POWER DEMAND FORECASTING

PROJECT REPORT SUBMITTED BY

SRI HARI PAVAN S
EE11B119

Submitted in partial fulfillment of the requirements for the degree of

Dual Degree - B.Tech in Electrical Engineering and M.Tech with specialization in Power Systems and Power Electronics

Under the guidance of

Dr.M.Thenmozhi

&

Dr.K.Shanti Swarup

Department of Electrical Engineering

INDIAN INSTITUTE OF TECHNOLOGY, MADRAS

CHENNAI

2011-2016

THESIS CERTIFICATE

This is to certify that the thesis titled '**Power Demand Forecasting'** submitted by **Sri Hari Pavan Suryadevara**, to the Indian Institute of Technology Madras, Chennai for the award of **Dual Degree in Electrical Engineering**, is a bonafide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Thenmozhi. M
Research Guide
Professor
Dept. of Management Studies
Place: Chennai
IIT-Madras, 600 036

Prof. Shanti Swarup K.
Co- Guide
Professor
Dept. of Electrical Engineering
Place: Chennai
IIT-Madras, 600 036

Date: 12th May 2016

ACKNOWLEDGEMENTS

I wish to express my deep sense of gratitude and indebtedness to my research guide **Dr. Thenmozhi M**, Professor in Department of Management Studies and my project guide **Dr. Shanti Swarup K**, Professor in Department of Electrical Engineering, for their most valuable guidance, discussions, suggestions and encouragement, from the conception to the completion of this project. I am extremely grateful to them for agreeing to take me on as their students.

I would also like to thank my professors and friends from my undergraduate studies while at Indian Institute of Technology, Madras. The preparation and experience I gained were invaluable.

I am also indebted to my parents who supported me in this journey.

SRI HARI PAVAN,

EE11B119.

ABSTRACT

Power System planning starts with Electric load (demand) forecasting. Accurate electricity load forecasting is one of the most important challenges in managing supply and demand of the electricity, since the electricity demand is volatile in nature; it cannot be stored and has to be consumed instantly. The optimal use of national energy resources requires long term energy forecasting. The consumption of electricity is the most sophisticated form of energy use. It is also one of the major energy consumption sectors. A methodology of forecasting long term electrical energy demand is discussed in detail in this thesis. The methodology is based on the forecasting of long term hourly demand. The demand is predicted using the past demand and price values. The demand data is modelled using multivariate **VAR** and **VECM** models to forecast the demand for a new instance.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The basis for Power system development is the forecast of future demand. The consumption of electricity in India has increased by around 10% every year in the first decade of the 21st century. It is the primary prerequisite for achieving the goal of optimal planning and operation of power systems. Electrical Energy is a primary source for industrial, social and economic development of all societies. The growth in energy generation and consumption and the gap between the two is often inextricably linked with the growth in economy and industry.

All India (Anticipated) Power Supply Position in FY2015-16[43]

Region	Energy			Peak Power		
	Requirement (MU)	Availability (MU)	Surplus(+)/Deficit(-)	Demand (MW)	Supply (MW)	Surplus(+)/Deficit(-)
Northern	355,794	354,540	-0.4%	54,329	54,137	-0.4%
Western	353,068	364,826	+3.3%	48,479	50,254	+3.7%
Southern	313,248	277,979	-11.3%	43,630	35,011	-19.8%
Eastern	124,610	127,066	+2.0%	18,507	19,358	+4.6%
North-Eastern	15,703	13,934	-11.3%	2,650	2,544	-4.0%
All India	***1,162,423***	***1,138,346***	***-2.1%***	***156,862***	***152,754***	***-2.6 %***

Figure 1.1 – India Anticipated Power Supply Position

As the figure clearly indicates, there is either a Surplus or Deficit in the Power Supply Position across all the regions. As we very well know that Power cannot be stored, the need of the hour is to gauge the power demand accurately so as to reduce the wastage, i.e. during a surplus or deficit power supply situation.

There are quite a few reasons why the demand for Electricity increases. Electricity demand increases due to the increase in population, higher per capita consumption, and rapid development of industrial, financial, agricultural & commercial sectors, higher Gross Domestic Product (GDP) growth, institutional and structural changes in the economy of India. Furthermore, future patterns of energy production and consumption will be influenced by government policies concerning the energy sector development in world energy markets. Power Demand forecasting is very important tool for the reliable, efficient and economical operation of the power system. Modeling and prediction of electricity consumption play a vital role in developed and developing countries for policy makers and related organizations which will further help them to cut down unnecessary costs on electricity that is wasted. The underestimation of the demand would lead to potential shortages in industrial, agricultural and financial sectors devastating the normal life and economy, whereas the overestimation would lead to unnecessary wastage which means wasted financial, human and economic resources. Therefore, it would be better to model electricity demand with good accuracy in order to avoid unnecessary wastage and cost.

The forecasting of electricity demand and consumption has become one of the major research fields in electrical engineering. The supply industry generally requires forecasts with lead times that range from the short term (a few minutes, hours, or days ahead) to the long term (up to 20 years ahead). Short-term forecasts, in particular, have become increasingly important since the rise of the dynamic competitive energy markets and privatisation. Many countries have recently privatised and deregulated their power systems, under the pressure of Multi National Corporations and electricity has been turned into a commodity to be sold and bought at market prices. Since the load forecasts play a crucial role in the composition of these prices, their accuracy has become vital for the supply industry to forecast consumption and demand for electricity.

CHAPTER 2

A HISTORICAL PERSPECTIVE

There are so many methods that are presently available for forecasting demand. Appropriate method is said to be chosen given the nature of data available, desired nature and level of detail of forecasts. In general more often used approach is to use more methods and compare the forecasts to arrive at a better forecast. Forecaster can use combination of different techniques that gives him the aggregate annual forecasts and also those which gives hour by hour demand prediction for electricity. This helps so much in tariff setting and designing the demand-side management programs. Here, let us discuss about the methods commonly used in literature on demand forecasting. Most can be used for both short term and long term forecasting.

2.1 Trend Approach:

This is a unique method which falls under non-causal demand forecasting that explains the variation of demand with respect to time only. In this method, we neglect the economic, demographic, policy and technological factors which have a significant role in determining the future value of demand. The function of time is realized to be the function that explains the available data and gives appropriate forecasts. The main reason for the popularity of this method lies in it's ease of use and it's simplicity. Having said that, there is a major disadvantage pertaining to the implementation of this method as it's not accurate enough since it does not take into account the various factors such as role of per capita income, prices, population rise, urbanization, policy changes etc. Since it does not provide any scope for internalizing the changes in such factors, this method is used in only in short term forecasts where those factors have minimal effect on the demand equation thereby approximating our forecast of demand to actual value.

2.2 End-Use Approach:

This method tries to capture impact of energy usage patterns of different systems and devices. The end use model for electricity demand focus on its different uses in residential, agriculture, industrial and commercial sectors of economy. For example, in residential sector electricity is used for lighting, cooking, refrigeration. In agriculture sector it is used for lift irrigation etc. This method is such that energy required is for service it delivers and not as final good.

The following equation defines end-use method for a sector:

$$\mathbf{E = S * N * P * H}$$

E = energy consumption of an appliance in kWh

S = penetration level in terms of number of such appliances per customer

N = number of customers

P = power required by the appliance in kW

H = hours of appliance use.

This, summed over various end-uses in a sector, gives aggregate energy demand. This method takes into account improvements in efficiency of energy use, utilization rates, inter-fuel substitution etc., in a sector as these are captured in the power required by an appliance, P. In the process the approach implicitly captures the price, income and other economic and policy effects as well.

2.3 Econometric Approach:

This method mixes economic theory with statistical methods to produce system of equations for forecasting demand. Taking cross sectional/pooled data or time series data, causal relations can be established between power demand and other economic variables. Here electricity demand (dependent variable) is expressed as function of different economic factors. These variables can be price of power, population, income per capita, proxies for penetration of appliances. Thus, the equation we have:

$$\mathbf{ED = f\ (Y, Pi, Pj, POP, T)}$$

where,

ED = electricity demand

Y = output or income

Pi = own price

Pj = price of related fuels

POP = population

T = technology

Different functional forms and combinations of these and other variables may have to be tried till basic assumptions of model are met and the relationship is found statistically significant
For example, the demand for energy in specific sectors could be explained as a function of the variables indicated in the right hand side of the following equations:

Residential ED = f (Y per capita, POP, Pi, Pj)

Industrial ED = g (Y of power intensive industries, GFKF or I, index of T, index of GP)

where,

GFKF = gross fixed capital formation

I = investment

GP = government policy, and

f and g represent functional forms.

Inserting the forecasts of independent variables into the equation would yield the projections of electricity demand. The coefficients and sign of each variable, estimated, would indicate the strength and direction of each of the right-hand-side variable in explaining the electricity demand in a sector.

2.4 Time Series Approaches:

A time series is said to be an ordered set of data values of certain variable. Time series models are econometric models where the only explanatory variables used are lagged values of the variable to be explained and predicted. The intuition underlying time-series processes is that the future behavior of variables is related to its past values, both actual and predicted, with some adaptation/adjustment built-in to take care of how past realizations deviated from those expected. Thus, the essential prerequisite for a time series forecasting technique is data for the last 20 to30 time periods. The difference between econometric models based on time series data and time series models lies in the explanatory variables used. It is worthwhile to highlight here that in an econometric model, the explanatory variables (such as incomes, prices, population etc.) are used as causal factors while in the case of time series models only lagged (or previous) values of the same variable are used in the prediction.

