MSc Research Report

Causality effect between Electricity Consumption and Gross Domestic Product in SA and the Effectiveness of the Predictive Techniques

By

Sheila Intamba

Supervisor

Dr C Chimedza

A research report submitted to the Faculty of Science, University of the Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Science

May 23, 2017

Contents

List of Tables

List of Figures

Dedication

I would like to dedicate this research report to God Almighty for allowing me to be part of His ultimate vision for myself, and to my husband, Fidel Rolse who were so patient and supportive throughout my studies. My parents and grandparents, thank you for your support and your endless faithfulness in me all of these years since childhood. For this reason, I dedicate this report to you.

Acknowledgements

I was able to successfully complete my research report due to the support of many people to whom I give my gratitude. I would like to thank my husband, for his endless patience, understanding and encouragement. Special thanks to my supervisor, Dr. Charles Chimedza for his great support, guidance and encouragement throughout my studies. Furthermore, I give thanks to the school of Statistics and Actuarial Science for offering me the opportunity to obtain a Master in Mathematical Statistics with research and coursework.

Abstract

The aim of this study was to investigate the relationship and direction between electricity consumption and gross domestic product including energy infrastructure as a third variable in South Africa using the time series data from 1993 to 2015. The relationship was modelled in South Africa focusing on the industry sectors that influence economic growth and using techniques such as ARIMA model, Multivariate Regression Analysis, Vector Autoregressive and Granger Causal Test. The Vector Autoregressive model performed better than Multivariate Regression analysis in modelling the relationship between consumption and economic growth in South Africa. The Granger causal effect illustrated a direction from consumption to economic growth and again Granger cause effect from infrastructure to economic growth.

The results from these models revealed that there was a relationship between electricity consumption and economic growth, as well as electricity infrastructure. South Africa supports a growth hypothesis meaning that South Africa is energy dependent.

The results of the study signals that the electricity consumption of South Africa have an effect on the economic growth.

Declaration

I declare that this research report is my own work. It is being submitted for the Degree of Master of Science to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.
23 May 2017

List of Acronyms

AR	Autoregressive
ADL	Autoregressive Distributed Lag
ARIMA	Autoregressive integrated moving average
CIA	Central Intelligence Agency
COMESA	Common Market for Eastern and Southern Africa
CPI	Consumer Price Index
PPI	Producer Price Index
EC	Electricity Consumption
EIU	Economic Intelligence Unit
EI	Electricity Infrastructure
ESKOM	Electricity Supply Commission
GDP	Gross Domestic Product
IEA	International Energy Agency
KWh	Kilowatt Hour
MA	Moving Average
NDP	National Development Plan
SARB	South African Reserve Bank
STATSSA	Statistics SA
US	United States
WB	World Bank

Chapter 1

Introduction

1.1 Introduction

According to the World Bank, one of the economic growth indicators for a country is electricity consumption. South Africa has been facing power shortages since 2008 and the majority power producer Eskom started projects that will help the energy status of the country. In the year 2000, South Africa agreed on goals which were initiated by the United Nations. The Millennium Development Goals will assist the country in planning and development. One of the eight goals that the country has to achieve is the improvement of infrastructure and adequate maintenance programs for these constructions projects National Development Plan 2030 (National Development Plan 2030 - The Presidency, 2009). The latter is the challenge that South Africa's energy industry is facing and the reason why the electricity consumption has been decreasing, since the generating capacity is decreasing (Economist Intelligence Unit, 2015).

This research will investigate the relationship between electricity consumption, economic growth and energy infrastructure in South Africa. The economic growth will consist of the different economic sectors (South African Reserve Bank, 2016).

According to Chouaibi and Abdessalem (2011), page 2) "*Energy is an essential part of life. It is indispensable to factories, commercial establishments, household. . . . Lack* *of energy causes not only difficulty, but also economic loss due to reduced industrial production*" . This is reason why we need to establish what the relationship and the direction of the causal effect between electricity consumption and economic growth is. The South African economy is dependent on mineral resources and manufacturing. According to the South African Reserve Bank (2016) this means we need more electricity generation in order to meet the electricity demand of companies in the mineral resources and manufacturing.

The research report is structured as follows: Section 1.2 outlines the background of South African economy and energy status. Section 1.3 describes the problem statement and Section 1.4 outlines the aims and objectives of this study. Section 1.5 outlines the limitations and assumptions and Chapter 2 reviews the literature concerning simple regression analysis, econometric modelling and the ARIMA models. Chapter 3 outlines the methodology of the regression model, econometric model, Box Jenkins modelling, the data and the evaluation of results. Finally, Chapter 4 and 5 outline the analysis and conclusion of the study.

1.2 Background

South Africa is part of Sub-Saharan Africa and has agreed to take part in the Millennium Development Goals which is an African forum initiative to improve most critical issues such as poverty, infrastructure, inequality and education. The purpose of this section is to illustrate what happened in South Africa between 2007 and 2015. It illustrate how power shortage affected the economic growth and outlines the electricity history.

The South African state owns the power institution ESKOM (Electricity Supply Commission) which has been the major provider for most of South Africa's electricity. The energy problem started in 2007 but it was at its worst from late 2007 to late 2008 and mid-2014 to 2015. Eskom's power is mostly generated from coal power stations and one nuclear power station. Africa (2012) announced that manufacturing and mining consume above 60% of the electricity produced in the country, and the addition of business takes this figure to nearly 75%. This proves that our mining sector and industrial sector make use of most of the power. Africa (2012) also stated that the mining sector was using 14.3% by 2011.

Eskom focused more on providing electricity to all citizens than improving energy infrastructure and maintaining the existing energy power stations. Studies such as (Solarin, 2011) have shown that African countries are energy dependent in order to sustain or increase the current economic growth and that is when all factors that influence economic growth are held constant. Why is economic growth so important? It measures the stability of the country's national accounts, the demographics of the citizens and businesses wellbeing.

Eskom had unstable power and power shortages from late 2007 to 2015. The most vulnerable years for South Africa in terms of electricity supply to sectors such as mining and manufacturing was 2008 and 2015. In 2008 there were many hikes in tariffs in order for Eskom to remain in business. These hikes in tariffs make electricity expensive and mining produces less output at the end of the day (Muller, 2008). Eskom also stopped exporting power to our neighbouring countries (Fin24, 2008) and power rationing was started in July (Muller, 2008). Both these last mentioned actions assisted Eskom in managing limited electricity generation. According to (News24, 2008b) South Africa's state-owned power utility was not able to supply mines and other industrial customers with more than 90% of their electricity needs until 2012. These were signals of the uncertain energy status in South Africa, and the slow pace of new developments made the 2015 energy status more difficult. When load shedding started, mining and manufacturing companies suffered as production decreased and they had to start using expensive alternative power supply systems like generators. Load shedding is a term that is used to describe rotation of electricity supply in South Africa.

1.3 Statement of the Problem

The importance of this study is quite evident from the events that occurred during the period 2007 to 2015 as it helps with planning, which will ameliorate power production planning. South Africa has been facing a number of blackouts in the past few years and this surely will affect its economic growth. Economic growth is a measure of a country's growth, that is, this is done by looking at the national accounts of the country, that is, all import, expenditures, exports and investments. The World Bank looks at other factors too among others its infrastructure, production levels (labour rate), and standard of living.

Considering the importance of electricity to GDP, we need to investigate how such factors as electricity consumption, energy infrastructure and GDP influence each other. This should help South Africa's electricity provider to strengthen policy and put more focus on energy infrastructure.

1.4 Aims and Objectives

1.4.1 Aims

The main aim of this investigation is to study the relationship between gross domestic product and independent variables which are electricity consumption and energy infrastructure.

1.4.2 Objectives

This study will answer the following questions;

1. Is there a relationship between electricity consumption, energy infrastructure and GDP?
2. Which direction does this relationship follow?
3. What is the current state of this relationship in SA?
4. How effective are the three methods for describing the relationship:
    (a) Regression Analysis,
    (b) Box Jenkins (ARIMA) models, and
    (c) Econometric models – Vector Autoregressive model

5. Which of the three methods in 4. is the most efficient that is, analise the required efficiency statistics?

1.5 Limitations and Assumptions

This section will outline the assumptions and limitations of this study.

1. This study is only limited to South Africa.

2. We assume the data is complete and the series is for the period 1993 to 2015.

3. Electricity infrastructure is an annual series divided by four because the other two series are in quarterly manner.

4. We assume the responses are normally distributed.

Chapter 2

Literature Review

2.1 Introduction

This section features the background of the problem, the history and developments in modelling the relationship between economic growth and electricity consumption. Time series analysis (econometric model and Box Jenkins model) is also reviewed. There are several studies on the relationship between economic growth and electricity, most of them analyse a specific country and focus primarily on the residential electricity consumption or both residential and industrial electricity consumption. This study focuses on the industry sector electricity consumption as I believe that industries such as mining and manufacturing are mostly influenced by electricity production.

2.2 Contributors of Economic Growth

The World Bank identified indicators that influence the economic growth of a country. The contributors are inflation, exchange rate, the living standards of the population, unemployment rate, transport, buildings and energy infrastructure, investments and savings. All of these factors influence economic growth, however, this study will focus on the energy infrastructure and consumption.

Inflation is determined by the consumer price index (CPI) and producer price index (PPI). The CPI and PPI determine the standard of life of each citizen and business in a country, respectively. The exchange rate is all about the strength of a country's currency against the global exchange rate usually the United States (US) dollar. This will influence the imports and exports of a country and in return it will impact on the economic growth. Living standards of a population have to do with the health standards and social services. Unemployment rate is also an indicator of economic growth. It indicates the state of the national accounts as the majority of the government's revenue comes from the working class taxes (Abel et al., 2008).

A country's infrastructure plays a role in their economic status. Energy infrastructure is very important for all sub-Saharan countries as it harvests our most valuable asset, mineral resources. It is better to manufacture the product inland since we can charge for the raw material, labour and the final product which we can export. Mineral resources are in the earth and to get to them we need machinery to extract and melt them. This machinery needs electricity to operate (Economist Intelligence Unit, 2015).

According to the South African Reserve Bank (2016) the sectors that contributes to economic growth are as follows:

1. Agriculture, forestry and fishing
2. Mining and quarrying
3. Manufacturing
4. Electricity and water
5. Construction (contractors)
6. Wholesale and retail trade, catering and accommodation
7. Transport, storage and communication
8. Finance and insurance, real estate and business services
9. Personal services

10. General government services

These are summarised in Figure 2.1. The major contributing sectors except the govern-

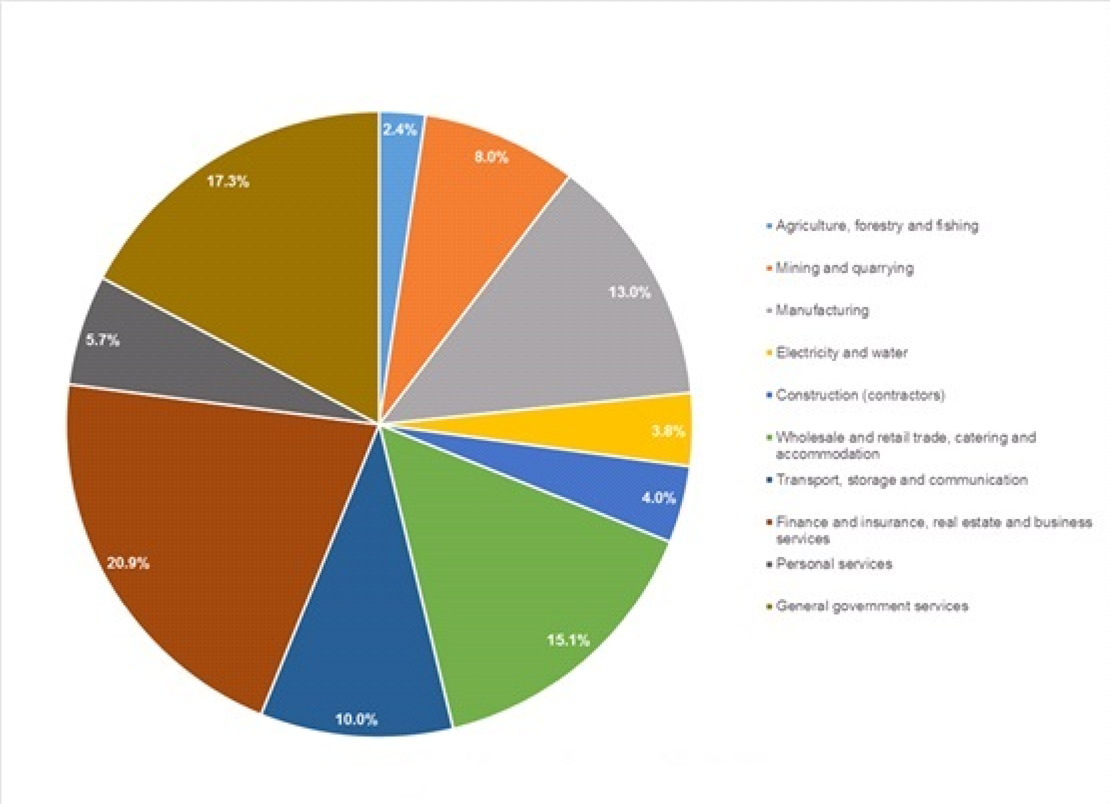


Figure 2.1: Segmentation 2015

mental services are finance and insurance, real estate and business services at 20.9%, Wholesale and retail trade, catering and accommodation at 15.1% and Manufacturing at 13%.

2.3 Events that occurred in South Africa between 2007 and 2015

In the 1990's Eskom supplied energy to over 97 percent of South Africa(SA) as a whole and only 40 percent was assigned to residents. After 1994 the state decided to provide energy to all residents in SA, they also exported a portion to neighbouring countries (Central Intelligence Agency, 2016). Eskom decided that some municipalities should stop supplying their own electricity because of a surplus of electricity (Vermeulen, 2015).

Maintenance and expansion of new projects did not occur as regularly as was expected. In late 2006 trouble started when one of the Koeberg station reactor units was shut down causing some blackouts (News24, 2006). Eskom was finally alerted in 2007 when demand of electricity peaked and electricity tariffs increased to help with the nuclear bill (Fin24, 2007). Eskom started planning and constructing new stations to provide power and by 2008. The electricity demand was outstripping electricity supply and Eskom had to implement load shedding to prevent a national blackout. Construction of Medupi and Kusile only began in 2007 and 2008, respectively when the power system was already under strain. Eskom stopped supplying electricity to neighbouring countries and power rationing impacted Eskom's industrial, commercial and residential customers (Muller, 2008).

The supply strains facing Eskom were well exposed, with the effects of load shedding having placed a significant amount of pressure on the entire South African economy. Business was worried about growth in the country. President Mbeki announced in the state of nation address that the government would focus on reducing consumer demand while investing in the longer-term in new generation capacity (News24, 2008c). The commercial capital Johannesburg took the hardest hit, and analysts warned of foreign investors taking flight as factory production was being affected. Gold and platinum production were halted when power to the mines could not be guaranteed (News24, 2008c) and businesses were forced to shut down for few days (News24, 2008a).

Private companies approached Eskom with new strategies to assist Eskom during its time of stress. There were initiatives and solutions to aid in the crisis but these solutions could only be achieved after a longer timeframe. According to the Economist Intelligence Unit (2015), power shortages played a big role in the year 2015 apart from the other factors. Mining contracted year-on-year and manufacturing stagnated. The government has injected more funds into the parastatal Eskom to improve and develop new power stations. Eskom were importing gas from Mozambique and hydropower from the Democrtic Republic of Congo. The mining sector struggled from the 2007 crisis and load shedding heavily affected industries until the year 2015. Power rationing was occurring frequently in South Africa and a fearof Eskom being downgraded to junk status would have worsened the country's energy crisis. Load shedding came under control when one silo of the new power stations came online(Economist Intelligence Unit, 2015). The country still faces a demanding market for electricity as our industries need sufficient energy to operate at its maximum.

This analysis is important because South Africa has been suffering from electricity shortages and one should understand the implications it has on the industry sectors and economic growth. Section 2.4 highlights the theoretical methods that are used to analyse the relationship of economic growth, electricity consumption and electricity infrastructure. It will also showcase the historical papers and the methods used and the conclusions they drew.

2.4 Background of Causality between Electricity Consumption and Economic Growth

There are several studies that test the causal effect of economic growth and electricity consumption. Chouaibi and Abdessalem (2011) in the Tunisian context studied the relationship between Electricity Consumption and the GDP. They used the Augmented Dickey Fuller and Phillips-Perron test to test for stationarity, they also used the Johansen and Juselius Methodology to test for cointegration. They concluded that there exists a unidirectional causality effect running from electricity consumption to economic growth when examined in a bivariate vector autoregressive framework. Nondo, Kahsai and Schaeffer (2010) have studied the relationship between electricity consumption and economic growth of the Common Market for Eastern and Southern Africa (COMESA). They used the Levin, Lin and Chu panel unit root test, the Pedroni's methodology for panel cointegration and investigate the direction of causal effect by using the error correction model. The conclusion they reached from this study was that a reduced electricity consumption could lead to a decline in the economic growth. Both these latter studies concluded that if there is an increase in electricity consumption then the economic growth will also increase and vice versa. All these countries are within Africa. This illustrates the causal relationship as expected since African countries are highly reliant on revenue generated from mineral resources to boost economic growth.

According to Kraft and Kraft (1978) conservative hypothesis is when Gross Domestic Product is less dependent on electricity consumption and growth hypothesis exist when a country's growth is energy dependent and a conservation hypothesis occur when a country's growth is less energy dependent.
Several countries such as South Africa, United Kingdom, Canada, Japan, China, Brazil, Italy, France and Turkey have studied the relationship between the GDP, electricity consumption, industrial output, industrial electricity consumption, electricity consumption per capita and gross domestic product per capita (Bildirici, Bakirtas and Kayikci, 2012). Autoregressive distributed lag methodology was used to examine stationarity and cointegration and it was concluded that there was a growth hypothesis for the following countries; US, China, Canada and Brazil. Countries such as India, Turkey, SA, Japan, UK, France and Italy supports a conservation hypothesis which means that economic growth is less dependent on the energy consumption (Bildirici et al., 2012). According to Bildirici et al. (2012) South Africa's economic growth is less affected by electricity consumption but it is still affected. Solarin (2011) in the context of Botswana studied the relationship between electricity consumption, economic growth and capital formation. The study used the augmented Dickey-Fuller and the Phillips-Perron tests to examine stationarity, the autoregressive distributed lag model to test for cointegration and the unrestricted error correction model to assess the long run relationship. The paper concluded that electricity consumption has a positive relationship with gross domestic product in the long run. Both Solarin (2011) and Bildirici et al. (2012) made use of the autoregressive distribute lag technique which include all series that are integrated in I(1) and I (0) and there is no need to test for unit roots.

Babatunde and Shuaibu (2009) studied the relationship between electricity consumption, real income and electricity price in the context of Nigeria. The technique used was the autoregressive distributed lag with an advantage that it does not require making use of prior unit root tests and the conclusion reached was that in the long run the real income, the price of the substitute appears as the main determinant of electricity demand in Nigeria, while electricity price is insignificant. Pradhan (2010) in the context of India studied the relationship between the electricity consumption, economic growth and transport infrastructure. This study used the Phillips-Perron test for stationarity, Johansen technique to test for cointegration and vector error correction representation technique to test for the direction of the causality effect. The study concluded that energy and transportation policies should recognise the transport- energy consumption- growth nexus in order to maintain sustainable economic growth in the country. The contrast between the last two mentioned studies is that Babatunde and Shuaibu (2009) focuses on demand of residential electricity and which variables are significant whereas Pradhan (2010) studies the causality relationship between the variables mentioned above.

There are similarities between all studies except in studies such as Solarin (2011), Nondo et al. (2010), Bildirici et al. (2012) and Babatunde and Shuaibu (2009) on applying the unit root tests such as Augmented Dickey and Phillip-Perron. Together the studies by Solarin (2011), Bildirici et al. (2012) and Babatunde and Shuaibu (2009) made use of the ADL technique and one of the advantage of this technique is that one don't need to test for unit roots. Nondo et al. (2010) analysed the COMESA countries and made use of the Levin, Lin and Chu panel unit root test to investigate the stationary status. There exist a common response variable, electricity consumption and one common predictor, economic growth throughout the studies with some studies having more than one predictor. The causality effect between electricity consumption and economic growth differs from country to country. According to the studies most of the African countries economic growth are energy dependent except in the study by Bildirici et al. (2012). It shows that South Africa supports a conservation hypothesis.

2.5 Theoretical Review

In this section we will review the theory behind these techniques that will be being used.

2.5.1 Multiple Linear Regression

Multiple linear regression is a technique that investigate the relationship of a phenomena with more than one independent variables. The dependent variable should be numeric and continuous. The independent variable(s) can be either categorical or numeric. Regression analysis make use of cross sectional data and identify the significant variable that shows a dependency between the dependent variable and independent variable. The basic form of a multiple regression model according to Santana (2009) is:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_p X_{p-1,i} + e_i \qquad i = 1, 2, \ldots, n\,; p = 1, 2, \ldots \tag{2.1}$$

Where
Y_i is the random response
$X_{k,i}$ is the k^{th} predictor variable and i^{th} observation
β_k is the k^{th} parameter
e_i is the i^{th} random error term (identical independent distributed (iid))

2.5.1.1 Parameters estimation

The significance of the estimates will be tested by using F-tests. The hypothesis that is tested is the following;

$$H_0 : \; \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$$

$$versus$$

$$H_A : Not\ all\ \beta_i\ is\ zero$$

Let the test statistic to test the significance of the coefficient b and the sum of squares be;

$$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{Y} \tag{2.2}$$

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} \tag{2.3}$$

where $\mathbf{b}' = \left[\hat{\beta}_1, .., \hat{\beta}_{1p}\right]$ the vector of the regression coefficients, $\mathbf{X}$ is the predictor vector, $\mathbf{Y}$ is the response vector. Equations (2.2) and (2.3) are respectively the regression sum of squares and the error sum of squares;

$$MSR = \frac{SSR}{p-1} \tag{2.4}$$

and

$$MSE = \frac{SSE}{n-p} \tag{2.5}$$

, where, MSR and MSE is the regression mean sum of squares and error mean sum of squares, p number of coefficients. The final equation (2.6) will be then the F-test. (Santana, 2009)

$$F_r = \frac{MSR}{MSE} \,. \tag{2.6}$$

H_0 is rejected if $F_r > F_{p-1,n-p}\,(1-\alpha)$

2.5.1.2 Diagnostic Checks of Residuals

In the multiple regression analysis we will assess the assumptions that need to hold that is normal error term, constant variance, no multicollinearity, influential values and no outliers.

- **Normality**

$$H_0 : Residuals\ are\ normally\ distributed$$

$$versus$$

$$H_A : Residuals\ are\ not\ normally\ distributed$$

One can assess this assumption by plotting residuals on a Quantile-Quantile Plot (QQ-Plot). QQ-Plot is a quantiles vs residuals of a sample plot, normality exists if the errors are along the straight line.

The **Shapiro-Wilk** test is defined as the ratio of the best estimator of the variance (based on the square of a linear combination of the order statistics) to the usual corrected sum of squares estimator of the variance (Santana, 2009).

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)} \tag{2.7}$$

where, $X_{(i)}$ is the i^{th} order statistics and $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. The constants a_i are defined as;

$$\mathbf{a} = [a_1, a_2, \ldots, a_n] = \frac{\mathbf{m}'\mathbf{V}^{-1}}{(\mathbf{m}'\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{m})^{1/2}}, \tag{2.8}$$

where m_i is the expected value of the i^{th} order statistic from a sample of n independent identically distributed standard normal random variables, $Z_1, Z_2, \ldots, Z_n$ and $\mathbf{V} = Cov(\ Z_{(1)}, Z_{(2)}, \ldots, Z_{(n)})$. The distribution of the test statistics is only known for $n = 3$. The following statistic however, has an approximate standard normal distribution and is used to estimate the p-values:

$$S_Z = \begin{cases} -\log\left(\gamma - \log\left(1 - W\right)\ \right)\ - \mu/\sigma, & if\ 4 \leq n \leq 11 \\ \log\left(1 - W\right)\ - \mu)/\sigma, & if\ 12 \leq n \leq 2000 \end{cases}, \tag{2.9}$$

where γ, μ and σ are functions of n and are obtained through simulation ((Santana, 2009); (Shapiro and Wilk, 1965)). Normality exists when the p-value is greater than the significant value alpha (0.05) where one does not reject the null hypothesis.

- **Homoscedasticity**

One can use the Breusch-Pagan Test, a large-sample test,

$$H_0 : Residuals\ are\ homoscedastic$$

$$versus$$

$$H_A : Residuals\ are\ not\ homoscedastic$$

assume that the error terms are independent and normally distributed and that the variance of the error term e_i denoted by s_i^2 is related to the level of X in the following way:

$$log_e s_i^2 = \gamma_0 + \gamma_1 X_i. \tag{2.10}$$

Note that equation (2.10) implies that s_i^2 either increases or decreases with level X depending on the sign of γ_1. Constant variance corresponds to $\gamma_1 = 0$. The test of

$H_0 : \gamma_1 = 0$ versus $H_a : \gamma_1 \neq 0$, is achieved by means of regressing the squared residuals e_i^2 against X_i and obtaining the regression sum of squares (SSR^*) which leads to the test statistics X_{BP}^2;

$$X_{BP}^2 = \frac{SSR^*}{2} \div \left(\frac{SSE}{n}\right)^2, \tag{2.11}$$

where, SSR^* is the regression sum of squares when regressing e^2 on X ($SSE = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2$) and SSE the error sum of squares when regressing Y on X ($SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$. $X_{BP}^2 \sim \chi_1^2$, follows a chi-squared distribution with one degree of freedom (Neter, Kutner, Nachtsheim and Wasserman, 1996). When p-values are greater than $alpha$ the significance level then constant variance exists.

