FORECASTING PETROL AND DIESEL PRICE IN TAMIL NADU USING ARIMA MODEL

Mini Project submitted to Manonmaniam Sundaranar University
in partial fulfilment of the requirement for the
award of the degree of

Master of Science in Statistics

Submitted by

VELKUMAR V

(Reg. No. 20214012529133)

Department of Statistics

Manonmaniam Sundaranar University

Tirunelveli – 627 012

APRIL 2022

BONAFIDE CERTIFICATE

Certified that this Project entitled **"FORECASTING PETROL AND DIESEL PRICE IN TAMIL NADU USING ARIMA MODEL"** is the Bonafide work of **VELKUMAR. V (RegNo.20214012529133)** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported here in does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Signature of the HOD	Signature of the Supervisor
Dr.K.SENTHAMARAI KANNAN	DR.P.ARUMUGAM
Senior Professor and Head Department of Statistics Manonmaniam Sundaranar University Tirunelveli – 627 012	Professor Department of Statistics Manonmaniam Sundaranar University Tirunelveli – 627012

Declaration

I hereby declare that the project entitled **" FORECASTING PETROL AND DIESEL PRICE IN TAMIL NADU USING ARIMA MODEL"** submitted to Manonmaniam Sundaranar University I partial fulfilment of the requirements for the award of the degree of Master of Science in Statistics is my original work and that it has not previously formed the basis for the award of any degree, diploma or similar title of any University or Institution.

Tirunelveli (VELKUMAR V)

April 2022

ACKNOWLEDGEMENT

I would thank God for being able to complete this project with success. Then I would like to special thanks of gratitude to my Guide **Dr.P.Arumugam,** Professor, Department of Statistics, for his guidance who gave me the golden opportunity to do this wonderful project on the topic "**FORECASTING PETROL AND DIESEL PRICE IN TAMIL NADU USING ARIMA MODEL**", which also helped me in doing a lot of analysis and I came to know about so many new things and as well as providing necessary information regarding project and also for his support to completing the project.

I wish to express my sincere thanks to **Dr.K.Senthamarai Kannan** Senior Professor and Head, Department of Statistics, **Dr.A.Loganathan** Professor, Department of Statistics, **Dr.A.Rajarathinam** Professor, Department of Statistics, **Dr.R.Sasikumar** Assistant Professor, Department of Statistics, **Dr.V.Deneshkumar** Assistant Professor, Department of Statistics, **Dr.K.Manoj** Assistant Professor, Department of Statistics, for their help rendered during my study period.

I wish to express my thanks to Research Scholar's, Department of Statistics. Then I would like to thank my parents and last but not least I would like to thank my friends who have helped me a lot with their valuable suggestions and guidance has been helpful in various phases of the completion of the project successful.

(VELKUMAR V)

TABLE OF CONTENTS

CHAPTER-I

Introduction

Petroleum is referred to as **“Black Gold.”** This name itself is an indication of its importance to humans. Crude oil is considered to be the **“mother of all commodities”** as it is used to manufacture various products such as pharmaceuticals, plastics, gasoline, synthetic fabrics, etc. Petroleum or oil has also been the world’s leading source of energy since the 1950s.

Petroleum is a liquid which occurs naturally in rock formations. This consists of a complex mixture of different molecular weights of hydrocarbons, plus other organic compounds. Some petroleum-produced chemical compounds are also obtained from other fossil fuels.

Petrochemicals are produced mainly at a few manufacturing sites around the world. Petroleum is also the raw material for many industrial products, including pharmaceuticals, solvents, fertilizers, pesticides, synthetic fragrances, and plastics.

Petroleum or mineral oil is India’s next biggest source of energy after coal. It supplies heat and lighting power, machinery lubricants, and raw materials for a variety of manufacturing industries. Petroleum refineries for synthetic textiles, fertilisers and numerous chemical industries act as a “nodal industry.” Most of India ‘s petroleum occurrences are associated with anticlines and fault traps in tertiary-age rock formations. It occurs in folding regions, anticlines, or domes, where oil is trapped in the unfold crest.

Uses of Petroleum

Refined products obtained from crude oil have a number of uses.

- Liquefied Petroleum Gas or LPG is used in households as well as in the industry.
- Diesel and petrol are used as fuels for vehicles. Diesel is generally preferred for heavy motor vehicles.
- Petrol is also used as a solvent for dry cleaning, whereas diesel is also used to run electric generators.
- Kerosene is used as a fuel for stoves and jet planes.
- Lubricating oil reduces wear and tear and corrosion of machines.

- Paraffin wax is used to make candles, ointments, ink, crayons, etc.
- Bitumen or asphalt is mainly used to surface roads.

When people consider petroleum, they first think of energy. Petroleum and other fossil fuels now provide more than 86% of the energy consumed by mankind. In addition, fossil resources, especially petroleum and natural gas, serve as the organic source of tens of thousands of consumer products, which enrich our daily lives.

To understand petroleum and the petroleum industry, one must be familiar with the technology used to find and recover crude oil and natural gas and transform them into useful products. These technologies can also be applied to gases and liquids from coal, shale, and renewable biomass. Research and development aimed at improving or modifying existing technologies and developing new ones usually require physical testing and chemical characterization.

Three-dimensional imaging exhibits geological formations most likely to contain oil and gas. Rigorous basin modeling optimizes exploration and production. Modern production technology includes enhancements in horizontal drilling and offshore platform design. The application of hydraulic fracturing to previously unrecoverable oil and gas from tight reservoirs has transformed the United States into the world's leading producer of oil and gas.

Midstream technology includes trading, shipping, and transportation, along with processing prior to transportation. Midstream processing includes froth treatment for upgrading bitumen from steam-assisted gravity drainage (SAGD) into synthetic crude oil (syncrude). Sophisticated planning models enable global energy companies to quickly decide logistics: which oils to buy, how to allocate them between numerous processing plants, and whether to resell them.

Crude oil goes to refineries, which use distillation, treating, conversion, extraction and blending processes to produce fuels, hydrogen, lubricants, waxes, coke products, asphalt, and sulfur. Some refinery streams are sent to petrochemical plants.

Ever-improving mathematical models enhance all aspects of petroleum technology. Model-predictive control (MPC) stabilizes operations and reduces product-quality giveaway, increasing profitability at relatively low cost. The simple return on investment for an MPC project can be 3–4 months. Process engineers rely on rigorous equipment and piping models to optimize designs, not just in the oil and chemical businesses, but in all process industries. With such models, energy consumption in processing plants has been reduced by up to 70% since the 1980s. Rigorous reaction models, based on molecular characterization, serve as the foundation for real-time online economic optimization, in some cases for entire refineries. Economic optimization uses an objective function to find the most profitable balance between equipment constraints, feed quality, product yields, product properties, and utilities costs.

One cannot over-emphasize the importance of safety and protection of the environment. Failure to understand technology fundamentals and process details is the root cause of many infamous industrial catastrophes. Lack of understanding of technology fundamentals occurs at all levels, from the control board to the board room. Corporate executives who insist that safety is Number One must invest in safety-enhancing infrastructure. They must ensure that operators are well-trained and equipment is well.

Petroleum will remain significant for decades to come. Hopefully, ever-advancing technology will continue to supply energy and raw materials while protecting workers and the environment.

DIESEL

The most common type of diesel fuel is a specific fractional distillate of petroleum fuel oil, but alternatives that are not derived from petroleum, such as biodiesel, biomass to liquid (BTL) or gas to liquid (GTL) diesel are increasingly being developed and adopted. To distinguish these types, petroleum-derived diesel is increasingly called petrol diesel in some academic circles.

In many countries, diesel fuel is standardised. For example, in the European Union, the standard for diesel fuel is EN 590. Diesel fuel has many colloquial names; most commonly, it is simply referred to as diesel. In the UK, diesel fuel for on-road use is commonly abbreviated DERV, standing for diesel-engined road vehicle, which carries a tax premium over equivalent fuel for non-road use. In Australia, diesel fuel is also known as distillate, and in Indonesia, it is known as Solar, a trademarked name from the country's national petroleum company Pertamina.

Ultra-low-sulfur diesel (ULSD) is a diesel fuel with substantially lowered sulfur contents. As of 2016, almost all of the petroleum-based diesel fuel available in the UK, mainland Europe, and North America is of a ULSD type.

Before diesel fuel had been standardised, the majority of diesel engines typically ran on cheap fuel oils. These fuel oils are still used in watercraft diesel engines. Despite being specifically designed for diesel engines, diesel fuel can also be used as fuel for several non-diesel engines, for example the Akroyd engine, the Stirling engine, or boilers for steam engines

Origin

Diesel fuel originated from experiments conducted by German scientist and inventor Rudolf Diesel for his compression-ignition engine he invented in 1892. Originally, Diesel did not consider using any specific type of fuel, instead, he claimed that the operating principle of his rational heat motor would work with any kind of fuel in any state of matter. However, both the first diesel engine prototype and the first functional Diesel engine were only designed for liquid fuels.

In addition to that, Diesel experimented with different types of lamp oil from various sources, as well as different types of petrol and ligroin, which all worked well as Diesel engine fuels. Later, Diesel tested coal tar creosote, paraffin oil, crude oil, gasoil, and fuel oil, which eventually worked as well. In Scotland and France, shale oil was used as fuel for the first 1898 production Diesel engines because other fuels were too expensive. In 1900, the French Otto society built a Diesel engine for the use with crude oil, which was exhibited at the 1900 Paris Exposition and the 1911 World's Fair in Paris. The engine actually ran on peanut oil instead of crude oil, and no modifications were necessary for peanut oil operation

During his first Diesel engine tests, Diesel also used illuminating gas as fuel, and managed to build functional designs, both with and without pilot injection According to Diesel, neither was a coal-dust–producing industry existent, nor was fine, high quality coal-dust commercially available in the late 1890s. This is the reason why the Diesel engine was never designed or planned as a coal-dust engine. Only in December 1899, did Diesel test a coal-dust prototype, which used external mixture formation and liquid fuel pilot injection. This engine proved to be functional, but suffered from piston ring failure after a very few minutes due to coal dust deposition.