In general, the most valuable applications of time series come from developing short-term forecasts, for example monthly models of demand for three years or less. Econometric models are usually preferred for long term forecasts. Another advantage of time series models is their structural simplicity. They do not require collection of data on multiple variables. Observations on the variable under study are completely sufficient. A disadvantage of these models, however, is that they do not describe a cause-and-effect relationship. Thus, a time series does not provide insights into why changes occurred in the variable.

CHAPTER 3
METHODOLOGY

It is common to use a combination of econometric and time series models to achieve greater precision in the forecasts. This has the advantage of establishing causal relationships as in an econometric model along with the dependency relationship.

In the following the combination of time series model and econometric models are used. The econometric parameter considered is Price of the power and time series parameter is the lag of the previous power demand values.

Using this economic parameter of Price along with time series parameter of demand, the forecasted equation of demand is formulated.

The price and demand data used for formulating the forecasted demand equation is obtained from AEMO (Australian Energy Market Operator) for the region of New South Wales. The data file contains all the values of price and demand values sampled at every half an hour from 2013 to 2015.

Flow Chart:

The following diagram shows the steps involved in the process of power demand forecasting.

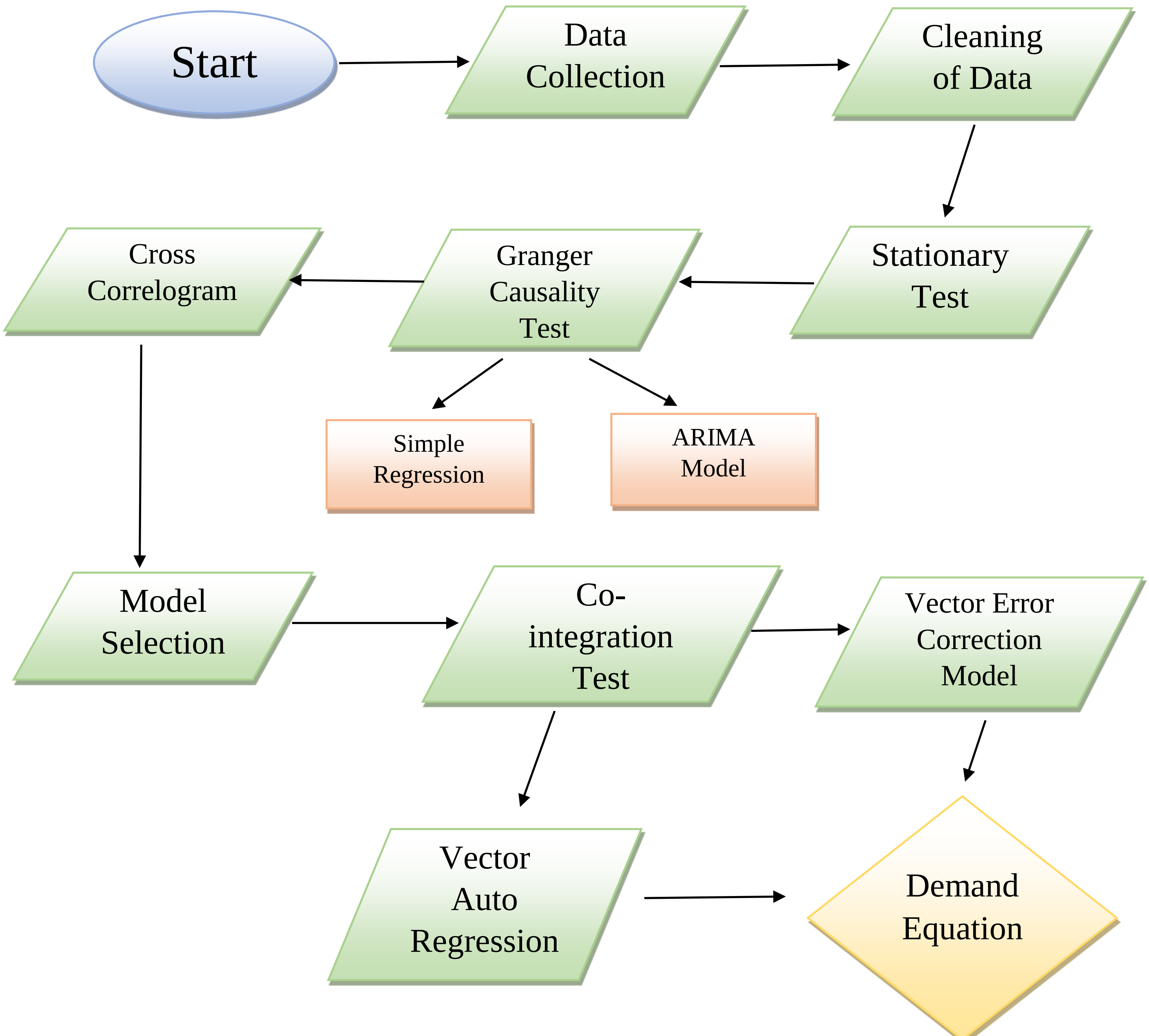


Figure 3.2 – Flow Diagram

3.1 Power Demand

3.1.1 Descriptive Statistics

The demand of power over a period of 2013 to 2015 of __________ has been collected. The given demand values are plotted as a histogram in the below figure.

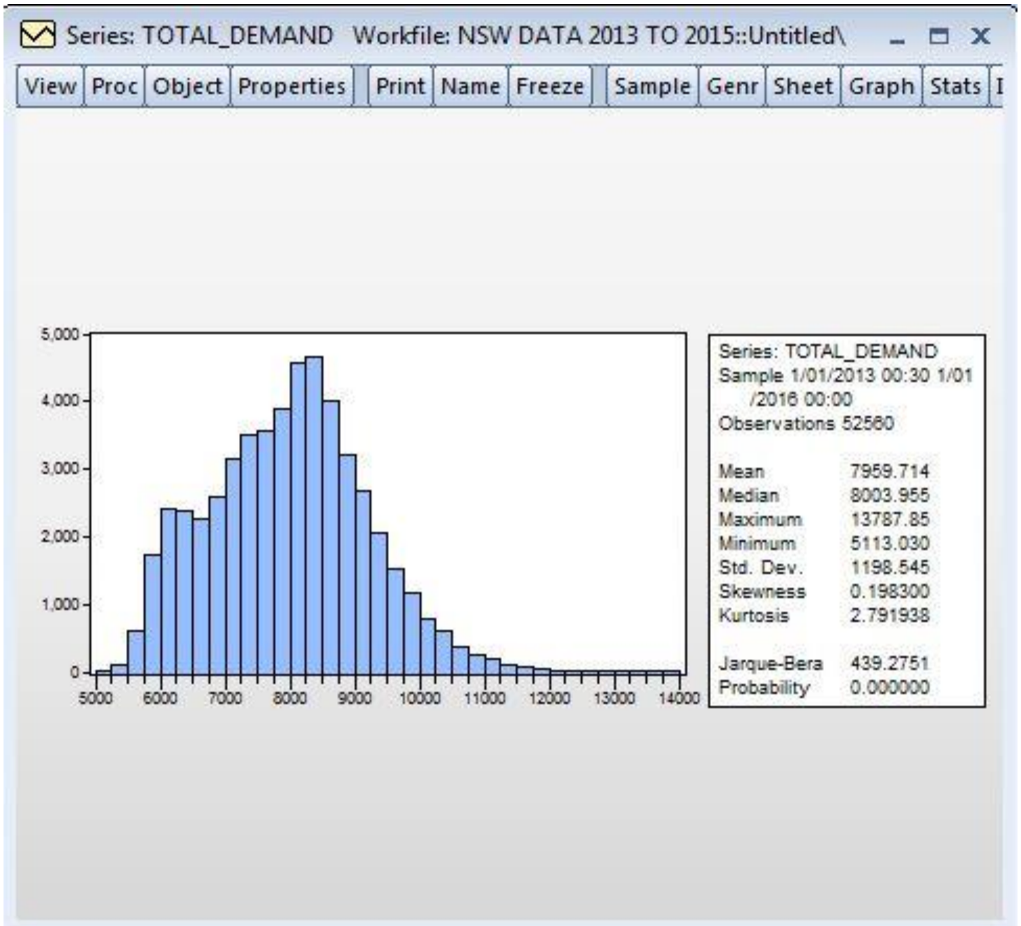


Figure 3.2 - Power Demand Histogram and Statistics (Raw Data)

3.1.2 Preprocessing

Confidence Interval = Mean ± Z statistic * Standard deviation

Assuming a confidence level, confidence interval for the demand data is calculated. Outliers are eliminated by pivoting the data between the confidence interval values.

Assume *95%* confidence level,

CI = 7959.714 ± 1.96*1198.545

CI = 7959.714 ± 2349.148

CI = {5610.5658, 10308.8622}

The demand value less than *5610.565* and greater than *10308.8622* are considered as outliers.

Assume 99% confidence level,

CI = 7959.714 ± 2.576*1198.545

CI = {4872.262, 11047.166}

The demand value less than 4872.262 and greater than 11047.166 are assumed to be outliers.

As given from the data,

Minimum value = 5113.03

Maximum value = 13787.85

3.1.3 Data Cleansing

Data Cleansing is done to remove the inconsistent values and make the data more consistent. So for *Demand value > 11047* for any period, the demand value for that period will be treated as *11047* so that the data will be consistent.

The cleansed data is plotted as a histogram in the below figure with minimum demand value as 5113.03 and maximum demand value as 11047.

The following histogram displays frequency of demand with cleaned data.

