the 0.05 level | | | | |
| **MacKinnon-Haug-Michelis (1999) p-values | | | | |

Results shown in above Table 6.1 and 6.2 describe the Johanson Cointegration Test using Eigen value statistic and trace statistic respectively. It can be seen that the both test results show the same results. The both test reject the two hypotheses namely, there exists at most one and two linear combinations of the six time series as the corresponding p-values are less than 5% level of significance. Therefore it can be concluded that there are 3 cointegrating linear combinations of the all six time series based on the two tests. Hence, it can be concluded that there is long run equilibrium relationship between LNDC and the 5 variables (LNNCPI, LNTA, LNEP, LNEX, and LNMI) with 95% confidence. That means all series move together for long term equilibrium.

Since the Table 6.1 and 6.2 confirm that there is a cointegration between the time series variables, the VEC model is carried out with the first differenced series for the estimation.

## 6.4 Optimal Lag Length

In order to use the VEC model it is required to find the optimum lag length of the model. Therefore, following lag length criteria is carried out. The lag length criterion is described in the Table 6.3

Table 6. 3 VAR Lag Order Selection Criteria

VAR Lag Order Selection Criteria
Endogenous variables: LNDC LNNCPI LNEP LNEX LNMI LNTA
Exogenous variables:
Sample: 1998M01 2017M07
Included observations: 233

| Lag | LogL | LR | FPE | AIC | SC | HQ |
|---|---|---|---|---|---|---|
| 1 | 2239.893 | NA | 2.45e-16 | -18.91753 | -18.38433 | -18.70252 |
| 2 | 2362.990 | 233.5153* | 1.16e-16* | -19.66515* | -18.59873* | -19.23512* |

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

According to the Table 6.3, it is evident that all the indicators show the minimum values at the lag 2. Therefore, lag 2 is selected as the optimum lag for long run equilibrium and the VEC model.

## 6.5 VEC Model

The following Table 6.4 describes the long run equilibrium of VECM for the diesel consumption (LNDC) with the five selected time series: LNNCPI, LNTA, LNEP, LNEX, and LNMI. As the optimum lag length is two, VECM(2) was considered.

Table 6. 4 Vector Error Correction Model VECM (2) Estimates

Vector Error Correction Estimates
Sample (adjusted): 1998M04 2017M07
Included observations: 232 after adjustments
Standard errors in ( ) & t-statistics in [ ]

| CointegratingEq: | CointEq1 |
|---|---|
| LNDC(-1) | 1.000000 |
| LNNCPI(-1) | 0.711751<br>(0.22236)<br>[ 3.20086] |
| LNEP(-1) | 1.565351<br>(0.44150)<br>[ 3.54551] |
| LNEX(-1) | -2.125965<br>(0.36413)<br>[-5.83845] |
| LNMI(-1) | -0.913717<br>(0.18672)<br>[-4.89354] |
| LNTA(-1) | -0.048959<br>(0.05719)<br>[-0.85614] |
| C | -8.697142 |

Table 6.4 (Continued) Vector Error Correction Model Estimates

| Error Correction: | D(LNDC) |
|---|---|
| CointEq1 | -0.163816 |
| | (0.05772) |
| | [-2.83818] |
| D(LNDC(-1)) | -0.145041 |
| | (0.07677) |
| | [-1.88926] |
| D(LNDC(-2)) | 0.045073 |
| | (0.07144) |
| | [ 0.63096] |
| D(LNNCPI(-1)) | 0.078360 |
| | (0.52718) |
| | [ 0.14864] |
| D(LNNCPI(-2)) | -0.650085 |
| | (0.52893) |
| | [-1.22905] |
| D(LNEP(-1)) | -0.411393 |
| | (0.17786) |
| | [-2.31303] |
| D(LNEP(-2)) | -0.336751 |
| | (0.16114) |
| | [-2.08979] |