Since the 20^{th} Century

Before diesel fuel had been standardised, diesel engines typically ran on cheap fuel oils. In the United States, these were distilled from petroleum, whereas in Europe, coal-tar creosote oil was used. Some diesel engines were fuelled with mixtures of several different fuels, such as petrol, kerosine, rapeseed oil, or lubricating oil, because they were untaxed and thus cheap The introduction of motor-vehicle diesel engines, such as the Mercedes-Benz OM 138, in the 1930s meant that higher quality fuels with proper ignition characteristics were needed. At first no improvements were made to motor-vehicle diesel fuel quality. After World War II, the first modern high quality diesel fuels were standardised. These standards were, for instance, the DIN 51601, VTL 9140-001, and NATO F 54 standards. In 1993, the DIN 51601 was rendered obsolete by the new EN 590 standard, which has been used in the European Union ever since. In sea-going watercraft, where diesel propulsion had gained prevalence by the late 1970s due to increasing fuel costs caused by the 1970s energy crisis, cheap heavy fuel oils are still used instead of conventional motor-vehicle diesel fuel. These heavy fuel oils (often called Bunker) can be used in diesel-powered and steam-powered vessels.

Chapter - II

METHODOLOGY

2.1 Time Series

A set of ordered observations of a quantitative variable taken at successive points in time is known as time series. In other words, the arrangement of statistical data in chronological order i.e., in accordance with occurrence of time, is known as time series. Time, in terms of years, months, days, or hours is simply a device that enables one to relate all phenomenon to set of common, stable reference points.

Applications of statistical methods enable to make decisions on several aspects of many real-world problems based on the given information. The information can be classified accordingly whether it is collected in periodical time intervals. If values of the characteristic under study are observed in a time interval, then the information may be called as a time series.

Analysis of time series helps to understand variations in the data over time periods and the pattern of variations in the data. A time series model fitted appropriately to the given data can be used 2 for forecasting. Predicting information to future time periods is useful to the administrators and policy makers.

2.2 Components of time series

The factors that are responsible for bringing about changes in a time series, also called the components of time series, are as follows:

1. Secular Trends (or) General Trends

2.Seasonal Movements

3.Cyclical Movements

4.Irregular Fluctuations

Secular Trends

The secular trend is the main component of a time series which results from long term effects of socio-economic and political factors. This trend may show the growth or decline in a time series over a long period. This is the type of tendency which continues to persist for a very long period. Prices and export and import data, for example, reflect obviously increasing tendencies over time.

Seasonal Trends

These are short term movements occurring in data due to seasonal factors. The short term is generally considered as a period in which changes occur in a time series with variations in weather or 4 festivities. For example, it is commonly observed that the consumption of ice-cream during summer is generally high and hence an ice-cream dealer's sales would be higher in some months of the year while relatively lower during winter months. Employment, output, exports, etc., are subject to change due to variations in weather. Similarly, the sale of garments, umbrellas, greeting cards and fire-works are subject to large variations during festivals like Valentine’s Day, Eid, Christmas, New Year's, etc. These types of variations in a time series are isolated only when the series is provided biannually, quarterly or monthly.

Cyclic Movements

These are long term oscillations occurring in a time series. These oscillations are mostly observed in economics data and the periods of such oscillations are generally extended from five to twelve years or more. These oscillations are associated with the well known business cycles. These cyclic movements can be studied provided a long series of measurements, free from irregular fluctuations, is available.

Irregular Fluctuations

These are sudden changes occurring in a time series which are unlikely to be repeated. They are components of a time series which cannot be explained by trends, seasonal or cyclic movements. These variations are sometimes called residual or random components. These variations, though accidental in nature, can cause a continual change in the trends, seasonal and cyclical oscillations during the forthcoming period. Floods, fires, earthquakes, revolutions, epidemics, strikes etc., are the root causes of such irregularities.

2.3 Uses of time series

The analysis of time series has been found useful to economists and business persons, in particular, and also the scientists, sociologist, etc. It has also found its utility in microbiology, seismology, oceanography, geomorphology, etc., in earth sciences: electrocardiograms in medical science and problem of estimating missile trajectories. Time series analysis helps in understanding the following phenomena.

- **(i)** It helps in knowing the real behavior of the past.
- **(ii)** It helps in predicting the future behavior like demand, production, weather conditions, prices, etc.
- **(iii)** It helps in planning the future operations.
- **(iv)** Analysis of time series helps to compare the present accomplishments with the past performances.
- **(v)** Two or more time series can be compared belonging to the same reference period.

2.4 Drawbacks of the Time series

The drawbacks of the time series analysis can be summarized as follows:

- **(i)** The conclusions drawn on the basis of the time series analysis are not cent percent true.
- **(ii)** Time series analysis is unable to fully adjust the influence affecting a time series like customs, climate, policy changes, etc.
- **(iii)** The complex forces affecting a time series existing at certain period may not have the same complex forces in future. Hence, the forecasts may not hold true.

The information about petrol and diesel on Tamil Nadu is considered in monthly wise. The observations in the time series may be auto correlated. Before fitting a time series model to the given data, the stationary of a time series should be verified. The Box-Jenkins methodology can be applied to fit the time series model. The methodology for determining the autoregressive integrated moving average models, simple exponential smoothing and double exponential smoothing, for a time series is described in the following subsections.

2.5 Examining Stationary of Time Series

Firstly, stationary of the time series should be examined. If the mean and variance of a time series are constants over time, the time series may be considered as stationary. Plotting of such data against time will be horizontal along the time axis. Stationary means that there is no growth or decline in the data. The following graphical and analytical procedures may be followed to examine the stationary of a time series

- Time plot (Plotting the observations against time)
- Plotting of autocorrelation function values against time and
- Plotting of partial autocorrelation function values against time.

Time Plot

In the time plot, various time periods during which the observations are collected are taken along the X-axis. The values observed for the forecast variable are taken along the Y-axis. Then, the data is plotted in a graph. Successive points are joined by line segments. The resultant graph represents the given time series. If the data fluctuates horizontally along with the time axis, then it indicates that the mean of the time series is constant. In other words, the data fluctuates around a constant mean which is independent of time without shift. Then the time series is said to be stationary in mean. If the plotted series reveals no obvious change in the pattern of variation over time, then the series is said to be stationary in variance. If a time series is stationary in both mean and variance, then it is called a stationary time series.

Autocorrelation Function

Autocorrelation in a time series is a correlation between values of the same variable at different time periods. Correlation between current observations(Y_1) and observation from P periods before the current one (Y_{t-p}). This is for a given (Y_t).Autocorrelation at lag

P=correlation (Y_t, Y_{t-p})and is given by

$$r_p = \frac{\sum_{t=1}^{n-p}(Y_1 - \bar{Y})(Y_{t-p} - \bar{\bar{Y}})}{\sum_{t=1}^{n}(Y_t - \bar{Y})}$$

It ranges from -1 to+1. Then $r_1$indicates how successive values of Y relate to each other and so on. The pattern of autocorrelations for lags 1, 2,.. Lag is known as the autocorrelation functions or ACF. A plot of the ACF against the lag is known as the correlogram. It is frequently used to identify whether or not seasonality is present in a given time series(and the length of that seasonality), to identify appropriate time series models for specific situations, and to determine if data are stationary.

Partial Autocorrelation Function

Partial autocorrelation function is used to measure the degree of association between Y_t andY_{t-k} , when the effect of the observations at time lags $t = 1, 2, 3.., k-1$ are removed. The partial autocorrelation function is another graphical tool used for testing stationary of time series. The partial autocorrelation co-efficient of order k can be calculated by regressing Y_t against Y_{t-1}……,Y_{t-k} as

$$Y_1 = b_0 + b_t Y_{t-1} + b_2 Y_{t-2} + \ldots\ldots + b_t Y_{t-k} \quad , \quad K = 1,2,3, \ldots$$

Partial autocorrelation co-efficient for different time lags $k = 1, 2, 3,$ ... are the values of k b which can be estimated by fitting the above model. These are the values of partial autocorrelation function (PACF). Values of PACF may be plotted against different time lags taken along the X-axis. As like ACF, the spikes corresponding to partial autocorrelation co-efficient should be close to zero and/or should lie within the upper and lower limits for stationary time series. Otherwise, the time series will be non-stationary.

Autoregressive Process

Let Y_t represent the time series model at the time period t, model Y_t, as

$$Y_t\text{-}\delta = \alpha_1(Y_{t-1}) + u_t$$

Where δ is the mean of Y

Where u_1 is an uncorrelated random error term with the zero mean and constant varianceσ^2

Thus Y_t follows a first order autoregressive or AR (1) Stochastic process. Here the value of Y at time t depends on its value in the previous time period and a random term; the Y values are expressed as deviations from their mean value . In other words, this model says that the forecast value of Y at time t is simply proportion (=α_1) of its value at ttime(t-1) plus a random shock or disturbance at time t; again the Y values are expressed around their mean values. But if consider this model,

$$Y_t\text{-}\delta = \alpha_1(Y_t - \delta) + \alpha_3(Y_t - \delta) + u_t$$

Then one can say that Y_t follows a second order autoregressive or AR (2) process. That is, the value of Y at time t depends on its value in the previous two time periods, the Y values being expressed around their mean value δ.

$$Y_t\text{-}\delta = \alpha_1(Y_{t-1} - \delta)\alpha_2(Y_{t-2} - \delta) + \ldots + \alpha_p(Y_{t-p} - \delta) + u_t$$

In which case Y_t is p^{th} order autoregressive or AR (p) process.