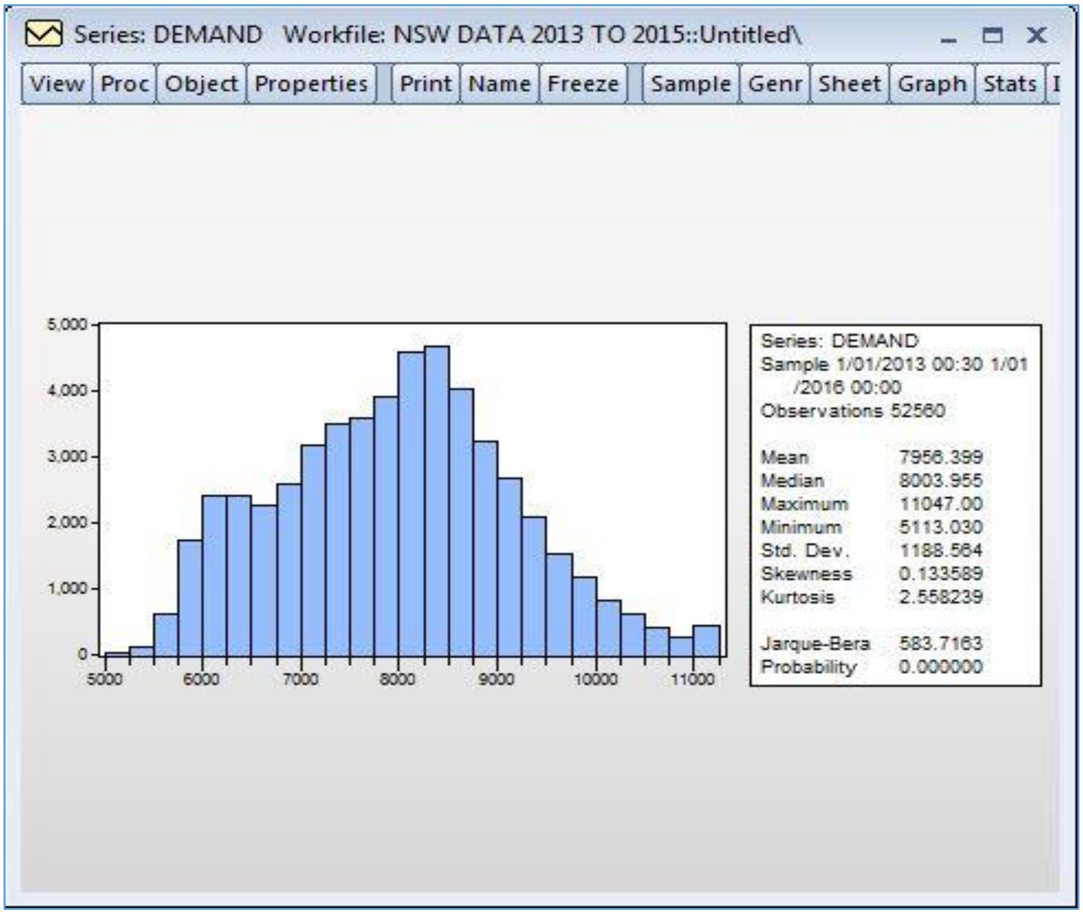


Figure 3.3 - Power Demand Histogram and Statistics (Cleansed Data)

3.1.4 Test for Stationarity

A time series is said to be *stationary*, if its statistical properties such as mean, variance and autocorrelation are all constant over time. Most statistical forecasting methods assume that the underlying time series is stationary. Hence checking for the stationarity of a time series is a necessary task before the forecasting. If the data is non stationary, it should be rendered approximately stationary using different mathematical transformations.

The **Unit Root Test** is used to determine whether a given time series is stationary or not. A non-stationary series is said to have a unit root. **Augmented Dickey-Fuller Test** is performed on the demand data. The null hypothesis for this test will be that **Demand has a unit root**. If null hypothesis is rejected, then demand has no unit root and hence the data is stationary. If null hypothesis is not rejected then demand has a unit root and the data is non stationary.

Series: DEMAND Workfile: NSW DATA 2013 TO 2015::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ide

Augmented Dickey-Fuller Unit Root Test on DEMAND

Null Hypothesis: DEMAND has a unit root
Exogenous: Constant
Lag Length: 56 (Automatic - based on SIC, maxlag=57)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-15.68664	0.0000
Test critical values:	1% level	-3.430305	
	5% level	-2.861404	
	10% level	-2.566738	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DEMAND)
Method: Least Squares
Date: 05/10/16 Time: 23:09
Sample (adjusted): 1/02/2013 05:00 1/01/2016 00:00
Included observations: 52503 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEMAND(-1)	-0.007874	0.000502	-15.68664	0.0000
D(DEMAND(-1))	0.500368	0.004359	114.7779	0.0000

Table 3.1 - Augmented Dickey-Fuller Unit Root Test on Demand

In the above Table 1, the Augmented Dickey-Fuller test statistic is more negative than the test critical values and the probability is not significant. Hence the null hypothesis of a unit root in the demand data is convincingly rejected.

Hence, **Power demand data is stationary.**

3.2 Retail Price

3.2.1 Descriptive Statistics

The given demand values are plotted as a histogram in the below figure 3.

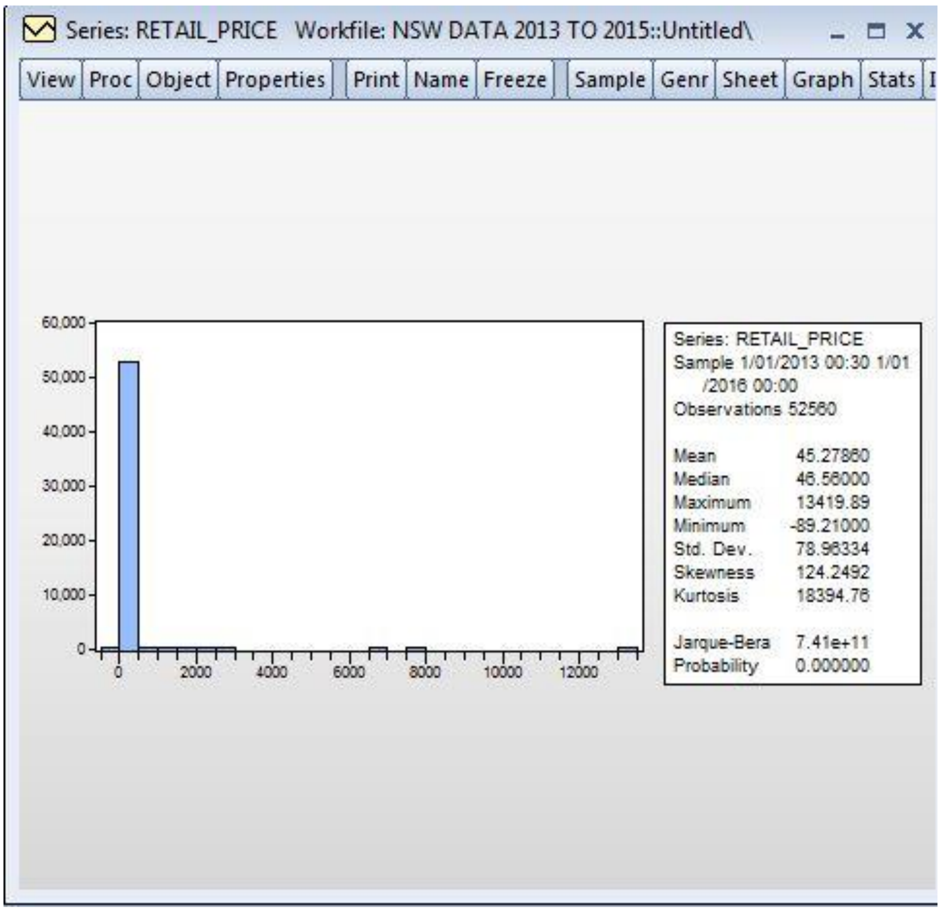


Figure 3.4 - Price Histogram and Statistics (Raw Data)

3.2.2 Preprocessing

Confidence Interval = Mean ± Z statistic * Standard deviation

Assuming a confidence level, confidence interval for the demand data is calculated. Outliers are eliminated by pivoting the data between the confidence interval values.

Assume 95% confidence level,

CI = 45.2786 ± 1.96*78.9633

CI = 45.2786 ± 154.768

CI = {0, 200.046}

The confidence interval is assumed to be greater than 0 as there is no sense for negative price values. The price values greater than 200.046 should be considered as outliers.

Assuming 99% confidence level,

CI = 45.2786 ± 2.576*78.9633

CI = {0, 248.688}

The confidence interval is assumed to greater than 0 as there is no sense for negative price values. The price values greater than *248.6* should be considered as outliers.

As given from the data

Minimum value = -89.21, changed to **2.37**(the minimum price value)

Maximum value = 13419.89

3.2.3 Price Data Cleansing

Data Cleansing is done to remove the inconsistent values and make the data more consistent. So for *Price value > 248.6* for any period, the demand value for that period will be treated as *248.6* so that the data will be consistent. There are two instances where the price value is less than zero. Since it is meaningless to have negative price values, those two values are replaced with the minimum price value, 2.37.

The cleansed data is plotted as a histogram in the below figure 4 with minimum price value as 2.37 and maximum price value as 247.9.

