Modelling and forecasting the costs of attending to electricity faults using univariate and multivariate time series forecasting models

UNIVERSITY OF
KWAZULU - NATAL

INYUVESI
YAKWAZULU-NATALI

Nkosiyapha Mthunzi Buthelezi

February, 2018

Modelling and forecasting the costs of attending to electricity faults using univariate and multivariate time series forecasting models

by

Nkosiyapha Mthunzi Buthelezi

A thesis submitted to the

University of KwaZulu-Natal

in fulfilment of the requirements for the degree

of

MASTER OF SCIENCE

in

STATISTICS

Thesis Supervisor: Dr. Oliver Bodhlyera

UNIVERSITY OF KWAZULU-NATAL

SCHOOL OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE

PIETERMARITZBURG CAMPUS, SOUTH AFRICA

Declaration - Plagiarism

I, Nkosiyapha Mthunzi Buthelezi, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowlegded as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then
   (a) their words have been re-written but the general information attributed to them has been referenced, or
   (b) where their exact words have been used, then their writing has been placed in italics and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the reference sections.

Nkosiyapha Mthunzi Buthelezi (Student) — Date: 3 August 2020

Dr. Oliver Bodhlyera (Supervisor) — Date: 3 August 2020

Disclaimer

This document describes work undertaken as a Masters programme of study at the University of KwaZulu-Natal (UKZN). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institution.

Acknowledgements

First and foremost, I would like to thank the Almighty God for giving me strength throughout the two years working on my Master's project.

I would like to extend my heartiest gratitude to my supervisor, Dr. Oliver Bodhlyera for his patience, continuous supervision, guidance, advice, and support in completing this thesis, you have been a tremendous mentor for me. He is actually the backbone of this project by giving wonderful suggestions and constructive criticism which gave this study a life. I would like to express my humble gratitude to my supervisor for his willingness to spare me his time and guide me to complete my study, your advices on both research as well as on my career have been so incredible. If not for his support and assistance, completing this dissertation would not be successful.

Mostly, I would like to dedicate this dissertation to my beloved parents, Mr. and Mrs. Buthelezi. They have always visionalised my success and sacrificed their life to raise me to be a better person. To my parents, I hope I have made both of you proud. I also would like to convey my appreciation to my lovely siblings, Nonkululeko, Nosipho and Mthobisi for their endless love,motivation, support and prayers as well as my pastors back home.

Finally, a huge thank you to everyone who has been a part of my life directly or indirectly supported and prayed for me along the way.

Contents

List of Figures

List of Tables

Abstract

Electricity price forecasting has turned into a very essential element for both public and private decision making. Both shortage of supply of electricity and electricity cost still remains the country's most biggest problems and needs to be addressed decisively. Apart from the demand and supply side of electricity, electricity cost is an important part of electricity delivery. Therefore, the accurate estimation of electricity cost and it's maintenance is an important part of the country's electricity supply strategy.

The main aim of this study is to forecast the cost of rectifying or attending to electricity faults. The study demonstrates that the AutoRegressive Integrated Moving Average (ARIMA), AutoRegressive Integrated Moving Average with exogenous variables (ARIMAX), Vector AutoRegressive (VAR) and Random Forest methods are capable of producing accurate forecasts of costs associated with attending to reported faults.

In this study, we analyse the costs of attending to electrical faults in the Bethlehem and Bloemfontein areas of the Free State region of South Africa, from 4 January 2012 to 3 June 2017, using univariate and multivariate ARIMA, ARIMAX, VAR and Random Forest models. ARCH and GARCH models are also used to model the volatility found in the daily costs data. The model developed based on these data can be used to forecast future faults costs and can help policy makers with planning decisions.

Chapter 1

Introduction

Electricity is the backbone for an economy's prosperity and progress because it plays an important role in socio economic development. Uses of electricity are rapidly increasing day by day, leading to a tremendous advancement in human civilization. The demand for electricity that leads to the increment of the cost of electricity is a vital topic to study. This chapter covers the initial and introductory parts of this study. It gives the basis for definitions used in the study. This research is inspired by Nakiyingi (2016), who did almost similar work with South Africa's daily electricity demand.