Autoregressive and Moving Average (ARMA)

It is quite that Y has characteristics of both AR and MA and is therefore ARMA. Thus,Y_t follows an ARMA (1,1) if it can be written as,

$$Y_t - \delta = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \ldots. + \alpha_p(Y_{t-p} - \delta) + u_t$$

Because there is one autoregressive and one moving average term . In the above equation θ represents the constant term.

In general, in an ARMA (p, q) process, there will be p Autoregressive and q Moving average terms.

Autoregressive Integrated Moving Average (ARIMA) Process

Auto-Regressive Integrated Moving Average (ARIMA) models are a broad class of time series models which generate forecasts based on past trends. Whilst they share many characteristics of simple trend extrapolation methods, they have a separate literature and set of methods, and are significantly more complex than trend extrapolation methods, hence their description in a separate section in this report. The current set of methods for creating ARIMA models was initially developed by Box and Jenkins (1970).

In algebraic notation an ARIMA model is written as ARIMA (p, d, q) where p denotes the number of autoregressive terms, d the order of differencing and q the number of moving average terms. Mostly p, d and q take values of 0, 1 or 2. Some specific examples of ARIMA models are presented to demonstrate their features. An autoregressive model is one in which the current value includes part of the preceding value (Makridakis et al. 1998 p 337):

$$Y(t) = \varphi Y(t-1) + c + \epsilon(\mathrm{t})$$

Where θ is the auto regressive parametric constant and random Error. This may be described as an ARIMA (1,0,0) model, or simply AR(1). It only has one autoregressive term, no integration and no 33 moving average terms. A moving average model includes the current error term and part of the previous error term (Makridakis et al. 1998 p 339):

$$Y(t) = c + \epsilon(\mathrm{t}) - \theta\epsilon(\mathrm{t} - 1)$$

Where θ is depend on the moving average parameter. This may be described as an ARIMA (0,0,1) model or simply MA(1). A model with just differencing is known as a random walk model and may be expressed as:

$$Y(t) = Y(t-1) + c + \epsilon(\mathrm{t})$$

The current value of the time series equals the previous value plus random error. Where the constant is non-zero it is termed a random walk with drift model. Various combinations of autoregressive, differencing and moving average parameters may be used. For example, the commonly-used ARIMA(1, 1,0) model is:

$$Y(t) = Y(t-1) + \emptyset(Y(t-1) - (Y(t-2) + c + \epsilon(\mathrm{t})$$

Whilst an ARIMA (0, 1, 1) is:

$$Y(t) = Y(t-1) + c + \epsilon(\mathrm{t}) - \theta\epsilon(\mathrm{t}-1)1$$

Stationary process

Stationary is defined as a quality of a process in which the statistical parameters (mean and standard deviation) of the process do not change with time. The most important property of Stationary process is that the auto-correlation function depends on lag alone and does not change with time at which the function was calculated.

Non-Stationary Time Series

If a time series is not stationary time series. In other words, a non stationary time series will have a time varying mean or a time- varying variance or both

CHAPTER III

DATA ANALYSIS AND INTERPRETATION

The data collected on price of petrol and diesel in Tamil Nadu for the period, 1st may 2017to 1st April 2022 is analyzed in this chapter applying methodology described in chapter II. The dataset was collected from https://www.petroldieselprice.com/. It consists information of 60 observations and three variables. One variable is "Month" which is independent variable and another variable is "Diesel price" which is independent variable and another variable is "petrol price". Since the data is collected at equal time interval (1Month), it is attempted to determine a suitable time series model.

3.1 Diesel price

Descriptive Analysis:

The basic statistical measures of the data are calculated to describe its properties.

Table 3.1.1

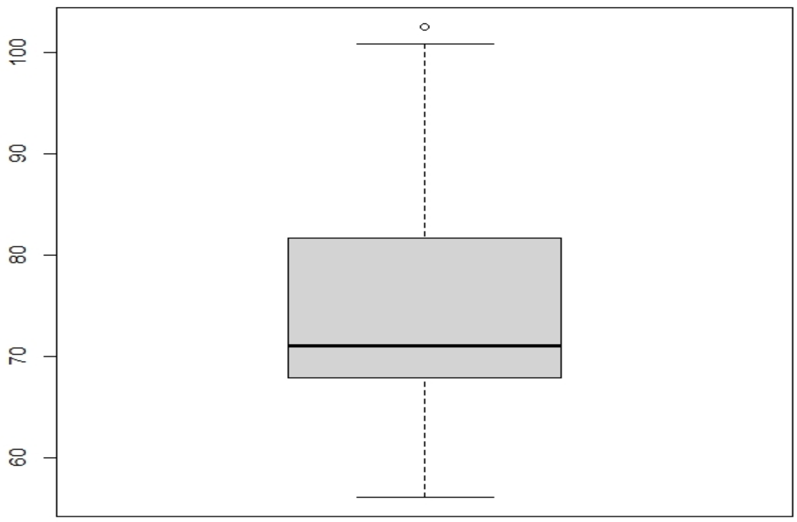
Descriptive statistics for Diesel price in Tamil Nadu

Statistical Measures	**Diesel**
Mean	75.2766
Median	71.1200
Maximum	102.59
Mode	91.43
Minimum	56.13
1st Quartile	67.8050
3rd Quartile	83.7300
Skewness	.645
Kurtosis	-.499

From this empirical distribution, values of some of the central positional and dispersion measures have been calculated using R software version 4.0.2. The values are presented in Table 3.1.1. It deviates much from the average Diesel price is75.2766. Among the Diesel price, maximum of 50% of them were less than that of 71.32 price. The maximum Diesel price is 102.59. and the minimum Diesel price is 56.13. The first and third quartiles values specified that the maximum of 25% of the Diesel price was less than 67.8050 and another 25% of them was more than 83.7300. It is positively skewed with long tail at right and the kurtosis is leptokurtic.

Fig 3.1.1

Box plot for Diesel price in Tamil Nādu



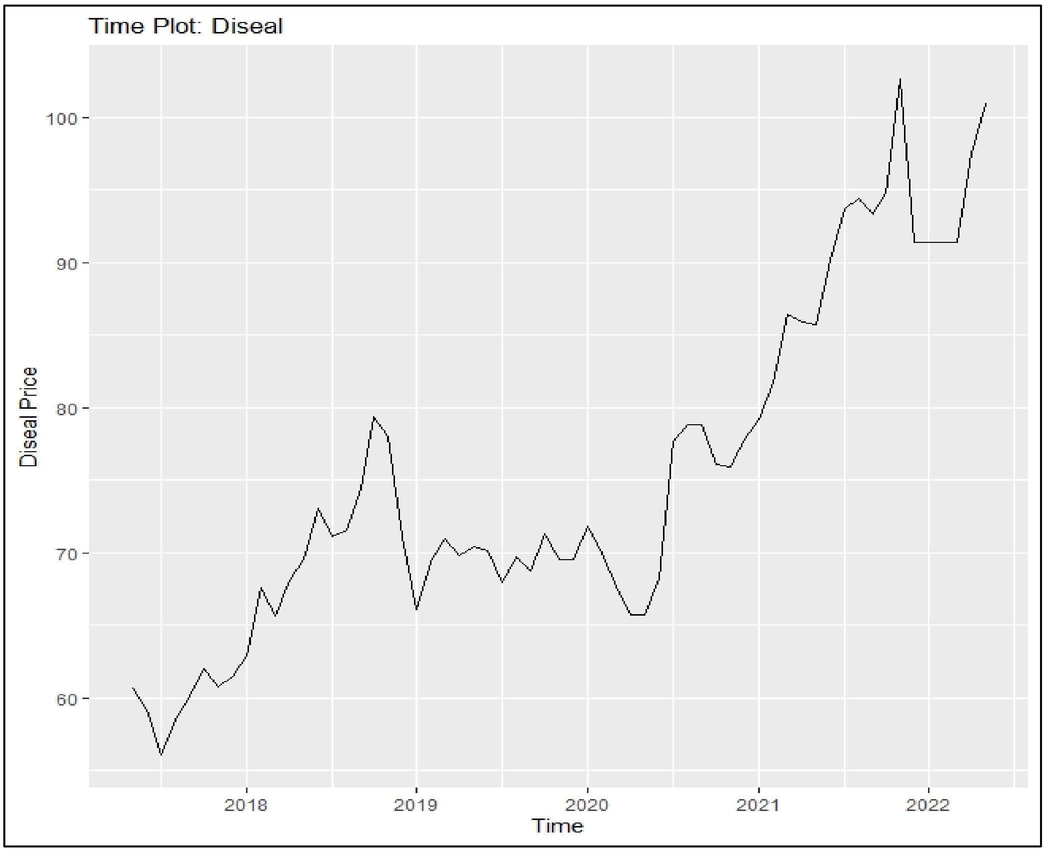
Box Whisker plot shows that the distance between third quartile and median is smaller than the distance between first quartile and median. There is variation in length of both the tails. There is one outlier found in the data by box plot it because of the suddenly hike of the diesel price. The spread of the price of diesel in the data width ranges of more than 80rs.

Testing the stationary

Time series models and the methodology used for fitting them to a given set of data can be chosen based on the stationary value of the time series. As described in the preceeding chapters, stationary time series can be analyzed using time plot.