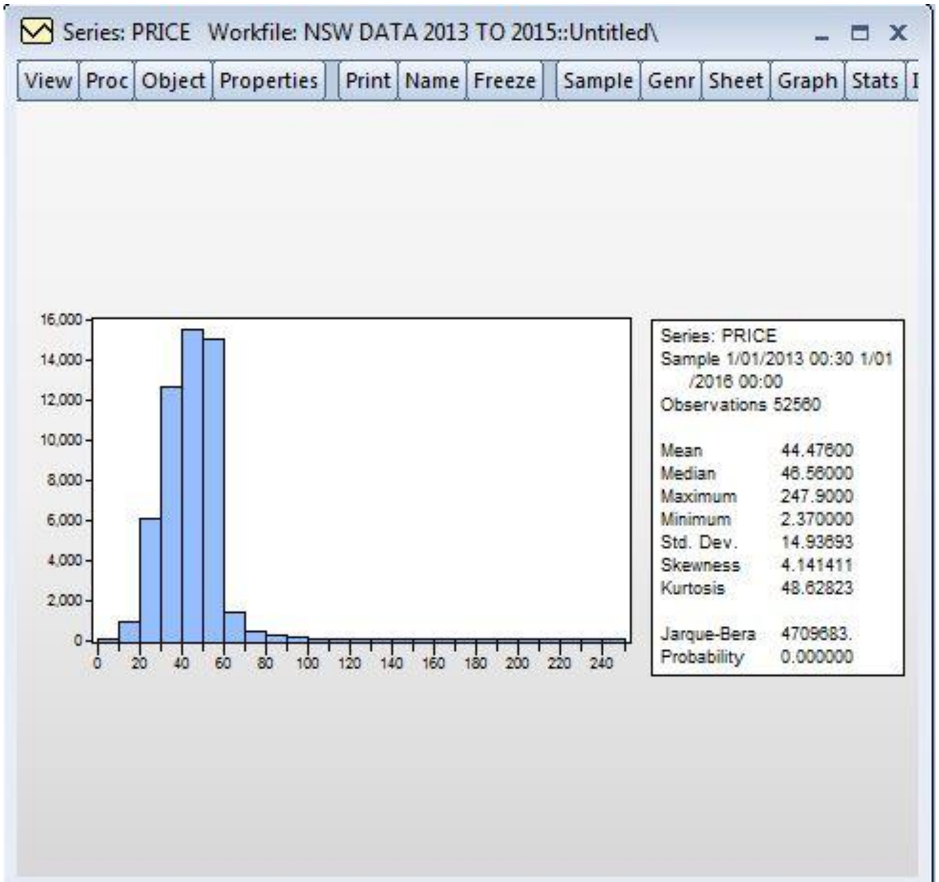


Figure 3.5 - Price Histogram and Statistics (Cleansed Data)

3.2.4 Test for Stationarity

Checking for the stationarity of a time series is a necessary task before the forecasting. The **Unit Root Test** is used to determine whether a given time series is stationary or not. A non-stationary series is said to have a unit root. **Augmented Dickey-Fuller Test** is performed on the price data. The null hypothesis for this test will be that **Price has a unit root**. If null hypothesis is rejected, then price has no unit root and hence the data is stationary. If null hypothesis is not rejected then price has a unit root and the data is non stationary.

Series: PRICE   Workfile: NSW DATA 2013 TO 2015::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | I

Augmented Dickey-Fuller Unit Root Test on PRICE

Null Hypothesis: PRICE has a unit root
Exogenous: Constant
Lag Length: 52 (Automatic - based on SIC, maxlag=57)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.060321	0.0000
Test critical values:	1% level	-3.430305	
	5% level	-2.861404	
	10% level	-2.566738	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PRICE)
Method: Least Squares
Date: 05/10/16   Time: 14:36
Sample (adjusted): 1/02/2013 03:00 1/01/2016 00:00
Included observations: 52507 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PRICE(-1)	-0.029115	0.003213	-9.060321	0.0000
D(PRICE(-1))	-0.328305	0.005319	-61.71859	0.0000

Table 3.2 - Augmented Dickey-Fuller Unit Root Test on Price

In the above figure the Augmented Dickey-Fuller test statistic is more negative than the test critical values and the probability is not significant. Hence the null hypothesis of a unit root in the demand data is convincingly rejected.

Hence, **Price data is stationary.**

3.3 Demand Correlogram

3.3.1 Autocorrelation

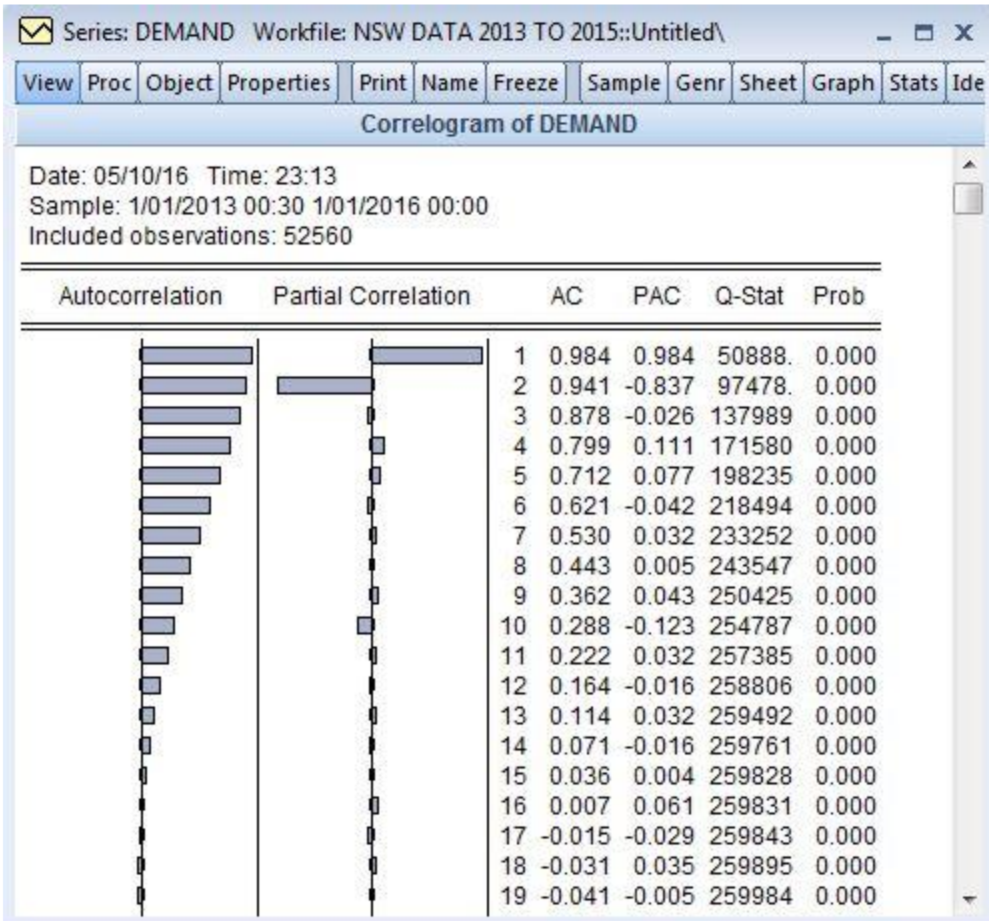
Autocorrelation is a mathematical representation of the degree of similarity between a given time series and a lagged version of itself over successive time intervals. It is the same as calculating the correlation between two different time series, except that the same time series is used twice - once in its original form and once lagged one or more time periods.

3.3.2 Partial Autocorrelation

In time series analysis, the partial autocorrelation function (PACF) gives the partial correlation of a time series with its own lagged values, controlling for the values of the time series at all shorter lags. It contrasts with the autocorrelation function, which does not control for other lags.

In the analysis of data, a **correlogram** is an image of correlation statistics. It is the plot of the sample autocorrelations and partial correlations against time lags. The correlogram is a commonly used tool for checking randomness in a data set. This randomness is ascertained by computing autocorrelations for data values at varying time lags. If random, such autocorrelations should be near zero for any and all time-lag separations.

The correlogram for the underlying demand data is shown in the below figure 5.



Correlogram of DEMAND

Date: 05/10/16 Time: 23:13
Sample: 1/01/2013 00:30 1/01/2016 00:00
Included observations: 52560

	AC	PAC	Q-Stat	Prob
1	0.984	0.984	50888.	0.000
2	0.941	-0.837	97478.	0.000
3	0.878	-0.026	137989	0.000
4	0.799	0.111	171580	0.000
5	0.712	0.077	198235	0.000
6	0.621	-0.042	218494	0.000
7	0.530	0.032	233252	0.000
8	0.443	0.005	243547	0.000
9	0.362	0.043	250425	0.000
10	0.288	-0.123	254787	0.000
11	0.222	0.032	257385	0.000
12	0.164	-0.016	258806	0.000
13	0.114	0.032	259492	0.000
14	0.071	-0.016	259761	0.000
15	0.036	0.004	259828	0.000
16	0.007	0.061	259831	0.000
17	-0.015	-0.029	259843	0.000
18	-0.031	0.035	259895	0.000
19	-0.041	-0.005	259984	0.000