- **Influential Values and Outliers**

The residual can be used to identify outlying observations. The residual e_i is defined as the difference between the observed value Y_i and the fitted value $\widehat{Y}_i$;

$$e_i = Y_i - \widehat{Y}_i. \tag{2.12}$$

The studentized deleted residuals are defined as:

$$Rstudent_i = \frac{e_i}{\sqrt{MSE_{(i)}\left(1 - h_{ii}\right)}} \tag{2.13}$$

where $i = 1, 2, 3, \ldots, n$, h_{ii} is the i^{th} diagonal of the $\hat{}$ matrix $\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ and $MSE_{(i)}$ is mean squared error calculated with the i^{th} observation deleted, $MSE_{(i)} = \frac{1}{n-p}\ SSE_{(i)}$. The rule of thumb is that an $Rstudent$ value greater than 2 in absolute value might indicate an influential outlier (Santana, 2009). Influential outlier is a datapoint that pulls the regression line towards it.
The difference between fitted values measures the distance between fitted values and the fitted values when i^{th} observation is deleted.

$$DFFITS_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{MSE_i h_{ii}}} \tag{2.14}$$

where $\hat{Y}_{i(i)}$ is the i^{th} predicted value where the i^{th}case was deleted, $\sqrt{MSE_i}$ is Mean Squared Error with i^{th} observation deleted and h_{ii} is the i^{th} diagonal of the 'Hat' matrix. The influential values are identified when the $DFFITS_i$ value is larger than $2\sqrt{p/n}$ where p is the number of parameters (Santana, 2009).

2.5.1.3 Assumptions of Normal Regression Analysis versus Regression Analysis of Time Series Data

There are certain assumptions one needs to test under a normal cross sectional data regression analysis, one also needs to test the same assumption under time series data for regression analysis. We need to focus on the large sample analysis in the time series context. The classical assumptions are;

1. The model needs to be linear in its parameters. The stochastic process $\{(X_{t1}, X_{t2}, \ldots, X_{tk}) : t = 1, 2, \ldots, n\}$ follows the linear model $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$ where $\{u_t : t = 1, 2, \ldots, n\}$ the sequence of errors or disturbances is $N\left(0, \sigma^2\right)$ and n is the number of observations (time period).

2. There exists no perfect collinearity which means there is no constant independent variable or a perfect linear combination of any of the other independent variables.

3. The expected value of the error u_t, given the predictors variables for all time periods, is zero. $E\left(u_t \mid X\right) = 0$ where $t = 1, 2, \ldots, n$.

4. Under the assumptions mentioned before, the OLS estimators are unbiased conditional on X.

5. The variance of u_t conditional on X, is constant for all t : $Var\left(u_t \mid X\right) = Var\left(u_t\right) = \sigma^2$, $t = 1,\ 2,\ \ldots,\ n.$

6. There exist no serial correlation in the u_ts.

The following assumptions need to be tested in the time series context. These assumptions are exactly the same but require a few adjustments since the large sample analysis is important in time series data.

1. The regression analysis with time series model is exactly the same as assumption 1 in the classical assumption but one need to assume that $\{(\mathbf{x}_t, y_t) : t = 1, 2, \ldots\}$ is stationary and weakly dependent, the law of large numbers and central limit theorem can be applied to sample averages.

2. There exists no perfect collinearity which means no independent variable is constant nor has a perfect linear combination of the others.

3. The explanatory variables $\mathbf{x}_t = (X_{t1}, X_{t2}, \ldots, X_{tk})$ are contemporaneously exogenous. Contemporaneously exogenous holds when the expected value of the error u_t, given the predictor variables for all time periods, is zero.

4. Under the latter assumptions (1,2,3), the OLS estimators are consistent: probability limit $\hat{\beta}_j = \beta_j,\ j = 0, 1, \ldots, k$.

5. The errors are contemporaneously homoscedastic, that is $Var\left(u_t \mid \mathbf{x}_t\right) = \sigma^2$ for all t.

To be able to apply regression analysis on time series data, one must ensure that the classic assumptions discussed above, are met. Weak dependences are necessary for applying the standard large sample results on the time series context (Wooldridge, 2015).

2.5.2 Background of Econometric Modelling

Econometrics is the combination of statistics, mathematics and economics. Econometrics was developed because economy theory needed to be quantified, necessary predictions of the economic trends needed to be explained to assist in decision making in the public sector and private sector.

According to Hansen (2000) ;(Frisch, 1933, p. 1-4) “The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics.... Its main objective shall be to promote studies that aim at a unification of the theoretical quantitative and the empirical-quantitative approach to economic problems”

A model is a simplified representation of a real world process. Models assist one to display the relationship between variables. A model should have at least one response variable and one independent variable to be classified as a model. Models can assist us in predicting, drawing conclusion from a real life situations and improving the world’s decision making. The model can only be created when one has an underlying phenomena to analyse or investigate. An economic model is a set of assumptions that describes the behaviour of an economy or more generally a phenomenon (Baltagi, 2008).

Econometrics primarily uses linear regression models. Regression modelling is the underlying model for analysing relationships and predicting behaviours of certain studies. In this study, we will explore either of the following techniques to build econometric models namely, 1. Autoregressive distributed lag model and the 2. Johansen and Juselius Methodology and the Engle and Granger Methodology. Both these approaches will help us to find the direction of the relationship if any exists, either between electricity consumption and economic growth or electricity infrastructure. The Granger causality effect tests for four hypotheses according to Kraft and Kraft (1978) seminal work. The neutral hypothesis is when none of the variables have a causal effect on one another. The growth hypothesis occurs when a country's growth is energy dependent and a conservation hypothesis occur when a country's growth is less energy dependent. The bi-directional hypothesis assumes an increase in energy consumption and stimulates economic growth, and vice-versa. The following section will explore the techniques on testing stationarity and cointegration.

2.5.2.1 Stationarity

Consider a finite set of random variables $\{Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n}\}$ from a stochastic process $\{Z(\omega, t) : t = 0, \pm 1, \pm 2, \ldots\}$. The n-dimensional distribution function of the Z_i is defined by;

$$F_{Z_{T_1},\ldots,Z_{T_n}}(x_1, \ldots, x_n) = P\{\omega : \ Z_{t_1} \leq x_1, \ldots, Z_{t_n} \leq x_n\}, \tag{2.15}$$

where $x_i, (\ i = 1, \ldots, n)$ are any real numbers.

A process is said to be first-order stationary in distribution if its one-dimensional distribution function is time invariant, that is if $F_{Z_{t_1}}(x_1) = F_{Z_{t_1+k}}(x_1)$ for any integers t_i, k and $t_1 + k$ and second-order stationary in distribution if $F_{Z_{t_1},Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k},Z_{t_2+k}}(x_1, x_2)$ for any integers $t_1, t_2, k, t_1 + k$ and $t_2 + k$; and n^{th}-order stationary in distribution if

$$F_{Z_{t_1},\ldots,Z_{t_n}}\left(x_1,\ldots,x_n\right)=F_{Z_{t_1+k},\ldots,Z_{t_n+k}}\left(x_1,\ldots,x_n\right), \tag{2.16}$$

for any n-tuple $(t_1,\ldots,t_n)$ and k integers. A process is said to be strictly stationary if equation (2.16) is true for any $n=1,2,\ldots$

A stochastic process is a set of time indexed random variables defined on a sample space. We usually suppress the variable ω and simply state $Z\left(\omega,t\right)$ as $Z\left(t\right)$. For a given real-valued process $\{Z_t : t=0,\pm1,\pm2,\ldots\}$, we define the mean function of the process;

$$\mu_t=E\left(Z_t\right). \tag{2.17}$$

The variance function of the process

$$\sigma_t^2=E\left(Z_t-\mu_t\right)^2. \tag{2.18}$$

The covariance function between $Z_{t_1}$and Z_{t_2} is given by;

$$\gamma\left(t_1,t_2\right)=E\left(Z_{t_1}-\mu_{t_1}\right)\left(Z_{t_2}-\mu_{t_2}\right) \tag{2.19}$$

And the correlation function between $Z_{t_1}$and Z_{t_2} given by;

$$\rho\left(t_1,t_2\right)=\frac{\gamma\left(t_1,t_2\right)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}} \tag{2.20}$$

For a strictly stationary process, the mean $\mu_t=\mu$ is constant given that $E\left(|Z_t|\right)<\infty$ and if $E\left(|Z_t^2|\right)<\infty$ then$\sigma_t^2=\sigma^2$, for all t hence also a constant. (Wei, 1994) Moreover, since $F_{Z_{t_1},Z_{t_2}}\left(x_1,x_2\right)=F_{Z_{t_1+k},Z_{t_2+k}}\left(x_1,x_2\right)$ for any integer t_1,t_2 and k we have

$$\gamma\left(t_1,t_2\right)=\gamma\left(t_1+k,t_2+k\right), \tag{2.21}$$

and

$$\rho\left(t_1,t_2\right)=\rho\left(t_1+k,t_2+k\right). \tag{2.22}$$

Letting $t_1=t-k$ and $t_2=t$, we get

$$\gamma\left(t_1,t_2\right)=\gamma\left(t,t+k\right)=\gamma_k, \tag{2.23}$$

and

$$\rho\left(t_1,t_2\right)=\rho\left(t,t+k\right)=\rho_k. \tag{2.24}$$

Therefore, for a strictly stationary process with first finite two moments the covariance and the correlation between Z_t and Z_{t+k} depends on the time difference k (Wei, 1994).

2.5.2.2 Autocorrelation Functions

For a stationary process $\{Z_t\}$, we have $E(Z_t) = \mu$ and $Var(Z_t) = E(Z_t - \mu)^2 = s^2$ which is constant and the covariance between Z_t and Z_{t+k} is;

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu). \tag{2.25}$$

We define autocorrelation function (acf) equation (2.26) and partial autocorrelation function equation (2.27) as follow;

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}, \tag{2.26}$$

where we note that $Var(Z_t) = Var(Z_{t+k}) = \gamma_0$. The estimate of the autocorrelation function can also be defined as $r_k = \hat{\rho}_k$,

$$r_k = \frac{\sum_{i=1}^{T-k}(y_i - \overline{y}_1)(y_{i+k} - \overline{y}_2)}{\sqrt{\sum_{i=1}^{T-k}(y_i - \overline{y}_1)^2 \sum_{i=1}^{T-k}(y_i - \overline{y}_2)^2}}, \tag{2.27}$$

where, $\overline{y}_1$is the mean of the first $T-k$ observations and $\overline{y}_2$ is the mean of the last $T-k$ observations. A simplified version of equation (2.27) $\overline{y}_1$ and $\overline{y}_2$ is replaced by $\overline{y}$ with the summation being taken over the full range $1, 2, \ldots, T$ so that;

$$r_k = \frac{\sum_{i=1}^{T-k}(y_i - \overline{y})(y_{i+k} - \overline{y})}{\sum_{i=1}^{T-k}(y_i - \overline{y})^2}, \tag{2.28}$$

The partial autocorrelation function (pacf) is the conditional correlation. The partial autocorrelation function between Z_t and Z_{t+k} is;

$$P_k = \frac{Cov\left[\left(Z_t - \hat{Z}_t\right), \left(Z_{t+k} - \hat{Z}_{t+k}\right)\right]}{\sqrt{Var\left(Z_t - \hat{Z}_t\right)}\sqrt{Var\left(Z_t - \hat{Z}_{t+k}\right)}}, \tag{2.29}$$

Where $Var\left(Z_t - \hat{Z}_{t+k}\right) = E\left[(Z_{t+k} - aZ_{t+k-1} - \ldots - a_{k-1}Z_{t+k})^2\right] =$ $Var\left(Z_t - \hat{Z}_t\right) = \gamma_0 - a_1\gamma_1 - \ldots - a_{k-1}\gamma_{t+1}$ and $a_i = \beta_i\,(1 \le i \le k-1)$ we have $Cov\left[\left(Z_t - \hat{Z}_t\right), \left(Z_{t+k} - \hat{Z}_{t+k}\right)\right] = \gamma_k - a_1\gamma_{k-1} - \ldots - a_{k-1}\gamma_1.$

2.5.2.3 Autoregressive Distributed lag (ADL)

The lag operator L or B is defined for a time series y_t by $Ly_t = y_{t-1}$. The operater can be defined for linear combinations by,

$$B\left(a_1 y_{t_1} + a_2 y_{t_2}\right) = a_1 y_{t_1 - 1} + a_2 y_{t_2 - 1}. \tag{2.30}$$

An alternative test for cointegration is to estimate the Autoregressive Distributed Lag (ADL) model where, we consider the simple $ADL(1,0)$ in equation (2.47) as model containing I(1) regressors and a linear trend,

$$\phi\left(L\right) y_t = \; a_0 + a_1 t + \beta' \mathbf{x}_t + \mu_t, \tag{2.31}$$

where, $t = 1, \ldots, T$, $\{y_t : t = 0, 1, \ldots\}$, $\phi\left(L\right) = 1 - \phi L$ with L being the one period lag operator, $\mathbf{x}_t$ is a $k \times 1$ vector of regressors assumed to be integrated of order 1: $\mathbf{x}_t = \; \mathbf{x}_{t-1} + \mathbf{e}_t$ and β is a $k \times 1$ vector of unknown parameters (Pesaran and Shin, 1998). One can choose the lag length such that the residual is a white noise. This model can be rewritten as an error correction model $y_t = \phi\left(1\right)^{-1} B\left(1\right)$ where $B\left(1\right) = a_0 + a_1 t + \beta' \mathbf{x}_t$ represents the long-run steady state solution of the model (Sjö, 2008). Cointegration exist when polynomial $\phi\left(L\right)$ do not contain any unit roots.

2.5.2.4 Autoregressive Integrated Moving Average Model (ARIMA)

Firstly we will define the autoregressive (AR) model and the moving average (MA) model. The Autoregressive $AR(p)$ model is known as;

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + u_t,$$

where u_t is iid $N(0, s_u^2)$, μ is assumed to be 0, u_t independent of $z_{t-1}, z_{t-2}, \ldots, z_{t-p}$ and $z_t = z_t - \mu$

The Moving average $MA(q)$ model is defined as;

$$y_t = \mu + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q a_{t-q} + u_t$$

The Box-Jenkins Approach is a combination of the AR and MA models which is integrated. Most economic series needs to be differenced once or several times before process is stationary in which case we will have an ARIMA (*p, d, q*) model with p^{th} order AR and q^{th} order MA and *d* number integrated. Therefore, ARIMA (*p, d, q*) can be written as $f(B)(1-B)^d z_t = \phi(B)u_t$, when we replace $z_t = z_t - \mu$ then, the model is as follows (Chimedza and Galpin, 2004);

$$\phi(B)[(1-B)^d(z_t-\mu)] = \theta(B)u_t. \tag{2.32}$$

2.5.2.5 Unit Root Tests

The two techniques one can use to test for stationarity are the Dickey Fuller test and Augmented Dickey Fuller test. When a unit root exist in a series it refer that the series is unpredictable and the series can exhibit non-stationary. The Augmented Dickey is used when series or models are more complex.

2.5.2.6 Dickey-Fuller Test

The Dickey Fuller test is based on the $AR(1)$ model, where,

$$AR\,(1) = y_t = \rho y_{t-1} + e_t,\ t = 1, 2, \ldots, \tag{2.33}$$

where y_0 is the observed initial value and e_t is the random error. Equation (2.33) has a unit root if $a = 1$. If $a = 0$ and $\rho = 1$ then y_t follows a random walk without a drift, and if $a \neq 0$ then y_t follows a random walk with a drift. The drift a will be left unspecified under the null hypothesis.
The null hypothesis is that $\{y_t\}$ has a unit root:

$$H_0 : \rho = 1, \tag{2.34}$$

$$versus$$

$$H_1 : \rho < 1. \tag{2.35}$$

The alternative $H_1 : \rho > 1$ will never be considered since it will state that ρ is explosive. When $|\rho| < 1$, then $\{y_t\}$ is a constant $AR\,(1)$ process which means it is weakly dependent. Thus, testing equation (2.34) in (2.33), with the alternative given by equation (2.35) is actaully testing of whether $\{y_t\}$ follows I(1) versus the alternative that $\{y_t\}$ follows $I(0)$. A time series $\{y_t\}$ is intergrated of order 1 $I(1)$ if $\{y_t\}$ is not stationary but its first difference $y_t - y_{t-1}$ is stationary. When carrying out the unit test one need to subtract y_{t-1} from both side of equation (2.33) and define $\theta = \rho - 1$:

$$\triangle y_t = a + \theta y_{t-1} + e_t \tag{2.36}$$

where, $\triangle \ = y_t - y_{t-1}$. Since the central limit theorem does not apply in testing $\theta = 1$, we will use the Dickey- Fuller distribution (Wooldridge, 2015). We reject the null hypothesis $H_0 : \theta = 1$ against $H_1 : \theta < 0$ if $t_{\hat{\theta}} < c$ where c is one of the negative values from an output table.

2.5.2.7 Augmented Dickey-Fuller Test

There is an extended version of the Dickey-Fuller test which is the augmented Dickey Fuller test. When $\{y_t\}$ follows equation (2.33) with $\rho = 1$ then $\triangle y_t$ is serially uncorrelated. One can allow $\{\triangle y_t\}$ to follow an autoregressive model by augmenting equation (2.36) with additional lags. For illustration, let

$$\triangle y_t = a + \theta y_{t-1} + \gamma_1 \triangle y_{t-1} + e_t, \tag{2.37}$$

where, $|\gamma_1| < 1$ this ensures that, under $H_0 : \ \theta = 0$, $\{\triangle y_t\}$ follows a constant autoregressive of order 1 model. Under the alternative $H_1 : \theta < 0$, it can be shown that $\{\triangle y_t\}$ follows a constant autoregressive of order 1 model. Adding p lags of $\triangle y_t$ to the equation to justify for dynamics in the process. The approach we take to test for the null hypothesis is the same as before. One will run the regression of

$$\triangle y_t \ , \ y_{t-1}, \triangle y_{t-1}, \ldots, \triangle y_{t-p}, \tag{2.38}$$

and carry out the t test on $\hat{\theta}$, the coefficient of y_{t-1} just as before. Equation (2.38) is the augmented Dickey-Fuller test because the regression has been augmented with lagged changes. The critical value and rejection rule is the same as in the Dickey Fuller test However, under H_0, y_t is non-stationary, so conventional normal asymptotics are invalid. An alternative asymptotic framework has been developed to deal with non-stationary data (Hansen, 2000).

In the above paragraph we mentioned the Augmented Dickey-Fuller test for the unit root test which is derived from the normal Dickey-Fuller test. These tests have the same asymptotic distribution as the corresponding Dickey-Fuller test and the same critical values can be used (Baltagi, 2008). The only difference is that with the Augmented Dickey-Fuller test one can make use of more lags in order to reduce the term e_t.

2.5.2.8 Cointegration

Cointegration is defined as follows: If $\{y_t : t = 0, 1, \ldots\}$ and $\{x_t : t = 0, 1, \ldots\}$ are two $I(1)$ processes then $y_t - \beta x_t$ is an I(0) integrated of order 0 process for number β. (Wooldridge, 2015) There are several techniques that can be used to test for cointegration, the next few sections will illustrate a few of these techniques.

The idea behind cointegration is to see whether there co-exists stationarity between the response variable and the predictor(s) (Baltagi, 2008). The same level of co integration is useful to validating the F-statistics showing that one can trust the regressed model and the long run relationship of the phenomena under study. According to (Baltagi, 2008) the theory of cointegration tries to estimate this long-run relationship using the nonstationary series themselves, rather than their first differences.

2.5.2.9 Johansen and Juselius Methodology

Johansen (1988); (Österholm and Hjalmarsson, 2007) make use of eigenvalues and trace statistics for testing cointegration. Johansen's methodology takes its starting point in the equation (2.39) of p^{th} order Österholm and Hjalmarsson (2007) that is;

$$\mathbf{y}_t = \mu + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{e}_t, \tag{2.39}$$

where, $\mathbf{y}_t$ is a $n \times 1$ vector of variables that are integrated of order one which is generally denoted I(1), $\mathbf{A}$ is a $n \times n$ matrix of parameters and $\mathbf{e}_t$ is a $n \times 1$ vector of errors. This vector autoregressive can be transcribed as

$$\mathbf{y}_t = \mu + \mathbf{\Pi} y_{t-1} + \sum_{i=1}^{p-1} \mathbf{\Gamma} \triangle \mathbf{y}_{t-i} + \mathbf{e}_t, \tag{2.40}$$

where,

$$\mathbf{\Pi} = \sum_{i=1}^{p} \mathbf{A}_i - \mathbf{I} \; and \; \mathbf{\Gamma} = \; - \sum_{j=i+1}^{p} \mathbf{A}_j, \tag{2.41}$$

There are two different likelihood ratio tests that are recommended by Johansen. There is a reduced rank for the $\mathbf{\Pi}$ matrix. The trace test and maximum eigen value test are shown in equations (2.42) and (2.43);

$$J_{trace} = -N \sum_{i=r+1}^{n} ln\left(1 - \hat{\lambda}_i\right), \tag{2.42}$$

$$J_{max} = -Nln\left(1 - \hat{\lambda}_{r+1}\right), \tag{2.43}$$

where, N is the sample size and $\hat{\lambda}_i$ is the i^{th} biggest canonical correlation. The trace test null hypothesis is that there exists *r* cointegrating vectors versus the alternative hypothesis of only *n* cointegrating vectors.The maximum eigenvalue test investigate the null hypothesis of a number *r* cointegrating vectors versus the alternative hypothesis of *r* +1 cointegrating vectors. According to Österholm and Hjalmarsson (2007) to be able to trust the Johansen test one needs to first test if the unit root test of the series is integrated order I(1) or I(0). When the test statistics is greater than some of the critical values you can reject the null at that significance level.
The Johansen tests are likelihood-ratio tests which is the maximum eigenvalue test and the trace test. For both test statistics, the initial Johansen test is a test of the null hypothesis of no cointegration against the alternative of cointegration.

2.5.2.10 Engle and Granger's Two-Step Procedure

Engle and Granger (1987) formulated one of the first tests of cointegration. This test has the advantage that it is intuitive and easy to perform. It has two steps namely the estimation of the co-integrating regression equation where we assume all variables are integrated to order 1, $I(1)$ and might integrate to from a stationary relationship and testing the unit root test in the residual process of the cointegrating regression in equation (2.44). Suppose the regression equation is

$$y_{1,t} = \beta_1 + \beta_2 x_{2,t} + \ldots + \beta_p x_{p,t} + e_t \tag{2.44}$$

where, p is the number of variables, $y_{1,t}$ is the response variables, e_t is the residual term. The equation represents an assumed economically steady state among the variables. If the variables are cointegrated then they follow a common trend and form a stationary relationship in the long run (Sjö, 2008).

If this is the case then, the estimates can be viewed as correct for the long run state and the residuals can be used as errors for the error correction term in the error correction model. The second step of this procedure is that one needs to test the unit root test on the residuals by using the Augmented Dickey-Fuller test (Sjö, 2008).

According to Sjö (2008) there are three main problems with the two-step procedure. Firstly one needs to determine the number of lags in the augmentation. Secondly, the test is based on the assumption of one cointegrating vector and care must be taken when applying the test to models with more than two variables. Lastly, the test assumes a common factor in the dynamics of the system. To test for cointegration, suppose we have this model;

$$\hat{y}_t = \hat{a} + \hat{\beta} x_t \tag{2.45}$$

where $\{\hat{y}_t : t = 0, 1, \ldots\}$ and $\{x_t : t = 0, 1, \ldots\}$ are two processes.
Under the null hypothesis;

$$H_0 : \; y_t \; and \; x_i \; are \; not \; cointegrated$$

$$versus$$

$$H_1 : y_t \; and \; x_i \; are \; cointegrated$$

When one runs a regression test of $\triangle\hat{\mu}_t = \hat{\mu}_t - \hat{\mu}_{t-1}$ on $\hat{\mu}_{t-1}$ and the critical value of model is below the critical values of Davidson, MacKinnon et al. (1993) then the series are cointegrated.

2.5.3 Information Criteria

Akaike's information criterion (AIC), Schwarz' Bayesian criterion and Hannan-Quinn are model selection criteria that measures the quality of a statistical model for a given data set. The smallest values of the criteria are chosen where these criteria are given by;

$$AIC_p = n \, ln \, SSE_p - n \, ln \, p + 2p \, ; \; BIC_p = n \, ln \, SSE_p - n \, ln \, p + [ln \, n]p \tag{2.46}$$

where p number of parameters. According to (Neter et al., 1996) if $n \geq 8$ then the value for BIC_p is greater than AIC_p.
For a certain lag order Hannan-Quinn is given by;

$$HQ_p = -2\,(\frac{lnL}{n}) + 2p\frac{ln(ln\,n)}{n} \tag{2.47}$$

where p number of parameters.

Chapter 3

Methodology

3.1 Introduction

The following section will discuss how the different techniques chosen are applied. Since the data is of time series nature and one of the variables is of an economic nature, the most appropriate techniques used to analyse the data are mentioned in the literature review. All these technique can describe the relationships in the data but the most applicable techniques are the econometric model and the ARIMA model.

3.2 Data

The data used in this study were collected from various sources from the year 1993 to 2015. The data was collected from the South African Reserve Bank and Statistics SA. The electricity consumption and economic growth, are measured in Kilowatt per hour (kWh) and local currency (Rand) respectively. GDP is the sum of gross value added by all local producers in the economy plus any product taxes and minus any grants which is exempt from the value of the products (World Bank, 2008). The amount invested in the energy infrastructure measured in rands was obtained from Eskom. The type of data that was used is time series data.

The variables that will be used are as follows:

1. Electricity consumption
2. Gross domestic product
3. Electricity infrastructure

Electricity consumption which will be the response variable and the predictors are the eleven sectors which sum up to gross domestic product and energy infrastructure.

The model that we need to fit is as follow:

$$Electricity\ Consumption = \beta_0 + \beta_1 gdp + \beta_2 Energy\ infrastructure + \varepsilon$$

3.3 Descriptive Statistics

The following statistics and descriptive plot will assist in describing the variables used.

In this study we will explore the maximum, minimum, median, kurtosis, skewness and mean of each variable. The boxplot will assist in this regard. A scatterplot of economic growth and electricity consumption to see the relationship between the two variables will be done. Histograms to determine the distribution of each variable will also be done. In addition times series plots will be shown.

3.4 Regression Model

In the multiple regression analysis we will assess the assumptions that need to hold that is normal error terms, constant variance, no multicollinearity, influential values and no outliers. Table 3.1 summarise the tests that will be used;

Table 3.1: Tests to be used

Assumption	**Graphical**	**Formal statistical test**
Normality	QQ Plots	Shapiro Wilk
Constant Variance	Residual Plots	Bartlett test
Outliers		Studentized deleted residuals /Residual
Influential Values		Difference between fitted values and fitted value where the i^{th} observation was deleted.

The ANOVA test analyses the significant parameters and the overall model evaluation. Hypothesis is shown below (Santana, 2009)

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0 \;\; vs \;\; H_i : \beta_i \neq \beta_j \;\; for \;\; atleast \;\; one \;\; pair \;\; i \neq j$$

Assumptions To Be Check

Regression model assumptions that would be tested in this study (Santana, 2009)

1. The error terms have zero expected value.
2. The error terms are normally distributed.
3. The error terms are independently distributed.
4. The error terms are homoscedastic (constant variance).