Fig 3.1.2

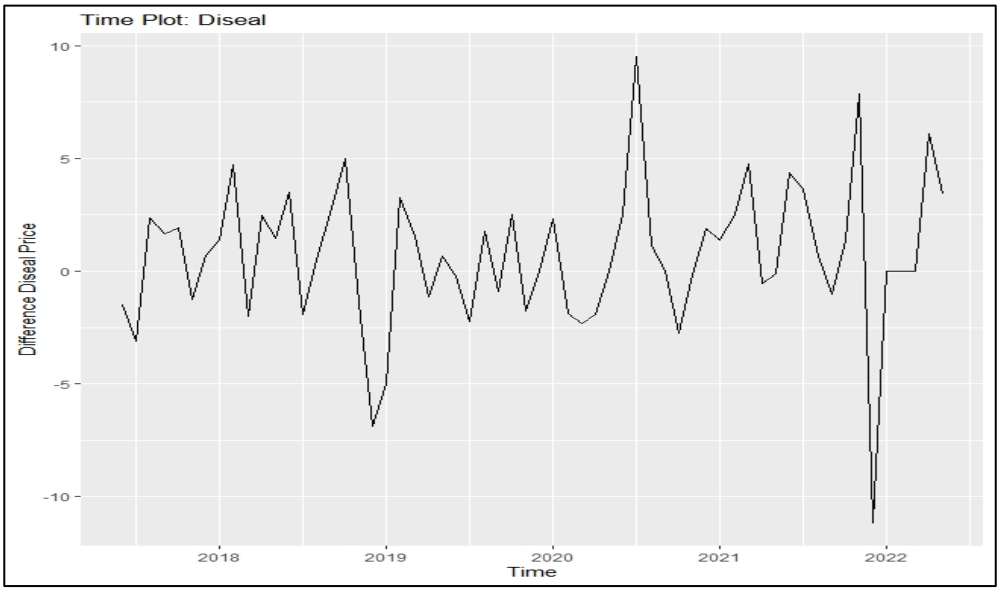
Diesel price in Tamil Nādu from 2017 to 2022



The time plot reveals the elements of the time series fluctuate moderately and there is an upward and downward movement in the fluctuations. As time passes, the pattern of variation changes. Additionally, a shift in the constant around which the observations oscillate is also observed. According to the study, the mean price of Diesel in a given month is not a fixed number and it varies over time. The data indicates that the time series of Monthly price of Diesel price in Tamil Nādu is non-stationary in mean and variance.

Fig 3.1.3

Second-Order differences Diesel price in Tamil Nadu from 2017 to 2022



Differencing method is applied to the actual data for removing the non-stationarity. The second order differences $Y'_t = Y_t - Y_{t-1} - Y_{t-2}$ are calculated to the time series. The time series of the Second differenced values of the monthly diesel price is stationary. Hence, the appropriate model can be determined to this stationary time series and the model can be used to forecasts.

Before forecasting the best model will be checked, among 59 models of ARIMA the Minimum AIC value is274.04and this model will be the best one. All the ARIMA model and the Akaike's information criterion Value has been listed in the Table 3.2.

Table 3.1.2

Model Comparison

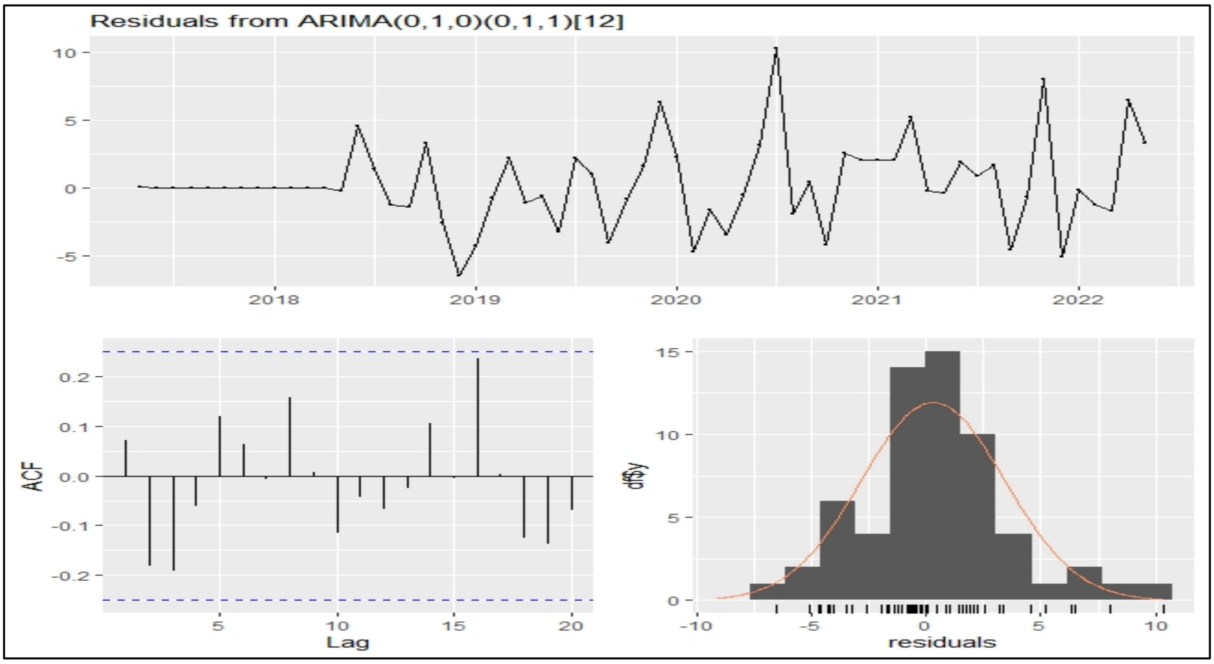
Model	AIC
ARIMA(0,1,0)(0,1,0)[12]	: 282.9323
ARIMA(0,1,0)(,1,1)[12]	: 274.3067
ARIMA(0,1,0)(10,1,0)[12]	: 276.9269
ARIMA(0,1,0)(1,1,1)[12]	: 276.5529
ARIMA(0,1,1)(0,1,0)[12]	: 284.6224
ARIMA(0,1,1)(0,1,1)[12]	: 276.5721
ARIMA(0,1,1)(1,1,0)[12]	: 279.0217

ARIMA(0,1,1)(1,1,1)[12]	: 278.9216
ARIMA(0,1,2)(0,1,0)[12]	: 284.3821
ARIMA(0,1,2)(0,1,1)[12]	: 277.479
ARIMA(0,1,2)(1,1,0)[12]	: 280.0981
ARIMA(0,1,2)(1,1,1)[12]	: 279.9772
ARIMA(0,1,3)(0,1,0)[12]	: 286.6721
ARIMA(0,1,3)(0,1,1)[12]	: 278.9812
ARIMA(0,1,3)(1,1,0)[12]	: 281.9933
ARIMA(0,1,3)(1,1,1)[12]	: 281.6009
ARIMA(0,1,4)(0,1,0)[12]	: 287.5588
ARIMA(0,1,4)(0,1,1)[12]	: 281.581
ARIMA(0,1,4)(1,1,0)[12]	: 284.229
ARIMA(0,1,5)(0,1,0)[12]	: 288.1499
ARIMA(1,1,0)(0,1,0)[12]	: 284.7893
ARIMA(1,1,0)(0,1,1)[12]	: 276.5761
ARIMA(1,1,0)(1,1,0)[12]	: 279.0727
ARIMA(1,1,0)(1,1,1)[12]	: 278.9263
ARIMA(1,1,1)(0,1,0)[12]	: 285.4632
ARIMA(1,1,1)(0,1,1)[12]	: 278.0781
ARIMA(1,1,1)(1,1,0)[12]	: 281.1386
ARIMA(1,1,1)(1,1,1)[12]	: Inf
ARIMA(1,1,2)(0,1,0)[12]	: 286.5378
ARIMA(1,1,2)(0,1,1)[12]	: 279.42
ARIMA(1,1,2)(1,1,0)[12]	: 282.1036
ARIMA(1,1,2)(1,1,1)[12]	: 282.0402
ARIMA(1,1,3)(0,1,0)[12]	: 289.026
ARIMA(1,1,3)(0,1,1)[12]	: 281.5966
ARIMA(1,1,3)(1,1,0)[12]	: 284.5367
ARIMA(1,1,4)(0,1,0)[12]	: 289.9264
ARIMA(2,1,0)(0,1,0)[12]	: 285.327
ARIMA(2,1,0)(0,1,1)[12]	: 277.8277
ARIMA(2,1,0)(1,1,0)[12]	: 280.2894
ARIMA(2,1,0)(1,1,1)[12]	: 280.3159
ARIMA(2,1,1)(0,1,0)[12]	: 286.3104
ARIMA(2,1,1)(0,1,1)[12]	: 278.8943
ARIMA(2,1,1)(1,1,0)[12]	: 281.6177
ARIMA(2,1,1)(1,1,1)[12]	: 281.5144
ARIMA(2,1,2)(0,1,0)[12]	: 288.7434
ARIMA(2,1,2)(0,1,1)[12]	: Inf

ARIMA(2,1,2)(1,1,0)[12]	: Inf
ARIMA(2,1,3)(0,1,0)[12]	: 291.1818
ARIMA(3,1,0)(0,1,0)[12]	: 287.0455
ARIMA(3,1,0)(0,1,1)[12]	: 278.2952
ARIMA(3,1,0)(1,1,0)[12]	: 281.4656
ARIMA(3,1,0)(1,1,1)[12]	: 280.9122
ARIMA(3,1,1)(0,1,0)[12]	: 288.5837
ARIMA(3,1,1)(0,1,1)[12]	: 280.6402
ARIMA(3,1,1)(1,1,0)[12]	: 283.6994
ARIMA(3,1,2)(0,1,0)[12]	: 290.8743
ARIMA(4,1,0)(0,1,0)[12]	: 286.3404
ARIMA(4,1,0)(0,1,1)[12]	: 280.4425
ARIMA(4,1,0)(1,1,0)[12]	: 283.08
ARIMA(4,1,1)(0,1,0)[12]	: 288.4166
ARIMA(5,1,0)(0,1,0)[12]	: 288.4232
Best model: ARIMA(0,1,0)(0,1,1)[12]	

Fig 3.1.4

Residual and ACF plot for ARIMA model



The figure 3.1.4 shows that the ARIMA method produces forecasts that appear to account for all available information. The mean of the residuals is close to zero and there is no significant correlation in the residuals series. The time plot of the residuals shows that the variation of the residuals stays much the same across the historical data, apart from the one outliers, and therefore the residual variance can be treated as constant. This can also be seen on the histogram of the residuals. The histogram suggests that the residuals may be normal — the Right tail seems a little too long, even when we ignore the outlier. Consequently, forecasts from this method will probably be quite good, but prediction intervals that are computed assuming a normal distribution may be accurate.