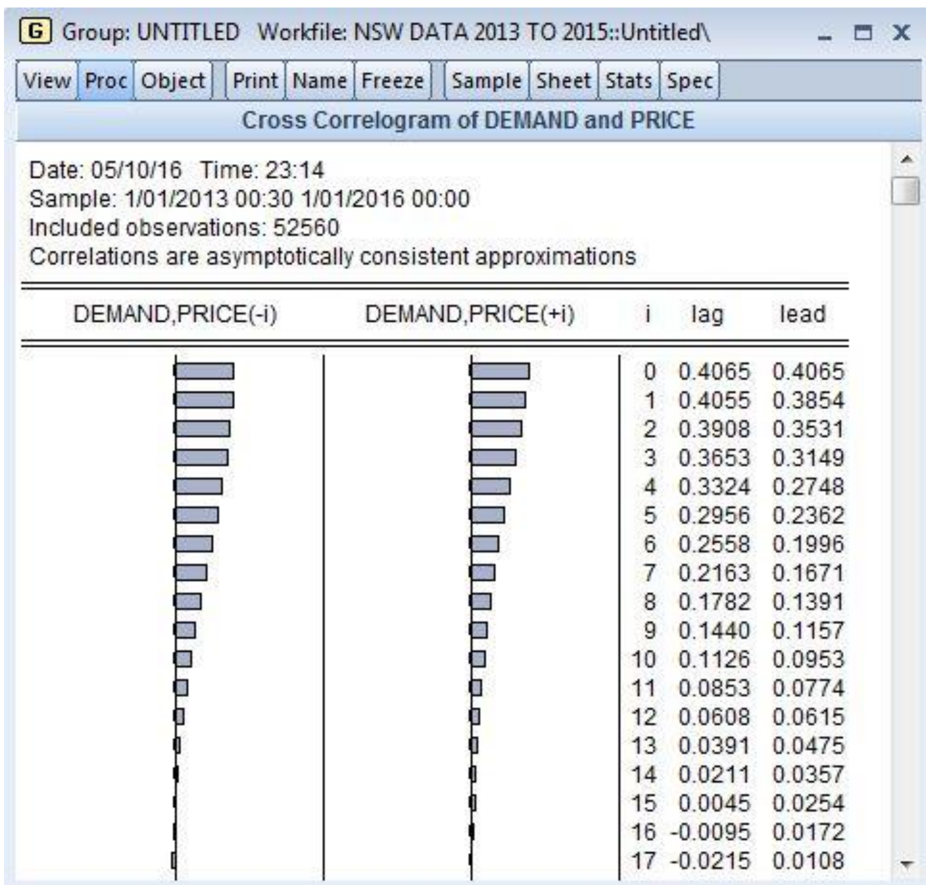
Figure 3.6 - Demand Correlogram

3.4 Demand and Price Cross Correlogram

3.4.1 Cross Correlation

Cross Correlation is the measure of the degree of the similarity between a time series and the lagged version of another time series. **Cross correlogram** gives the values of cross correlation between two different time series. It shows the number of lags of one series that depends on the other series with significant coefficient values.

The cross correlogram for the underlying demand and price data is shown in the below figure.



G Group: UNTITLED Workfile: NSW DATA 2013 TO 2015::Untitled\

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

Cross Correlogram of DEMAND and PRICE

Date: 05/10/16 Time: 23:14
Sample: 1/01/2013 00:30 1/01/2016 00:00
Included observations: 52560
Correlations are asymptotically consistent approximations

DEMAND,PRICE(-i)	DEMAND,PRICE(+i)	i	lag	lead
		0	0.4065	0.4065
		1	0.4055	0.3854
		2	0.3908	0.3531
		3	0.3653	0.3149
		4	0.3324	0.2748
		5	0.2956	0.2362
		6	0.2558	0.1996
		7	0.2163	0.1671
		8	0.1782	0.1391
		9	0.1440	0.1157
		10	0.1126	0.0953
		11	0.0853	0.0774
		12	0.0608	0.0615
		13	0.0391	0.0475
		14	0.0211	0.0357
		15	0.0045	0.0254
		16	-0.0095	0.0172
		17	-0.0215	0.0108

Figure 3.7 - Cross Correlogram of Demand and Price

Correlation between Demand and Price at Zero lag is ***0.4065.*** From the above figure, we can infer that the current demand value is determined by *14* lags of price value.

3.5 Granger Causality Test

The *Granger Causality Test* is a statistical hypothesis test for determining whether one time series is useful in forecasting another. A time series *X* is said to G*ranger-cause Y* if it can be shown that the lagged values of *X* provide statistically significant information about future values of *Y*. The following are considered as the null hypotheses for the Granger Causality Test:

- Price does not *Granger-cause* Demand.
- Demand does not *Granger-cause* Price.

The Table 3 shows the result of the Granger causality test on the price and demand data.

Group: UNTITLED Workfile: NSW DATA 2013 TO 2015::Untitled\

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

Pairwise Granger Causality Tests
Date: 05/10/16 Time: 23:40
Sample: 1/01/2013 00:30 1/01/2016 00:00
Lags: 14

Null Hypothesis:	Obs	F-Statistic	Prob.
PRICE does not Granger Cause DEMAND	52546	58.1022	3E-163
DEMAND does not Granger Cause PRICE		266.034	0.0000

Table 3.3 - Pairwise Granger Causality Test

From the above table, we can observe that the probability of the hypotheses is significantly low and the null hypotheses are rejected. Hence, we can infer that the price and demand are *bicausal*, i.e, each should be used to predict the future value of the other. Since the demand values that are to be predicted depends on the *past demand values* and the *price values*, a simple linear model will not be able to predict the future demand values. Hence we use the multivariate models like **VAR** and **VECM** to model the demand data in the next section.

3.6 Seasonality

Seasonality is a characteristic of a time series in which the data experiences regular and predictable changes which recur every calendar year. Any predictable change or pattern in a time series that recurs or repeats over a one-year period can be said to be **seasonal.**

Seasonality may be caused by various factors, such as weather, vacation, and holidays and usually consists of periodic, repetitive, and generally regular and predictable patterns in the levels of a time series. Seasonality can repeat on a weekly, monthly or quarterly basis, these periods of time are structured and occur in a length of time less than a year. Seasonal fluctuations in a time series can be contrasted with cyclical patterns. The latter occur in a period of time that extends beyond a single year, these fluctuations are usually of at least two year and do not repeat over fixed periods of time.

Now let us look into the demand graph.

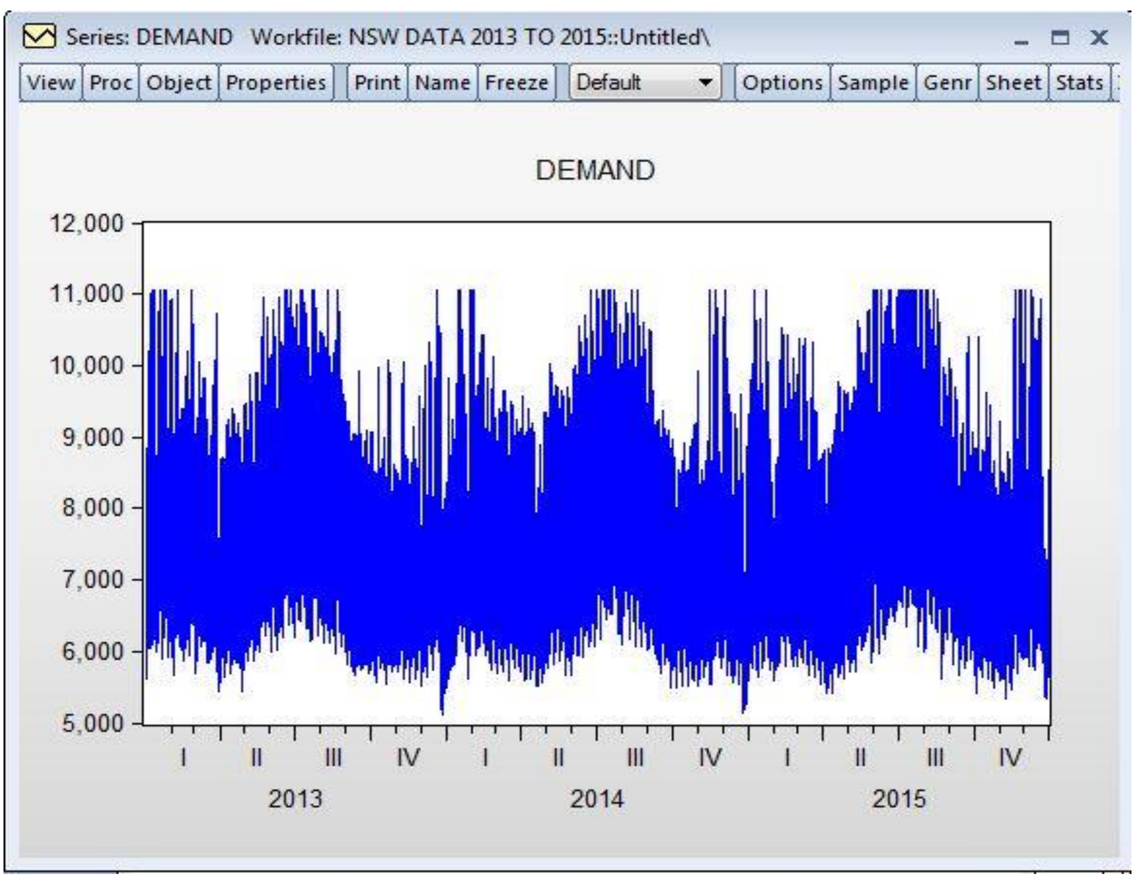


Figure 3.8 - Demand line graph

From the above figure 7, it can infer that there exists seasonality in the data. There is repetitive pattern happening in the levels of time series data.

Here the seasonality is in relation with the weather

Given in Australia the seasons are as follow:

- **Summer**: December to February
- **Autumn**: March to May
- **Winter**: June to August
- **Spring**: September to November

For consideration of the seasonality effect in modelling the demand data, dummy variables are created

The variables allotted for the months are as follow:

- Dummy1 – For the months December, January, February value is 1 remaining months 0
- Dummy2 – For the months March, April, May value is 1 remaining months 0
- Dummy3 – For the months June, July, August value is 1 remaining months 0
- Dummy4 – For the months September, October, November value is 1 remaining months0

3.7 Structural Modeling

From Granger Causality test, we inferred that there exists bi-causal relationship between demand and price. We need to use Multi-variate models.

We have two multi-variate models

1) VAR – Vector Auto Regression
2) VECM – Vector Error Correction Model

Which model to be used is decided by test for co-integration

3.8 Co-integration

Co-integration is a property of two or more variables moving together through time, and despite following their own individual trends will not drift too far apart since they are linked together in some sense.

If two series are co-integrated there exists long run relationship between them.