| Error Correction: | D(LNDC) |
|---|---|
| D(LNEX(-1)) | -0.969245 |
| | (0.67731) |
| | [-1.43103] |
| D(LNEX(-2)) | -0.587271 |
| | (0.69637) |
| | [-0.84333] |
| D(LNMI(-1)) | -0.325504 |
| | (0.06815) |
| | [-4.77644] |
| D(LNMI(-2)) | -0.116076 |
| | (0.06563) |
| | [-1.76858] |
| D(LNTA(-1)) | 0.047198 |
| | (0.03789) |
| | [ 1.24565] |
| D(LNTA(-2)) | 0.026756 |
| | (0.03702) |
| | [ 0.72280] |
| C | 0.017913 |
| | (0.00806) |
| | [ 2.22329] |
| R-squared | 0.321760 |
| Adj. R-squared | 0.281314 |

The p-values of the coefficients of the above VECM(2) model is described in the Table 6.5 below.

Table 6. 5 Description of the Coefficients of the model VECM (2)

| | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | -0.163816 | 0.057719 | -2.838178 | 0.0050 |
| C(2) | -0.145041 | 0.076771 | -1.889264 | 0.0602 |
| C(3) | 0.045073 | 0.071435 | 0.630964 | 0.5287 |
| C(4) | 0.078360 | 0.527182 | 0.148639 | 0.8820 |
| C(5) | -0.650085 | 0.528932 | -1.229052 | 0.2204 |
| C(6) | -0.411393 | 0.177859 | -2.313031 | 0.0217 |
| C(7) | -0.336751 | 0.161141 | -2.089794 | 0.0378 |
| C(8) | -0.969245 | 0.677307 | -1.431026 | 0.1539 |
| C(9) | -0.587271 | 0.696371 | -0.843331 | 0.4000 |
| C(10) | -0.325504 | 0.068148 | -4.776440 | 0.0000 |
| C(11) | -0.116076 | 0.065632 | -1.768584 | 0.0784 |
| C(12) | 0.047198 | 0.037890 | 1.245653 | 0.2142 |
| C(13) | 0.026756 | 0.037017 | 0.722802 | 0.4706 |
| C(14) | 0.017913 | 0.008057 | 2.223287 | 0.0272 |
| R-squared | 0.321760 | | | |
| Adjusted R-squared | 0.281314 | | | |

According to the Table 6.4 the coefficient of the cointegration equation of the model VECM (2) for LNDC time series is -0.163816. Table 6.5 shows that the corresponding p-value is 0.005 and it is less than 0.05. Thus the coefficient of the VECM (2) model is significant at 5% level of significance and hence the model is significant. This implies that the VECM (2) model corrects 16.38% of deviations from long run equilibrium each month by changing in the variable LNDC. Moreover, all the coefficients of the co-integration equation are less than 0.05 confirming that the coefficients are significant at 5% level.

By observing the Table 6.4 further, it can be noticed that the coefficients of LNNCPI and LNEP are negative while the coefficients of LNEX, LNMI, and LNTA are positive. Since the coefficients of the cointegration equation of the model VECM(2) is significant at 5% level, it can be concluded that LNEX, LNMI, and LNTA show significant long run positive associations with LNDC while LNNCPI and LNEP indicate significant long run negative associations with LNDC.

## 6.6 Diagnostics Tests for Residuals of the VECM(2) Model

### 6.6.1 Randomness

The following Figure 6.1 represents the correlogram of the residuals of the VEC model described above.