The lower AIC and BIC values are the higher Accuracy of model and the Residuals plot also given in the Table 3.1.3

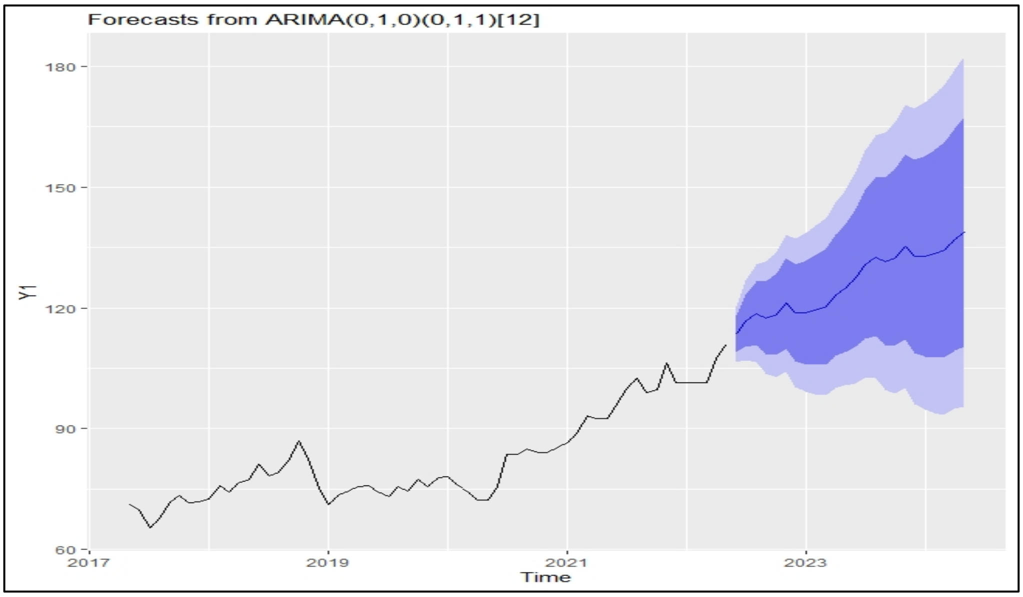
Table 3.1.4

AIC and BIC Values

AIC	BIC
274.04	277.78

Fig 3.1.5

Diesel price Forecasting using ARIMA (0,1,0)(0,1,1)[12]model



The Fig 3.1.5 indicate that forecasts are shown as a blue line, with the 80% prediction intervals as a dark shaded area, and the 95% prediction intervals as a light shaded area. This is the overall process by which we can analyze time series data and forecast values from existing series using ARIMA.

Lastly the Production is forecasted for the upcoming 24months, the forecasted Production of Diesel price as given in Figure 3.1.5. The error factors such as ME, RMSE, MAE, MPE, MAPE and MASE are used to determine the accuracy of the model. The Mean Error (ME) is 0.1744464, Root Mean Square Error (RMSE) is 3.339549, Mean Absolute Error (MAE) is 2.11134, Mean Percentage Error (MPE) is 0.1015712, Mean Absolute Percentage Error (MAPE) is 2.677937, Mean Absolute Scaled Error (MASE) is 0.2149237 and the p-value is 0.008952579, which indicates the accuracy is good for prediction.

Table 3.1.5

Training set error measures

	ME	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**P value**
Training Set	0.1744464	3.339549	2.11134	0.1015712	2.677937	0.2149237	0.008952579

Table 3.1.6

Forecasting value for Diesel price in Tamil Nadu

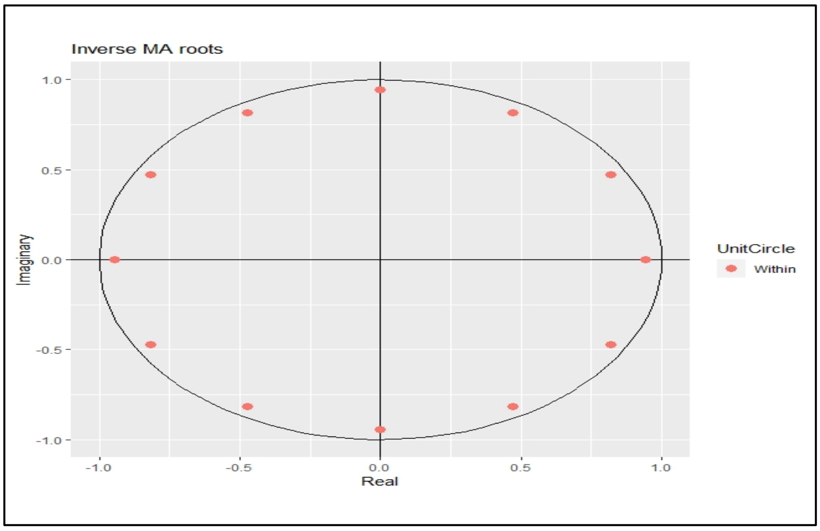
Diesel			
Month	Point Forecast	Lo 95	Hi 95
Jun-22	103.3478	95.85946	110.8361
Jul-22	105.9748	95.38472	116.5648
Aug-22	107.0991	94.12896	120.0692
Sep-22	107.1114	92.13484	122.088
Oct-22	108.1896	91.44529	124.934
Nov-22	110.3477	92.00519	128.6902
Dec-22	106.0243	86.21214	125.8365
Jan-23	106.2671	85.08701	127.4472
Feb-23	107.4668	85.00194	129.9318
Mar-23	108.1999	84.51985	131.88
Apr-23	110.0339	85.19799	134.8698
May-23	111.4623	85.52207	137.4026
Jun-23	113.8701	86.03427	141.7059
Jul-23	116.4971	86.88682	146.1074

Aug-23	117.6214	86.33713	148.9057
Sep-23	117.6338	84.76065	150.5069
Oct-23	118.712	84.32334	153.1006
Nov-23	120.87	85.02991	156.7101
Dec-23	116.5466	79.31157	153.7817
Jan-24	116.7894	78.20983	155.369
Feb-24	117.9892	78.11032	157.868
Mar-24	118.7223	77.58518	159.8593
Apr-24	120.5562	78.19825	162.9141
May-24	121.9846	78.44004	165.5292

Table 3.1.6 present the results of the price forecasts that we obtained by applying our model ARIMA (0,1,0) (0,1,1) [12] for the next 24months from june2022 to may2024. In this table the lower and upper prediction limits produce a prediction interval for each forecast. The prediction interval is a range of likely values of forecasts. For example, with a 95% prediction interval, it can be 95% confident that the prediction interval contains the forecast at the next 24 months from june2022 to May 2024.

Fig 3.1.6

Accuracy of Diesel price in Tamil Nadu



In Fig 3.1.6, Inverse characteristic roots for the ARIMA (0, 1, 0) (0, 1, 1) Model fitt ed to the seasonally adjusted Diesel price the red dots in the plot correspond to the roots of the polynomials (represent the$\phi(B)$). They are all inside the unit circle, as we would expect beca use R ensures the fitted model is both stationary and invertible. Any roots close to the unit cir cle may be numerically unstable, and the corresponding model will not be good for forecastin g. In this predicted model none of the roots are closed and indexed in the circle, so the corresp onding model will be good fit.

3.2 Petrol price

Descriptive Analysis:

The basic statistical measures of the data are calculated to describe its properties.