This study focuses most heavily on the time series forecasting models that are capable of predicting the future of South Africa's electricity status.

1.1 Background

For all societies, electricity has become an indispensable commodity. Commonly, it is traded in a market in which, due to the process of deregulation and the existence of competition within markets, it's price oscillates according to supply and consumer demand, the prediction of price has been discovered as being important, not only for energy companies (such as suppliers, power transmission operators and retailers) but also for all types of market participants that include traders and investors. Energy consumption is increasing all around the world due to growth of the economy, population and industrialisation. South Africa has also gone through considerable economic and social growth in past years and this growth has caused an increase in energy costs, especially electricity cost (Sigauke and Chikobvu, 2011).

According to the World Health Organization (2011), around three billion people do

not have access to modern fuels for cooking, heating, use of traditional stoves burning biomass (animal dung, wood and crop waste) and coal resulting in about four million pre-mature deaths every year. The impact of electricity on human life is very strong and therefore, various studies have been made in different directions related to this sector.

Forecasting electricity cost, forms a very important part of the energy policy of a country, morely for a developing country like South Africa whose electricity costs have increased gradually over the past few years. There are some factors that impact the daily energy costs, among them are interest rates, inflation, calendar effects, economic factors, meteorological factors and grid maintenance costs. Electricity cost is affected by different factors in different countries. For example, highly industrialized countries have a higher cost for electricity than low industrialized countries. Countries with minimal seasonal weather changes, usually have an almost similar cost of electricity across all seasons (Saab et al. 2001).

Other factors affecting electricity costs include, the supply of electricity, the faults in different areas, social factors and human activities. Since the cost of electricity keeps changing continuously in time, we consider it to be a time series. Therefore, quantitative methods are preferred in making forecasts about it, specifically, time series techniques, depending on the naturalness of the available data.

1.2 Problem Statement

South Africa is at the moment going through an electricity crisis. The shortage of electricity supply and cost remains one of the country's most critical challenges going forward. Costs for electricity in South Africa is forever going up. A rapid increase in the population, economic expansion, industrialization and other factors have led to an increment in the cost for electricity in South Africa. This study looks at costs of electricity in as far as attending to electrical faults is concerned. If costs of rectifying electrical faults can be modelled and properly estimated beforehand then planners will have clear picture of what might happen hence strategise accordingly.

1.3 Aim and Objectives

The aim of this study is to,

- Come up with the most accurate univariate and/or multivariate time series model(s) best suited for electricity fault response costs, modelling, forecasting

and assess if such model(s) provide accurate results and forecast the electricity maintenance cost in South Africa.

The objectives of this study are to,

- Compare various time series models that are suitable for forecasting the electricity maintenance data available in order to come up with a model with the best forecasting capabilities.
- Model and forecast the costs of attending to electricity faults using time series models and other methods.

Electricity cost is affected by different factors in different countries. For example, highly industrialised countries have a higher cost for electricity than low industrialised countries. Other factors affecting electricity cost include, the supply of electricity, social factors and human activities. Since electricity cost keeps changing continuously in time, we consider it to be a time series dataset. Therefore, quantitative methods are choosen in making forecasts about it, specifically, time series techniques, depending on the nature of the available data.

This study will focus on modelling and forecasting the costs of attending to electricity faults using univariate, multivariate time series forecasting models and series volatility forecasting models. Multivariate time series models will be used because of the available data that is of multivariate nature with more that two variables. For that reason, we shall follow studies such as Bruce et al., (2013) who found it more appropriate to use multivariate forecasting techniques to reach the objectives of their studies and nakiyingi (2016) found it more appropriate to use univariate forecasting methods to reach the objectives of her studies.

Hence, since the aim is come up with the most accurate univariate and/or multivariate time series model(s), we start with (univariate) time series models, then focus on the multivariate models.