There are two methods for testing co-integration:

1) Johansen System Co-integration test
2) Single Equation Co-integration test

Here we use Johansen System Co-integration test

Group: UNTITLED Workfile: NSW DATA 2013 TO 2015::Untitled\

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

Johansen Cointegration Test

Date: 05/11/16 Time: 00:48
Sample (adjusted): 1/01/2013 08:00 1/01/2016 00:00
Included observations: 52545 after adjustments
Trend assumption: Linear deterministic trend
Series: DEMAND PRICE
Lags interval (in first differences): 1 to 14

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.035153	2584.965	15.49471	1.0000
At most 1 *	0.013320	704.5895	3.841466	0.0000

Trace test indicates 2 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.035153	1880.375	14.26460	1.0000
At most 1 *	0.013320	704.5895	3.841466	0.0000

Max-eigenvalue test indicates 2 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 3.4 3- Johansen Co-integration Test

From the above Table 4, it can infer that the Trace Statistic, Max-Eigen Statistic values are greater than the critical value. So there exists co-integration between the series.

There exists a long run relationship between demand and price.

So **Vector Error Correction Model (VECM)** should be applied

3.9 Vector Error Correction Model (VECM)

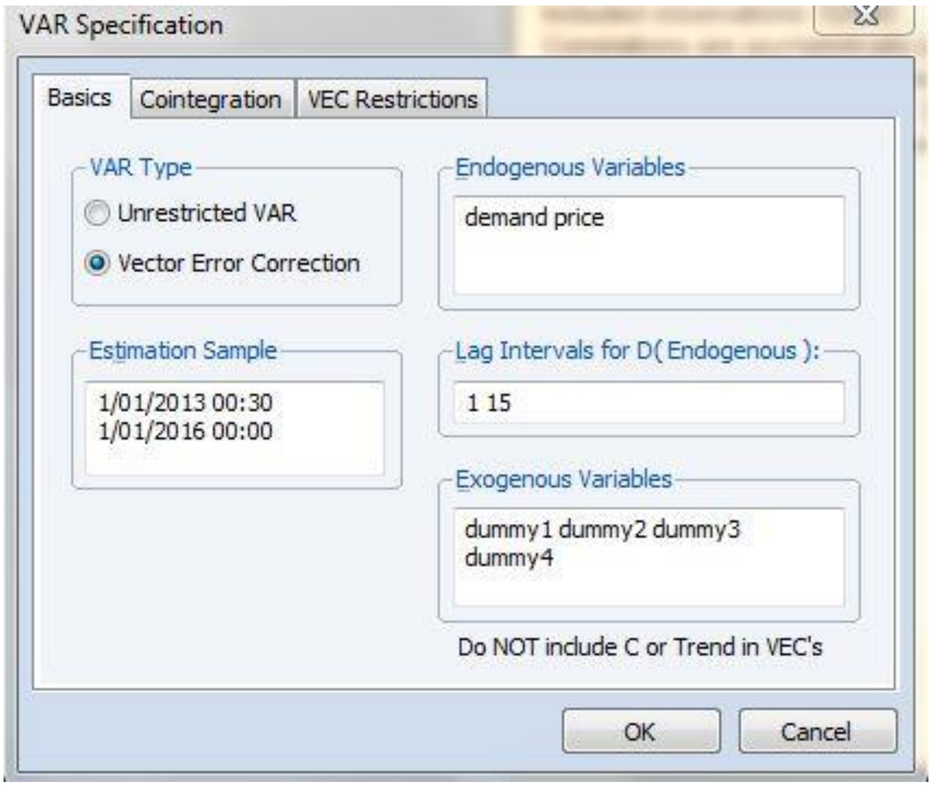


Figure 3.9 - VAR specification

Vector Error Correction Model results are as follow: (Table 5)

Var: UNTITLED Workfile: NSW DATA 2013 TO 2015::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Stats | Impulse | Resids

Vector Error Correction Estimates

Vector Error Correction Estimates
Date: 05/11/16 Time: 23:29
Sample (adjusted): 1/01/2013 08:30 1/01/2016 00:00
Included observations: 52544 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1
DEMAND(-1)	1.000000
PRICE(-1)	-3.788303 (1.43587) [-2.63833]
C	-7788.469

Error Correction:	D(DEMAND)	D(PRICE)
CointEq1	-0.029198 (0.00067) [-43.8674]	7.69E-05 (4.6E-05) [1.67133]
D(DEMAND(-1))	0.837339 (0.00435) [192.491]	0.010748 (0.00030) [35.7639]
D(DEMAND(-2))	0.115095 (0.00568) [20.2693]	0.001053 (0.00039) [2.68405]
D(DEMAND(-3))	-0.041763 (0.00570) [-7.32691]	0.000135 (0.00039) [0.34310]
D(DEMAND(-4))	-0.121745 (0.00569) [-21.3841]	-0.002335 (0.00039) [-5.93608]
D(DEMAND(-5))	0.079613 (0.00572) [13.9270]	0.001880 (0.00039) [4.76078]
D(DEMAND(-6))	-0.003644 (0.00572) [-0.63697]	0.000234 (0.00040) [0.59273]
D(DEMAND(-7))	0.010870 (0.00570) [1.90814]	6.03E-05 (0.00039) [0.15309]
D(DEMAND(-8))	-0.137674 (0.00565) [-24.3758]	-0.001927 (0.00039) [-4.93966]
D(DEMAND(-9))	0.142967 (0.00568) [25.1825]	0.002445 (0.00039) [6.23393]
D(DEMAND(-10))	-0.030459 (0.00571) [-5.33438]	0.000400 (0.00039) [1.01494]
D(DEMAND(-11))	0.030122 (0.00570) [5.28039]	0.000862 (0.00039) [2.18843]
D(DEMAND(-12))	-0.046029 (0.00567) [-8.11696]	4.58E-05 (0.00039) [0.11690]
D(DEMAND(-13))	0.030293 (0.00567) [5.34679]	0.000406 (0.00039) [1.03619]

D(DEMAND(-14))	0.040024 (0.00565) [7.08367]	0.000655 (0.00039) [1.67759]
D(DEMAND(-15))	-0.052295 (0.00444) [-11.7891]	-0.000675 (0.00031) [-2.20162]
D(PRICE(-1))	-1.718507 (0.06420) [-26.7670]	-0.370698 (0.00444) [-83.5741]
D(PRICE(-2))	-1.099030 (0.06831) [-16.0892]	-0.306318 (0.00472) [-64.9085]
D(PRICE(-3))	-0.689683 (0.07083) [-9.73691]	-0.252248 (0.00489) [-51.5469]
D(PRICE(-4))	-0.544176 (0.07240) [-7.51638]	-0.194495 (0.00500) [-38.8849]
D(PRICE(-5))	-0.210930 (0.07320) [-2.88145]	-0.139520 (0.00506) [-27.5875]
D(PRICE(-6))	-0.563058 (0.07345) [-7.66637]	-0.170416 (0.00507) [-33.5853]
D(PRICE(-7))	-0.196205 (0.07390) [-2.65487]	-0.137565 (0.00511) [-26.9430]
D(PRICE(-8))	-0.320617 (0.07408) [-4.32774]	-0.113486 (0.00512) [-22.1728]
D(PRICE(-9))	-0.144909 (0.07388) [-1.96146]	-0.108486 (0.00510) [-21.2550]
D(PRICE(-10))	-0.372643 (0.07340) [-5.07675]	-0.108713 (0.00507) [-21.4376]
D(PRICE(-11))	0.024384 (0.07317) [0.33325]	-0.089473 (0.00506) [-17.6996]
D(PRICE(-12))	-0.154024 (0.07235) [-2.12886]	-0.085431 (0.00500) [-17.0914]
D(PRICE(-13))	-0.276618 (0.07080) [-3.90689]	-0.080509 (0.00489) [-16.4587]
D(PRICE(-14))	0.156817 (0.06816) [2.30058]	-0.054268 (0.00471) [-11.5237]
D(PRICE(-15))	-0.078892 (0.06391) [-1.23438]	-0.044621 (0.00442) [-10.1056]

C	-10.92108	0.025573
	(1.01078)	(0.06983)
	[-10.8046]	[0.36621]
DUMMY1	10.88746	-0.020773
	(1.41174)	(0.09753)
	[7.71210]	[-0.21298]
DUMMY2	5.387518	-0.010548
	(1.38740)	(0.09585)
	[3.88317]	[-0.11004]
DUMMY3	27.28339	-0.067605
	(1.51518)	(0.10468)
	[18.0067]	[-0.64583]
R-squared	0.722734	0.166702
Adj. R-squared	0.722555	0.166163
Sum sq. resids	6.61E+08	3154178.
S.E. equation	112.1835	7.750438
F-statistic	4025.663	308.9554
Log likelihood	-322554.0	-182136.1
Akaike AIC	12.27882	6.934042
Schwarz SC	12.28472	6.939950
Mean dependent	0.013199	0.000133
S.D. dependent	212.9808	8.487612
Determinant resid covariance (dof adj.)		728954.4
Determinant resid covariance		727983.6
Log likelihood		-503733.8
Akaike information criterion		19.17653
Schwarz criterion		19.18868

Table 3.5 - Vector Error Correction Model Result

Considering the lags for which t statistic value is greater than 1.96 (95% confidence level).

Demand till 5 lags and Price till 10 lags are to be considered.