Table 3.2.1

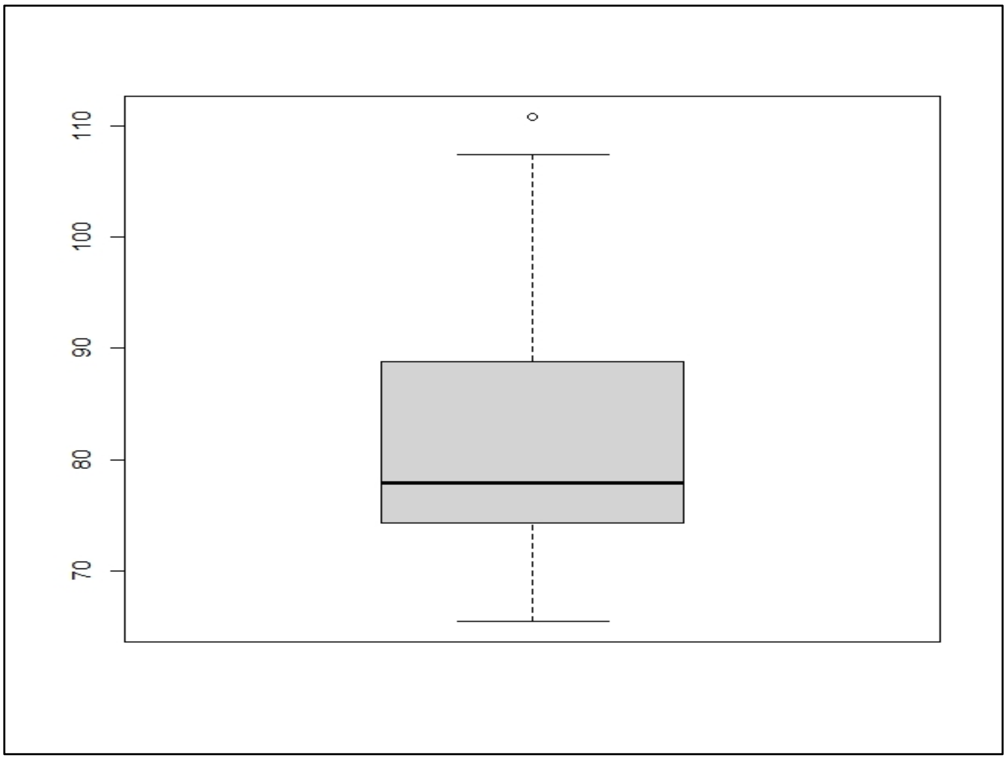
Descriptive statistics for petrol price in Tamil Nādu

Statistical Measures	Petrol
Mean	82.6148
Median	77.9100
Maximum	110.85
Mode	101.40
Minimum	65.46
1st Quartile	74.2450
3rd Quartile	90.6250
Skewness	.864
Kurtosis	-.404

From this empirical distribution, values of some of the central positional and dispersion measures have been calculated using R software version 4.0.2. The values are presented in Table 3.2.1. It deviates much from the average Petrol price is 82.6148. The maximum Petrol price is 110.85.and the minimum Petrol price is 65.46. The first and third quartiles values specified that the maximum of 25% of the Petrol price is 74.2450 and another 25% of them are 90.6250. The skewness value is 0.864 which is positive and implies that the data is positively skewed and skewed right. The co-efficient of kurtosis is -0.404, point out that the distribution is play kurtic. This kind of distribution has a tail that’s thinner than a normal distribution.

Fig 3.2.1

Box plot for Petrol price in Tamil Nadu



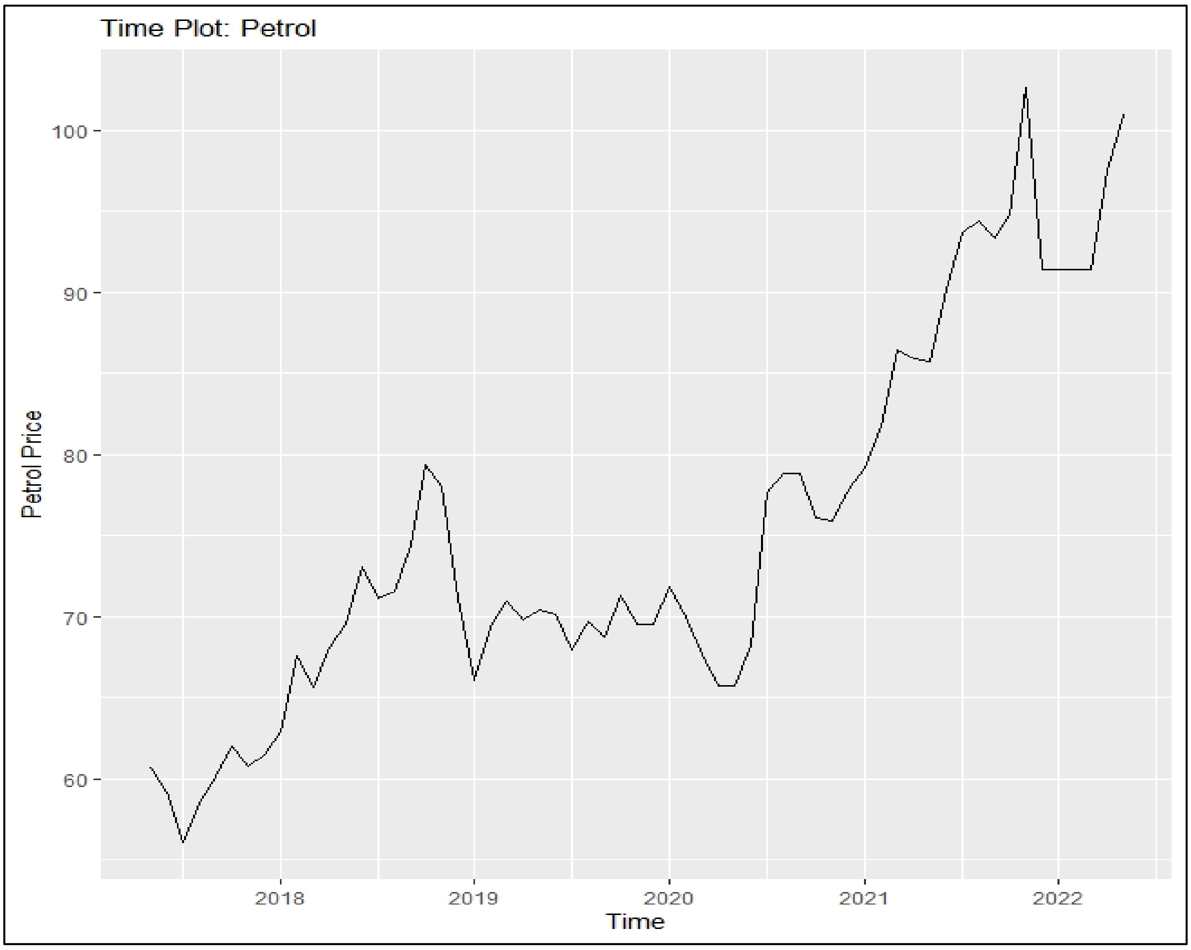
Box Whisker plot shows that the distance between third quartile and median is smaller than the distance between first quartile and median. There is variation in length of both the tails. That there is no outlier in the data. The prices of the petrol in the data width range of more than 78rs.

Testing the stationary

Time series models and the methodology used for fitting them to a given set of data can be chosen based on the stationary value of the time series. As described in the preceeding chapters, stationary time series can be analyzed using time plot. Fig 3.2 illustrates the monthly price of petrol in Tamil Nadu over time.

Fig 3.2.2

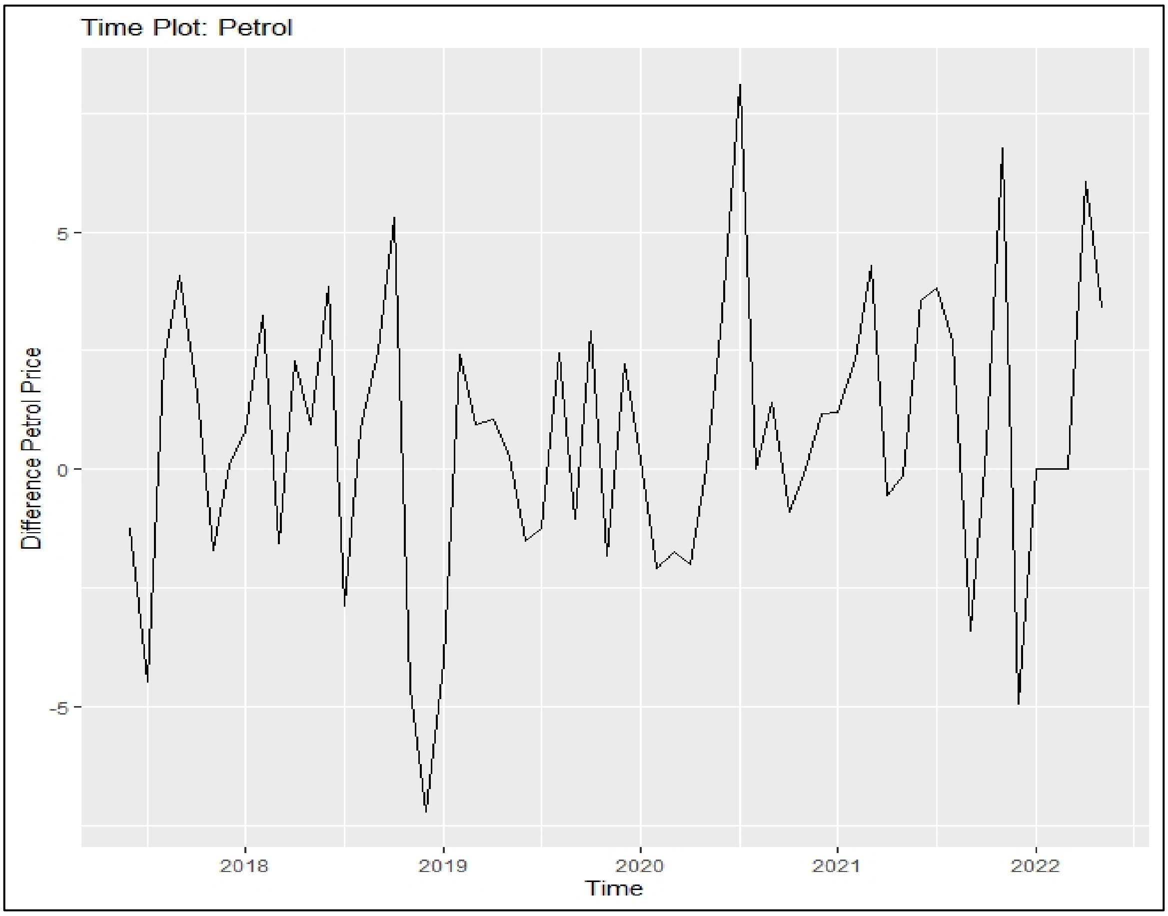
Petrol price in Tamil Nadu from 2017 to 2022



The time plot reveals the elements of the time series fluctuate moderately and there is an upward and downward movement in the fluctuations. As time passes, the pattern of variation changes. Additionally, a shift in the constant around which the observations oscillate is also observed. According to the study, the mean price of petrol in a given month is not a fixed number and it varies over time. The data indicates that the time series of monthly price of petrol in Tamil Nadu is non-stationary in mean and variance.