**Correlogram of Residuals**

Date: 02/02/20 Time: 10:04
Sample: 1998M01 2017M07
Included observations: 232
Q-statistic probabilities adjusted for 13 dynamic regressors

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob* |
|---|---|---|---|---|---|---|
| | | 1 | 0.021 | 0.021 | 0.1074 | 0.743 |
| | | 2 | 0.031 | 0.031 | 0.3385 | 0.844 |
| | | 3 | -0.052 | -0.053 | 0.9702 | 0.808 |
| | | 4 | -0.133 | -0.132 | 5.1664 | 0.271 |
| | | 5 | -0.012 | -0.003 | 5.1990 | 0.392 |
| | | 6 | 0.052 | 0.060 | 5.8424 | 0.441 |
| | | 7 | -0.012 | -0.027 | 5.8751 | 0.554 |
| | | 8 | -0.034 | -0.057 | 6.1477 | 0.631 |
| | | 9 | -0.043 | -0.037 | 6.5922 | 0.679 |
| | | 10 | 0.089 | 0.110 | 8.5079 | 0.579 |
| | | 11 | -0.070 | -0.081 | 9.7029 | 0.557 |
| | | 12 | 0.085 | 0.062 | 11.495 | 0.487 |
| | | 13 | -0.069 | -0.070 | 12.666 | 0.474 |
| | | 14 | 0.033 | 0.060 | 12.942 | 0.531 |
| | | 15 | -0.096 | -0.110 | 15.253 | 0.433 |
| | | 16 | -0.004 | 0.000 | 15.257 | 0.506 |
| | | 17 | -0.024 | -0.027 | 15.409 | 0.566 |
| | | 18 | 0.060 | 0.068 | 16.310 | 0.571 |
| | | 19 | -0.037 | -0.065 | 16.653 | 0.613 |
| | | 20 | 0.073 | 0.060 | 18.009 | 0.587 |
| | | 21 | -0.135 | -0.127 | 22.690 | 0.361 |
| | | 22 | 0.022 | 0.030 | 22.810 | 0.413 |
| | | 23 | 0.015 | 0.026 | 22.869 | 0.468 |
| | | 24 | 0.063 | 0.041 | 23.915 | 0.466 |
| | | 25 | 0.071 | 0.069 | 25.247 | 0.449 |
| | | 26 | 0.116 | 0.096 | 28.782 | 0.321 |
| | | 27 | -0.081 | -0.053 | 30.511 | 0.292 |
| | | 28 | -0.027 | -0.040 | 30.699 | 0.331 |
| | | 29 | -0.091 | -0.047 | 32.902 | 0.282 |

Figure 6. 1 Correlogram of the Residual with (Q Statistic) of VECM (2) Model

According to the Figure 6.1, the autocorrelation of the residuals is not significant up to lag 29 in Q-statistic. Thus it can be concluded that the there is no significant autocorrelations among the residuals in the model as none of the lags shows lower p-value than 5%. Therefore, it can be concluded that the errors are random.

### 6.6.2 ARCH Effect

Figure 6.2 describes the correlogram of squared residuals of the VEC model above.

**Correlogram of Residuals Squared**

Date: 02/02/20 Time: 10:12
Sample: 1998M01 2017M07
Included observations: 232

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.033 | 0.033 | 0.2498 | 0.617 |
| | | 2 | -0.026 | -0.027 | 0.4109 | 0.814 |
| | | 3 | 0.085 | 0.087 | 2.1095 | 0.550 |
| | | 4 | -0.002 | -0.009 | 2.1105 | 0.715 |
| | | 5 | 0.002 | 0.008 | 2.1118 | 0.833 |
| | | 6 | -0.055 | -0.064 | 2.8417 | 0.828 |
| | | 7 | -0.023 | -0.018 | 2.9712 | 0.888 |
| | | 8 | 0.027 | 0.025 | 3.1515 | 0.924 |
| | | 9 | -0.041 | -0.034 | 3.5555 | 0.938 |
| | | 10 | 0.057 | 0.065 | 4.3491 | 0.930 |
| | | 11 | -0.045 | -0.058 | 4.8557 | 0.938 |
| | | 12 | -0.006 | 0.006 | 4.8646 | 0.962 |
| | | 13 | -0.000 | -0.018 | 4.8647 | 0.978 |
| | | 14 | -0.085 | -0.074 | 6.6840 | 0.946 |
| | | 15 | -0.062 | -0.061 | 7.6494 | 0.937 |
| | | 16 | -0.062 | -0.058 | 8.6025 | 0.929 |
| | | 17 | 0.076 | 0.093 | 10.043 | 0.902 |
| | | 18 | 0.043 | 0.037 | 10.516 | 0.914 |
| | | 19 | -0.131 | -0.116 | 14.906 | 0.729 |
| | | 20 | -0.001 | -0.023 | 14.906 | 0.782 |
| | | 21 | 0.045 | 0.029 | 15.429 | 0.801 |
| | | 22 | -0.031 | -0.020 | 15.670 | 0.832 |
| | | 23 | 0.105 | 0.120 | 18.549 | 0.727 |
| | | 24 | 0.047 | 0.049 | 19.118 | 0.746 |
| | | 25 | 0.031 | 0.018 | 19.363 | 0.779 |
| | | 26 | 0.014 | -0.008 | 19.417 | 0.818 |
| | | 27 | -0.088 | -0.108 | 21.460 | 0.764 |
| | | 28 | 0.018 | 0.007 | 21.542 | 0.802 |
| | | 29 | -0.011 | -0.001 | 21.576 | 0.837 |