There should be an investigation of influential values, outliers and regressor variable that may correlated with one another.

3.5 Econometric Model

The econometric model will be assessed in the follow manner;

1. Check the stationarity status using Augmented Dickey-Fuller test. The null hypothesis is the series is not stationary.

2. The next step is to analyse the cointegration between the series. We will make use of the Johansen and Juselius Methodology. If the test statistic is greater than critical values, we reject the null hypothesis at that level of significance.

3. Granger causality effect will be tested for and if present one of the four hypothesis will be chosen.

3.6 Box-Jenkins Model

The Autoregressive Integrated Moving Average model will be determined as follows;

Stationary will be investigated using plots such as Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). If the series is not stationary, the data would be differenced or transformed and the series are again plotted. These plots can also assist in identifying the possible model that the series depicts.

The different models under ARIMA will be assessed and the model with the smallest Akaike Information Criterion chosen.

The future values for the time series is predicted using the forecast package (Hyndman, 2008) in R.

3.7 Evaluation

The most efficient model will be the one that exhibits the following traits:

1. The highest coefficient of determination that is the adjusted R-squared compares the explanatory power of regression models that contain different numbers of predictors.

2. The smallest information criteria either Schwartz Bayesian Criterion or Akaike information Criterion.

3. The model that gives a sensible and realistic predictive values.

Chapter 4

Analysis

4.1 Introduction

This chapter presents the results of the techniques discussed in Chapter 2, that is, the Box Jenkins model, the regression model and the econometric model.

4.2 Description of the Data

In this section the data is described using bar graphs and summary tables that showcase the mean, maximum, minimum, 75th percentile and 25th percentile. Figure 4.1 presents the pairs plot of each variable against another. There is a positive linear relationship between GDP and electricity infrastructure (CapExpenditure) whereas GDP and electricity consumption have no linear relationship but more of a quadratic relationship. There seems to be a weak quadratic the relation between electricity consumption and electricity infrastructure.

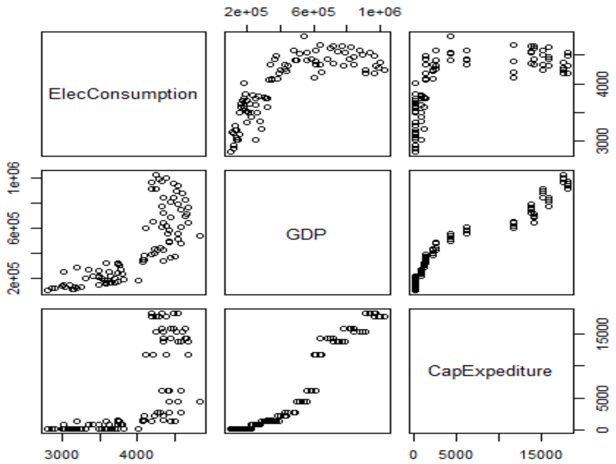


Figure 4.1: Scatterplot of all pairs

The following figures that is, Figure 4.2, 4.3 and 4.4 are the histogram graphs of electricity consumption, economic growth and electricity infrastructure respectively. They will showcase the distribution of these variables. Electricity consumption is skewed to the left, economic growth is skewed to the right as well as the variable electricity infrastructure which are shown in Figure 4.2, 4.3 and 4.4 respectively.

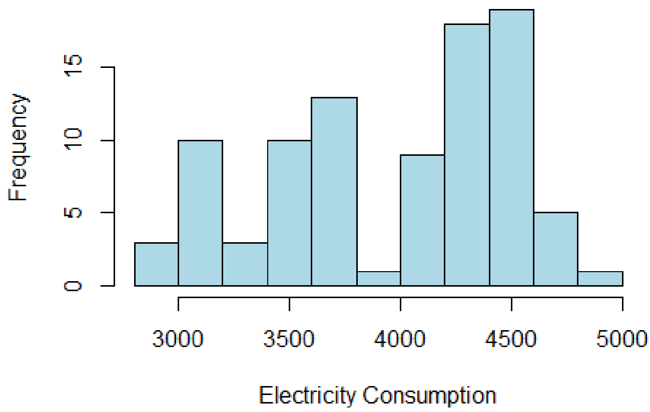


Figure 4.2: Histogram for Electricity Consumption

Figure 4.2 shows the histogram of the electricity consumption variable and it is skewed to the left. The most frequent electricity consumption amount is 4 500 KWh units.

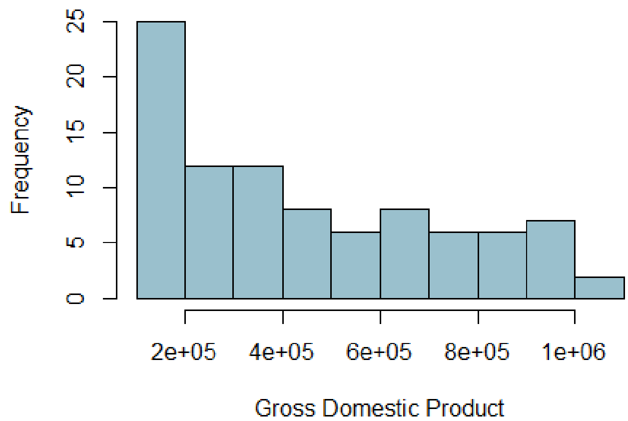


Figure 4.3 Histogram for Gross Domestic Product

Figure 4.3 shows that the histogram of the gross domestic product variable is skewed to the right. Note that the most frequent gross domestic product value is 200 000 million rands.

Figure 4.4 below shows the histogram of the electricity infrastructure variable and it is skewed to the right.

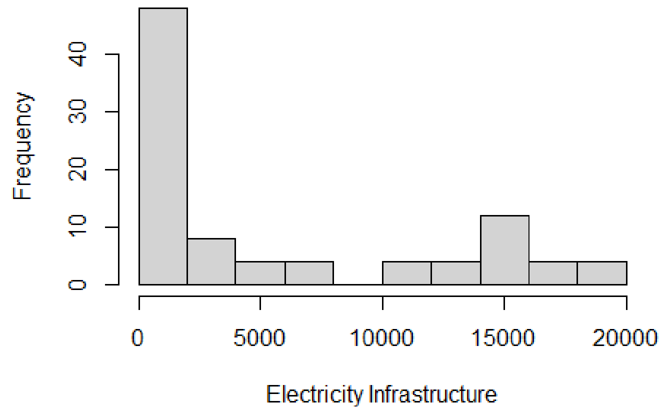


Figure 4.4: Histogram for Electricity Infrastructure

Table 4.1 depicts the mean for electricity consumption, gross domestic product and electricity Infrastructure are 3 965 KWh, R 448 565 million and R 5 655.20 million respectively. The latter are the average amount South Africa's quarterly GDP and electricity infrastructure since year 1993.

Table 4.1: Descriptive Statistics of variables

Variable	Min	1^{st} Quartile	Median	Mean	3^{rd} Quartile	Max
Electricity Consumption	2,802	3,579	4,146	3,965	4,419	4,833
Gross Domestic Product	100,787	195,366	369,613	448,565	659,443	1,027,026
Electricity Infrastructure	202.0	263.8	1,519.5	5,655.2	13,864.2	18,179

Note in Table 4.2 the GDP and electricity infrastructure distribution is skewed to the right where electricity consumption is skewed to the left. All of these variables are kurtosis value is less than 3, therefore the distribution of the data is platykurtic.

Table 4.2: Kurtosis and Skewness

Variable	Kurtosis	Skewness
Electricity Consumption	1.99	-0.467
Gross Domestic Product	1.979	0.55
Electricity Infrastructure	1.928	0.8197

4.3 Time Series Analysis

Time series analysis primarily assists in investigating the trends of a phenomena and what the effect of time has on trends. The following analysis will assist in determining the causal effect between consumption/electricity consumption, electricity infrastructure and growth/economic growth (GDP).

4.3.1 Analysing the Three Variables Separately

4.3.1.1 Electricity Consumption

Figure 4.5 is a graphical display of the quarterly electricity consumption (EC) series from 1993 to 2016. Figure 4.5 shows that the series is not stationary in the mean. In the earlier years the consumption was much lower as the normal mean.

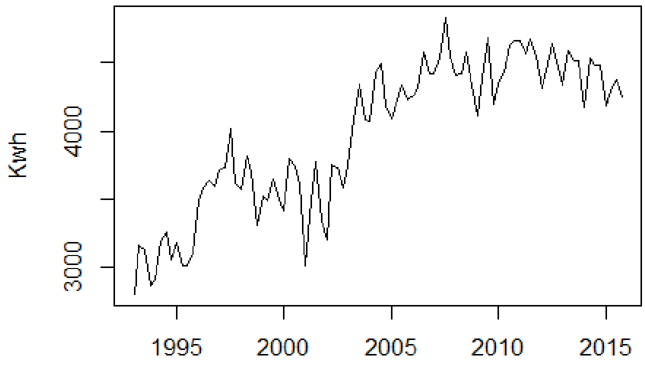


Figure 4.5: Time series of electricity Consumption

Figure 4.6 explores the stationarity status of the series and illustrate the correlation between the different lags. Remember that the ACF is the Autocorrelation function and PACF is the Partial Autocorrelation Function which just show us the correlation between the lags. Note that the ACF in Figure 4.6 shows a slowly decaying trend and seems to have a seasonal component and PACF lags off with a seasonal component.

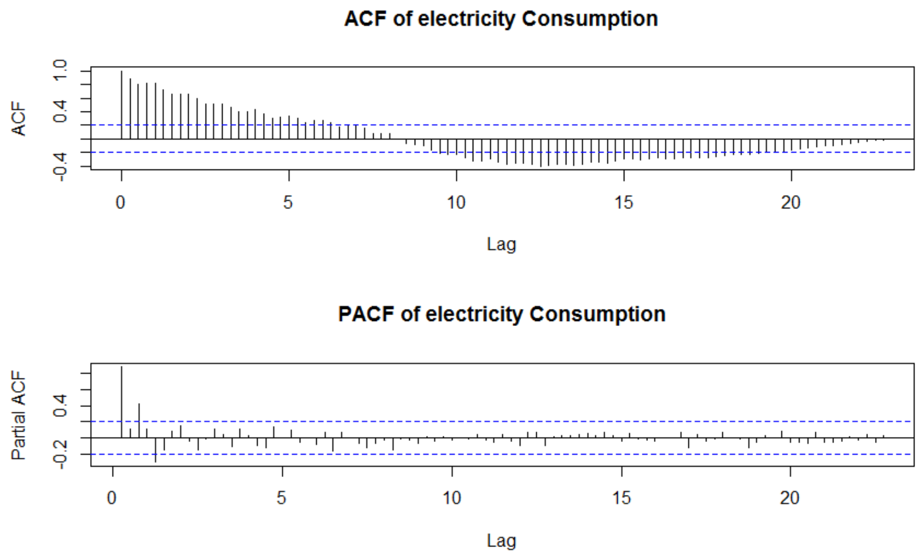


Figure 4.6: The ACF and PACF of electricity Consumption

A change either in the variance or the mean of the series will assist in determining the stationarity. Firstly, an attempt of seasonal difference will be performed. Applying a seasonal differencing on the original time series to see whether there is any change in the stationarity status showing that the differenced series changed the stationarity-status from nonstationary to stationary in the mean.

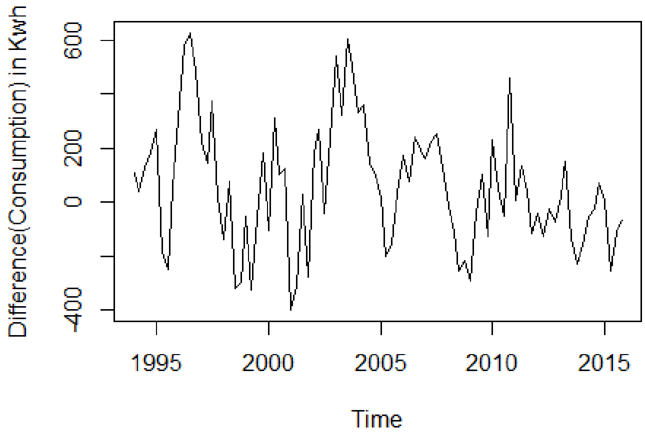


Figure 4.7: Differenced electricity Consumption

Figure 4.7 show that semi - stationarity has been reach after a seasonal difference has been taken.

Note that the ACF has a decaying trend which is the first indication of working with an autoregressive model and there is seasonal component. The PACF cut off at lag 1 and the seasonality effect cuts off at lag 2. The possible model that can be derived from Figure 4.8 is $ARIMA(1,0,0)(P,1,Q)$ or another consideration is $ARIMA(1,0,0)(2,1,0)_4$.

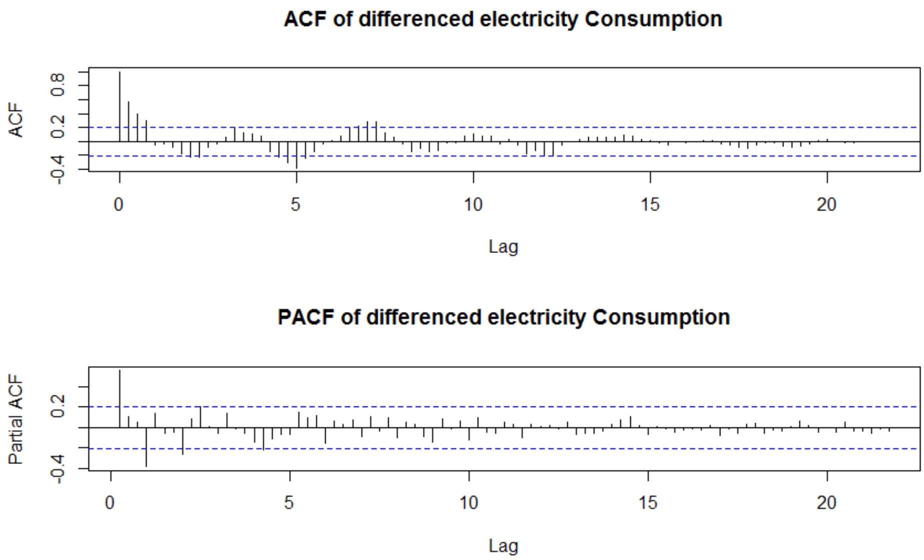


Figure 4.8: The ACF and PACF of differenced electricity Consumption

Note that when exploring a regular differenced seasonal differenced series, the stationary status has been reach as shown on Figure 4.9.

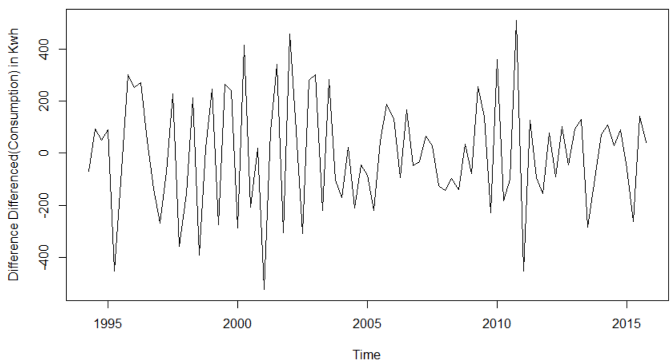


Figure 4.9: Differenced (difference(electricity Consumption))

Figure 4.9 shows a constant mean over time a low seasonal variation, between 2005 and 2009 is low as well as for 2012 to 2015. The ACF and PACF of the seasonally and regularly differenced electricity consumption are shown in Figure 4.10.

The ACF show a cut off at lag 1 with a seasonal component cutting at lag 2 and the PACF has a decaying trend. The possible model that Figure 4.10 depicts is a $ARIMA(0,1,1)(0,1,2)_4$ model.

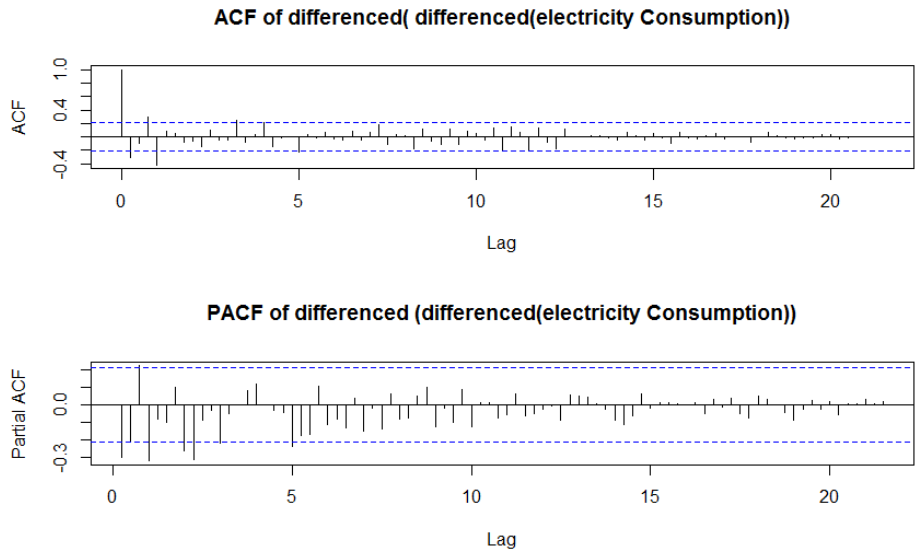


Figure 4.10: The ACF and PACF of differenced (differenced(electricity Consumption))

4.3.1.1.1 Estimating the model fit, estimates and forecasting - Electricity Consumption

This section will showcase the different possible models that the consumption time series data follow, present the graphical and table format forecast of the next three quarters and the optimal model for consumption series.

Exploring the possible models that were manually identified and the other two will be identified by the auto.arima (Hyndman, 2008) built-in function in R software. Note that these two model were manually identified, ARIMA(1,0,0)(2,1,0) and ARIMA(0,1,1)(0,1,2). These fitted model are shown in Appendix A, A1 to A6. The model with the smallest Akaike Information Criteria will be the optimal model which was produced by the auto arima function (Hyndman, 2008).
Note that the snapshot of all the possible models identified are in Table 4.3. Note that Θ denotes the use of seasonal.

Table 4.3: Possible Models - Electricity Consumption

	Description	AIC
1	ARIMA(0,1,1)(0,0,1)	1143.45
2	ARIMA(3,1,0)(2,0,0)	1163.41
3	ARIMA(1,0,0)(2,1,0)	1155.62
4	ARIMA(0,1,1)(0,1,2)	1148.66

Note that Table 4.3 show all the possible model that were studied for a possible description model of the consumption series. The optimal model is the model that present the smallest Akaike Information criteria which is model $ARIMA(0,1,1)(0,0,1)_4$ with criteria 1143.45.

Table 4.4: ARIMA(0,1,1)(0,0,1) model

	Value	Standard Error	z value	p value
θ_1	-0.3636	0.099	-3.672727	0.0002399756
$\Theta\ \theta_1$	-0.8677	0.1007	-8.616683	6.892147e-18
AIC	1143.45			

The model $ARIMA(0,1,1)(0,0,1)_4$ has a AIC of 1143.45 which is shown in Table 4.4. The criterion is the smallest among the several model identified. The significant values on model $ARIMA(0,1,1)(0,0,1)_4$ are 0. -0.3636 and -0.8677 since the p value is smaller than alpha of 0.05 therefore the model is $y_t = -0.3636\theta_1 - 0.8677 \text{seasonal}\theta_1$

Forecasting the next three values of the consumption series which displayed in Figure 4.11. There are two methods to forecast these values namely Holt-Winters and ARIMA function. Note that Table 4.5 depicts these values of the different methods and compares them to the actual consumption.

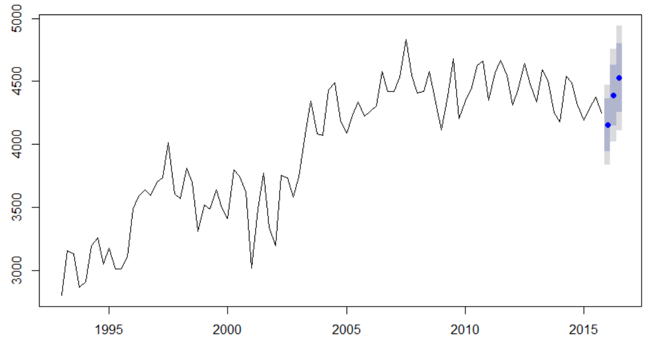


Figure 4.11: Forecasting - Consumption next three values

The Holt-Winter (Hyndman, 2008) predicted the following estimates for the model, depicted in Table 4.5.

Table 4.5: HoltWinters Estimates

	α	β	γ	**a**	**b**	Θ_1	Θ_2	Θ_3	Θ_4
Value	0.604	0	0.237	4323.798	25.588	-195.289	14.108	126.487	-84.975

Note that Table 4.6 shows the predicted values of consumption using ARIMA and Holtwinter models for the next three quarter. Comparing the two methods, both methods overestimate the consumption of the next two quarters. The average difference between actuals and ARIMA predictions is 108 whereas the average difference between actuals values and HoltWinters is 168 which higher than the average difference of the ARIMA prediction. The ARIMA method has shown to be the method that is closer to the actual values. In this ease , more observations are needed to have a more accurate conclusion.

Table 4.6: Predictors for the next three quarters in year 2016 - Electricity Consumption

	Quarter 1	**Quarter 2**	**Quarter 3**
Holt-Winters	4154.097	4389.082	4527.048
ARIMA	4154.077	4269.499	4299.411
Actual	4056	4151	N/A

Figure 4.12 show the graphical forecast of the model $ARIMA(0,1,1)(0,0,1)_4$. The rest of the graphical Figures and Tables for electricity consumption are in Appendix A, A1. that is, Table A.1 to A.3.

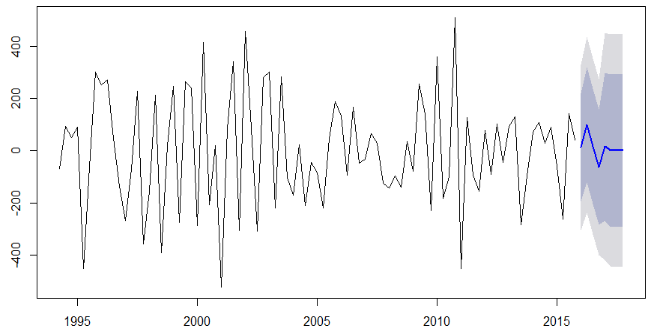


Figure 4.12: Forecasts from ARIMA(0,1,1)(0,0,1)

Figure 4.13 shows the goodness of fit for the chosen model $ARIMA(0,1,1)(0,0,1)_4$. The residuals are randomly around zero and the ACF show only one significance at lag 0 proofing that the residuals are not related. The Ljung box statistics have a null hypothesis of the model does not exhibit a lack of fit. There is no significance lags which at lag 0, 7 and 8 and the rest of the lags depicts are lack of fit as shown at the Ljung statistics.

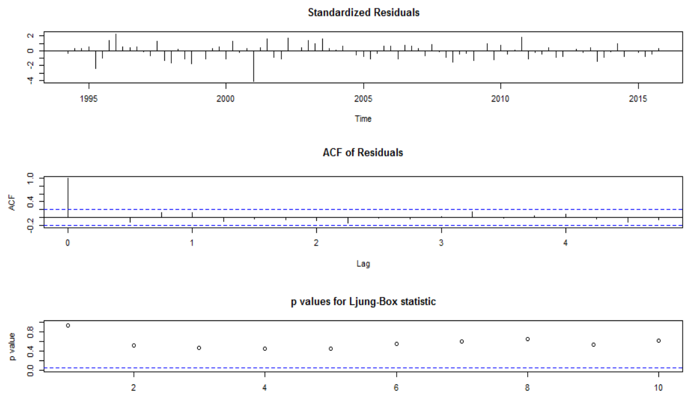


Figure 4.13: Diagnostics of Electricity Consumption ARIMA(0,1,1)(0,0,1)

4.3.1.2 Electricity Infrastructure

The second variable that is being investigated is the electricity infrastructure variable. Figure 4.14 shows a trend and shift in mean of the series. This shows that the series is not stationary and the stationarity status of electricity infrastructure needs to be investigated. Note how variability starts off almost constant and increases over time.

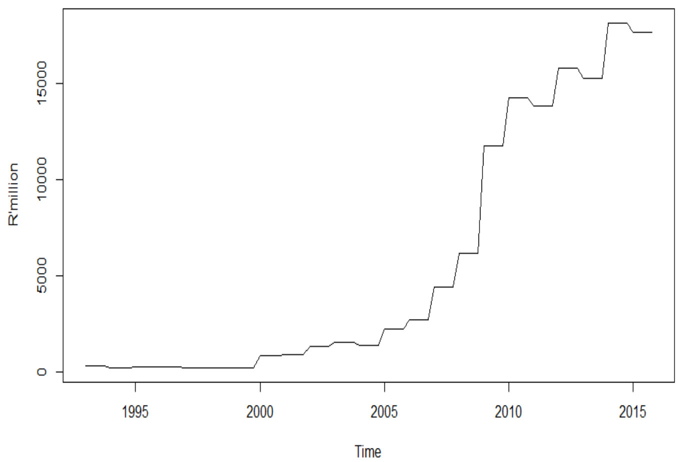


Figure 4.14: Electricity infrastructure(Capital infrastructure)

Figure 4.15 shows that the ACF has a slow decaying trend showing that Electricity infrastructure is not stationary.

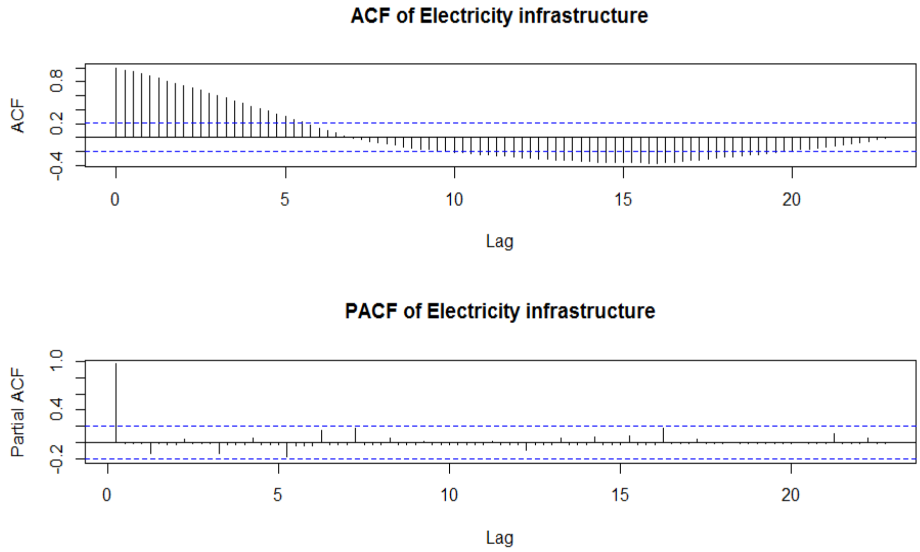


Figure 4.15: The ACF and PACF of electricity infrastructure

The variability of the series is investigated to see whether the stationarity status will change. Figure 4.15 shows that the variability has stabilise after taking logs not change but the series is still not stationary. The next step is to stabilise the mean by differencing the series in order to reach stationarity.