Fig 3.2.3

Second-Order differences petrol price in Tamil Nadu from 2017 to 2022



Differencing method is applied to the actual data for removing the non-stationarity. The second order differences $Y't = Yt - Yt-1 - Yt-2$ are calculated to the time series. The time series of the Second differenced values of the monthly petrol price is stationary. Hence, the appropriate model can be determined to this stationary time series and the model can be used to forecasts.

Before forecasting the best model will be checked, among 63 models of ARIMA the Minimum AIC value is 264.38 and this model will be the best one.

All the ARIMA model and the Akaike's information criterion Value has been listed in the Table 3.2.2

Table 3.2.2

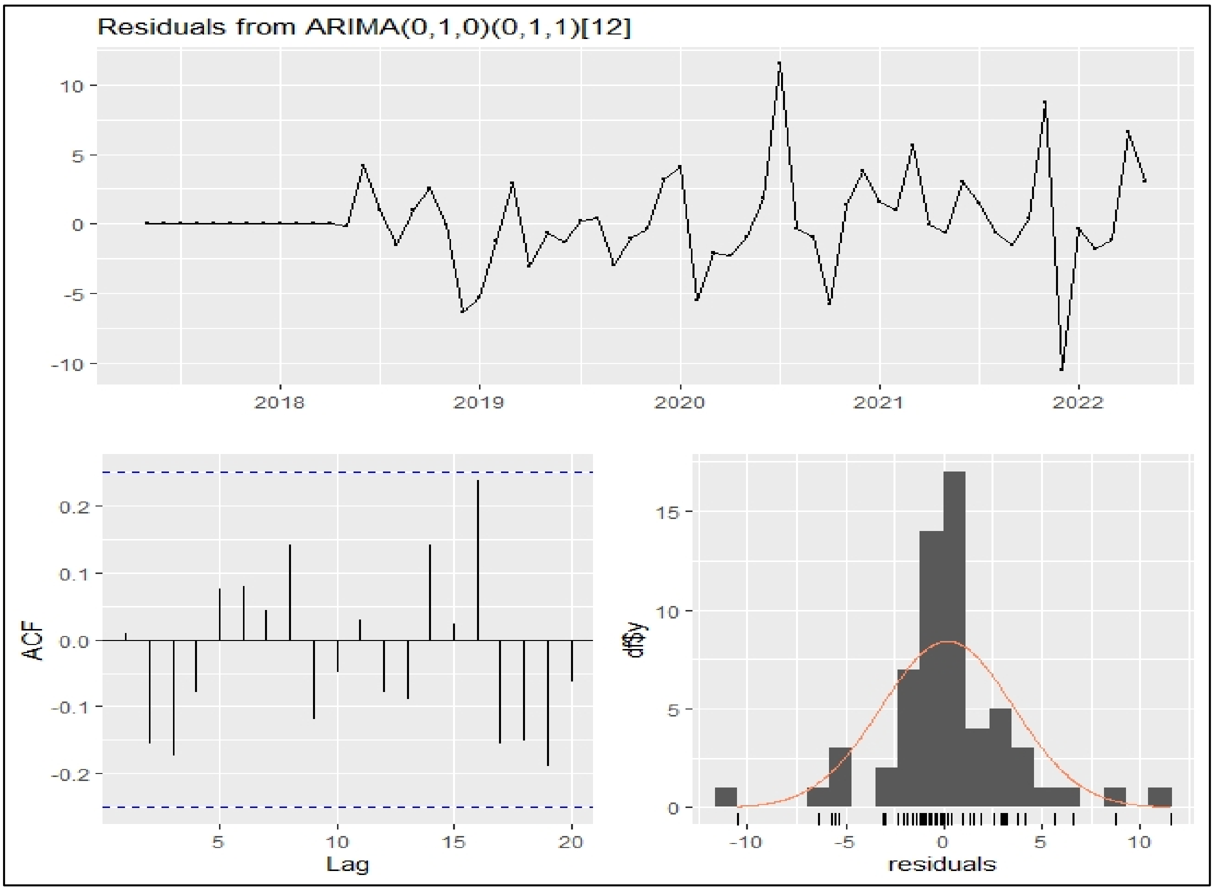
Model Comparison

Model	AIC
ARIMA(0,1,0)(0,1,0)[12]	: 270.8032
ARIMA(0,1,0)(0,1,1)[12]	: 264.6452
ARIMA(0,1,0)(1,1,0)[12]	: 265.418
ARIMA(0,1,0)(1,1,1)[12]	: 266.8849
ARIMA(0,1,1)(0,1,0)[12]	: 271.9232
ARIMA(0,1,1)(0,1,1)[12]	: 266.4537
ARIMA(0,1,1)(1,1,0)[12]	: 267.3513
ARIMA(0,1,1)(1,1,1)[12]	: 268.8316
ARIMA(0,1,2)(0,1,0)[12]	: 273.0433
ARIMA(0,1,2)(0,1,1)[12]	: 267.5677
ARIMA(0,1,2)(1,1,0)[12]	: 269.0611
ARIMA(0,1,2)(1,1,1)[12]	: 269.9438
ARIMA(0,1,3)(0,1,0)[12]	: 274.6568
ARIMA(0,1,3)(0,1,1)[12]	: 269.1135
ARIMA(0,1,3)(1,1,0)[12]	: 270.8993
ARIMA(0,1,3)(1,1,1)[12]	: Inf
ARIMA(0,1,4)(0,1,0)[12]	: 275.4384
ARIMA(0,1,4)(0,1,1)[12]	: 271.6751
ARIMA(0,1,4)(1,1,0)[12]	: 273.0125
ARIMA(0,1,5)(0,1,0)[12]	: 277.2144
ARIMA(1,1,0)(0,1,0)[12]	: 272.2093
ARIMA(1,1,0)(0,1,1)[12]	: 266.5946
ARIMA(1,1,0)(1,1,0)[12]	: 267.4265
ARIMA(1,1,0)(1,1,1)[12]	: 270.4481
ARIMA(1,1,1)(0,1,0)[12]	: 273.7686
ARIMA(1,1,1)(0,1,1)[12]	: 269.0326
ARIMA(1,1,1)(1,1,0)[12]	: 269.5893
ARIMA(1,1,1)(1,1,1)[12]	: 271.1019
ARIMA(1,1,2)(0,1,0)[12]	: 274.4751
ARIMA(1,1,2)(0,1,1)[12]	: 269.5171
ARIMA(1,1,2)(1,1,0)[12]	: 271.0912
ARIMA(1,1,2)(1,1,1)[12]	: Inf
ARIMA(1,1,3)(0,1,0)[12]	: 276.7933

ARIMA(1,1,3)(0,1,1)[12]	: 271.7249
ARIMA(1,1,3)(1,1,0)[12]	: 273.424
ARIMA(1,1,4)(0,1,0)[12]	: 277.7286
ARIMA(2,1,0)(0,1,0)[12]	: 273.0415
ARIMA(2,1,0)(0,1,1)[12]	: 267.3389
ARIMA(2,1,0)(1,1,0)[12]	: 268.993
ARIMA(2,1,0)(1,1,1)[12]	: Inf
ARIMA(2,1,1)(0,1,0)[12]	: 273.7897
ARIMA(2,1,1)(0,1,1)[12]	: 268.7373
ARIMA(2,1,1)(1,1,0)[12]	: 270.6481
ARIMA(2,1,1)(1,1,1)[12]	: Inf
ARIMA(2,1,2)(0,1,0)[12]	: Inf
ARIMA(2,1,2)(0,1,1)[12]	: Inf
ARIMA(2,1,2)(1,1,0)[12]	: Inf
ARIMA(2,1,3)(0,1,0)[12]	: Inf
ARIMA(3,1,0)(0,1,0)[12]	: 274.0483
ARIMA(3,1,0)(0,1,1)[12]	: 268.0986
ARIMA(3,1,0)(1,1,0)[12]	: 270.2726
ARIMA(3,1,0)(1,1,1)[12]	: Inf
ARIMA(3,1,1)(0,1,0)[12]	: 275.7951
ARIMA(3,1,1)(0,1,1)[12]	: 270.6783
ARIMA(3,1,1)(1,1,0)[12]	: 272.7071
ARIMA(3,1,2)(0,1,0)[12]	: Inf
ARIMA(4,1,0)(0,1,0)[12]	: 274.8545
ARIMA(4,1,0)(0,1,1)[12]	: 270.6226
ARIMA(4,1,0)(1,1,0)[12]	: 272.3648
ARIMA(4,1,1)(0,1,0)[12]	: 277.1725
ARIMA(5,1,0)(0,1,0)[12]	: 277.1189
Best model: ARIMA(0,1,0)(0,1,1)[12]	

Fig 3.2.4

Residual and ACF plot for ARIMA model



The figure 3.2.4 shows that the ARIMA method produces forecasts that appear to account for all available information. The mean of the residuals is close to zero and there is no significant correlation in the residuals series. The time plot of the residuals shows that the variation of the residuals stays much the same across the historical data, apart from the one outliers, and therefore the residual variance can be treated as constant. This can also be seen on the histogram of the residuals. The histogram suggests that the residuals may be normal — the Right tail seems a little too long, even when we ignore the outlier. Consequently, forecasts from this method will probably be quite good, but prediction intervals that are computed assuming a normal distribution may be accurate.