1.4 Forecasting procedure

Forecasting process is a statistical planning tool that aids the management to cope with the uncertainty of the future, relying on data from the past and present, and the analysis of trends. Forecasting can also be described as the procedure of making predictions of the future based on the past and present data and most commonly by

the analysis of trends (Nakiyingi, 2016). Forecasting begins with some assumptions based on the management's knowledge, experience and judgement. These estimates are projected to the coming months or years using one or more statistical tools such as the Delphi method, Box-Jenkins models, moving averages, regression analysis, exponential smoothing and trend projection.

Ku (2002) mentioned that forecasting the market price of electricity maintenance cost is a key factor for the decision makers in defining the short term operating schedules and strategies in the electricity markets. For instance, a transmission company wants to know the value of the future price of electricity maintenance cost to strategically bid into the market. Another example, a demand response market participant wants to know whether the price of electricity is low or high to optimize operation.

The process of forecasting is mainly used to plan, make budgets and estimate future growth. Businesses use forecasting procedure to define how to allocate their budgets or plan for anticipated expenses for an upcoming period of time. Investors utilise forecasting to determine if the events that affect a company such as sales expectation, will rapidly increase or decline the price of shares in that certain company. Forecasting process provides a significant benchmark for forms, which needs a long term perspective of operations. Stocks analysts use forecasting to abstract how trends, such as unemployment or GDP will change in the coming period, quarter or year (Box et al., 2015).

According to previous studies such as Hyndman and Athanasopoulos (2014), a forecast variable has never reached 100% accuracy. Various studies show that group forecasts are much better than individual forecasts, for instance, forecasting the average performance of the whole class using results from their mid term exams is much better than forecasting one student's performance using their previous mark. There is a common approach used when forecasting, it depends at the problem at hand, it can also be applied to support their strength and reduce their individual weaknesses.

The steps to be followed when forecasting as described by Hyndman and Athanasopoulos (2014) include :

- Defining the problem of the study : One needs to carefully define the problem according to who needs the forecasts, how the forecasts will be used, how the determined model fits the data at hand. At this stage it is essential to be aware that any decision made based on the results from the forecasts will affect the

future of the organization.

- Time : One needs to know how much lead into the future the forecast should cover. Short term forecasts typically cover a period of less than 1 year. Forecasts in the electricity sector are useful in estimating load flows in order to make decisions that will prevent load shedding. Medium forecasts take 1 to 3 years, and are for determining and planning for future resource requirement. Long term forecast take longer than 3 years and are for strategic planning and development.
- Collection of data : Data collection needs a lot of time and, after collecting the data one requires to know the nature or behaviour of that data. A time series plot is the best way to check if there are any patterns, like trend, seasonality, cycles and their significance. Plotting also helps pick out outliers and their meaning. The data needs to be cleaned first before it is used for any analysis or, in the development of forecasting models from which the best is chosen using accuracy measurements.

In this case, we do not choose the best model depending on its AIC, AICc, or BIC because, comparing models using these information criteria requires that all models have similar orders of differencing, which is not the case for the seasonal differencing in this study.

To estimate the model order, we look at the behaviour of the ACF and PACF.

1.5 Literature review

The study on electricity maintenance cost forecasting has been there for years and years, and a numbers of research results have been used by electricity companies. After realizing the continuous increase in electricity costs, developed countries have opted for deregulation which encourages using other power sources like solar and wind energy. Through deregulation, users get a variety of options to purchase and use electricity, for example, the use of panels and turbines. However, the use of these other sources makes the load forecasting problem harder because the production of these power sources cannot be predicted easily since they mainly depend on the weather (Hong and Dickey, 2014).

The process of forecasting electricity cost is hard for developed countries but harder for developing countries because of various factors like lack of necessary historical

data, inadequate expertise and institutions to carry out the process with appropriate models. Developed countries mainly face problems like inappropriate assumptions made by experts while constructing the models. Due to such situations, the deviation between predicted and actual electricity cost seems to be a worldwide problem, irrespective of the level of development (Yasmeen and Sharif, 2014). In this chapter, we look at various studies that have been carried out in order to forecast electricity cost with deviation from the actual cost being as minimum as possible.