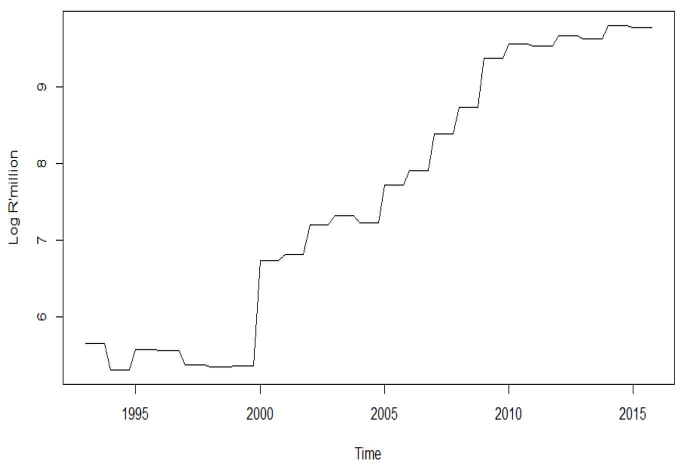


Figure 4.16: Time series of log(electricity infrastructure)

Note that Figure 4.16 has made the seasonal fluctuation more regular over time The series is still not stationary.

Note that Figure 4.17 shows the ACF and PACF of the logged electricity infrastructure(EI). It represent a $ARIMA(1,0,0)$ but this observation is not trustworthy since the series is not stationary.

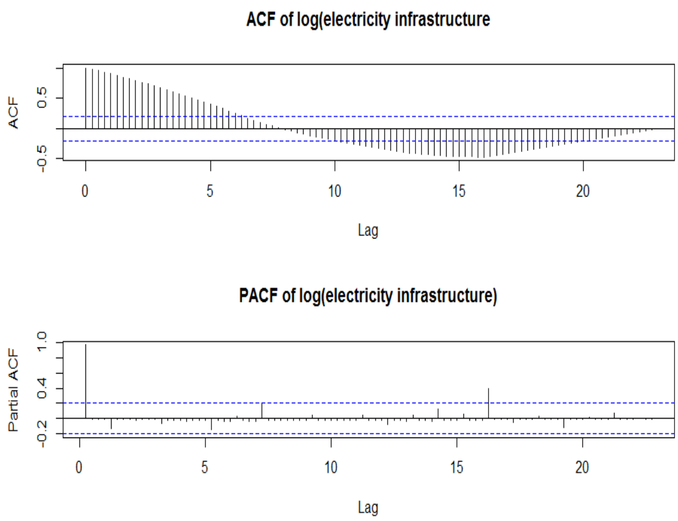


Figure 4.17: The ACF and PACF of log(electricity infrastructure)

Applying seasonal differencing on the logged series to see if there is any change in the stationarity status.
Figure 4.18 shows a constant mean and stationarity status is reached.

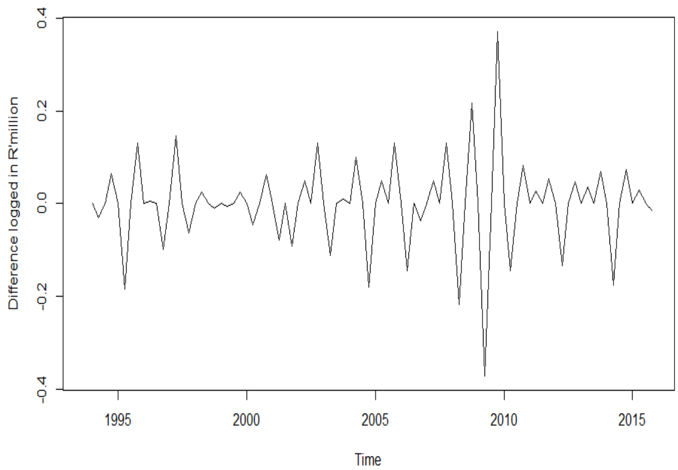


Figure 4.18: Differenced logged electricity infrastructure

The ACF cut off after lag 2 and the PACF shows a slow decaying trend in Figure 4.19. The auto.arima() (Hyndman, 2008) estimated an $SARIMA(2,1,2)(P,D,Q)$. The moving averaged by two period, differenced logged series has reach constant mean and variability.

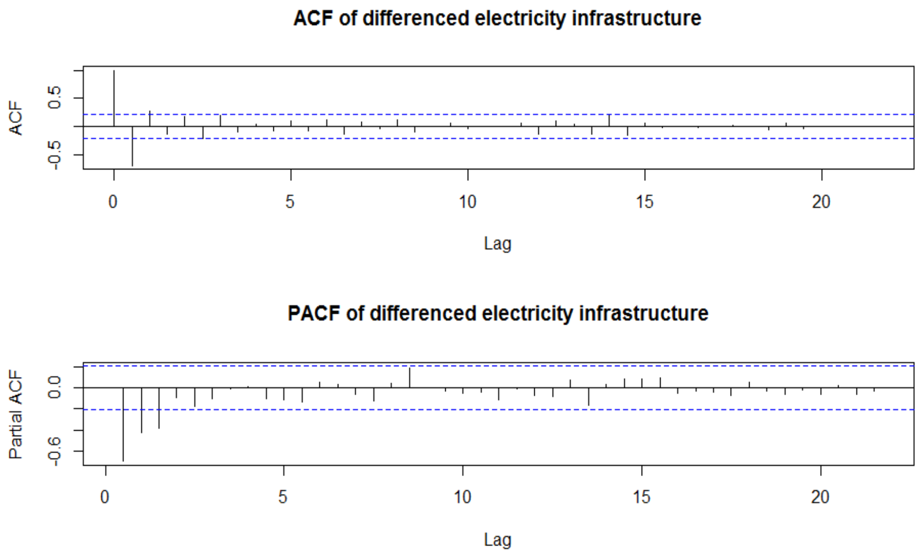


Figure 4.19: The ACF and PACF of differenced logged electricity infrastructure

4.3.1.2.1 Estimating the Model Fit, Estimates and Forecasting - Electricity Infrastructure

This section will present the different possible models that the electricity infrastructure time series data follow, present the graphical and table format forecast of the next three quarters and the optimal model for electricity infrastructure series. The possible model for this series are shown in Table 4.7.

Note that Table 4.7 show all the possible model that were studied for a possible description model of the infrastructure series. The optimal model is the model that present the smallest Akaike Information criteria which is model $SARIMA(2,1,2)(0,1,0)_4$ with criteria -254.05.

Table 4.7: Possible Models - Infrastructure

	Description	**AIC**
1	SARIMA(2,1,2)(0,1,0)	-254.05
2	ARIMA(0,0,0)(0,1,1)	-162.63

Note Table 4.8 show optimal model with AIC of -254.05. The significant parameters are all of the above because they are smaller than alpha 0.05 and the null hypothesis is rejected. The null hypothesis is $\beta_0 = \beta_1 = \beta_r = 0$.

Table 4.8: SARIMA(2,1,2)(0,1,0) model

	Value	**Standard Error**	**z value**	**p value**
φ_1	0	0.1064	0	1
φ_2	-0.4132	0.1058	-3.905	9.4038e-05
θ_1	0	0.0823	0	1
θ_2	-0.8113	0.0835	-9.716	2.57283e-22
AIC	-254.05			

The graphical forecast is shown in Figure 4.23. It shows the prediction of next three values.

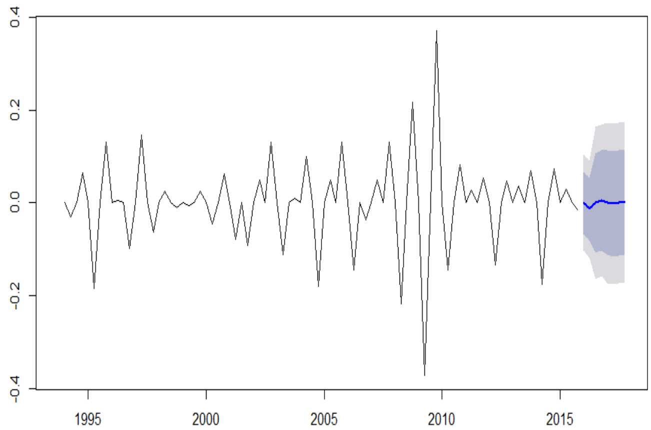


Figure 4.20: Forecasts from ARIMA(2,1,2)

The Holt winter (Hyndman, 2008) predicted the following estimates for the model, depicted in Table 4.9.

Table 4.9: HoltWinters Estimates - Electricity Infrastructure

	α	β	γ	**a**	**b**	Θ_1	Θ_2	Θ_3	Θ_4
Value	0.88	0.053	0.957	17932.507	274.918	685.726	357.749	-41.556	-263.941

Note that Table 4.10 show the prediction values of consumption through using methods such as ARIMA and Holt-Winter for the next three quarter. Comparing the two methods prediction with the actual values and both methods underestimate the infrastructure of the next quarters. The average difference between actual and ARIMA predictions is 1158.74 whereas the average difference between actual values and Holt-Winters is

1011.68 which is smaller than the average difference of the ARIMA prediction. Holt-Winters method has shown to be the method that predicts the estimates which are closer to the actual values.

Table 4.10: Predictors for the next three quarters in year 2016 - Infrastructure

	Quarter 1	**Quarter 2**	**Quarter 3**
HoltWinters	18893.152	18840.093	18715.707
ARIMA	18669.264	18669.264	18669.264
Actual	19828	19828	19828

Figure 4.21 shows the goodness of fit for the chosen model $SARIMA(2,1,2)(0,0,0)_4$. The residuals are randomly around zero and the ACF show only one significance at lag 0 proofing that the residuals are not related. The Ljung box statistics have a null hypothesis of the model that does not exhibit a lack of fit. All the lag are significant since all p values are greater than the 0.05 therefore the model does fit the data.

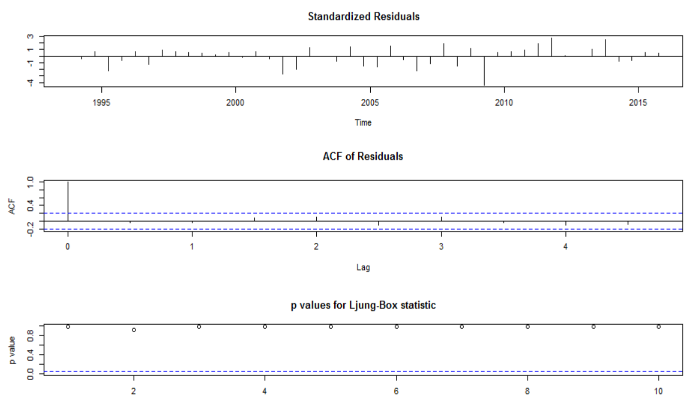


Figure 4.21: Diagnostics of Electricity Infrastructure - ARIMA(2,1,2)

4.3.1.3 Economic Growth (GDP)

The third variable to analyse is the quarterly GDP (economic growth) variable. There is an increasing trend shown in Figure 4.22 which means that the series is not stationary. There is also evidencce of increasing variability over time.

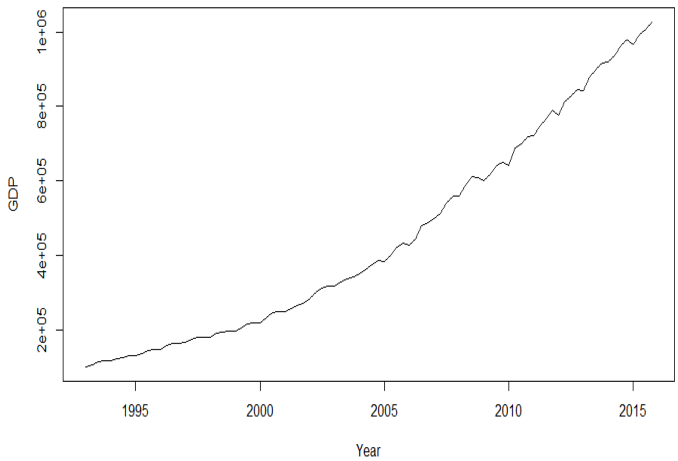


Figure 4.22: Time Series of Quarterly Economic Growth

Figure 4.23 shows evidence of the series not being stationary as the ACF has a slow decay trend.

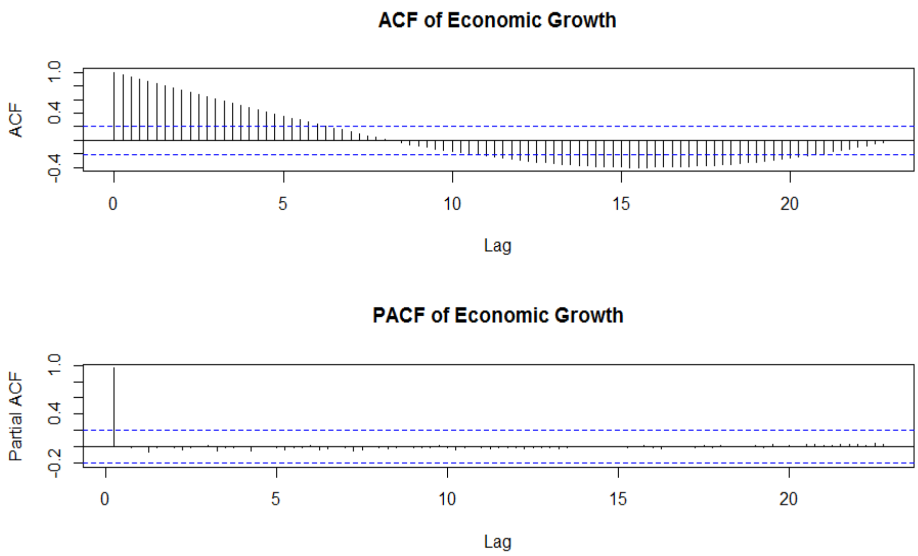


Figure 4.23: The ACF and PACF of Economic Growth

The sereis is logged to stabilise the variability but the series is still not stationary in the mean as shown in Figure 4.24 and the variability of the series seems to increase over time.

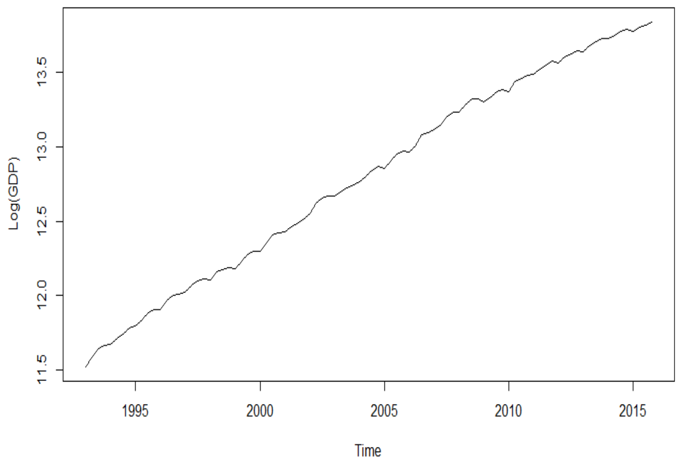


Figure 4.24: Time series of Log Economic Growth

Another way to reach stationarity in the mean is to difference the series. The ACF and PACF suggest a ARIMA(1,0,0) model in Figure 4.25.

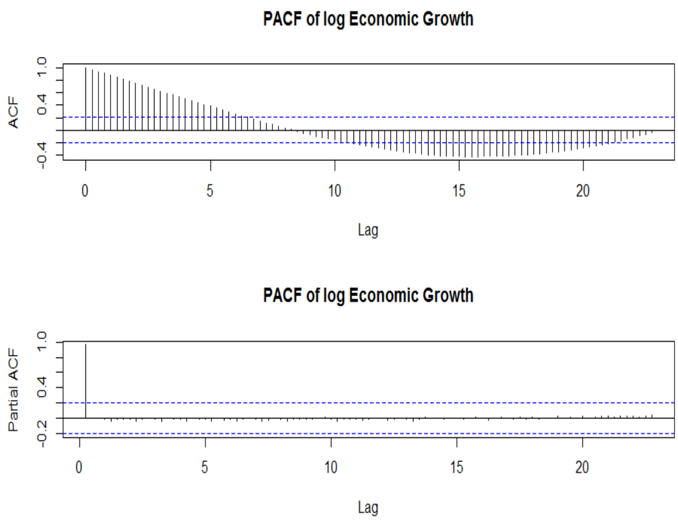


Figure 4.25: The ACF and PACF of log Economic Growth

Note that Figure 4.26 show the differenced logged economics series which has constant mean and variabilty overtime. The Figure 4.26 presenting a stationary time series.

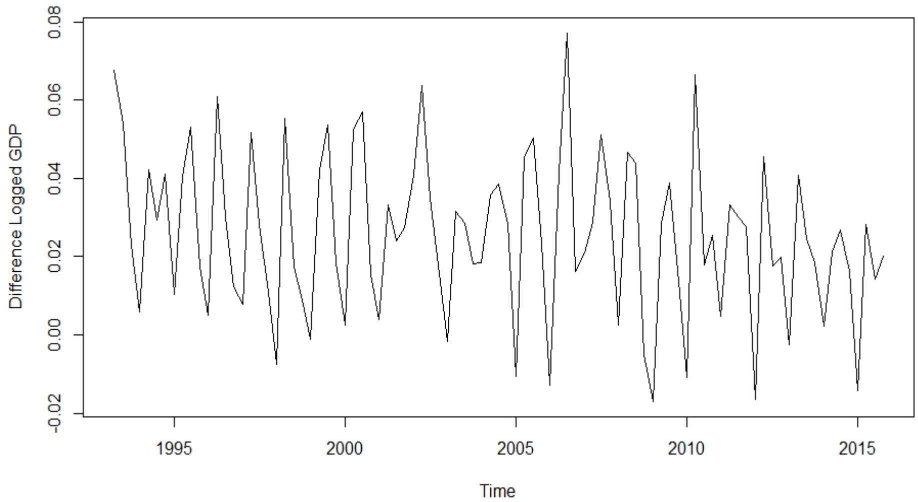


Figure 4.26: Differenced Logged Economic Growth

Note that Figure 4.27 shows the ACF and PACF of the differenced logged series and suggest a $ARIMA(0,1,0)(0,0,1)$ model. The model shows a strong seasonal pattern but the series is still stationary. There periodical changes within the series due to certain economic events.

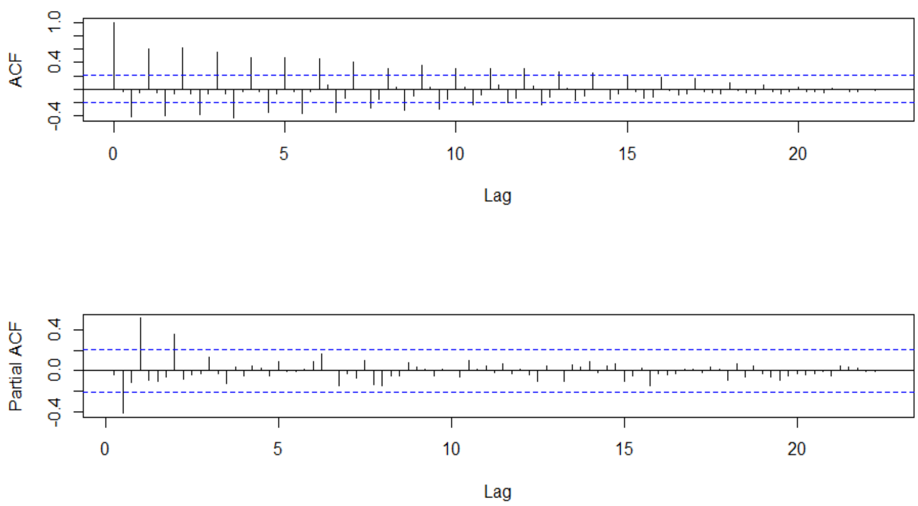


Figure 4.27: The ACF and PACF of differenced Logged Economic Growth

Note that the ACF show alternative decaying trend and PACF cut off at lag 2.

4.3.1.3.1 Estimating the Model Fit, Estimates and Forecasting -Economic Growth

This section will present the optimal model that the economic growth time series data follow, present the graphical and table format of the predictives capabilities of the next three quarters and the optimal model for economic growth series. The possible model for this series are shown in Table 4.11.

Table 4.11: $ARIMA(0,1,0)(0,0,1)_4$ model

	Value	**Standard Error**	**z value**	**p value**
$\Theta\theta_1$	-0.8078	0.0871	-9.274	1.786261e-20
AIC	-493.25			

Note that it is found $ARIMA(0,1,0)(0,0,1)_4$ to be the optimal model shown Table 4.11. The final model is $y_t = -0.8078\Theta\theta_1$ because all of these coefficients are significant base on the p values being smaller than significant level 0.05.

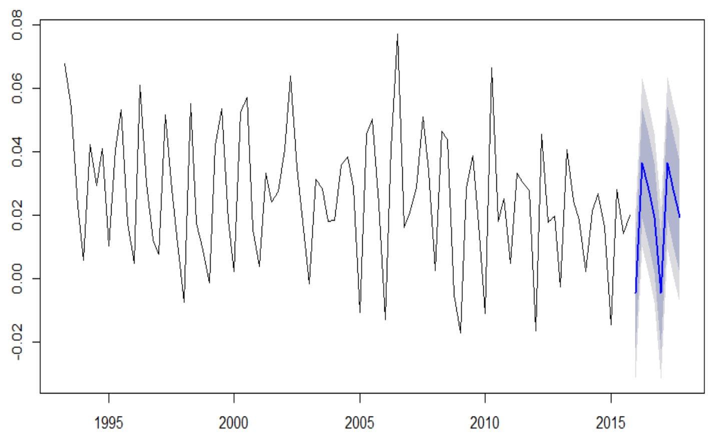


Figure 4.28: Forecasts from ARIMA(0,0,0)(0,1,1)

Figure 4.28 shows the forecast of few quarters of GDP from $ARIMA(0,0,0)(0,1,1)_4$ model. Note that the graphical prediction follows the past trend.

$ARIMA(1,0,0)(1,0,0)$ is another possible model for the economic series shown in Table 4.12.

Table 4.12: $ARIMA(1,0,0)(1,0,0)$ model

	Value	**Standard Error**	**z value**	**p value**
φ_1	-0.4547	0.0932	-4.879	1.0675e-06
$\Theta\varphi_2$	0.7092	0.0725	9.782	1.344339e-22
AIC	-440.81			

Model $ARIMA(1,0,0)(1,0,0)$ have a significant model of ,

$$y_t = -0.4547\varphi_1 + 0.7092\Theta\varphi_1 \tag{4.1}$$

since all these p values are smaller than alpha 0.05 percent. The null hypothesis is that all $\varphi_1 = \varphi_2 = 0$.
The possible models that this series depicts is shown in Table 4.13 and the chosen model has an AIC criteria of -493.25.

Table 4.13: Possible Models - Economic growth

	Description	**Model**	**AIC**
1	ARIMA(0,1,0)(0,0,1)	$y_t = -0.8078\Theta\theta_1$	-493.25
2	ARIMA(1,0,0)(1,0,0)	$y_t = -0.4547\varphi_1 + 0.7092\Theta\varphi_1$	-440.81

The Holt winter (Hyndman, 2008) predicted the following estimates for the model, depicted in Table 4.13.

Table 4.14: HoltWinters Estimates - Economic growth

	α	β	γ	**a**	**b**	Θ_1	Θ_2	Θ_3	Θ_4
Value	0.744	0.165	0.964	1012996	14096	-9060	5262	11164	-13982

Note that Table 4.15 show the prediction values of consumption through using methods such as ARIMA and Holt-Winter for the next three quarter. Comparing the two methods prediction with the actual values and both methods underestimate the economic growth of the next quarters. The average difference between actual and ARIMA predictions is -21849.11 whereas the average difference between actual and Holt-Winters is -23590.73 which is greater than the average difference of the ARIMA prediction. ARIMA method has shown to be the method that predicts the estimates closer to the actuals.

Table 4.15: The Predictors for the next three quarters in year 2016 - Economic growth

	Quarter 1	**Quarter 2**	**Quarter 3**
HoltWinters	1018031.586	1046450.142	1066448.382
ARIMA	1020273.924	1047691.046	1066640.465
Actual	1043132	1068531.1801	N/A

Figure 4.29 shows the goodness of fit for the chosen model $ARIMA(0,1,0)(0,0,1)_4$. The residuals are randomly around zero and the ACF show only one significance at lag 0 proofing that the residuals are not related. The Ljung box statistics have a null hypothesis of the model does not exhibit a lack of fit. All the lag are significant since all p values are greater than the 0.05 therefore the model does fit the data.

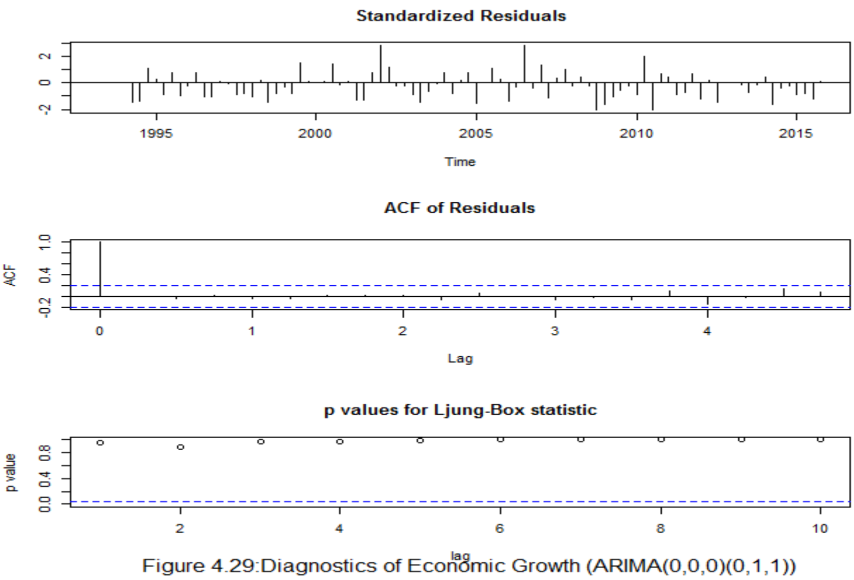


Figure 4.29:Diagnostics of Economic Growth (ARIMA(0,0,0)(0,1,1))

4.3.2 Summary: Univariate Time Series

We have reach stationarity in all variable and the optimal models for GDP, Electricity infrastructure and Electricity consumption are $ARIMA(0,1,0)(0,0,1)_4$, $SARIMA(2,1,2)(0,0,0)_4$ and $ARIMA(0,1,1)(0,0,1)_4$ respectively. The predictive capabilities are quite good and each respective models fits the data well.