3.10 Power Demand Equation

The demand equation of power based on the above tests is as follows.

$$
\begin{aligned}
\boldsymbol{D_t} = \; & \mathbf{0.83739}\boldsymbol{D_{t-1}} + \mathbf{0.115095}\boldsymbol{D_{t-2}} - \mathbf{0.041763}\,\boldsymbol{D_{t-3}} - \mathbf{0.121745}\,\boldsymbol{D_{t-4}} \\
& + \mathbf{0.079613}\boldsymbol{D_{t-5}} - \mathbf{1.718507}\boldsymbol{P_{t-1}} - \mathbf{1.099030}\,\boldsymbol{P_{t-2}} - \mathbf{0.689683}\,\boldsymbol{P_{t-3}} \\
& - \mathbf{0.544176}\boldsymbol{P_{t-4}} - \mathbf{0.210930}\boldsymbol{P_{t-5}} - \mathbf{0.563058}\boldsymbol{P_{t-6}} - \mathbf{0.196205}\boldsymbol{P_{t-7}} \\
& - \mathbf{0.320617}\boldsymbol{P_{t-8}} - \mathbf{0.144909}\boldsymbol{P_{t-9}} - \mathbf{0.372643}\boldsymbol{P_{t-10}} \\
& + \mathbf{10.88746}\boldsymbol{dummy}\mathbf{1} + \mathbf{5.387518}\boldsymbol{dummy}\mathbf{2} + \mathbf{27.28339}\boldsymbol{dummy}\mathbf{3} \\
& - \mathbf{10.92108}\boldsymbol{dummy}\mathbf{4}
\end{aligned}
$$

In the similar terms, Price equation can be found.

3.11 Variance Decomposition

Variance Decompostion is used to aid in the interpretation of a multi variate model once it has been fitted. The variance decomposition indicates the amount of information each variable contributes to the other variables in the autoregression. It determines how much of the forecast error variance of each of the variables can be explained by exogenous shocks to the other variables.

Var: UNTITLED Workfile: NSW DATA 2013 TO 2015::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Stats | Impulse | Resids

Variance Decomposition

Variance Decomposition of DEMAND:

Period	S.E.	DEMAND	PRICE
1	112.2982	100.0000	0.000000
2	230.2610	99.70766	0.292341
3	362.6450	99.38447	0.615527
4	498.4020	99.13163	0.868371
5	622.7823	98.91637	1.083631
6	734.8100	98.79287	1.207127
7	831.5092	98.67764	1.322360
8	913.0764	98.60067	1.399328
9	974.9968	98.52840	1.471598
10	1022.027	98.48699	1.513006

Variance Decomposition of PRICE:

Period	S.E.	DEMAND	PRICE
1	7.706056	3.530511	96.46949
2	9.251464	7.590200	92.40980
3	10.09707	12.38669	87.61331
4	10.74552	17.36810	82.63190
5	11.26393	20.95508	79.04492
6	11.75995	23.90247	76.09753
7	12.11631	26.25575	73.74425
8	12.40618	27.94350	72.05650
9	12.62612	28.72906	71.27094
10	12.80401	29.15801	70.84199

Cholesky Ordering: DEMAND PRICE

Table 3.6 - Variance Decompostion

From the above table 3.6 we can infer that

1) The value of Demand is 98.5% depended on the previous demands and only 1.5% depended on Price values
2) The value of Price depends 29.2% on the demand values and 70.8% on the previous price values.

3.12 Impulse Response

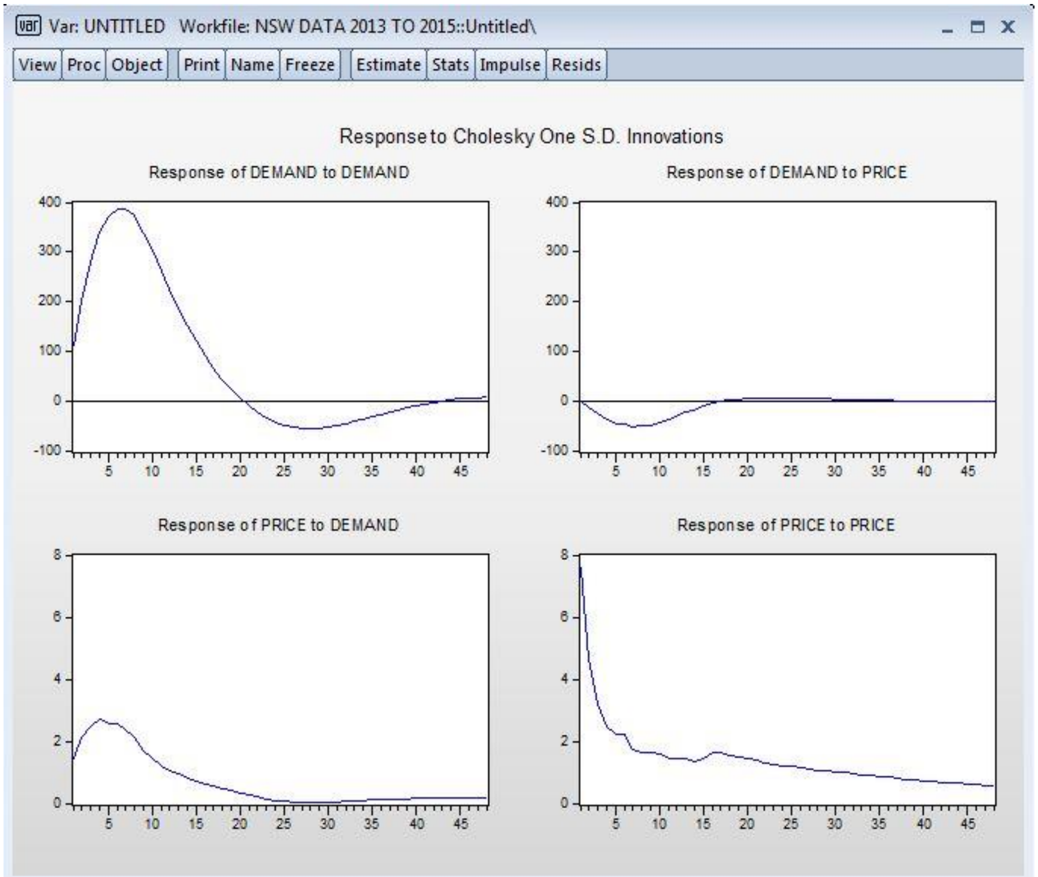


Figure 3.10 - Impulse Response

The above graph (figure 9) represents

- For a unit change in demand, the impact of demand values for the subsequent future periods
- For a unit change in price, the impact of demand values for the subsequent future periods
- For a unit change in demand, the impact of demand values for the subsequent future periods
- For a unit change in price, the impact of price values for the subsequent future periods

CHAPTER 4

TEST RESULTS

4.1 Testing of Demand Equation

Forecast for the year 2015 is done using the demand equation, keeping the values in 2013 and 2014 the same.

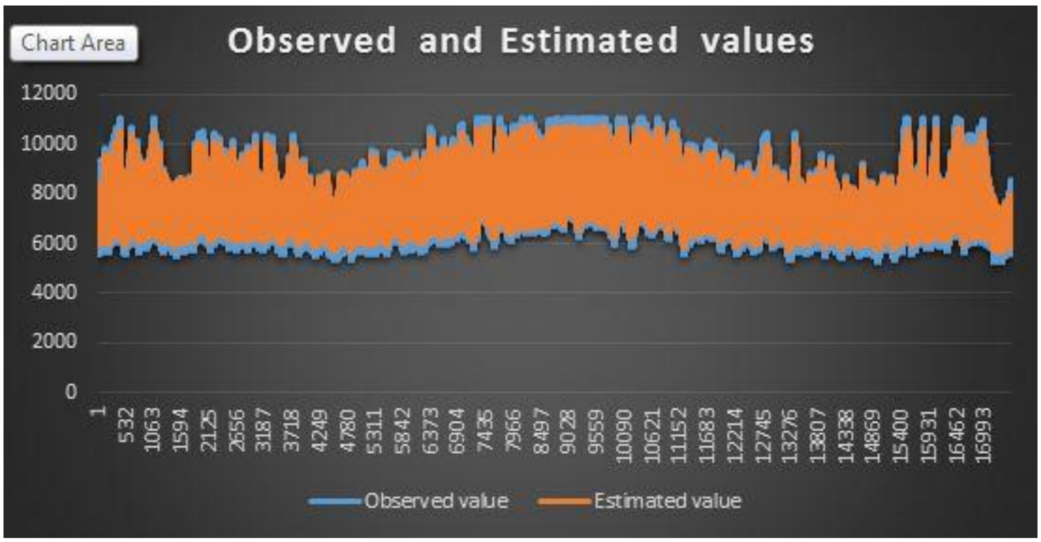


Figure 4.1 - Observed and Estimated Demand Values

Error = Forecasted value – Actual value

The Error statistics turned out to be:

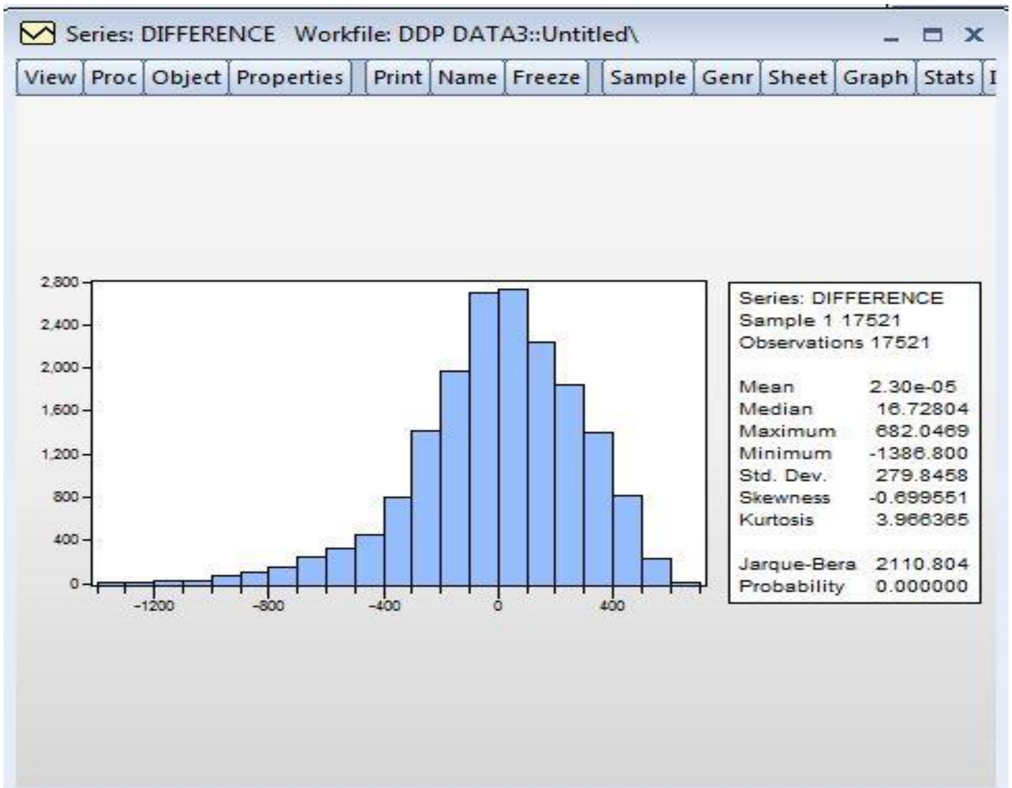


Figure 4.2 - Error Statistics – Difference

The error variation is {-1386.8, 682.07}

Error Percentage = (Forecasted reading – Actual reading)/ Actual reading

Error percentage stats are given below:

The error percentage boundary is [-14.74%, 10.71%]

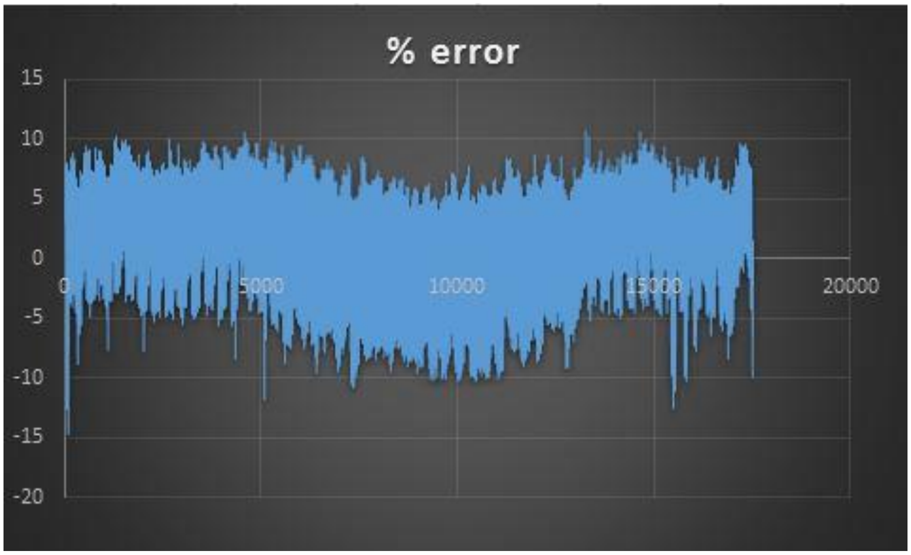


Figure 5.3 Error Percentage Boundary

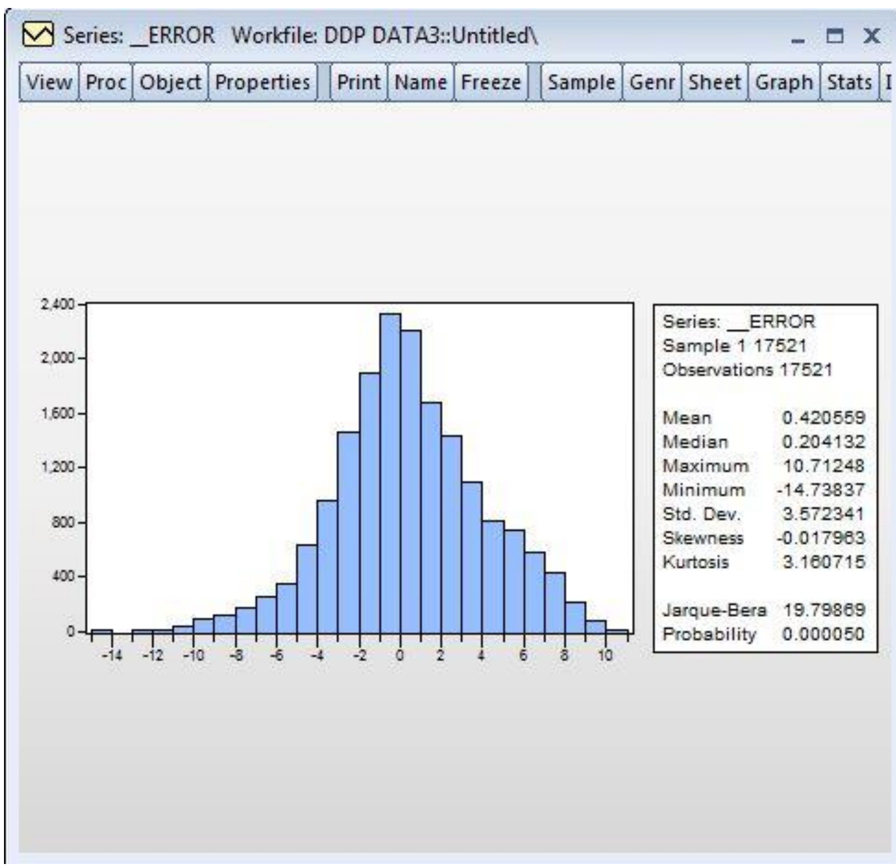


Figure 4.4 – Error Percentage Statistics

The error percentage boundary is [-14.74%, 10.71%]

CHAPTER 5

CONCLUSION

In view of the fast changing power scenarios across the world with increased industrialization and the need for more generation of power, the need for development and use of more sophisticated and relevant tools, technology and methods for estimating demands has emerged post World War 2 as technology has made significant advances. Time and again, the importance accorded to these exercises by governmental and international non-governmental organizations have remained rhetorical and on the periphery and never really got translated into the action in the true sense because they lacked conviction and dedication in coming up with models which will help us power forecast. Clearly, it is time that not just a pro-active approach is adopted but also high time, we realize the need for cutting down on the power wastage at a time when ecology and environment are being destructed to generate electricity. We need to initiate work with better methods, which would also provide the much needed impetus for data collection, analyze the data and feed effectively into the electricity reforms processes.

Specifically, drawing upon the discussions presented in the foregoing sections of the paper, the following are some recommendations which are proposed in the paper:

With respect to the annual forecasts, because of excessive reliance on simple extrapolation of past rates of growths and trends, power forecasting in India is way below the mark, compared to other developing and developed countries, both in terms of rigor, accuracy and precision. The prevailing conditions in the power sector in India are supposed to be radically altered to make our power system forecast a reliable and accurate. It is advisable to make a slow and simple beginning with the use of simple time-series method or econometric methods or we can combine both the above techniques to predict the forecast. We can also utilize more extensive end-use approaches for purpose of forecasting, to say the least. Independent regulation, pricing and related policy reforms will have an ever increasing bearing on demand for power. Thus, given the data and other constraints, it might be preferable to depend on a combination of the techniques discussed above to suit individual requirements rather than either one of them as that might lead to a compromise with precision and accuracy.

With renewed political and economic focus on demand side management and role of new technologies in aiding to forecast the same, the need for determining the shape of the load curve and predicting the impact of new technologies has gained additional importance. Thus, new innovative methods will have to be deployed to estimate the demand variations precisely and accurately across different time intervals such as hours, weeks, months and spatial geographical regions.

References:

1. Brooks Introductory econometrics for finance CHRIS BROOKS
2. Demand Forecasting for Electricity : Meeta Mehra
3. Neural Networks for Short-Term Load Forecasting : A Review and Evaluation: Henrique Steinherz Hippert, Carlos Eduardo Pedreira, and Reinaldo Castro Souza
4. Electric load forecasting: Literature survey and classification of methods - Hesham K. Alfares & Mohammad Nazeeruddin
5. Monthly Energy Forecasting Using Decomposition Method with Application of Seasonal ARIMA - P. Damrongkulkamjorn, and P. Churueang
6. Model Selection Criteria for Short-Term Microgrid-Scale Electricity Load Forecasts - Sankaranarayanan Subbayya, Jorjeta G. Jetcheva, and Wei-Peng Chen
7. D. W. Bunn, "Forecasting loads and prices in competitive power markets,"*Proc. IEEE*, vol. 88, no. 2, pp. 163–169, 2000.
8.]D. W. Bunn and E. D. Farmer, Eds., Comparative Models for Electrical Load Forecasting: John Wiley & Sons, 1985.
9. W. Charytoniuk, M. S. Chen, and P. Van Olinda, "Nonparametric regression based short-term load forecasting," *IEEE Trans. Power Systems* vol. 13, no. 3, pp. 725–730, 1998.
10. Analyzing demand behavior: a study of energy elasticities. Douglas R. Bohi (1981) John Hopkins University Press, Baltimore.
11. Demand forecasting for electric utilities. C.W. Gellings, PE (1991) The Fairmont Press, Inc.
12. Power Diary. Uttar Pradesh Electricity Regulatory Commission. Lucknow. August 2000.
13. A Base paper on demand projection. Central Electricity Authority, Ministry of Power, Government of India. New Delhi. December 1999.
14. Note on the "Base paper on demand projections". Submitted by TERI to the CEA.