The lower AIC and BIC values are the higher Accuracy of model and the Residuals plot also given in the Table 3.2.3

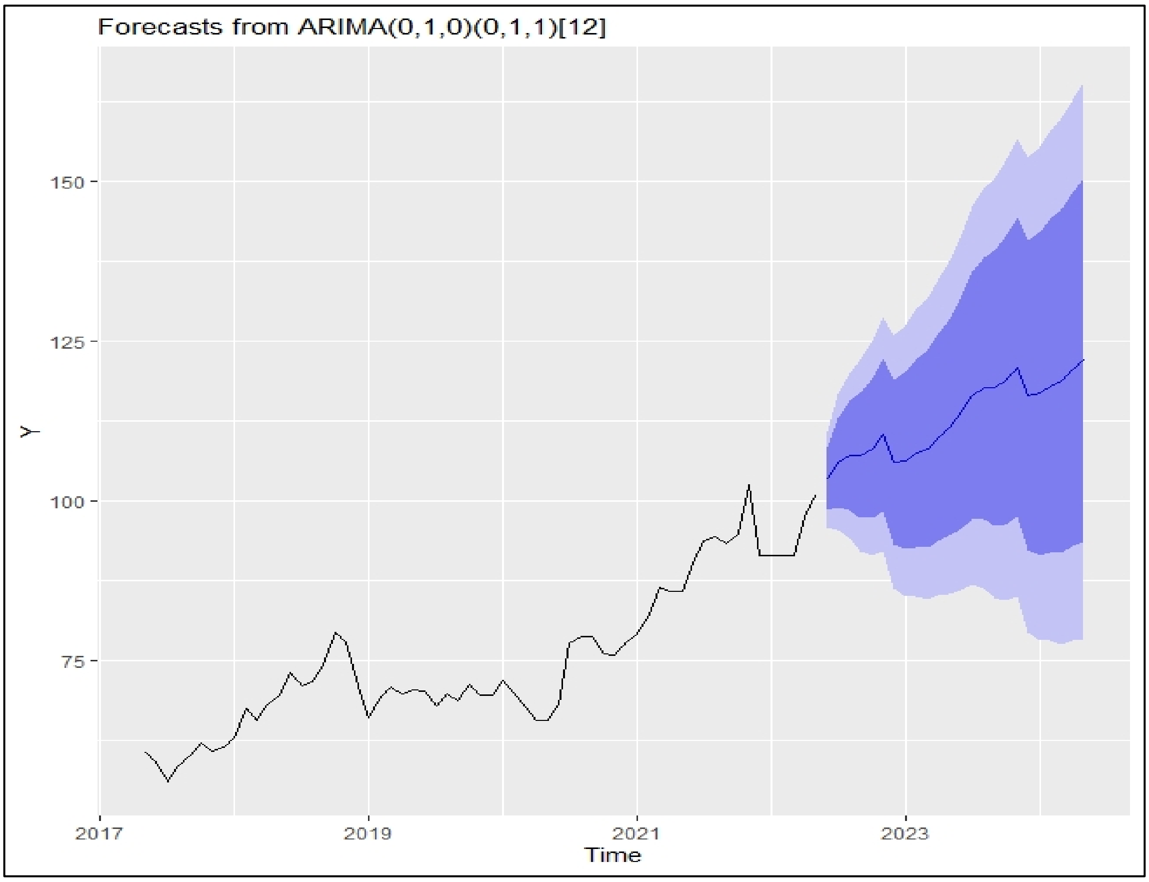
Table 3.2.3

AIC and BIC Values

AIC	BIC
264.65	268.12

Fig 3.2.5

Petrol price Forecasting using ARIMA (0,1,0)(0,1,1)[12]model



The Fig 3.2.5 indicate that forecasts are shown as a blue line, with the 80% prediction intervals as a dark shaded area, and the 95% prediction intervals as a light shaded area. This is the overall process by which we can analyze time series data and forecast values from existing series using ARIMA.

Lastly the Price is forecasted for the upcoming 24 monthly, the forecasted petrol price value as given in Figure 3.2.5. The error factors such as ME, RMSE, MAE, MPE, MAPE and

MASE are used to determine the accuracy of the model. The Mean Error (ME) is0.2913861, Root Mean Square Error (RMSE) is 3.118618, Mean Absolute Error (MAE) is 2.178288, Mean Percentage Error (MPE) is 0.2438244, Mean Absolute Percentage Error (MAPE) is 2.571594, Mean Absolute Scaled Error (MASE) is 0.2270208 and the p-value is 0.072031 which indicates the accuracy is good for prediction.

Table 3.2.4

Training set error measures

	ME	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**P value**
Training Set	0.2913861	3.118618	2.178288	0.2438244	2.571594	0.2270208	0.072031

Table 3.2.5

Forecasting value for Petrol price in Tamil Nadu

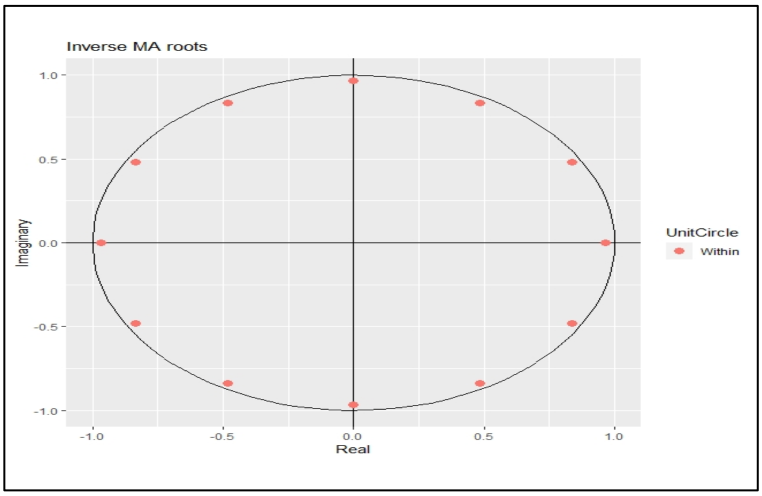
Point Forecast	Lo 95	Hi 95
113.4696	106.5035	120.4357
116.836	106.9844	126.6876
118.6654	106.5998	130.7311
117.5544	103.6222	131.4867
118.4085	102.8318	133.9852
121.1487	104.0852	138.2121
118.7516	100.321	137.1823
118.8322	99.12901	138.5354
119.4699	98.57154	140.3683
120.3172	98.28838	142.346
123.1292	100.0251	146.2332
124.8549	100.7236	148.9863
127.4745	101.1829	153.7662
130.8409	102.5534	159.1285
132.6704	102.5189	162.8219

131.5593	99.65254	163.4661
132.4134	98.84296	165.9838
135.1536	99.99818	170.3091
132.7566	96.08456	169.4285
132.8371	94.70883	170.9654
133.4748	93.94389	173.0058
134.3221	93.4366	175.2077
137.1341	94.93744	179.3307
138.8598	95.39162	182.3281

Table 3.2.5 present the results of the sales forecasts that we obtained by applying our model ARIMA **(0,1,0) (0,1,1) [12]** for the next 24 months from june2022 to May 2024.In this table the lower and upper prediction limits produce a prediction interval for each forecast. The prediction interval is a range of likely values of forecasts. For example, with a 95% prediction interval, it can be 95% confident that the prediction interval contains the forecast at the next 24months from june2022 to May 2024.

Fig 3.2.6

Accuracy of Petrol price in Tamil Nadu



In Fig 3.2.6, Inverse characteristic roots for the ARIMA **(0,1,0)(0,1,1)[12]**model fitted to the seasonally adjusted Petrol price . The red dots in the plot correspond to the roots of the polynomials (represent the $\phi(B)$). They are all inside the unit circle, as we would expect because R ensures the fitted model is both stationary and invertible. Any roots close to the unit circle may be numerically unstable, and the corresponding model will not be good for forecasting. In this predicted model none of the roots are closed and indexed in the circle, so the corresponding model will be good fit.

CHAPTER IV

CONCLUSION

We looked at the effectiveness of One method and different ARIMA Models for making time series forecasts. In the early stages of time series modelling the selection of models was very subjective. Since then, many techniques and methods have been suggested to add mathematical rigor to the search process of an ARIMA model, including Akaike's information criterion (AIC), and the Bayes information criterion (BIC). Often these criteria come down to minimizing (in sample) one step-ahead forecast errors, with a penalty term for over fitting. It should be noted that these model comparison techniques are useful for selecting the best model of similar structure. For instance, if the model is two datasets to choose from, AIC or BIC can be used to select from those models. It is for this reason that measures of forecast accuracy like MAE, MAPE, and MASE are used to compare models of different structures.

For each model and 24 Year production forecast, six error statistics were calculated: Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error, Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE). The results are shown in Table 3.1.4 and Table 3.2.4

The results from Table 3.1.5 and Table 3.2.5 suggest that ARIMA acts as a better predictor for the Diesel and Petrol. This forecasting of Diesel and petrol will be helpful for the citizens of Tamil Nadu.

REFERENCES:

1. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung, Time Series Analysis: Forecasting and Control.
2. Gebhard Kirchgässner, Jürgen Wolters, Introduction to Modern Time Series Analysis.
3. Jonathan D. Cryer, Kung-Sik Chan, Time Series Analysis: With Applications in R.
4. Andrew V. Metcalfe, Paul S.P. Cowpertwait (auth.), Introductory Time Series with R.
5. Fildes, R., Hibon, M., Makridakis, S., & Meade, N. (1998). Generalising about univariate forecasting methods: further empirical evidence. International journal of Forecasting, 14(3), 339-358.