Note that there were other possible models for each variable that were derived. These are shown in the Appendix A, A3 as Figure 4.13.1.A, 4.13.2.B and 4.13.3.C for consumption; Figure 4.21.1.A for infrastructure and Figure 4.29.1.A for economic growth. Figure 4.13.2.B and Figure 4.13.3.C show that the model do not fit the data. The second model on infrastructure series do not show a good fit as well as the second model on economic growth do not show a good fit.

4.4 Multivariate Time Series

The next section will discuss the results on all the three variables. A similar thought process to the one used in Section 4.3 will be applied, where one will determine stationarity of the data. When the series is not stationary, differencing or transforming

the data is done. Thereafter, the model is determined and forecasting will be applied to predict the values.
Figure 4.30 illustrate the three variables trend over a period of time.

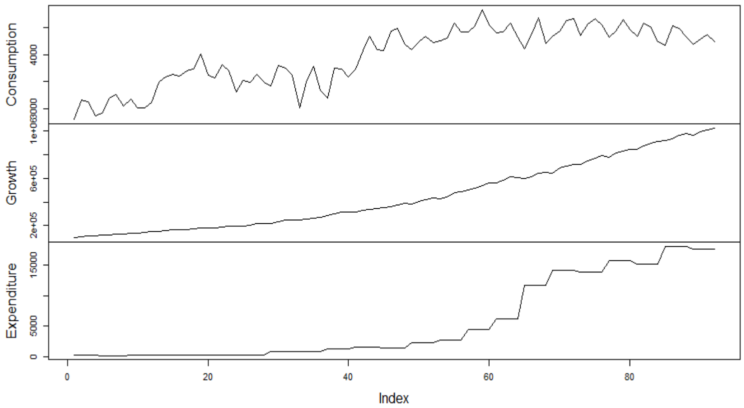


Figure 4.30: Time Series of Electricity Consumption, Electricity Infrastructure and Economic Growth

Figure 4.30 illustrate the three variables trend over a period of time. The stationarity check was done in the univariate tests and stationarity was reach when all three series were differenced. Figure 4.31 is a plot of the three variables from using the *zoo* package (Zeileis and Grothendieck, 2005). The series is plotted against the time index observations.

Figure 4.31 is a plot of the three variables from using the *mvtsplot* package (Peng, 2008). The series is plotted against the vector time series. Showing the three time series over time the of 94 quarters.

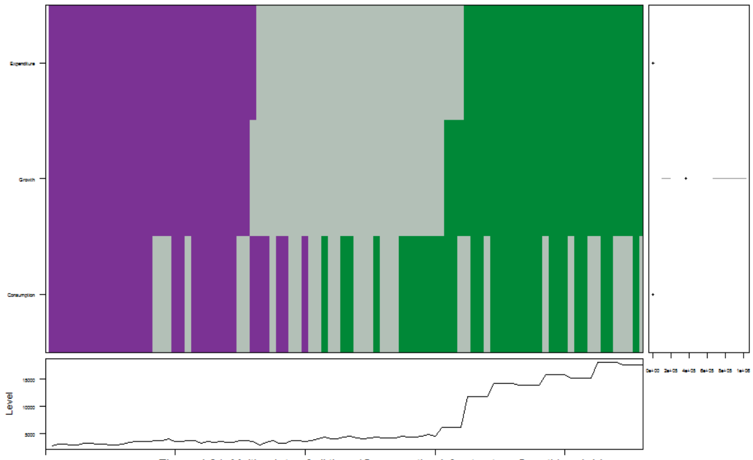


Figure 4.31: Multivariate of all three(Consumption,Infrastructure,Growth) variables

The purple colour represents the low values, grey colour middle values and green the high values. The Figure 4.31 shows where all the values of the three times series lie if they categorized. Economic growth and electricity infrastructure categories do not overlap with each other over time and the consumption categories do overlap over time, showing that consumption does change over time.

The stationarity status was reached when the three series were differenced and transformed respectively. Applying the vector autoregressive function will assist in the study of time series model which will be presented later. The *Varselect* (Pfaff, 2008) function in R will give the most optimal lag in the different information criteria.

Table 4.16 show that the optimal lag for a VAR model will be 9 using the AIC and 5 for both Hannan-Quin and final prediction error criteria. The optimal lag is the lag at which each criteria illustrate the smallest information criteria.

Table 4.16: The optimal Lag

AIC(n)	HQ(n)	SC(n)	FPE(n)
9	5	1	5

The Akaike information criteria says that the optimal lag is 9 whereas Hannan-Quin and Final Prediction Error say is 5. The Bayesian Schwarz criteria shows the smallest criteria at lag 1.
Note that Table 4.17 shows the detailed minimum values of each information criteria which is shown in lag 9 for AIC criteria.

Table 4.17: Information Criteria Values - Test for Optimal Lag

Criteria	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
AIC(n)	42.344	42.203	41.941	41.948	41.675	41.743	41.907	41.766	41.663
HQ(n)	42.485	42.449	42.293	42.404	42.237	42.411	42.68	42.644	42.646
SC(n)	42.694	42.815	42.816	43.084	43.074	43.404	43.831	43.952	44.111

An investigate whether the residuals are uncorrelated using the Portmanteau test was done for the null hypothesis of no serial correlation in residuals is not rejected in $VAR(5)$ and rejected in $VAR(9)$. Since serial correlation is not a desired trait $VAR(5)$ model estimators will have the following properties such as unbiasedness and efficiency when there is no autocorrelation. There is a fine line between bias error and the power of the test when it comes to the lenght of the lag. When the lag length is too small, the remaining serial correlation will bias the test and when the lag is too large then the power of the test will suffer.
Note that model 1: $\widehat{EC}_t = -0.8395EC_{t-1} + 0.8302EI_{t-1} + 0.02821GDP_{t-2} - 0.03345GDP_{t-4}$ with lag 5 does not reject the null hypothesis since the p value, 0.3166 is greater than significance level $alpha$ 0.05 and model 2 $\widehat{EC}_t = 0.6216EC_{t-1} -$

Table 4.18: Portmanteau (asymptotic) Test

	Chi-Squared	**df**	**p-value**
$EC_t = -0.8395EC_{t-1} + 0.8302EI_{t-1} + 0.02821GDP_{t-2} - 0.03345GDP_{t-4}$ with lag 5	105.18	99	0.3166
$EC_t = 0.6216EC_{t-1} - 0.009052GDP_{t-5} - 0.07709EI_{t-5}$with lag 9	83.335	63	0.04413

$0.009052GDP_{t-5} - 0.07709EI_{t-5}$with lag 9 will reject the null hypothesis of no serial correlation.

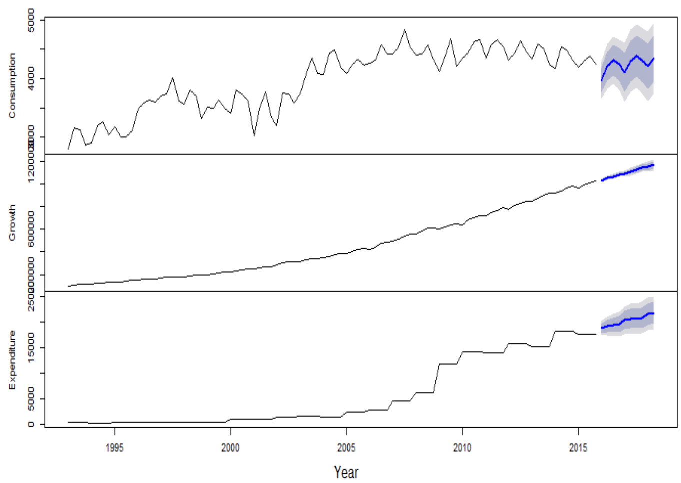


Figure 4.32: Forecasts from Model 1 with lag 5

Figure 4.32 display the forecasted values of the electricity consumption, economic growth and electricity infrastructure.

The model where consumption is the dependent variable is of interest and will be presented in the economic modelling section and the other two models are depicted in Table A.6 and Table A.7 in Appendix A (A.4).

4.5 Econometric Models

In this section the econometric model will be discussed. The relationship between the three variables causal effects on each other, the direction of this causal effect is also studied and will be investigated for. The model that were derived from the multivariate approach will now be tested. The ADF (Augmented Dickey Fuller) test will be used to test for stationarity. The null hypothesis of ADF is that the series is not stationary.

4.5.1 ADF for each of the Variables; Electricity Consumption, Electricity Infrastructure and Economic Growth

The stationarity status of the three variables electricity consumption, electricity infratructure and economic growth series are investigated using the Augmented Dickey Fuller test.

Table 4.19 shows the differenced consumption series is stationary because the p value 0.01 is smaller than significance level of 0.05, rejecting the null hypothesis of no stationary.

Table 4.19: Augmented Dickey-Fuller Test - Consumption

	Value
Critical Value	-5.3922
Lag order	4
p-value	0.01
Alternative Hypothesis	stationary

Table 4.20 shows the differenced economic growth series is stationary because of the p value 0.01 is smaller than significance level 0.05.

Table 4.20: Augmented Dickey-Fuller Test - Economic Growth

	Value
Critical Value	-3.7301
Lag order	4
p-value	0.02622
Alternative Hypothesis	stationary

Table 4.21 shows that the infrastructure series is stationary because of the p value 0.01 is smaller than significance level 0.05.

Table 4.21: Augmented Dickey-Fuller Test - Infrastructure

	Value
Critical Value	-6.2104
Lag order	4
p-value	0.01
Alternative Hypothesis	stationary

4.5.2 Cointegration Granger Engle-Test

We make use of the linear model function in R to determine the cointegration of the three variables. In the following paragraphs, cointegration are being tested through two step. Firstly the relationship is determined and then the residuals are investigated whether they constant overtime.

Since the relationship between consumption and economic growth is quadratic, the economic growth variable are squared. The appropriate model for consumption is

$\widehat{consumption} = 2559 + 0.0055\,GDP - 3.837e - 09\,GDP^2$ with a adjusted R-squared of 83 percent. In Table 4.22 the overall model is significant because the p-value is smaller than 0.05. The null hypothesis of there is no relationship between consumption and infrastructure is rejected. The residual plots in Figure 4.33 shows that the residuals are stationary. The latter proves that economic growth is cointegrated with consumption.

Table 4.22: Granger Engle Test - Consumption ˜ Economic growth

	Estimates	**Standard Error**	**t value**	**p-value(>|t|**
Intercept	2559	80.68	31.72	$<$ 2e-16 ***
GDP	0.005529	0.0003821	14.47	$<$ 2e-16 ***
GDP^2	-3.837e-09	3.549e-10	-10.81	$<$ 2e-16 ***
Adjusted R-squared	0.8309			
Overall Model P-value	$<$ 2e-16			
Signif. codes	0 '***'	0.001 '**'	0.01 '*'	0.05

The consumption vs economic growth variables are investigated for cointegration. Figure 4.33 shows the residuals are stationary therefore there exist a cointegration between the two variables. The ADF test illustrated that the cointegration was possible at $I(0)$ and $I(1)$.

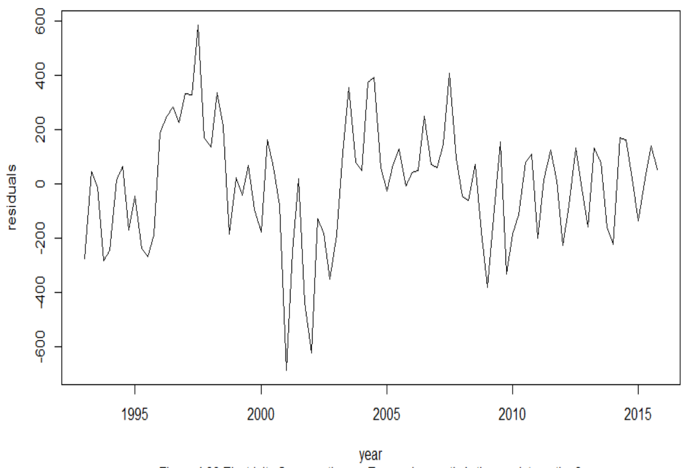


Figure 4.33 Electricity Consumption vs. Economic growth: Is there cointegration?

Table 4.23 shows that the overall model is significant since the p-value is smaller than 0.05. The null hypothesis of there is no relationship between consumption and infrastructure is rejected.

Table 4.23: Granger Engle Test - Consumption ~ Energy Infrastructure

	Estimates	**Standard Error**	**t value**	**p-value(>|t|**
Intercept	4095	76.95	53.217	< 2e-16 ***
Electricity infrastructure	0.02345	0.006757	3.47	0.000804 ***
1/Electricity infrastructure	167000	24070	-6.938	6.15e-10 ***
Adjusted R-squared	0.6288			
Overall Model P-value	< 2e-16			
Signif. codes	0 '***'	0.001 '**'	0.01 '*'	0.05

Figure 4.34 is used to investigate if there is cointegration between consumption and electricity infrastructure. It shows the relationship residuals are not stationary therefore, according the Granger Engle test the two variables are not cointegrated.

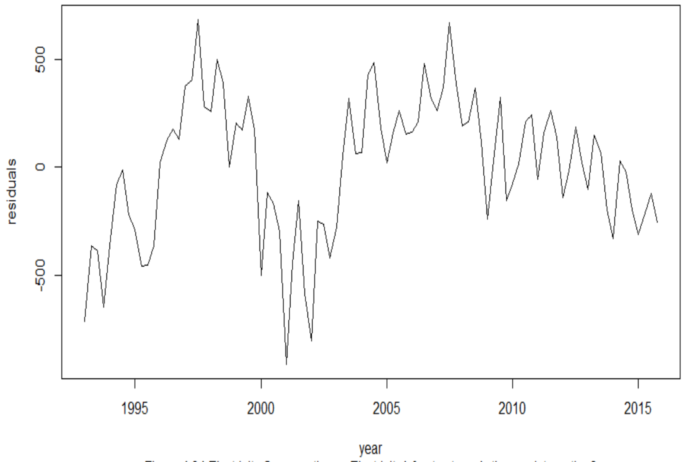


Figure 4.34 Electricity Consumption vs.Electricity Infrastructure : Is there cointegration?

4.5.3 Cointegration: Johansen Test

The Johansen test is another cointegration test, the next function will illustrate it. The output that Johansen procedure investigate 3 types of hypothesis, one to test whether the needs r = 0,1,2 unit root to be cointegrated with each other.
Johansen procedure shows that the series is cointegrated in r = 0 since the test statistic of 223.91 exceeds the 1 percent significance level. This is also true for r<= 1 and r<=2 since the test statistics is greater than the 1 percent significance level. Cointegration exist between the series.

Table 4.24: Johansen-Procedure

	test	**10pct**	**5pct**	**1pct**
r <= 2	55.27	6.5	8.18	11.65
r <= 1	115.23	15.66	17.95	23.52
r = 0	223.91	28.71	31.52	37.22

4.6 Granger Causality Test

The Granger Causality effect tells us whether there is a causal effect between the two variables. Using the package lmtest (Zeileis and Hothorn, 2002);Note that EC = Electricity Consumption, GDP is Economic growth GDP and EI is Electricity infrastructure. The output shows the two models; Model 1 the unrestricted model that includes the Granger-causal terms, Model 2 the restricted model where the Granger-causal terms are omitted. The Wald test that assesses whether the restricted Model 2 is the same as unrestricted Model 1.

In order to proceed with the Granger test, the series need to be stationary or cointegrated which was proven by the previous paragraphs.
The null hypothesis states that consumption does not Granger cause economic growth. Table 4.25 analyses the causality relationship of electricity consumption and GDP. The significance level chosen is 5 percent. The null hypothesis will be rejected since the p value is less than 0.05. Therefore there exists a Granger causal relationship between electricity consumption and economic growth (GDP).

Table 4.25: Granger causality test - Electricity consumption ˜ GDP

	Res.Df	**Df**	**F**	**Pr(>F)**
Model 1: EC ˜ Lags(EC, 1:1) + GDP, 1:1)	88			
Model 2: EC ˜ EC, 1:1)	89	-1	4.1228	0.04533 *

Table 4.26 analyse whether GDP Granger causality Electricity consumption. The null hypothesis is that economic growth does not granger cause consumption and the p value is greater than alpha thus we will not reject the null hypothesis. The latter is evident that no granger causal relationship exists between economic growth and electricity consumption.

Table 4.26: Granger causality test - GDP ˜ Electricity consumption

	Res.Df	Df	F	Pr(>F)
Model 1: GDP ˜ Lags(GDP, 1:1) + Lags(EC, 1:1)	88			
Model 2: GDP ˜ Lags(GDP, 1:1)	89	-1	0.8911	0.3478

Table 4.27 analyses whether there esixts a granger causal relationship between GDP and electricity infrastructure. The p-value is greater than significance level 0.05 thus we will not reject the null hypothesis. Therefore no Granger causal relationhip exists between GDP and electricity infrastructure.

Table 4.27: Granger causality test - GDP ˜ Electricity infrastructure

	Res.Df	Df	F	Pr(>F)
Model 1: GDP ˜ Lags(GDP, 1:1) + Lags(EI, 1:1)	88			
Model 2: GDP ˜ Lags(GDP, 1:1)	89	-1	0.542	0.4636

Table 4.28 analyses whether there is a relationship between electricity infrastructure and GDP. The p-value is smaller than alpha thus we will reject the null hypothesis. Thus there exists a granger causal relationship between electricity infrastructure and economic growth.

Table 4.28: Granger causality test - Electricity Infrastructure ˜ GDP

	Res.Df	Df	F	Pr(>F)
Model 1: EI ˜ Lags(EI, 1:1) + Lags(GDP, 1:1)	88			
Model 2: EI ˜ Lags(EI, 1:1)	89	-1	8.5413	0.004412 **

Table 4.29 analyses whether there exists a Granger causality electricity consumption and electricity infrastructure. The p-value is greater than alpha thus we will not reject the null hypothesis. Electricity consumption does not have a causal relationhip with electricity infrastructure.

Table 4.29: Granger causality test - Electricity Consumption ˜ Electricity Infrastructure

	Res.Df	**Df**	**F**	**Pr(>F)**
Model 1: EC ˜ Lags(EC, 1:1) + Lags(EI, 1:1)	88			
Model 2: EC ˜ Lags(EC, 1:1)	89	-1	3.0435	0.08455

Table 4.30 analyses whether there exists a Granger causal relationship between electricity infrastructure and electricity consumption. The p-value is greater than significance level thus we will not reject the null hypothesis. Electricity infrastructure does not have a granger causal relationship with electricity consumption.

Table 4.30: Granger causality test - Electricity Infrastrucutre ˜ Electricity Consumption

	Res.Df	**Df**	**F**	**Pr(>F)**
Model 1: EI ˜ Lags(EI, 1:1) + Lags(EC, 1:1)	88			
Model 2: EI ˜ Lags(EI, 1:1)	89	-1	1.545	0.2172

According to Jinke, Hualing and Dianming (2008) when one tests whether electricity consumption causes economic growth then we would achieve a unidirectional growth hypothesis. According to Table 4.25 there exists a causal effect running from electricity consumption to economic growth. Table 4.29 shows that there also exists a causal effect running from electricity consumption to electricity infrastructure. It can be concluded that there exists a growth hypothesis in South Africa.

The optimal econometric model was derived in the multivariate Section 4.4. The model is shown in Table 4.31, it only shows the significant coefficients. The consumption lag 4 coefficient is significant on significant level of 0.1 and the rest are significant base on significance level of 0.05. The models where the growth and electricity infrastructure variable are the dependent variable will be shown in the Appendix A (A.4), Table A.5. The significant variables when electricity consumption is the dependent variable are electricity consumption in lag 1,3,4 and 5 and growth in lag 4 and 5 since the p-value is smaller than the significant value alpha. The final model is as follow

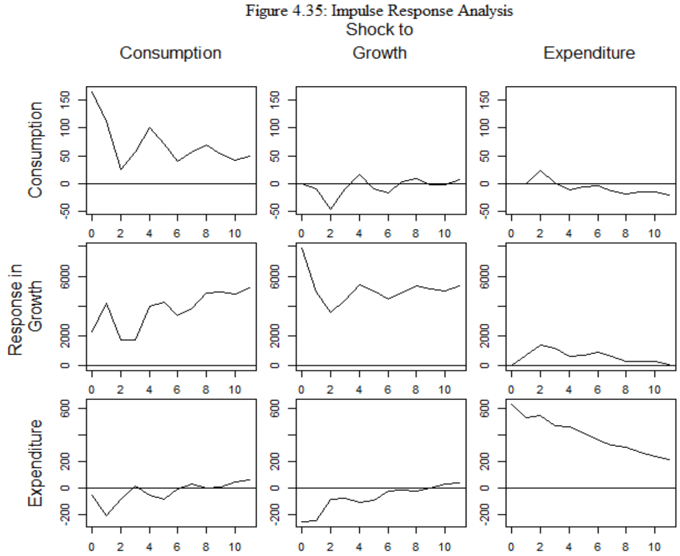
$\widehat{EC} = 0.647817EC_{t-1} + 0.261885EC_{t-3} + 0.008378GDP_{t-4} + 0.221899EC_{t-4} - 0.250573EC_{t-5} - 0.008329GDP_{t-5}$.

Table 4.31: Optimal model of Econometric model

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
$Consumption_1$	0.647817	0.110498	5.863	1.32e-07 ***
$Consumption_3$	0.261885	0.123830	2.115	0.03795 *
$Growth_4$	0.008378	0.003076	2.723	0.00812 **
$Consumption_4$	0.221899	0.126812	1.750	0.08447 .
$Consumption_5$	-0.250573	0.111585	-2.246	0.02784 *
$Growth_5$	-0.008329	0.002796	-2.979	0.00396 **
Adjusted R-squared	0.8911			
AIC	1149.379			

The impulse response analysis is an exogenous shock to one variable not only directly affects this variable but is also transmitted to all of the other endogenous variables through the dynamic (lag) structure of the VAR. An impulse response function traces the effect of a one standard shock to one of the innovations on current and future values of the endogenous variables. In order to interpret the model over time we need to investigate the impulse responses (Brandt and Appleby, 2007) of each variable overtime. The Figure 4.35 shows each response shock experience by the variables. The shock of consumption with itself is positive and has a declining trend overtime. Note that after quarter 4 it started to diminish. The shock effect between consumption and GDP flunctuate are around zero. South Africa growth is dependent on consumption and it appear to slighty affect the growth depending specific times. It appears that growth has a alternate shock effect with consumption and quarter 4 and 8 have a positive shock effect with consumption. The electricity infrastructure starts off with a positive shock effect and decline from quarter 3 and onwards.

Figure 4.35: Impulse Response Analysis



4.7 Regression Analysis

The overall aim for analysing time series data is to investigate the time effect on a study. This study also looks at the effect of economic growth and electricity infrastructure. In this section regression analysis is used to model the nature of the relationship between the variables electricity consumption, GDP and electricity infrastructure.

A scatterplot of the relationship between electricity consumption and economic growth is shown in Figure 4.36. The relationship is not linear and suggest a quadratic model may best describe the relationship or a change of regime model.

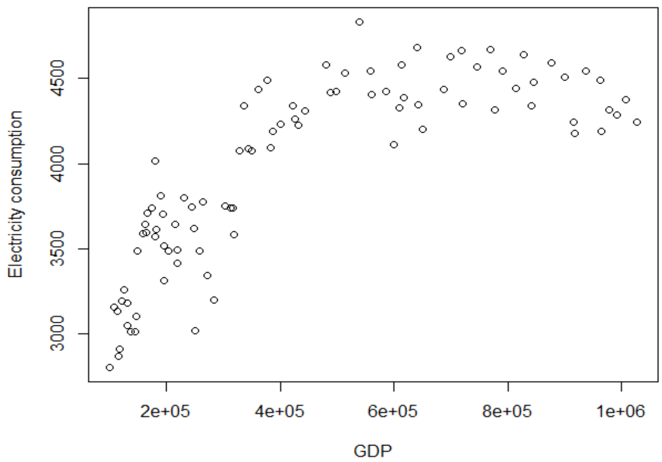


Figure 4.36: Electricity Consumption vs growth

Figure 4.37 shows a scatterplot analysis of electricity consumption versus electricity infrastructure which suggest a hyperbolic model.

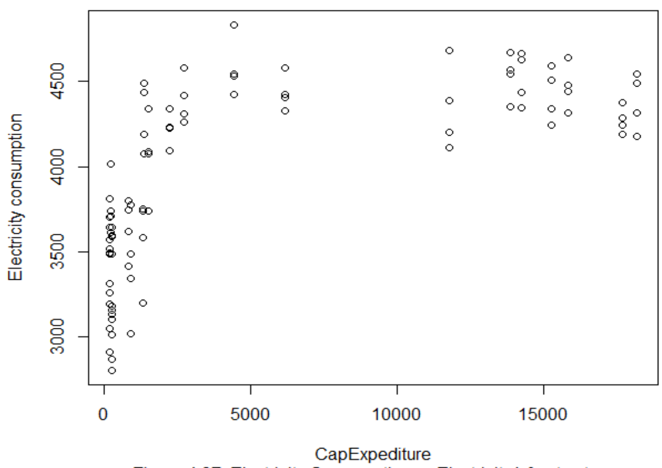


Figure 4.37: Electricity Consumption vs Electricity Infrastructure

The model results are shown in Table 4.32. The significant variables are the intercept and economic growth, that is, GDP is significant at 0.05 significant level.
We however noted that the model has an adjusted R-squared of 39 percent, meaning the model does not fit the data very well.