1.5.1 Overview of electricity maintenance cost forecasting methods

According to Bhattacharyya and Timilsina (2009), electricity maintenance cost forecasting is now one of the most important aspects in the electricity sector, especially for supply planning. Positive world economic trends have attracted lot of stakeholders to invest time and money for the development of new algorithm for precise prediction of electricity cost. This financial aspect has drawn enormous interest to many researchers, and has yielded lot of important research and contribution in electricity maintenance cost forecasting. Precise knowledge about the future electricity maintenance cost will help consumers to plan their consumption and suppliers to plan their production based on user behaviour. This research explains some of the most known algorithms, tools and research proposed to date, in the field of electricity maintenance cost forecasting.

Short term electricity maintenance cost forecasting (STPF)

Short term forecasting is essential for fast decision making processes. Markets can plan their strategies using the forecast cost to maximize the profit in deregulated markets. Consumers can make decision of power consumption based on the current and predicted future costs. Short term prediction involves next hour maintenance cost predictions or day ahead maintenance cost predictions (Kodogiannis, 2002).

Medium term electricity maintenance cost forecasting (MTPF)

Medium term maintenance cost forecasting involves predictions for week's, month's up to a year's lead time. In pricing schemes, MTPF is being affected by seasonal effects, like an increase in electricity maintenance cost in winter or decline during holidays or summer. In the smart grid, the deployment of other sources of energy might also influence electricity maintenance cost. Medium term maintenance cost forecasts can be used by suppliers to maximize their production cost by planning the resource allocation for the generation of electricity (Dudek, 2010).

Long term electricity maintenance cost forecasting (LTPF)

The time horizon for long term maintenance cost forecasting varies from years to decades. Long term maintenance cost trends are muchly used by policy makers to plan pricing schemes and for the management of resources. Investors use it for analysing recovery of investment in transmission, power plant construction and production. Based on the type of production, maintenance cost forecasts can be used to preserve resources. For example, a nuclear plant can plan on how and when to purchase uranium and hydropower plants can consider the construction of reservoirs while solar and wind farms can make advance analysis on the development of new plants based on cost analysis (maintenance cost and Sharp, 1986).

1.6 Existing forecasting techniques

Due to the significance of accurate maintenance cost forecasting in electricity market, a number of techniques have been demonstrated in the literature review. These techniques range from traditional time series analysis to machine learning techniques for forecasting future maintenance cost. ARIMA and GARCH models are examples of traditional techniques. Artificial Neural Network (ANN), ARIMA models enhanced with wavelet transforms, Markov models, random forests, fuzzy inferred neural networks, and support vector regression are some examples of machine learning methods that have been studied (Hamilton and Douglas, 1994). We will analyse few of these methods.

1.6.1 ARIMA

In other studies, the authors used ARIMA model based on wavelet transformation for electricity maintenance cost forecasting. Historical data was splitted using some transformation before the applying ARIMA modelling. The forecast results were obtained by application of inverse transformations. An AutoRegressive Integrated Moving Average (ARIMA) model are also called Box-Jenkins models because they were developed by Box and Jenkins (1976). They used this model to forecast one day ahead electricity maintenance cost for the German market.

1.6.2 Artificial Neural Network(ANN)

Artificial Neural Network (ANN) is a very popular approach for short term forecasting, we use ANN model for forecasting cost in the next hour using the input data with designed features. The ANN generally gives accurate results provided that the ANN is trained with the correct set of input features and enough input data points

(Dudek, 2013). The authors presented the ANN model for maintenance cost prediction by using history and other factors estimated in the future to fit and abstract the maintenance cost and quantities.

The authors present a combined model that include orthogonal experimental, probability and neural networks designs for electricity maintenance cost prediction. They implemented the PNN model combined as a classifier that is consist of advantages of being a fast learning process as it needs a single pass network training stage for adjusting weights. Orthogonal experimental design was used to get the optimal smoothing parameter which helps to increase prediction accuracy (Dudek, 2010).