Figure 6. 2 Correlogram of Squared Residuals in VEC model

According to Figure 6.2, it is clear that all the squared residuals lie between the 95% confidence interval limits in the correlogram. Thus it can be concluded that residual has no ARCH effect.

### 6.6.3 Normality

Figure 6.3 represents the test for normality of the residuals.

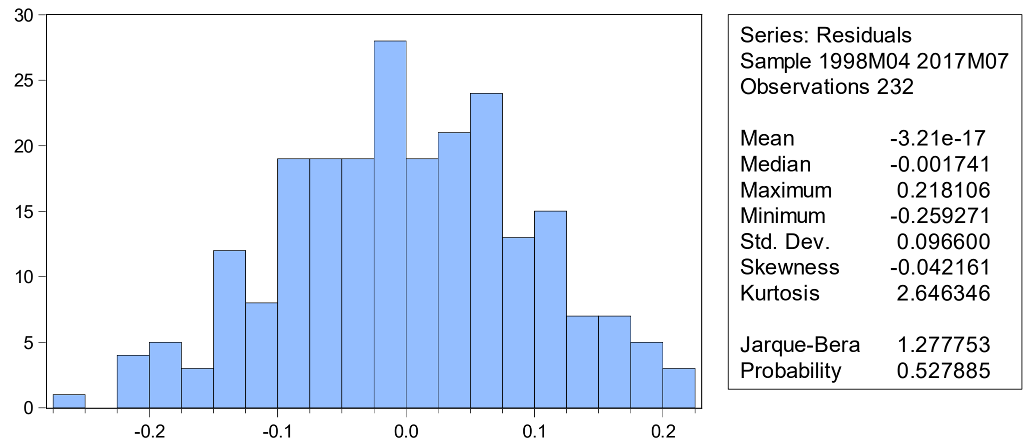


Figure 6. 3 Normality Test for the Residuals

According to Jarque-Bera test statistic in the Figure 6.3, it is clear that the errors are normally distributed at 5% level of significant since the test statistics is 1.277 with p-value 0.52, which is greater than 0.05.

### 6.6.4 Serial Correlation

Test for serial correlation of the residual of the model is described in the Table 6.6 below.

Table 6. 6 Test for Serial Correlation

Breusch-Godfrey Serial Correlation LM Test:

| | | | |
|---|---|---|---|
| F-statistic | 1.030520 | Prob. F(2,216) | 0.3586 |
| Obs*R-squared | 2.192785 | Prob. Chi-Square(2) | 0.3341 |

According to the Table 6.6, it is evident that the F-Statistic is 1.031 with the corresponding p-value of 0.3586 confirming that the there is no significant serial correlation in the residuals of the model at 5% level of significance, as the p-value is greater than 5%.