Table 4.32: Regression Analysis of Model 1 - Electricity Consumption ˜ GDP + Electricity Infrastructure

	Estimates	**Standard Error**	**t value**	**p-value(>|t|**
Intercept	3014.882	187.4073	16.087323	4.437728e-28
GDP	0.002923763	0.0007741674	3.776655	0.0002865813
Electricity Infrastructure	-0.05537457	0.03291619	-1.68229	0.0960182
Adjusted R-squared	0.3921537			

We now explore nonlinear models, the quadratic and hyperbolic model. We first transform the variables economic growth by squaring and electricity infrastructure by applying multiplicative inverse. Thereafter a model selection is executed and the optimal model with a minimised Akiake Criterion Information is chosen. Two models are fitted that is, Model 2 is

$$EC = \beta_0 + \beta_1 GDP + \beta_2 GDP^2 + \beta_3 EI + e \tag{4.2}$$

in Table 4.32 and Model 3

$$EC = \beta_0 + \beta_1 GDP + \beta_2 GDP^2 + \beta_3 EI + \beta_4 EI^{-1} + e \tag{4.3}$$

shown in Table 4.33.

Model 2 in Table 4.33 has an AIC of 991.35 with a very high adjusted R-squared showing that the model fits the data well explaining most of the electricity consumption.

Model 3 in Table 4.34 has an AIC of 978.86 and the overall Adjusted R-squared 87 percent indicating that the model fits the data very well.

The model assumptions were analysed in Figure 4.38, Table 4.35 and 4.36.
Figure 4.38 shows that the residuals are normally distributed as seen in the Normal

Table 4.33: Model 2 - Electricity Consumption ~ GDP + I(GDP^2) + Electricity Infrastructure

	Estimates	**Standard Error**	**t value**	**p-value(>|t|**
Intercept	2501	81.97	30.506	<2e-16***
GDP	0.005773	0.000386	14.954	<2e-16***
GDP^2	-3.312e-09	3.917e-10	-8.455	5.42e-13 ***
Electricity Infrastructure	-0.03344	0.01353	-2.472	0.0154 *
Adjusted R-squared	0.8451			
AIC	991.35			

Table 4.34: Model 3 - Electricity Consumption ~ GDP + I(GDP^2) + Electricity Infrastructure + I($ElectricityInfrastructure^{-}1$)

	Estimates	**Standard Error**	**t value**	**p-value(>|t|**
Intercept	1837	188.4	9.751	1.30e-15 ***
GDP	0.007783	0.0006333	12.29	<2e-16 ***
GDP^2	-4.484e-09	4.744e-10	-9.451	5.36e-15 ***
Electricity Infrastructure	-0.04542	0.01295	-3.506	0.000721 ***
Electricity Infrastructure^-1	100400	26060	3.853	0.000223 ***
Adjusted R-squared	0.8662			
AIC	978.86			

Q-Q plot and have a constant variance as seen in the residuals vs fitted plot. The influential datapoints are 33, 19 and 37 as shown in the Residuals vs Leverage plot in Figure 4.38.

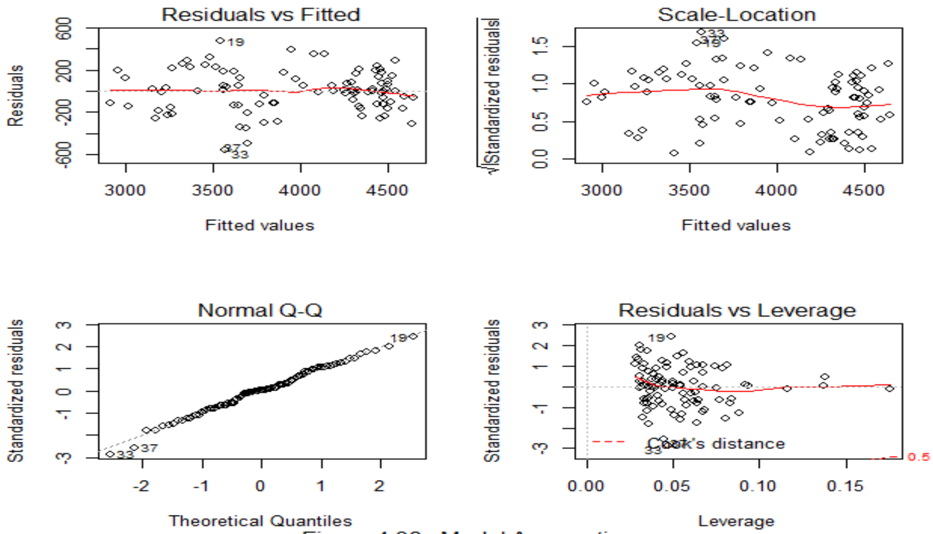


Figure 4.38 : Model Assumptions

The outlier that have a large negative residual is observation 33 as seen on Figure 4.38. This outlier do not affect the regression slope coefficient since the slope is not pull towards the datapoint, see Figure 4.39.

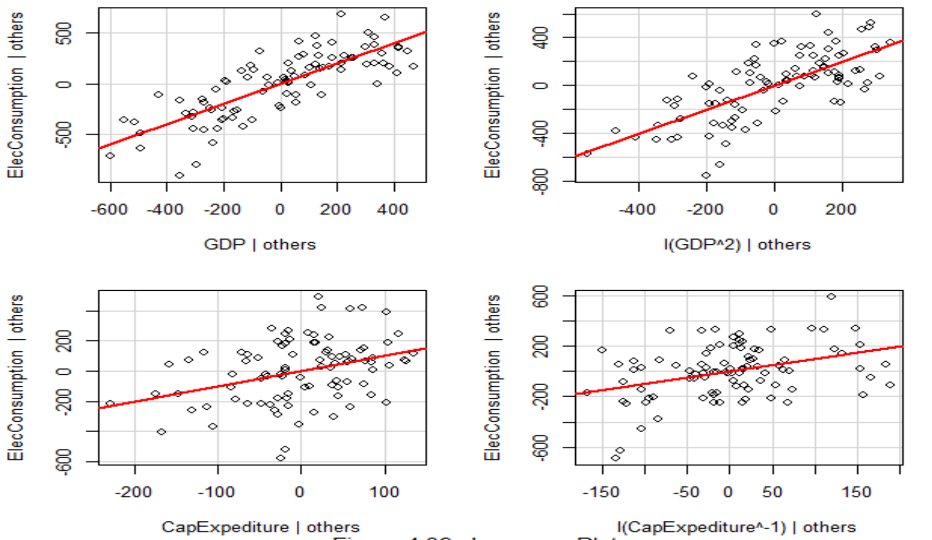


Figure 4.39 : Leverage Plot

The Breusch Pagan test in Table 4.35 shows that the variance can be assumed to be constant as the null hypothesis is not rejected since p value is greater than 0.05. Although

Model 3 is a better model in terms of the AIC and R^2, we support the justification of the extra term EI^{-1} as the increases by 3 percent. We ideally would be happier with a parsimonious model. Note that the Breush-Pagan test measures how errors increase across the explanatory variable. The test assumes the error variances are due to a linear function of one or more explanatory variables in the model.

Table 4.35: Breusch-Pagan

	Test statistic	**Degree of freedom**	**p-value**
Values	6.3067	4	0.1774

The Shapiro Wilk test in Table 4.36 shows that the residuals are normally distributed as the null hypothesis is not rejected since p value is greater than significance value 0.05.

Table 4.36: Shapiro-Wilk

	Test statistic	**p-value**
Values	0.99225	0.9356

Note that in Table 4.37 show that there exist multicollinearity and removing the variables with high variance inflation factors made the model trustworthy. The rule of thumb to assess whether multicollinearity exist is sqaure root of the variance inflation factor is greater than 2. Multicollinearity do not exist in the model and it is shown (only showing the signifcant values) with an adjusted r of 63% as, $\widehat{EC} = 4081 + 5.194e - 10GDP^2 - 165900EI^{-1}$. The coefficient of the the GDP has a positive direction toward the consumption whereas coeffiecient of the multiplcative inverse of infrastructure have a negative direction from consumption.

Table 4.37: Variance inflation factors

	GDP	**GDPˆ2**	**EI**	**EIˆ-1**
GDP + I(GDPˆ2) + EI + I(EIˆ-1)	73.42	47.77	16.9	5.42
I(GDPˆ2)+I(EIˆ-1)	N/A	1.725141	N/A	1.725141

In order to trust the new model, the assumptions such as normality and constant variance are again tested. Table 4.38 shows that both assumptions exist.

Table 4.38: Assess Residual normality and constant variance

	Test statistic	**p-value**
Breusch-Pagan	5.9855	0.05015
Shapiro Wilk	0.97567	0.08273

4.8 Conclusion

There is a growth hypothesis as electricity consumption Granger cause economic growth. The growth hypothesis just states that electricity consumption granger cause economic growth which means the country's economy is energy dependent. It is as expected that South Africa is energy dependent. There was also found that the electricity infrastructure causal effect on electricity consumption which was expected. The country needs appropriate structures to ensure sustainable electricity supply and consumers should be more prudent in their consumption of electricity since strains the available structures. The economic model derived from multivariate time series have an Akaike information criteria of 1149.379.

The possible models for the univariate economic growth, electricity infrastructure and electricity consumption are $ARIMA(0,1,0)(0,0,1)_4$,$SARIMA(2,1,2)$ and $ARIMA(0,1,1)(0,0,1)_4$ respectively. According to the AIC these last mentioned models are the most optimal. The Akaike information criteria for the electricity consumption, electricity infrastructure and economic growth are 1143.45, -254.05 and -493.25 respectively.

The regression analysis depicts a model with AIC of 1068.17 with economic growth squared and the reciprocal of electricity infrastructure as significant variables and perfect fit of 62 percent. When a unit of EI^{-1} is present then EC decreases whereas a GDP^2 increases.

The overall model that will be chosen is the economic model which was derive from the multivariate time series. The model takes in account the seasonality factor and it has the capabilities to answer the main question of the report appropriately. The model also present the second best Akaike information criteria but with best adjusted R-squared of 89 percent.

Chapter 5

Summary and Conclusion

5.1 Introduction

This study investigated the use of the statistical techniques to model the causal effect or relationship between consumption, economic growth and electricity infrastructure by using time series data, which was collected by Statistics South Africa. The interest for the relationship investigation originated from the fact that South Africa lack of adequate electricity supply for the past years. The models that were used for this study are the ARIMA models, vector autoregressive models and multiple regression models. This chapter presents the summary, discussion, recommendations and conclusion of the study.

5.2 Summary

All available times series data were extracted from the Statistics South Africa and Eskom databases. Data credibility is guaranteed as the data is governed and monitored according to the internal, national and international standards. Data preparation included preparing the data in a quarterly format since electricity consumption was a monthly series and the electricity infrastructure was an annual series. Electricity infrastructure annual series was divided by 4 thus each quarter received equal allocations.

Descriptive analysis was used to visualise each variable and investigate each distribution. All variables consumption, economic growth and infrastructure are skewed to the left. The descriptive plots also showed that the relationship between electricity consumption and economic growth is a non-linear relation which is a quadratic form. The relation between electricity consumption and electricity infrastructure indicate a non-linear relation which is reciprocal form.

The ARIMA approach has been used to predict each variable with possible models that describe the series. The partial autocorrelation function and autocorrelation function graph assisted in the manual investigation of the possible models each variable followed.

Consumption had an optimal ARIMA model shown in Table 4.3 which produced an Akaike Information Criteria of 1143.45. The other possible models produced Akaike information criteria greater than the optimal model. The forecast capability of the ARIMA model produce an average deviation of 108. The model did not exhibit a lack of fit means the model fits the data.

Economic growth has an optimal ARIMA model in Table 4.12 which produced an Akaike Information Criteria of -493.25. The other possible produced an Akaike information criteria greater than the optimal model. The forecast capability of the ARIMA model produced an average deviation of 21 849. This suggested that the model fits the data well.

Electricity infrastructure has an optimal ARIMA model in Table 4.7 which produced an Akaike Information Criteria of -254.05. The other possible model produced an Akaike information criteria greater than the optimal model. The forecast capability of the additive seasonal model produced an average deviation of 1012. The chosen model did not exhibit a lack of fit on the data.

The descriptive analysis from the multiple regression analysis showed a quadratic relationship between electricity consumption and economic growth whereas the reverse hyperbolic relationship between electricity consumption and electricity infrastructure.

This is clearly a non-linear relationship with an adjusted R-squared of 62 percent which is evidence that the model fits the data well and that the new term improves the model more than would be expected by chance. The Akaike information criteria is 978.56 with the assumptions being valid. The normality assumption of residuals was confirmed in Table 4.38 by using the Shapiro-Wilk test as well as the constant variance by the Breusch-Pagan test.

The econometric model was created using the vector autoregressive model. The Augmented Dickey Fuller in Tables 4.19, 4.20 and 4.21 showed that all variables are stationary. The second step of the analysis was to investigate cointegration and the techniques used were the Granger Engle and Johansen tests. Granger Engle test exhibited a cointegration between electricity consumption and economic growth. The Johansen test concluded that the series is cointegrated. The final stage of the procedure was to investigate the Granger causality effect. There was a causal effect from electricity consumption to economic growth and also a granger causal effect from electricity consumption to infrastructure shown in Table 4.25 and Table 4.29 respectively. There was also a causal effect from electricity infrastructure to GDP which was shown in Table 4.28. The vector autoregressive model has an Akaike information criteria of 1149.4 and the model fits data well as shown by the adjusted R-squared of 89.9%.

The ideal technique that is chosen is the vector autoregressive model that has coefficient of determination of 62%. The latter shows that the model closely fits the data and also the model can be trusted since the technique describes the time series data the best. The chosen model is sensible and describe the relationship between electricity consumption and economic growth as expected.

5.3 Recommendation

The South African economy primary drivers are mining and quarrying and manufacture. These sectors need sufficient electricity supply to operate to their maximum therefore, Eskom needs to have the necessary structures in place for an uninterrupted electricity supply.

The rules and policies needs to be tighten to ensure a regular and continuous maintenance of the current and future infrastructure of electricity. Continuous development of sustainable and environment friendly structure need to be built to ensure electricity supply since South African supports a growth hypothesis.

Eskom needs their analytics team to monitor the frequency of maintenance and investigations in order for them to meet Eskom's targets. The team should also liaise with Stats SA to understand the impact sufficient electricity has on the South Africa economy. Since the study has proven that there is a relationship the next step can be for the team to study each sector in order to understand characteristics of this relationship.

There are other factors that influence the South African economy these include politics and employment rates but the wellbeing of the country is also important **??**. Therefore, frequent analysis or monitoring needs to be a priority to recognise the dangers/problems before they get out of hand.

Monitoring the consumption of each sector and doing research on what the consumption limits are per sector in order not to abuse the electricity consumption. These recommendations will surely be a start to steer our economy in the right direction. The impact of electricity consumption may be a small but it is significant.

5.4 Conclusion

The results of this study confirmed that both the time series and regression analysis approach in modelling of time series performs well when the series are stationary. In the event of non- stationary series, time series will be able to analyse the relationship over a long period of time. Also that the large sample analysis is important in the time series data. Researches consider to use ARIMA models, vector autoregressive models and Granger causal test to analyse time series model when they want to understand the developments overtime and the causal effect, respectively.

Therefore, the vector autoregressive model in conjunction with Granger causal effect will answer the question adequately. It shows the second lowest Akaike Information Criteria when comparing the models and the best Coefficient of determination proving the model is the best in predicting capability. Although AIC is considered as a decision measure, note that the paper weigh more on the adjusted R-squared for final decision of measure. The Granger causal test concluded that South Africa supports a growth hypothesis.

The growth hypothesis is when electricity consumption granger causal effect on economic growth meaning that South Africa is energy dependent. This is true because of South Africa's primary sectors such as mining and manufactures are energy dependent. The results also illustrate that electricity infrastructure has a Granger causal effect on economic growth. This effect is as expected since electricity consumption and energy infrastructure go hand in hand.

As mentioned before, the most effective method is the vector autoregressive technique in conjunction with Granger causal effect. The technique cover the time factor and the Granger causal test support the causal effect study well. Further area of study can be of each sector in order to understand characteristics of this relationship. The policy makers should make use of the vector autoregressive method in conjunction with Granger causal test.

We conclude that there is a relationship between electricity consumption and economic growth. The economic growth is significant in lag 4 and 5 and the economic growth coefficient has alternate contributions towards consumption as seen in the impulse response graphs. From the result the study has provided useful information for possible policy amendments to ensure that South Africa is equip to assist in the country's well-being.

References

Abel, A. B., Bernanke, B. S. and Croushore, D. (2008). Macroeconomics. 6th. 12

Africa, G. (2012). The eskom factor: Power politics and the electricity sector in south africa, *South Africa: Greenpeace Africa* . 7, 8

Babatunde, M. A. and Shuaibu, M. I. (2009). The demand for residential electricity in nigeria: a bound testing approach, *proceedings of Second International Workshop on Empirical Methods in Energy Economics, University of Alberta, Edmonton, AB, Canada*, Vol. 2829. 17

Baltagi, B. (2008). *Econometric analysis of panel data*, John Wiley & Sons. 24, 30

Bildirici, M. E., Bakirtas, T. and Kayikci, F. (2012). Economic growth and electricity consumption: Auto regressive distributed lag analysis, *Journal of Energy in Southern Africa* **23**(4): 29–45. 16, 17

Brandt, P. T. and Appleby, J. (2007). Msbvar: Bayesian vector autoregression models, impulse responses and forecasting, *R package version 0.3* **1**.
URL: *http://CRAN.R-project.org/package=MSBVAR* 76

Central Intelligence Agency, C. (2016). The world factbook, *Retrieved August* **20**. 14

Chimedza, C. and Galpin, J. (2004). Time series analysis. 28

Chouaibi, N. and Abdessalem, T. (2011). Causality between electricity consumption and economic growth in tunisia: policy implications, *International Journal of Economic Policy in Emerging Economies* **4**(3): 211–226. 6, 15

Davidson, R., MacKinnon, J. G. et al. (1993). Estimation and inference in econometrics. 32

Economist Intelligence Unit, E. (2015). Country report sa 2015. 6, 12, 14, 15

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing, *Econometrica: journal of the Econometric Society* pp. 251–276. 31

Fin24, . J. (2007). Taxpayers to foot nuclear bill.
URL: *http://www.fin24.com/Economy/Taxpayers-to-foot-nuclear-bill-20070620* 14

Fin24, . J. (2008). Eskom stops exporting power.
URL: *http://www.fin24.com/Economy/Eskom-stops-exporting-power-20080120* 8

Frisch, R. (1933). Ragnar frisch, editor of econometrica 1933-1954, *Econometrica: journal of the Econometric Society* pp. 1–4. 23

Hansen, B. (2000). Economics:study guide. 23, 29

Hyndman, R. J. (2008). Forecast package in r.
URL: *http://CRAN.R-project.org/package=forecast* 37, 46, 48, 53, 55, 61

Jinke, L., Hualing, S. and Dianming, G. (2008). Causality relationship between coal consumption and gdp: difference of major oecd and non-oecd countries, *Applied Energy* **85**(6): 421–429. 75

Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of economic dynamics and control* **12**(2): 231–254. 30

Kraft, J. and Kraft, A. (1978). Relationship between energy and gnp, *J. Energy Dev.;(United States)* **3**(2). 16, 24

Muller, R. (2008). Power rationing until july.
URL: *http://mybroadband.co.za/news/general/2704-power-rationing-until-july.html* 8, 14

National Development Plan 2030 - The Presidency (2009). Our future - make it work, p. 15.
URL: *http://wwwrs.resbank.co.za/webindicators/EconFinDataForSA.aspx* 6

Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. (1996). *Applied linear statistical models*, Vol. 4, Irwin Chicago. 21, 33

News24, . F. (2008a). Eskom's buyback plan in motion.
URL: *http://www.news24.com/SouthAfrica/PowerCrisis/Eskoms-buyback-plan-in-motion-20080211-2* 14

News24, . F. (2008b). Limited power for mines to 2012.
URL: *http://www.news24.com/SouthAfrica/PowerCrisis/Limited-power-for-mines-to-2012-20080214-2* 8

News24, . F. (2008c). Mbeki apologises for power cuts.
URL: *http://www.news24.com/SouthAfrica/News/Mbeki-apologises-for-power-cuts-20080208* 14

News24, . N. (2006). Koeberg reactor shut down.
URL: *http://www.news24.com/SouthAfrica/PowerCrisis/Koeberg-reactor-shut-down-20061106* 14

Nondo, C., Kahsai, M. and Schaeffer, P. V. (2010). Energy consumption and economic growth: evidence from comesa countries, *Research paper* **1**. 15, 17

Österholm, P. and Hjalmarsson, E. (2007). *Testing for cointegration using the Johansen methodology when variables are near-integrated*, number 7-141, International Monetary Fund. 30, 31

Peng, R. D. (2008). Multivariate time series plot.
URL: *http://CRAN.R-project.org/package=mvtsplot* 64

Pesaran, M. H. and Shin, Y. (1998). An autoregressive distributed-lag modelling approach to cointegration analysis, *Econometric Society Monographs* **31**: 371–413. 27

Pfaff, B. (2008). Var modelling.
URL: *http://CRAN.R-project.org/package=vars* 65

Pradhan, R. P. (2010). Transport infrastructure, energy consumption and economic growth triangle in india: cointegration and causality analysis, *Journal of Sustainable Development* **3**(2): 167. 17

Santana, L. (2009). *Statistical Analysis Using the Computer Packages SAS and R*, Andcork Publishers. 18, 19, 20, 21, 36

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples), *Biometrika* **52**(3/4): 591–611. 20

Sjö, B. (2008). Testing for unit roots and cointegration, *Unpublished note - Linkoping University* . 27, 31, 32

Solarin, S. A. (2011). Electricity consumption and economic growth: Trivariate investigation in botswana with capital formation, *International Journal of Energy Economics and Policy* **1**(2): 32. 8, 16, 17

South African Reserve Bank (2016). Economic and financial data for south africa.
URL: *http://wwwrs.resbank.co.za/webindicators/EconFinDataForSA.aspx* 6, 7, 12

Vermeulen, J. (2015). Eskom power stations from 1926 to 2015.
URL: *http://mybroadband.co.za/news/energy/122478-eskom-power-stations-from-1926-to-2015.html* 14

Wei, W. W.-S. (1994). *Time series analysis*, Addison-Wesley publ Reading. 25

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*, Nelson Education. 23, 29, 30

World Bank (2008). World development indicator.
URL: *http://data.worldbank.org/data-catalog/world-development-indicators* 34

Zeileis, A. and Grothendieck, G. (2005). zoo: S3 infrastructure for regular and irregular time series (z's ordered observations).
URL: *http://CRAN.R-project.org/package=zoo* 64

Zeileis, A. and Hothorn, T. (2002). Diagnostic checking in regression relationships, *R News* **2**(3): 7–10.
URL: *http://CRAN.R-project.org/doc/Rnews/* 73

Appendix A

Extra Results

A.1 Electricity Consumption Series – Possible Models

Table A.1: ARIMA(3,1,0)(2,0,0) model

	Value	**Standard Error**	**z value**	**p value**
φ_1	0.5393	0.1109	4.86294	1.156552e-06
φ_2	0.1271	0.1255	1.012749	0.3111801
φ_3	0.1906	0.1061	1.796418	0.07242799
$\Theta\varphi_1$	-0.5076	0.1041	-4.876081	1.082144e-06
$\Theta\varphi_2$	-0.3761	0.1122	-3.35205	0.0008021556
AIC	1163.41			

Table A.2: ARIMA(1,0,0)(2,0,0) model

	Value	Standard Error	z value	p value
φ_1	-0.3276	0.1064	-3.078947	0.002077334
$\Theta\varphi_1$	-0.5223	0.1008	-5.181548	2.200524e-07
$\Theta\varphi_2$	-0.3917	0.1059	-3.698772	0.0002166447
AIC	1155.62			

Table A.3: ARIMA(0,1,1)(0,0,2) modell

	Value	Standard Error	z value	p value
θ_1	-1	0.0469	-21.32196	7.101842e-101
$\Theta\theta_1$	-0.7629	0.1344	-5.676339	1.376077e-08
$\Theta\theta_2$	-0.1495	0.1137	-1.314864	0.1885557
AIC	1148.66			

A.2 Electricity infrastructure – Possible models

Table A.4: ARIMA(0,1,1)(0,0,0)

	Value	Standard Error	z value	p value
$\Theta\theta_1$	-1	0.0287	-34.84	5.396097e-266
AIC	-162.63			

A.3 Diagnostic Plots

These are the diagnostic plots for all different models of consumption.

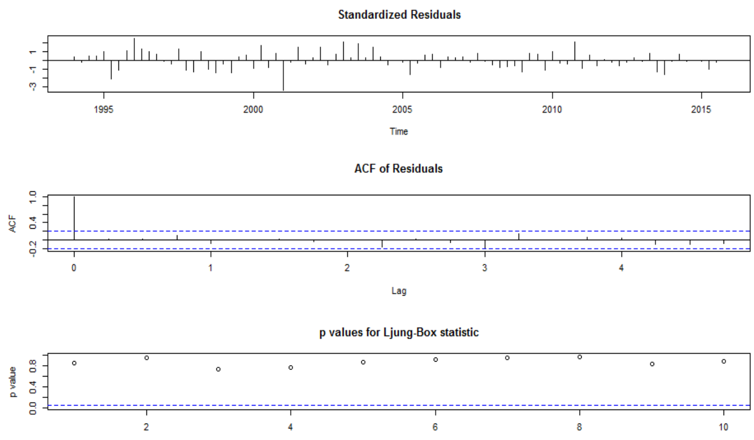


Figure 4.13.1.A: Diagnostics of Electricity Consumption ARIMA(3,1,0)(2,0,0)

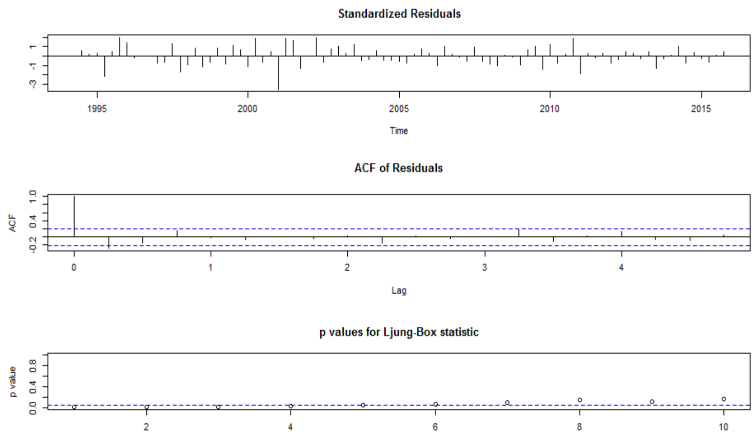


Figure 4.13.2.B: Diagnostics of Electricity Consumption ARIMA(1,0,0)(2,0,0)

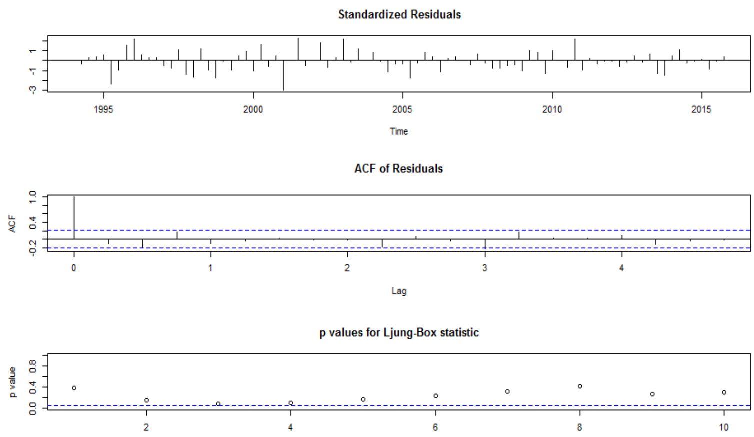


Figure 4.13.3.C: Diagnostics of Electricity Consumption ARIMA(0,1,1)(0,0,2)

These are the diagnostic plots for all different models of electricity infrastructure.