### 6.6.5 Heteroskedasticity

Test for Heteroskedasticity of the model is represented in the Table 6.7 below.

Table 6. 7 Test for Heteroskedasticity

Heteroskedasticity Test: Breusch-Pagan-Godfrey

| | | | |
|---|---|---|---|
| F-statistic | 1.173215 | Prob. F(18,213) | 0.2855 |
| Obs*R-squared | 20.92684 | Prob. Chi-Square(18) | 0.2831 |
| Scaled explained SS | 15.21009 | Prob. Chi-Square(18) | 0.6475 |

According to the Table 6.7, it is evident that the F-Statistic is 1.17 with the p-value of 0.2855, which is greater than 0.05, confirming that the there is no significant heteroskedasticity in the residuals of the model at 5% level of significance. Thus the residuals are of constant variance.

### 6.6.6 Errors of the Fitted Model

The VECM(2) model is fitted to the training data set and calculated Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), which are displayed in the Figure 6.4.

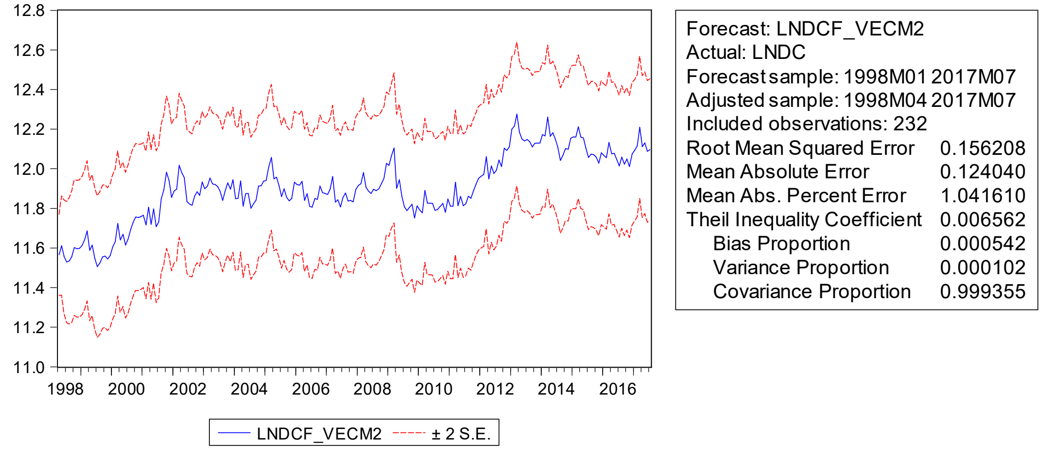


Figure 6. 4 Error of VECM(2) model

According to the Figure 6.4, RMSE, MAE and MAPE of the fitted model to the training data set is 0.156, 0.124 and 1.04 respectively. Thus it can be concluded that the predicting errors of the training data are fairly low.

### 6.7 VECM(2) Model

On the view of the above results, the equation of the model is as follows.

$$d(LNDC) = -0.163\times[CE]+ [e] \qquad \text{(Model 2)}$$

$$CE = LNDC_{t-1}+0.71\times(LNNCPI_{t-1})+1.56\times(LNEP_{t-1})- 2.12\times(LNEX_{t-1}) - 0.91\times(LNMI_{t-1})-0.04\times(LNTA_{t-1})-8.697$$

$$e = -0.14\times d(LNDC_{t-1})+0.04\times d(LNDC_{t-2})+0.07\times d(LNNPI_{t-1})- 0.65\times d(LNNCPI_{t-2})-0.41\times d(LNEP_{t-1})-0.33\times d(LNEP_{t-2})- 0.96\times d(LNEX_{t-1})-0.58\times d(LNEX_{t-2})-0.32\times d(LNMI_{t-1})- 0.11\times d(LNMI_{t-2})+0.04\times d(LNTA_{t-1})+0.026\times d(LNTA_{t-2})+0.01$$

Where d(X) is the first difference of the time series X

### 6.8 Validation of the VEC Model

Above Model (2) is validated using monthly data specified in the Table 3.1 (Section 3.1). The results of the forecasted diesel consumption and percentages of the errors are summarized in the Table 6.8 below.