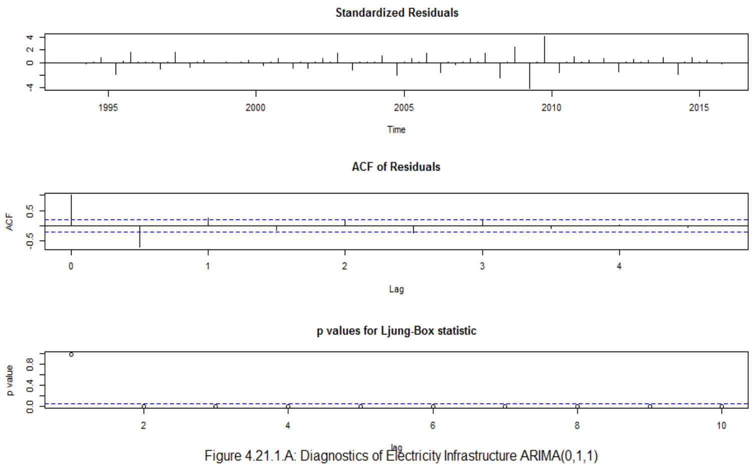


Figure 4.21.1.A: Diagnostics of Electricity Infrastructure ARIMA(0,1,1)

The diagnostic plots for all extra model of economic growth.

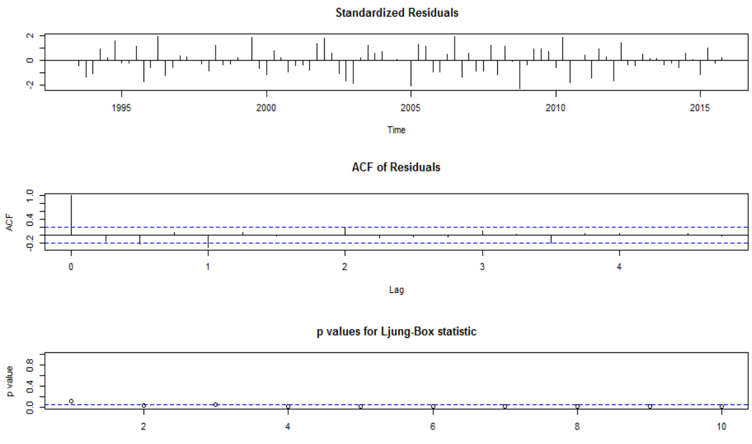


Figure 4.29.1.A: Diagnostics of Economic Growth (ARIMA(1,0,0)(1,0,0))

A.4 Economic models

Table A.5: Optimal model of Econometric model

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
Consumption.l1	0.647817	0.110498	5.863	1.32e-07 ***
Growth.l1	0.001514	0.002481	0.610	0.54378
Expenditure.l1	0.004897	0.028024	0.175	0.86177
Consumption.l2	-0.052569	0.127578	-0.412	0.68154
Growth.l2	-0.003288	0.002809	-1.171	0.24570
Expenditure.l2	0.011820	0.036560	0.323	0.74741
Consumption.l3	0.261885	0.123830	2.115	0.03795 *
Growth.l3	0.001920	0.002960	0.649	0.51855
Expenditure.l3	-0.015578	0.036854	-0.423	0.67379
Consumption.l4	0.221899	0.126812	1.750	0.08447 .
Growth.l4	0.008378	0.003076	2.723	0.00812 **
Expenditure.l4	0.033863	0.036838	0.919	0.36108
Consumption.l5	-0.250573	0.111585	-2.246	0.02784 *
Growth.l5	-0.008329	0.002796	-2.979	0.00396 **
Expenditure.l5	-0.043554	0.028374	-1.535	0.12923
constant	568.554006	263.075678	2.161	0.03406 *
Adjusted R-squared	0.8911			
AIC	1149.379			

Table A.6: Optimal model of Econometric model - Electricity Infrastructure

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
Consumption.l1	-0.8395	0.4773	-1.759	0.0829
Growth.l1	-0.007446	0.01072	-0.695	0.4895
Expenditure.l1	0.8302	0.1211	6.858	2.14e-09 ***
Consumption.l2	0.3455	0.5511	0.627	0.5327
Growth.l2	0.02821	0.01213	2.324	0.0230 *
Expenditure.l2	0.232	0.1579	1.469	0.1462
Consumption.l3	-0.2956	0.5349	-0.553	0.5822
Growth.l3	-0.001335	0.01278	-0.104	0.9171
Expenditure.l3	-0.1857	0.1592	-1.167	0.2472
Consumption.l4	0.1751	0.5478	0.32	0.7502
Growth.l4	-0.03345	0.01329	-2.517	0.0141 *
Expenditure.l4	0.1295	0.1591	0.814	0.4183
Consumption.l5	0.6419	0.482	1.332	0.1872
Growth.l5	0.01619	0.01208	1.34	0.1844
Expenditure.l5	-0.09922	0.1226	-0.81	0.4209
constant	-446	1136	-0.392	0.6959
Adjusted R-squared	0.989			
AIC	1403.971			

Table A.7: Optimal model of Econometric model - Economic Growth

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
Consumption.l1	9.933	4.99	1.99	0.050397
Growth.l1	0.7161	0.1121	6.39	1.51e-08 ***
Expenditure.l1	0.6852	1.266	0.541	0.589958
Consumption.l2	-4.403	1.266	-0.764	0.447292
Growth.l2	0.1296	0.1269	1.021	0.310668
Expenditure.l2	-0.2225	1.651	-0.135	0.893214
Consumption.l3	1.316	5.593	0.235	0.814613
Growth.l3	0.04135	0.1337	0.309	0.757933
Expenditure.l3	-0.0578	1.664	-0.035	0.972397
Consumption.l4	7.12	5.727	1.243	0.217867
Growth.l4	0.4949	0.1389	3.562	0.000662 ***
Expenditure.l4	-2.734	1.664	-1.643	0.104798
Consumption.l5	-5.404	5.04	-1.072	0.287187
Growth.l5	-0.3856	0.1263	-3.053	0.003184 **
Expenditure.l5	2.75	1.281	2.146	0.035304 *
constant	-22530	11880	-1.896	0.061995
Adjusted R-squared	0.9993			
AIC	1812.368			

Appendix B

Data

Consumption	GDP	Infrastructure
2802	100786.9	287
3156	107839.2	287
3130	113811	287
2869	116447	287
2909	117123.5	202
3194	122177.4	202
3261	125829.2	202
3050	131102.6	202
3179	132452	263.75
3011	137940.2	263.75
3015	145466	263.75
3105	148011.4	263.75
3486	148737.7	262.25
3589	158072.8	262.25
3641	162896.8	262.25
3593	164903.6	262.25
3706	166191.4	216.75
3737	174990.7	216.75
4017	179978.2	216.75
3611	181956.1	216.75
3569	180609.8	211.25
3812	190870.9	211.25
3700	194211.2	211.25
3313	195967.9	211.25
3519	195750.7	212.5
3489	204169.7	212.5
3642	215428.3	212.5
3494	219403.7	212.5
3414	219943.1	841.75
3800	231799.5	841.75

Consumption	GDP	Infrastructure
3747	245382.7	841.75
3620	249198.2	841.75
3017	250192.1	910.75
3486	258654.4	910.75
3776	264961.8	910.75
3343	272334	910.75
3198	283675	1346.25
3753	302366.2	1346.25
3736	313138.9	1346.25
3582	318084.2	1346.25
3740	317548	1519.5
4076	327703.3	1519.5
4342	337179.3	1519.5
4086	343335.1	1519.5
4073	349753.8	1373.5
4433	362514.7	1373.5
4490	376710.2	1373.5
4189	387645.3	1373.5
4092	383581.9	2249.75
4233	401384.9	2249.75
4338	422127	2249.75
4225	432160.6	2249.75
4262	426598.3	2716.75
4309	444554.1	2716.75
4580	480215.2	2716.75
4418	488033.4	2716.75
4422	498347.9	4426.75
4534	512943.8	4426.75
4833	539780.5	4426.75
4546	558428.9	4426.75

Consumption	GDP	Infrastructure
4407	559861.1	6191
4423	586568.6	6191
4581	612848.9	6191
4330	609784.4	6191
4114	599459.3	11774.75
4385	616828.2	11774.75
4682	641247	11774.75
4202	650142.3	11774.75
4345	643098.1	14250.75
4434	687346.6	14250.75
4630	699822.2	14250.75
4661	717741.3	14250.75
4352	721162.3	13864.25
4567	745483.4	13864.25
4669	768367.9	13864.25
4546	789936.9	13864.25
4314	777098.8	15838.5
4439	813259.1	15838.5
4642	827779.6	15838.5
4475	844405	15838.5
4337	842183.7	15261.5
4591	877188.9	15261.5
4510	899130.6	15261.5
4246	915822.6	15261.5
4179	917849.2	18179
4541	937528.7	18179
4488	962934.1	18179
4315	978755.6	18179
4189	964816	17668.25
4288	992388	17668.25