Table 6. 8 Validation of the VECM (2) Model

| Time | Actual Diesel Consumption (MT) | Predicted Diesel Consumption (MT) | Percentage of Error (%) |
|---|---|---|---|
| 2017-August | 194,671 | 178,083.4 | 8.52 |
| 2017-September | 183,343 | 171,111.5 | 6.67 |
| 2017-October | 172,930 | 181,769.4 | -5.11 |
| 2017-November | 178,297 | 180,660.7 | -1.32 |
| 2017-December | 183,808 | 183,177.4 | 0.34 |
| 2018-January | 178,589 | 190,448.3 | -6.64 |
| 2018-February | 183,788 | 192,960.0 | -4.99 |
| 2018-March | 204,830 | 207,412.1 | -1.26 |
| 2018-April | 191,143 | 198,505.4 | -3.85 |
| 2018-May | 190,501 | 206,511.2 | -8.40 |
| 2018-June | 185,649 | 199,671.0 | -7.55 |
| 2018-July | 190,676 | 196,645.8 | -3.13 |

According to the Table 6.8, it is clear that the forecasted diesel consumption do not deviate considerably with the true diesel consumption as all the percentages are less than 10%. Thus it can be concluded that the errors are fairly low.

### 6.9 Summary of Chapter 6

Of the ten explanatory variables, five variables namely, LNNCPI, LNEP, LNEX, LNMI, and LNTA were selected based on the correlation coefficient with LNDC to develop a VECM (2) model. The model was trained using data January 1998 to July 2017 VECM(2) model was validated using the data from August 2017 to July, 2018. All the diagnostic tests were satisfied by the error series of the VECM(2) model. The percentage error of the model for the trained set and validation set varies between -

8.4% and +8.5%. It was also found that LNEX, LNMI, and LNTA show significant long run positive associations with LNDC while LNNCPI and LNEP indicate significant long run negative associations with LNDC.

# CHAPTER 7

# COMPARISON OF VECM(5) AND VECM(2)

In this chapter, the model developed for DC in the chapter 5: VECM(5) and the model developed for DC in Chapter 6: VECM(2) are evaluated to decide the best fitted model. .

## 7.1 Similarities of the Two Models

The residuals of the both models show that the errors are random, the distributions of errors are normal, errors shows no ARCH effect, and the errors are of constant variance. Moreover, it is evident that the coefficients of MI and EX in both models are positive while the coefficients of NCPI in both models are negative.

## 7.2 Model Comparison

The common statistics such as (i) root mean square error, (ii) mean absolute error, and (iii) mean absolute percentage error of the 2 models are presented in the Table 7.1.

Table 7. 1 Comparison of the Models

| **Model Description** | Model (1) – VECM(5) | Model (2) – VECM(2) |
|---|---|---|
| **Root Mean Square Error (RMSE)** | 15,072.15 | **10,340.45** |
| **Mean Absolute Error (MAE)** | 12,154.03 | **8,969.22** |
| **Mean Absolute Percentage of Error (MAPE)** | 6.71 | **4.82** |

Based on the results in the Table 7.1, it is evident that the model (2) shows the lowest values in all 3 error measures (RMSE, MAE, and MAPE). The corresponding values are 10,340.45, 8,969.22, and 4.82. Thus it can be concluded that the model (2) is the best fitted model out of the two VEC models.

## 7.3 Discussion

This study was set out to describe the observed variation of diesel consumption in Sri Lanka using VEC models. The discussion on major findings of the study is presented in this section.