Consumption	GDP	Infrastructure
4377	1006723	17668.25
4246	1027026.1	17668.25

Appendix C

Code

C.1 R Code

The following functions are written as part of the research project.

```
#Descriptive Analysis
#Load Data
require(Gmisc)
dataCd<- read.table("C:/Time Series/All.csv",sep = ",",header = T)
Cons<- dataCd[,1:3]
#Multiple Scatterplots and histograms per variable
pairs(Cons,main = "")
title(sub = "Figure 4.1: Scatterplot of all pairs")
hist(Cons[,1],col = "lightblue",xlab = "Electricity Consumption",main ="")
title(sub = "Figure 4.2: Histogram for Electricity Consumption")
hist(Cons[,2],col = "lightblue3",xlab = "Gross Domestic Product",main ="")
title(sub = "Figure 4.3 Histogram for Gross Domestic Product")
hist(Cons[,3],col = "lightgray",xlab = "Electricity Infrastructure",main ="")
title(sub = "Figure 4.4: Histogram for Electricity Infrastructure")
#Summary Statistics
htmlTable(summary(Cons),header = c("Electricity Consumption","Gross Domestic
    Product","Electricity Infrastructure"),caption = paste("Table 4.1 :
    Descriptive Statistics of variables"))
#kurtosis and skewness
require(moments)
#consumption
```

```
skewness(Cons[,1])
kurtosis(Cons[,1])
#gross domestic product
skewness(Cons[,2])
kurtosis(Cons[,2])
#EI
skewness(Cons[,3])
kurtosis(Cons[,3])
# Electricity Consumption
dataC<- read.table("C:/Time Series/All.csv",sep = ",",header = T)
myvars <- c("ElecConsumption")
Cons<- dataC[myvars]
consts<- ts(Cons,start=c(1993,1),frequency = 4)
plot.ts(consts,xlab = "Time",ylab="Consumption in Kwh",main ="")
title(sub = "Figure 4.5: Time series of electricity Consumption")
plot.new()
par(mfrow=c(2,1))
acf(consts,main="ACF of electricity Consumption",lag.max = 100)
pacf(consts,main="PACF of electricity Consumption",lag.max = 100)
title(sub = "Figure 4.6: The ACF and PACF of electricity Consumption",outer =
    TRUE)
mtext("Figure 4.6: The ACF and PACF of electricity Consumption",outer = T,side =
     1,line = -1)
#Seasonal difference
par(mfrow=c(1,1))
DiffConsts<- diff(consts,4)
plot.ts(DiffConsts,xlab = "Time",ylab="Difference(Consumption) in Kwh")
title(sub = "Figure 4.7: Differenced electricity Consumption")
par(mfrow=c(2,1))
acf(DiffConsts,main="ACF of differenced electricity Consumption",lag.max = 100)
pacf(DiffConsts,main="PACF of differenced electricity Consumption",lag.max =
    100)
title(sub = "Figure 4.8: The ACF and PACF of differenced electricity Consumption
    ",outer = TRUE)
mtext("Figure 4.8: The ACF and PACF of differenced electricity Consumption",
    outer = T,side = 1,line = -1)
#Difference on top of seasonal difference
par(mfrow=c(1,1))
DiffDiffConsts<- diff(DiffConsts)
plot.ts(DiffDiffConsts,xlab = "Time",ylab="Difference Differenced(Consumption)
    in Kwh",main="")
```

```
title(sub = "Figure 4.9: Differenced (difference(electricity Consumption))")
par(mfrow=c(2,1))
acf(DiffDiffConsts,main="ACF of differenced( differenced(electricity Consumption
    ))",lag=100)
pacf(DiffDiffConsts,main="PACF of differenced (differenced(electricity
    Consumption))",lag =100)
title(sub = "Figure 4.10: The ACF and PACF of differenced (differenced(
    electricity Consumption))",outer = TRUE)
mtext("Figure 4.10: The ACF and PACF of differenced (differenced(electricity
    Consumption))",outer = T,side = 1,line = -1)
#Estimating the model fit, estimates and forecasting - Consumption
require(forecast)
fit <- HoltWinters(DiffDiffConsts, beta=FALSE, gamma=FALSE)
fitD <- HoltWinters(DiffConsts, beta=FALSE, gamma=FALSE)
fit0 <- HoltWinters(consts)
AfitC = auto.arima(consts)
#tabled the predicted values
t42=matrix(c(round(forecast(fit0, 3)$mean,3),round(forecast(AfitC, 3)$mean,3)
    ,0,0,0),byrow = T,nrow = 3,dimnames = list(c("HoltWinters","ARIMA","Actual")
    ))
htmlTable(t42,header = c("Quarter 1","Quarter 2","Quarter 3"),caption = paste("
    Table 4.2 :Predictors for the next three quarters in year 2016"))
#holtwinter
hw = matrix(c( 0.6042009,0,0.2367775,4323.79837,25.58750,-195.28920,14.10826
 ,126.48665,-84.97475),nrow = 9,ncol = 1,byrow = T,
dimnames = list(c("alpha","beta","gamma","a","b",
"seasonal1","seasonal2","seasonal3","seasonal4"),c("Value")))
htmlTable(hw,caption = paste("HoltWinters model"))

require(forecast)
#predict next three future values
par(mfrow=c(1,1))
plot(forecast(fit0, 3),main = "")
title(sub = "Figure 4.11: Forecasting - Consumption next three values")
#Goodness of fit
require(forecast)
plot.new()
par(mfrow=c(1,1))
Autofit<-auto.arima(DiffDiffConsts)
plot(forecast((Autofit)),main="")
title(sub = "Figure 4.12: Forecasts from ARIMA(0,1,1)(0,0,1)")
```

```
AutofitD<-auto.arima(DiffConsts)
plot(forecast((AutofitD)),main="")
title(sub = "Figure 4.12.1: Forecasts from ARIMA(3,1,0)(2,0,0)")
#Diagnostics
#Appendix - A,B and C
tsdiag(Autofit,main="")
title(sub = "Figure 4.13: Diagnostics of Electricity Consumption ARIMA(0,1,1)
    (0,0,1)")
tsdiag(AutofitD,main="")
title(sub = "Figure 4.13.1.A: Diagnostics of Electricity Consumption ARIMA
    (3,1,0)(2,0,0)")

#Difference seasonal
a3 = matrix(c( 0.5393,0.1109,4.86294,1.156552e
    -06,0.1271,0.1255,1.012749,0.3111801,0.1906,0.1061,1.796418,0.07242799
,-0.5076,0.1041,-4.876081,1.082144e
    -06,-0.3761,0.1122,-3.35205,0.0008021556,1163.41,"","",""),nrow = 6,ncol =
    4,
byrow = T,dimnames = list(c("phi_1","phi_2","phi_3","seasonal phi_1","seasonal
    phi_2","AIC"),
 c("Value","Standard Error","z value","p value")))
htmlTable(a3,caption = paste("Table 4.3.1.A: ARIMA(3,1,0)(2,0,0) model"))
#Diff on topDifference seasonal
a4 = matrix(c( -0.3636 ,0.0990,-3.672727,0.0002399756,-0.8677,0.1007,-8.616683,
    6.892147e-18,1143.45,"","",""),nrow = 3,ncol = 4,byrow = T,dimnames = list(c
    ("theta_1","seasonal theta_1","AIC"),c("Value","Standard Error","z value","p
     value")))
htmlTable(a4,caption = paste("Table 4.3: ARIMA(0,1,1)(0,0,1) model"))
#Manual Difference seasonal
a = matrix(c( -0.3276 ,0.1064,-3.078947,0.002077334,-0.5223,0.1008,-5.181548,
    2.200524e-07,-0.3917,0.1059,-3.698772,0.0002166447,1155.62,"","",""),nrow =
    4,ncol = 4,byrow = T,dimnames = list(c("phi_1","seasonal phi_1","seasonal
    phi_2","AIC"),c("Value","Standard Error","z value","p value")))
htmlTable(a,caption = paste("Table 4.3.2.A: ARIMA(1,0,0)(2,0,0) model"))
a1=arima(DiffDiffConsts,order =c(0,1,1),seasonal = c(0,0,2) )
a2=arima(DiffConsts,order =c(1,1,0),seasonal = c(2,0,0))
#Appendix
par(mfrow=c(1,1))
tsdiag(a1,main="")
title(sub = "Figure 4.13.2.B: Diagnostics of Electricity Consumption ARIMA
    (1,0,0)(2,0,0)")
```

```
par(mfrow=c(1,1))
tsdiag(a2,main="")
title(sub = "Figure 4.13.3.C: Diagnostics of Electricity Consumption ARIMA
    (0,1,1)(0,0,2)")
#Manual diff on top of seasonal difference
b = matrix(c(-1.0000,0.0469,-21.32196,7.101842e-101,-0.7629,0.1344,
    -5.676339,1.376077e-08,-0.1495,0.1137,-1.314864,0.1885557,1148.66,"","",""),
    nrow = 4,ncol = 4,byrow = T,dimnames = list(c("theta_1","seasonal theta_1","
    seasonal theta_2","AIC"),c("Value","Standard Error","z value","p value")))
htmlTable(b,caption = paste("Table 4.3.3.A: ARIMA(0,1,1)(0,0,2) model"))
optimalmodel = matrix(c("ARIMA(0,1,1)(0,0,1)",0,0,"ARIMA(3,1,0)(2,0,0)",0,0,"
    ARIMA(1,0,0)(2,0,0)",0,0,"ARIMA(0,1,1)(0,0,2)",0,0),nrow = 4,ncol = 3,byrow
    = T,dimnames = list(c("1","2","3","4"),c("Description","Model","AIC")))
htmlTable(optimalmodel,caption = paste("Table 4.4: Possible Models"))

# Electricity infrastructure
par(mfrow=c(1,1))
Consi=scan("C:/Time Series/ElecInfrastructure.csv",skip = 1,sep = ",")
constsi<- ts(Consi,frequency = 4,start=c(1993),end =c(2015,4))
plot.ts(constsi,xlab = "Time",ylab="R'million",main="")
title(sub = "Figure 4.13: Electricity infrastructure(Capital infrastructure)")
par(mfrow=c(2,1))
acf(constsi,main="ACF of Electricity infrastructure",lag.max = 100)
pacf(constsi,main="PACF of Electricity infrastructure",lag.max = 100)
title(sub = "Figure 4.14: The ACF and PACF of electricity infrastructure",outer
    = TRUE)
mtext("Figure 4.14: The ACF and PACF of electricity infrastructure",outer = T,
    side = 1,line = -1)
#Logarithm transformation
par(mfrow=c(1,1))
logconstsi<- log(constsi)
plot.ts(logconstsi,xlab = "Time",ylab="Log R'million",main="")
title(sub = "Figure 4.15: Time series of log(electricity infrastructure)")
par(mfrow=c(2,1))
acf(logconstsi,main="ACF of log(electricity infrastructure",lag.max = 100)
pacf(logconstsi,main="PACF of log(electricity infrastructure)",lag.max = 100)
title(sub = "Figure 4.16: The ACF and PACF of log(electricity infrastructure)",
    outer = TRUE)
mtext("Figure 4.16: The ACF and PACF of log(electricity infrastructure)",outer =
     T,side = 1,line = -1)
# Difference logged series
```

```
par(mfrow=c(1,1))
DiffConstsi<- diff(logconstsi)
v = as.matrix(DiffConstsi)
which.max(na.omit(v))
m2015=mean(na.omit(v))
v[28]<-m2015
which.min(na.omit(v))
min_=mean(na.omit(v))
v[4]<-min_
vr = ts(v,frequency = 4,start=c(1993),end =c(2015,4))
#Apply moving average of 2
require("TTR")
vr2 = SMA(diff(vr),n=2)
a = which(vr2 < -0.15)
a1 = as.matrix(a)

vr2[a1[1,1]] = -0.08

vr2[a1[2,1]] = -0.08

vr2[a1[3,1]] = -0.08

b = which(vr2 > 0.15)
b1 = as.matrix(b)
vr2[b1[1,1]] = -0.13

vr2[b1[2,1]] = -0.13

vr2[b1[3,1]] = -0.13

c = which(vr2 > 0.05)
c1 = as.matrix(c)
vr2[c1[1,1]] = 0.05

vr2[c1[2,1]] = 0.05

vr2[c1[3,1]] = 0.05

vr2[c1[4,1]] = 0.05

vr2[c1[5,1]] = 0.05
```

```
vr2[c1[6,1]] = 0.05

vr2[c1[7,1]] = 0.05

vr2[c1[8,1]] = 0.05
#Plot the final dataset
d = diff(vr2,2)
plot.ts(d,xlab = "Time",ylab="Difference logged in R'million")
#ACF and PACF of differenced logged
title(sub = "Figure 4.18: Differenced logged electricity infrastructure")
par(mfrow=c(2,1))
acf(na.omit(d),main="ACF of differenced electricity infrastructure",lag.max =
    100)
pacf(na.omit(d),main="PACF of differenced electricity infrastructure",lag.max =
    100)
title(sub = "Figure 4.19: The ACF and PACF of differenced logged electricity
    infrastructure",outer = TRUE)
mtext("Figure 4.19: The ACF and PACF of differenced logged electricity
    infrastructure",outer = T,side = 1,line = -1)

# Estimating the model fit, estimates and forecasting - Infrastrucuture
require(forecast)
require(htmlTable)
par(mfrow=c(1,1))
Autofiti<-auto.arima(d)
plot(forecast((Autofiti)),main="")
title(sub = "Figure 4.20: Forecasts from ARIMA(2,1,2)")
#Diagnostic
tsdiag(Autofiti,main="")
title(sub = "Figure 4.21: Diagnostics of Electricity Infrastructure - ARIMA
    (2,1,2)")
#Results
di = matrix(c(0,0.1064,0,1,-0.4132,0.1058,-3.905482,9.403774e-05,0,0.0823,0,1
,-0.8113,0.0835, -9.716168,2.57283e-22,-254.05,"","",""),nrow = 5,ncol = 4,
byrow = T,dimnames = list(c("phi_1","phi_2","theta_1","theta_2","AIC")
,c("Value","Standard Error","z value","p value")))
htmlTable(di,caption = paste("Table 4.5: ARIMA(2,1,2) model"))
#Manual identification
d1=arima(d,order =c(0,1,1),seasonal = c(0,0,0) )
di3 = matrix(c(-1,0.0287,-34.84321,5.396097e-266,-162.63,"","",""),nrow = 2,
```

```
ncol = 4,byrow = T,dimnames = list(c("theta_1","AIC")
,c("Value","Standard Error","z value","p value")))
htmlTable(di3,caption = paste("Table 4.5.1.A: ARIMA(0,1,1)(0,0,0) model"))
tsdiag(d1,main="")
title(sub = "Figure 4.21.1.A: Diagnostics of Electricity Infrastructure ARIMA
    (0,1,1)")
#Forecast
optimalmodel1 = matrix(c("ARIMA(2,1,2)",0,0,"ARIMA(0,1,1)(0,0,0)",0,0),nrow = 2,
ncol = 3,byrow = T,dimnames = list(c("1","2"),c("Description","Model","AIC")))
htmlTable(optimalmodel1,caption = paste("Table 4.6: Possible Models"))
fitInfrastructure<-auto.arima(constsi)
fitInH<-HoltWinters(constsi)
t46=matrix(c(round(forecast(fitInH, 3)$mean,3),round(forecast(fitInfrastructure,
     3)$mean,3),0,0,0)
,byrow = T,nrow = 3,dimnames = list(c("HoltWinters","ARIMA","Actual")))
htmlTable(t46,header = c("Quarter 1","Quarter 2","Quarter 3"),
caption = paste("Table 4.7 :Predictors for the next three quarters in year 2016
    - Infrastructure"))
#holtwinter
hwi = matrix(c( 0.8796561,0.05256674,0.9574244,
17932.5072,274.9185,685.7262,357.7488,-41.5558,-263.9407),
nrow = 9,ncol = 1,byrow = T,dimnames = list(c("alpha","beta",
"gamma","a","b","seasonal1","seasonal2",
"seasonal3","seasonal4"),c("Value")))
htmlTable(hwi,caption = paste("HoltWinters model - Infrastructure"))

#  Economic growth (GDP)
par(mfrow=c(1,1))
dataC2<- read.table("C:/Time Series/All.csv",sep = ",",header = T)
#GDP-Univariate Time series Quarterly
myvars <- c("GDP")
Cons2<- dataC[myvars]
consts2<- ts(Cons2,start=c(1993,1),frequency = 4)
plot.ts(consts2,xlab= "Year",main="")
title(sub = "Figure 4.22: Time Series of Quarterly Economic Growth")
#PACF and ACF
par(mfrow=c(2,1))
acf(consts2,main="ACF of Economic Growth",lag.max = 100)
pacf(consts2,main="PACF of Economic Growth",lag.max = 100)
title(sub = "Figure 4.23: The ACF and PACF of Economic Growth",outer = TRUE)
mtext("Figure 4.23: The ACF and PACF of Economic Growth",outer = T,side = 1,line
```

```
    = -1)
#Logarithm transformation
par(mfrow=c(1,1))
logconsts2<- log(consts2)
plot.ts(logconsts2,ylab="Log(GDP)",main="")
title(sub = "Figure 4.24: Time series of Log Economic Growth")
par(mfrow=c(2,1))
acf(logconsts2,main="PACF of log Economic Growth",lag.max = 100)
pacf(logconsts2,main="PACF of log Economic Growth",lag.max = 100)
title(sub = "Figure 4.25: The ACF and PACF of log Economic Growth",outer = TRUE)
mtext("Figure 4.25: The ACF and PACF of log Economic Growth",outer = T,side = 1,
    line = -1)
#Difference logged transformation
par(mfrow=c(1,1))
DiffConsts5 <-  diff(logconsts2)
plot.ts(DiffConsts5,ylab="Difference Logged GDP")
title(sub = "Figure 4.26: Differenced Logged Economic Growth")
par(mfrow=c(2,1))
acf(DiffConsts5,main ="",lag.max = 100)
pacf(DiffConsts5,main="",lag.max = 100)
title(sub = "Figure 4.27: The ACF and PACF of differenced Logged Economic Growth
    ",outer = TRUE)
mtext("Figure 4.27: The ACF and PACF of differenced Logged Economic Growth",
    outer = T,side = 1,line = -1)

#Goodness of fit and Forecast
require(forecast)
par(mfrow=c(1,1))
Autofitgdp<-auto.arima(DiffConsts5)
plot(forecast((Autofitgdp)),main = "")
title(sub = "Figure 4.28: Forecasts from ARIMA(0,0,0)(0,1,1)")
#True value of next three quarters
Autofitgdp<-auto.arima(consts2)
#Auto.arima
tsdiag(Autofitgdp,main="")
title(sub = "Figure 4.29:Diagnostics of Economic Growth (ARIMA(0,0,1)(0,1,1))")
e = matrix(c(-0.8078,0.0871,-9.274397,1.786261e-20,-493.25,"","",""),nrow = 2,
    ncol = 4
,byrow = T,dimnames = list(c("Theta_theta","AIC"),c("Value","Standard Error","z
    value","p value")))
htmlTable(e,caption = paste("Table 4.8: ARIMA(0,0,0)(0,1,1) model"))
```

```
#Manual

f1=arima(DiffConsts5,order =c(1,1,0) ,seasonal = c(1,0,0))
#Appendix
tsdiag(f1,main="")
title(sub = "Figure 4.29.1.A: Diagnostics of Economic Growth (ARIMA(1,0,0)
    (1,0,0))")
#Tabled results
f = matrix(c(-0.4547,0.0932,-4.878755,1.067574e-06,
0.7092,0.0725,9.782069,1.344339e-22,-440.81,"","",""),
,nrow = 3,ncol = 4,byrow = T,dimnames = list(c("phi_1","sphi_1","AIC"),
c("Value","Standard Error","z value","p value")))
htmlTable(f,caption = paste("Table 4.9: ARIMA(1,0,0)(1,0,0) model"))
optimalmodel1 = matrix(c("ARIMA(0,0,1)(0,1,1)",0,0,"ARIMA(2,0,0)",0,0),nrow = 2,
    ncol = 3,
 byrow = T,dimnames = list(c("1","2"),c("Description","Model","AIC")))
htmlTable(optimalmodel1,caption = paste("Table 4.10: Possible Models"))
#Forecast
fitgdpvalue<-auto.arima(consts2)
fitgdpH<-HoltWinters(consts2)
htmlTable(t411,header =
c("Quarter 1","Quarter 2","Quarter 3"),
caption = paste("Table 4.11 :Predictors for the next three quarters in year 2016
     - Economic growth"))

##  Multivariate time series

dataC3<- read.table("C:/Time Series/All.csv",sep = ",",header = T)
myvars <- c("ElecConsumption","GDP","CapExpediture")
Cons3<- dataC3[myvars]
consts3<- ts(Cons3,start=c(1993,1),frequency = 4)
consMat<-as.matrix(consts3)
zt<-cbind(consMat[,1],consMat[,2],consMat[,3])
colnames (zt) <-c("Consumption", "Growth","Expenditure")
consMat<-as.matrix(Cons3)
#Multiple graph
require (zoo)
name.zoo <-zoo(cbind(consMat[, 1], consMat[, 2], consMat[, 3]))
colnames (name.zoo) <-c("Consumption", "Growth","Expenditure")
plot(name.zoo,main = "")
```

```
title(sub = "Figure 4.30: Time Series of Electricity Consumption,
Electricity Infrastructure and Economic Growth")
#Descriptive plot
require(mvtsplot)
par(mfrow=c(1,1))
mvtsplot (name.zoo,main = "")
title(sub = "Figure 4.31: Multivariate of all three(Consumption,Infrastructure,
    Growth) variables")

#Optimal Lag
#Choosing the number of lag with information criteria
ztF <- na.omit(zt)
require(vars)
require(Gmisc)
htmlTable(VARselect(ztF, lag.max=9, type="const")$selection,caption = paste("
    Table 4.12: The optimal Lag"),header = c("AIC(n)","HQ(n)","SC(n)","FPE(n)"))
htmlTable(round(VARselect(ztF, lag.max=9, type="const")$criteria,3),caption =
    paste("Table 4.13: The optimal Lag(Values)"),header = c("1","2","3","4","5",
    "6","7","8","9"))#,rowlabel = c("AIC(n)","HQ(n)","SC(n)","FPE(n)"))
#Var - Choosing optimal model
#Model1
require(vars)
zvar1<-VAR(ztF, p = 5, type = "const")
#testing whether residuals are correlated
serial.test(zvar1, lags.pt=16, type="PT.asymptotic")
#Model2
zvar2<-VAR(ztF, p =9, type = "const")
#testing whether residuals are correlated
serial.test(zvar2, lags.pt=16, type="PT.asymptotic")
#summary(zvar2)
#put in the appendix
#stepAIC(zvar1)
g = matrix(c(105.18, 99,0.3166,83.335,63,0.04413),nrow = 2,ncol = 3,byrow = T,
    dimnames = list(c("Model1 with lag 5","Model2 with lag 9"),c("Chi-Squared","
    df","p-value")))
htmlTable(g,caption = paste("Table 4.14: Portmanteau (asymptotic) Test"))
#Forecast

require(forecast)
fcst <- forecast(zvar1)
plot(fcst, xlab="Year",main="")
```

```
title(sub = "Figure 4.32: Forecasts from Model 1 with lag 5")

#explaining the results
require(MSBVAR)
rf.var <- reduced.form.var(ztF, p=3)
plot(irf(rf.var, nsteps = 12))

##  Econometric Model

require(vars)

dataC4<- read.table("C:/Time Series/All.csv",sep = ",",header = T)
myvars <- c("ElecConsumption","GDP","CapExpediture")
Cons4<- dataC4[myvars]
consts4<- ts(Cons4,start=c(1993,1),frequency = 4)
consMat<-as.matrix(consts4)
zt1<-cbind(diff(na.omit(diff(consMat[, 1],4))), diff(log(consMat[, 2]),1),na.
    omit(d))
colnames (zt1) <-c("Consumption", "EcononmicGrowth","CapExpediture")

#ADF for each of the variables; electricity consumption, electricity
    infrastructure and economic growth
#Consumption
require(tseries)
#adf.test(na.omit(zt1[,1]),k=4)
j = matrix(c(-5.3922,4,0.01,"stationary"),nrow = 4,ncol = 1,byrow = T,dimnames =
     list(c("Critical Value","Lag order","p-value","Alternative Hypothesis"),c("
    Value")))
htmlTable(j,caption = paste("Table 4.15: Augmented Dickey-Fuller Test -
    Consumption"))
#0.01 smaller than 0.05 so the series is stationary

#EcononmicGrowth
#adf.test(na.omit(zt1[,2]))
m = matrix(c(-3.7301,4,0.02622,"stationary"),nrow = 4,ncol = 1,byrow = T,
    dimnames = list(c("Critical Value","Lag order","p-value","Alternative
    Hypothesis"),c("Value")))
htmlTable(m,caption = paste("Table 4.16: Augmented Dickey-Fuller Test - Economic
     Growth"))
#P-value = 0.05, 0.08<0.05 then Series is stationary
```

```
#CapExpediture
#adf.test(na.omit(zt1[,3]))
s = matrix(c(-6.2104,4,0.01,"stationary"),nrow = 4,ncol = 1,byrow = T,dimnames =
     list(c("Critical Value","Lag order","p-value","Alternative Hypothesis"),c("
    Value")))
htmlTable(s,caption = paste("Table 4.17: Augmented Dickey-Fuller Test -
    Infrastructure"))

#Cointegration Granger Engle-Test
Engle<-lm(ElecConsumption ~ GDP + I(GDP^2) ,data = consts4)
#summary(Engle)
r = matrix(c(2.559e+03,8.068e+01,31.72,"< 2e-16 ***",5.529e-03,3.821e-04,14.47,"
    <2e-16 ***",-3.837e-09,3.549e-10,-10.81,"<2e-16 ***",
0.8309,"","","","< 2.2e-16","","","","0 '***'","0.001 '**'","0.01 '*'",0.05),
    nrow = 6,ncol = 4,byrow = T,
dimnames = list(c("Intercept","GDP","I(GDP^2)","Adjusted R-squared","Overall
    Model Pvalue","Signif. codes")
,c("Estimates","Standard Error","t value","p-value(>|t|")))
htmlTable(r,caption = paste("Table 4.18: Granger Engle Test - Consumption ~
    Economic growth"))

#Overall model is significant because of p-value that smaller than 0.05
residual<- resid(Engle)
year<-ts(start=c(1993,1),end=c(2015,4),frequency = 4)
ts.plot(year,residual, gpars=list(main="", xlab="year", ylab="residuals"))
title(sub="Figure 4.33 Electricity Consumption vs. Economic growth: Is there
    cointegration?")

Engle1<-lm(ElecConsumption~CapExpediture+I(CapExpediture^-1),data = consts4)
#summary(Engle1)
t = matrix(c(4.095e+03,7.695e+01,53.217,"< 2e-16 ***",2.345e-02,6.757e-03,3.470,
    "0.000804 ***"
,-1.670e+05,2.407e+04,-6.938 ,"6.15e-10 ***",
0.6288,"","","","< 2.2e-16","","","","0 '***'","0.001 '**'","0.01 '*'",0.05),
    nrow = 6,ncol = 4,byrow = T
,dimnames = list(c("Intercept","EI","1/EI","Adjusted R-squared","Overall Model
    Pvalue","Signif. codes"),c("Estimates","Standard Error","t value","p-value
    (>|t|")))
htmlTable(t,caption = paste("Table 4.19: Granger Engle Test - Consumption ~
    Energy Infrastructure"))
#Overall model is significant because of p-value that smaller than 0.05
```

```
residual1<- resid(Engle1)
plot.new()
year<-ts(start=c(1993,1),end=c(2015,4),frequency = 4)
ts.plot(year,residual1, gpars=list(main="", xlab="year", ylab="residuals"))
title(sub="Figure 4.34 Electricity Consumption vs.Electricity Infrastructure :
    Is there cointegration?")
#Cointegration: Johansen Test
require(urca)
coi<- ca.jo(zt1,type="trace",K=2,ecdet ="none",spec ="longrun")
#summary(coi)
w = matrix(c(55.27,6.50,8.18,11.65,
    115.23,15.66,17.95,23.52,223.91,28.71,31.52,37.22),nrow = 3,ncol = 4,byrow =
     T,dimnames = list(c("r <= 2","r <= 1","r = 0"),c("test","10pct","5pct","1
    pct")))
htmlTable(w,caption = paste("Table 4.20: Johansen-Procedure"))

#Granger Causality Test
require(lmtest)
#Test 1
#grangertest(consMat[,1]~consMat[,2],order=1)
t1 = matrix(c(88,"","","",89,-1,4.1228,"0.04533 *"),nrow = 2,ncol = 4,byrow = T,
    dimnames = list(c("Model 1: consMat[, 1] ~ Lags(consMat[, 1], 1:1) + Lags(
    consMat[, 2], 1:1)","Model 2: consMat[, 1] ~ Lags(consMat[, 1], 1:1)"),c("
    Res.Df","Df","F","Pr(>F)")))
htmlTable(t1,caption = paste("Table 4.21: Granger causality test - Electricity
    consumption ~ GDP"))
#Test 2
#grangertest(consMat[,2]~consMat[,1],order=1)
t2 = matrix(c(88,"","","",89,-1,0.8911,0.3478),nrow = 2,ncol = 4,byrow = T,
    dimnames = list(c("Model 1: consMat[, 2] ~ Lags(consMat[, 2], 1:1) + Lags(
    consMat[, 1], 1:1)","Model 2: consMat[, 2] ~ Lags(consMat[, 2], 1:1)"),c("
    Res.Df","Df","F","Pr(>F)")))
htmlTable(t2,caption = paste("Table 4.22: Granger causality test - GDP ~
    Electricity consumption"))
#Test 3
#grangertest(consMat[,2]~consMat[,3],order=1)
t3 = matrix(c(88,"","","",89,-1,0.542,0.4636),nrow = 2,ncol = 4,byrow = T,
    dimnames = list(c("Model 1: consMat[, 2] ~ Lags(consMat[, 2], 1:1) + Lags(
    consMat[, 3], 1:1)","Model 2: consMat[, 2] ~ Lags(consMat[, 2], 1:1)"),c("
    Res.Df","Df","F","Pr(>F)")))
htmlTable(t3,caption = paste("Table 4.23: Granger causality test - GDP ~
```

```
    Electricity infrastructure"))
#Test 4
#grangertest(consMat[,3]~consMat[,2],order=1)
t4 = matrix(c(88,"","","",89,-1,8.5413,"0.004412 **"),nrow = 2,ncol = 4,byrow =
    T,dimnames = list(c("Model 1: consMat[, 3] ~ Lags(consMat[, 3], 1:1) + Lags(
    consMat[, 2], 1:1)","Model 2: consMat[, 3] ~ Lags(consMat[, 3], 1:1)"),c("
    Res.Df","Df","F","Pr(>F)")))
htmlTable(t4,caption = paste("Table 4.24: Granger causality test - Electricity
    Infrastrucutre ~ GDP"))
#Test 5
#grangertest(consMat[,1]~consMat[,3],order=1)
t5 = matrix(c(88,"","","",89,-1,3.0435,"0.08455 ."),nrow = 2,ncol = 4,byrow = T,
    dimnames = list(c("Model 1: consMat[, 1] ~ Lags(consMat[, 1], 1:1) + Lags(
    consMat[, 3], 1:1)","Model 2: consMat[, 1] ~ Lags(consMat[, 1], 1:1)"),c("
    Res.Df","Df","F","Pr(>F)")))
htmlTable(t5,caption = paste("Table 4.25: Granger causality test - Electricity
    Consumption ~ Electricity Infrastrucutre"))
#Test 6
#grangertest(consMat[,3]~consMat[,1],order=1)
t6 = matrix(c(88,"","","",89,-1,1.545,0.2172),nrow = 2,ncol = 4,byrow = T,
    dimnames = list(c("Model 1: consMat[, 3] ~ Lags(consMat[, 3], 1:1) + Lags(
    consMat[, 1], 1:1)","Model 2: consMat[, 3] ~ Lags(consMat[, 3], 1:1)"),c("
    Res.Df","Df","F","Pr(>F)")))
htmlTable(t6,caption = paste("Table 4.26: Granger causality test - Electricity
    Infrastrucutre ~ Electricity Consumption"))
require(Gmisc)

#Optimal Model-Var
zvar1V = matrix(c(-8.395e-01,4.773e-01,-1.759,"0.0829 .",-7.446e-03,1.072e
    -02,-0.695,0.4895,8.302e-01,1.211e-01,6.858,"2.14e-09 ***",3.455e-01,5.511e
    -01,0.627,0.5327,2.821e-02,1.213e-02,2.324,"0.0230 *",2.320e-01,1.579e
    -01,1.469,0.1462,-2.956e-01,5.349e-01,-0.553,0.5822,-1.335e-03,1.278e
    -02,-0.104,0.9171,-1.857e-01,1.592e-01,-1.167,0.2472,1.751e-01,5.478e
    -01,0.320,0.7502,-3.345e-02,1.329e-02,-2.517,"0.0141 *",1.295e-01,1.591e-01
    ,0.814,0.4183,6.419e-01,4.820e-01,1.332,0.1872,1.619e-02,1.208e
    -02,1.340,0.1844,-9.922e-02,1.226e-01,-0.810,0.4209,-4.460e+02,1.136e
    +03,-0.392,0.6959,0.989,"","","",1149.379,"","",""),nrow = 18,ncol = 4,byrow
    = T,dimnames = list(c("Consumption.l1","Growth.l1","Expenditure.l1","
    Consumption.l2","Growth.l2","Expenditure.l2","Consumption.l3","Growth.l3","
    Expenditure.l3","Consumption.l4","Growth.l4","Expenditure.l4","Consumption.
    l5","Growth.l5","Expenditure.l5","constant","Adjusted R-squared","AIC"),c("
```

```
    Estimate","Std. Error","t value","Pr(>|t|) ")))
htmlTable(zvar1V,caption = paste("Table 4.27 Optimal model of Econometric model"
    ))

#Appendix
zvar1A = matrix(c(-8.395e-01,4.773e-01,-1.759,"0.0829 .",-7.446e-03,1.072e
    -02,-0.695,0.4895,8.302e-01,1.211e-01,6.858,"2.14e-09 ***",3.455e-01,5.511e
    -01,0.627,0.5327,2.821e-02,1.213e-02,2.324,"0.0230 *",2.320e-01,1.579e
    -01,1.469,0.1462,-2.956e-01,5.349e-01,-0.553,0.5822,-1.335e-03,1.278e
    -02,-0.104,0.9171,-1.857e-01,1.592e-01,-1.167,0.2472,1.751e-01,5.478e
    -01,0.320,0.7502,-3.345e-02,1.329e-02,-2.517,"0.0141 *",1.295e-01,1.591e
    -01,0.814,0.4183,6.419e-01,4.820e-01,1.332,0.1872,1.619e-02,1.208e
    -02,1.340,0.1844,-9.922e-02,1.226e-01,-0.810,0.4209,-4.460e+02,1.136e
    +03,-0.392,0.6959 ,0.989,"","","",1403.971,"","",""),nrow = 18,ncol = 4,
    byrow = T,dimnames = list(c("Consumption.l1","Growth.l1","Expenditure.l1","
    Consumption.l2","Growth.l2","Expenditure.l2","Consumption.l3","Growth.l3","
    Expenditure.l3","Consumption.l4","Growth.l4","Expenditure.l4","Consumption.
    l5","Growth.l5","Expenditure.l5","constant","Adjusted R-squared","AIC"),c("
    Estimate","Std. Error","t value","Pr(>|t|) ")))
htmlTable(zvar1A ,caption = paste("Table 4.27.1 Optimal model of Econometric
    model - Electricity Infrastructure"))
zvar1G = matrix(c(9.933e+00,4.990e+00,1.990,"0.050397 . ",7.161e-01,1.121e
    -01,6.390,"1.51e-08 ***",6.852e-01,1.266e+00,0.541,0.589958,-4.403e+00,1.266
    e+00,-0.764,0.447292,1.296e-01,1.269e-01,1.021,0.310668,-2.225e-01,1.651e
    +00,-0.135,0.893214, 1.316e+00,5.593e+00,0.235,0.814613,4.135e-02,1.337e
    -01,0.309,0.757933, -5.780e-02,1.664e+00,-0.035,0.972397,7.120e+00,5.727e
    +00,1.243,0.217867,4.949e-01,1.389e-01,3.562, "0.000662 ***",-2.734e
    +00,1.664e+00,-1.643,0.104798,-5.404e+00,5.040e+00,-1.072,0.287187,-3.856e
    -01,1.263e-01,-3.053,"0.003184 **",2.750e+00,1.281e+00,2.146,"0.035304 *"
    ,-2.253e+04,1.188e+04,-1.896,"0.061995 .",0.9993,"","","",1812.368,"","","")
    ,nrow = 18,ncol = 4,byrow = T,dimnames = list(c("Consumption.l1","Growth.l1"
    ,"Expenditure.l1","Consumption.l2","Growth.l2","Expenditure.l2","Consumption
    .l3","Growth.l3","Expenditure.l3","Consumption.l4","Growth.l4","Expenditure.
    l4","Consumption.l5","Growth.l5","Expenditure.l5","constant","Adjusted R-
    squared","AIC"),c("Estimate","Std. Error","t value","Pr(>|t|) ")))
htmlTable(zvar1G,caption = paste("Table 4.27.2 Optimal model of Econometric
    model - Economic Growth"))
summary(zvar1)
#AIC(zvar1$varresult$Consumption)
#AIC(zvar1$varresult$Growth)
#AIC(zvar1$varresult$Expenditure)
```

```
## Regression Analysis

dataC1<- dataCd[,1:3]
#model
fitreg <- lm(ElecConsumption ~ GDP + CapExpediture, data=dataC1)
#plotting the EC vs GDP and passing a straight through the datapoints to
    estbablish a relationship
plot(ElecConsumption ~ GDP, data=dataC1,
ylab="Electricity consumption",main="")
title(sub = "Figure 4.39: Electricity Consumption vs growth")
plot(ElecConsumption ~ CapExpediture, data=dataC1,
ylab="Electricity consumption",main = "")
title(sub = "Figure 4.40: Electricity Consumption vs Electricity Infrastructure"
    )
#Showcase the model estimates and model statistics
#l=summary(fitreg)
l1 = matrix(c(3.014882e+03,1.874073e+02,16.087323,4.437728e-28,2.923763e
    -03,7.741674e-04,3.776655,2.865813e-04,-5.537457e-02,3.291619e
    -02,-1.682290,9.601820e-02,0.3921537,"","",""),nrow = 4,ncol = 4,byrow = T,
    dimnames = list(c("Intercept","GDP","Electricity Infrastructure","Adjusted R
    -squared"),c("Estimates","Standard Error","t value","p-value(>|t|")))
htmlTable(l1,caption = paste("Table 4.28: Regression Analysis of model-
    ElecConsumption ~ GDP + CapExpediture"))
#polyniam model
require(MASS)
require(stargazer)
fitl = lm(ElecConsumption ~ GDP + I(GDP^2) + CapExpediture,data = dataC1)
#infrastructure look like a hyperbolic reltionship
fitl2 = lm(ElecConsumption ~ GDP + I(GDP^2) + CapExpediture + I(CapExpediture
    ^-1),data = dataC1)
#AICmin = stepAIC(fitl)
#AICmin2 = stepAIC(fitl2)
fl1 = matrix(c(2.501e+03,8.197e+01,30.506,"<2e-16***",5.773e-03,3.860e
    -04,14.954,"<2e-16***",-3.312e-09,3.917e-10,-8.455,"5.42e-13 ***",-3.344e
    -02,1.353e-02,-2.472,"0.0154 *",0.8451,"","","",991.35,"","",""),nrow = 6,
    ncol = 4,byrow = T,dimnames = list(c("Intercept","GDP","GDP^2","Electricity
    Infrastructure","Adjusted R-squared","AIC"),c("Estimates","Standard Error","
    t value","p-value(>|t|")))
htmlTable(fl1,caption = paste("Table 4.29: ElecConsumption ~ GDP + I(GDP^2) +
    CapExpediture"))
```

```
fl2 = matrix(c(1.837e+03,1.884e+02,9.751,"1.30e-15 ***",7.783e-03,6.333e
    -04,12.290,"<2e-16 ***",-4.484e-09,4.744e-10,-9.451,"5.36e-15 ***",-4.542e
    -02,1.295e-02,-3.506,"0.000721 ***",1.004e+05,2.606e+04,3.853,"0.000223 ***"
    ,0.8662,"","","",978.86,"","",""),nrow = 7,ncol = 4,byrow = T,dimnames =
    list(c("Intercept","GDP","GDP^2","Electricity Infrastructure","Electricity
    Infrastructure^-1","Adjusted R-squared","AIC"),c("Estimates","Standard Error
    ","t value","p-value(>|t|")))
htmlTable(fl2,caption = paste("Table 4.30: ElecConsumption ~ GDP + I(GDP^2) +
    CapExpediture + I(CapExpediture^-1)"))
#Testing the assumptions
par(mfcol=c(2,2))
plot(fitl2)
title(sub = "Figure 4.41: Model Assumptions",outer = TRUE)
mtext("Figure 4.41: Model Assumptions",outer = T,side = 1,line = -1)
#constant variance
#bptest(fitl2)
#DNR null hypothesis = constant variance hold
bp = matrix(c(6.8163,4,0.1459),nrow = 1,ncol = 3,byrow = T,dimnames = list(c("
    Values"),c("Test statistic","Degree of freedom","p-value")))
htmlTable(bp,caption = paste("Table 4.31: Breusch-Pagan"))
#normality
#shapiro.test(fitl2$residuals)
#DNR Null hypothesis= normality holds
sw = matrix(c(0.99348,0.9356),nrow = 1,ncol = 2,byrow = T,dimnames = list(c("
    Values"),c("Test statistic","p-value")))
htmlTable(sw,caption = paste("Table 4.32: Shapiro-Wilk"))

#Multicollinearity
require(car)
require(lmtest)
require(forecast)
#outliers
outlierTest(fitl2)

leveragePlots(fitl2,main = "")
title(main ="",sub = "Figure 4.39: Leverage Plots",outer = TRUE)
mtext("Figure 4.39 : Leverage Plots",outer = T,side = 1,line = -1)
#Influential

cutoff <- 4/((nrow(dataC1)-length(fitl2$coefficients)-2))
plot(fitl2, which=4, cook.levels=cutoff)
```

```
vif(fitl2) # variance inflation factors
sqrt(vif(fitl2)) > 2
#redo it after multicollinearity with appropriate model
fitl3 = lm(ElecConsumption ~ I(GDP^2) + I(CapExpediture^-1),data = dataC1)
vif(fitl3) # variance inflation factors
sqrt(vif(fitl3)) > 4
summary(fitl3)
bptest(fitl3)
shapiro.test(fitl3$residuals)
extractAIC(fitl3)
stepAIC(fitl3)
```

Rcode.R