By examining the behavior of the explanatory variables in the both models it is evident that Merchandize Imports (MI), Exchange Rates (EX) indicate significant positive relationships with diesel consumption (DC) while National Consumer Price Index (NCPI) shows significant negative relationship in the both models. Thus it can be concluded that relationships of the three variables are invariant on the order of the VECM models. Furthermore, it can be concluded that increment in MI and EX will increase DC while increment of NCCPI will cause DC to decrease.

The error series of both models found to be white noise. The percentage errors in both models for the training set as well as for the validation set are almost the same.

In this study, monthly time series data is used for the period from January 1998 to July 2018. The models developed in this study will be biased towards the past observations since the models are based on explaining the observed variation of the past data and therefore the data will contain more historical information heavily than newer information. As a result this will usually cause low prediction performance in the model. Further the biasness of the model toward the past data is a problem when developing models that plan to forecast the future. Therefore, using fewer variables and lags in a model, such as Model 2, is usually beneficial in a forecasting point of view since models that over-fit often have small errors within the sample, however do not lead to favorable forecasts.

This study is performed at an aggregate level in view of Sri Lanka as a one cluster without considering the clusters in the country and hence the developed models are of aggregate bias, which reduce the forecasting performance. However, it has been considered that there exist three sectors (urban sector, estate sector and rural sector) in Sri Lanka (Sri Lanka Census of Population and Housing, 2011). Therefore, it would have been more suitable to execute the study at a segregate level based on the three sectors to minimize the aggregate biasness and to improve predicting performance of the model.

Moreover, latest information is included rapidly on diesel consumption due to market participants can access easily this commodity. Markets have a tendency to adjust to a new equilibrium based on the new information within a short time and often more rapidly than monthly frequency. Therefore, using monthly data do not allow the developed models to measure how the diesel consumption response to some new information, since the modifications have already occurred and been consumed by the time it is forecasted. As a result, weekly or daily data is possibly more reliable than monthly data when predicting the diesel consumption in Sri Lanka.

## 7.4 Summary of Chapter 7

VECM(2) is recommended to model DC and to study the impact of other variables on DC.

# CHAPTER 8

# CONCLUSIONS, RECOMMENDATIONS AND SUGGESTIONS

Based on the results of the statistical analysis and the model developed, the following conclusions, recommendations and suggestions can be made.

## 8.1 Conclusion

- Different order of VECM models were developed to model usage of diesel and both models were tested for errors and were validated for both training and validation sets.
- Of the two models, VECM(2) is much better than VECM(5) with respect to statistical aspects as well non statistical aspects.
- National Consumer Price Index, Total Tourist Arrivals, Electricity Power Generation, Monthly Average Exchange Rates, and Merchandise Imports show highest impact to diesel consumption.
- Merchandize Exports, Distance Operated in Public Transport, Distance Operated in Rail Transport, Number of Vehicle Registered, and Wages rates of Employees show lowest impact to diesel consumption in Sri Lanka
- Of the five variables that show highest impact to diesel consumption, Total Tourist Arrivals, Monthly Average Exchange Rates, and Merchandise Imports show significant long run positive associations while National Consumer Price Index and Electricity Power Generation indicate significant long run negative relationship with diesel consumption.

## 8.2 Recommendation

- VECM(2) model is recommended to forecast monthly diesel consumption in Sri Lanka in one year ahead.
- Model provides the analyst with the ability to make decisions using various predicted intervals with different membership values by controlling the explanatory variables.
- According to the VECM(2), increase of the variables (Merchandize Imports, Monthly Average Exchange Rates and Total Tourist Arrivals) will cause to increase the consumption of diesel and hence it is required to increase the production of diesel in order to evade the scarcity of the fuel.
- Further, VECM(2) implies that the variables such as National Consumer Price Index, and Electricity Power Generation show significant negative long run relationship, increase of these factors will tend to decrease the consumption of diesel in Sri Lanka.

## 8.3 Suggestions

- The Model needs to be developed using daily or weekly data so that the model acquires market changes more rapidly and increase the predicting accuracy.
- The model need to be further developed so that the long term predictions can be made.

# REFERENCE


Abdel-Aal, R. E., Al-Garni, A. Z., & Al-Nassar, Y. N. (1997). Modelling and forecasting monthly electric energy consumption in eastern Saudi Arabia using abductive networks. *Energy*, *22*(9), 911-921.

Alam, M. S. (2006). Economic growth with energy.

Ayeni, B.J., Mohn, S.W., and Rollins, D.K., (2001). A Time Series Investigation of Cast Web Caliper Instability, *3M Internal Publication*.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting*. springer.

Central Bank of Sri Lanka. (2019). Retrieved November 9, 2019, from Cbsl.gov.lk website: https://www.cbsl.gov.lk/en/statistics/data/economic-data-library

Chai, J., Wang, S., Wang, S., &Guo, J. E. (2012). Demand forecast of petroleum product consumption in the Chinese transportation industry. *Energies*, *5*(3), 577-598.

Clemente, J. (2016, August 29). Global Oil Demand Can Only Increase. *Forbes*. Retrieved from https://www.forbes.com/sites/judeclemente/2016/08/28/global-oil-demand-can-only-increase/#a8610f931a05

Cerny, B. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate behavioral research*, *12*(1), 43-47.

Energy consumption in Sri Lanka. (2014). Retrieved from Worlddata.info website: https://www.worlddata.info/asia/sri-lanka/energy-consumption.php

Kumar, U., & Jain, V. K. (2010). Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy*, *35*(4), 1709-1716.

Liu, L. M. (1991). Dynamic relationship analysis of US gasoline and crude oil prices. *Journal of Forecasting*, *10*(5), 521-547.

Liu, L. M., & Lin, M. W. (1991). Forecasting residential consumption of natural gas using monthly and quarterly time series. *International Journal of Forecasting*, *7*(1), 3-16.

Loungani, P. (1986). Oil price shocks and the dispersion hypothesis, 1900-1980. *Rochester Center for Economic Research Working Paper*, *33*.

Lütkepohl, H., Krätzig, M., & Phillips, P. C. (Eds.). (2004). *Applied Time Series Econometrics*. Cambridge university press.

Mork, K. A. (1994). Business cycles and the oil market. *The Energy Journal*, 15-38.

Muda, N., & Mohamed, Z. (2017). The Modelling of Malaysia Crude Oil Production Using Autoregressive Integrated Moving Average (ARIMA) Model. *Terengganu International Finance and Economics Journal (TIFEJ)*, *1*(2), 1-9.

National Consumer Price Index for Sri Lanka - Department of Census and Statistics (DCS). (2020). Retrieved on February 26, 2020, from DCS website: http://www.statistics.gov.lk/price/NCPI/NCPI%20Tecnical%20%20Note.pdf

Pao, H. T. (2009). Forecasting energy consumption in Taiwan using hybrid nonlinear models. *Energy*, *34*(10), 1438-1446.

Prentice, A. M., &Poppitt, S. D. (1996). Importance of energy density and macronutrients in the regulation of energy intake. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity*, *20*, S18-23.

Sri Lanka Census of Population and Housing, - Department of Census and Statistics (DCS). (2011). Retrieved on February 26, 2020, from DCS website: http://www.statistics.gov.lk/PopHouSat/CPH2011/index.php?fileName=ConceptsandDefinitions&gp=StudyMaterials&tpl=2

Use of diesel - U.S. Energy Information Administration (EIA). (2016). Retrieved

November 9, 2019, from Eia.gov website:
https://www.eia.gov/energyexplained/diesel-fuel/use-of-diesel.php

Warr, B. S., & Ayres, R. U. (2010). Evidence of causality between the quantity and quality of energy consumption and economic growth. *Energy*, *35*(4), 1688-1693.

Wikipedia Contributors. (2019, April 23). World energy consumption. Retrieved April 28, 2019, from Wikipedia website:
https://en.wikipedia.org/wiki/World_energy_consumption