1.6.3 Support Vector Machine (SVM)

Authors proposed SVM model for maintenance cost forecasting by using historical maintenance cost, demand data and Projected Assessment of System Adequacy (PASA) data as input variables. They did their experiment using Australian National Electricity Market (NEM) using the New South Wales regional data at year 2002. The authors also proposed model for time series prediction using Fish Swarm Algorithm (AFSA) for choosing the parameter of SVM model. This model used Optimized Support Vector Machine for electricity maintenance cost forecasting (Dudek, 2010).

1.6.4 ARCH AND GARCH models

ARCH and GARCH models have become important tools in the analysis of time series data. These models are especially useful when the goal of the study is to analyze and forecast volatility. The simpler ARCH models will be considered first to provide a systematic framework for volatility modelling. It was developed in 1982 by economist Robert F. Engle (Engle,1982). The acronym ARCH stands for AutoRegressive Conditional Heteroscedasticity. The AR comes from the fact that this model is a type of autoregressive model. Heteroscedasticity means non constant variance. However, with an ARCH model, it is not the variance itself that changes with time, rather, the conditional variance.

Due to the limitations presented by the ARCH models, a better model was proposed by Bollerslev in 1986 (Bollerslev, 1986) in order to solve the problem of requiring many parameter to adequately describe any given data while using an ARCH model. It is called the Generalised AutoRegressive Conditional Heteroskedasticity (GARCH) model. Multivariate GARCH models have been used to investigate

volatility and correlation transaction and spillover effects studies.

1.6.5 Random Forest

Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of features to split each node yields error rates that compare favorably to boost, but are more robust with respect to noise. Internal estimates monitor error, strength, and correlation and these are used to show the response to increasing the number of features used in the splitting. Internal estimates are also used to measure variable importance. An example is random split selection where at each node the split is selected at random from among the K best splits (Freund and Schapire, 1996).

1.7 Relevant studies on electricity consumption and it's costs

Saab et al. (2001) modelled and forecasted electricity consumption and cost in Lebanon using univariate approaches, three univariate techniques were used to model and forecast electricity cost namely, AutoRegressive (AR), ARIMA models and an aggregation of an AR(1) with a high pass filter (AR(1)/HPF). The main purpose of their study was to look into different univariate models and use them to forecast one month ahead electricity consumption and cost in Lebanon. The interest was in identifying a forecasting method that would perform best on the mixed data with missing values that was available.

This was a vital study because electricity had become the main source of energy in all the economic sectors of Lebanon. It was critical to forecast cost to help in the development of that sector and the country at large. Monthly average electricity cost data was used, covering January 1970 until May 1999. A time series plot revealed an evident non-continuous behaviour between January 1975 to December 1989. This was attributed to the civil war that took place during that time in Lebanon. However, since this civil war brought about random fluctuations in the power sector, which caused an unusual pattern in the consumption, data during that period was ignored. There was uncertainty about the war happening again, therefore, data used run from January 1990 to May 1999. Due to the odd stochastic characteristics of the data, an adequate model was vital to carry out the forecasting exercise (Yasmeen and

Sharif, 2014).

A non-linear deterministic model was used to represent the trend in data after the war, since from the ACF plot, the data was a non-stationary random process. After the analysis, there were insignificant, almost uniform correlations in the ARIMA model, for all the positive lags, with the 39th lag having a 0,057 standard deviation and maximum correlation of 0.129. For the AR(1)/HPF model, there were diverse correlations in all positive lags, with the 4th lag having 0,091 standard deviation and 0.625 maximum correlation. Since it is necessary for residuals to be statistically uncorrelated for a reliable ARIMA model, and there were uncorrelated residuals for both the ARIMA and AR(1)/HPF models, the ARIMA model was ideal enough (Yasmeen and Sharif, 2014).

Assessment of each of the models was performed using sum of absolute errors (SAE), percentage mean absolute error (PMAE), sum of squared errors (SSE) and the percentage mean squared error (PMSE). Model performances were compared with the actual values and this indicated better forecasts from the AR(1)/HPF model, compared to both the AR and ARIMA models (Yasmeen and Sharif, 2014).

An application of linear models applied in this study was carried out by Kumar and Anand (2014). They used time series ARIMA forecasting models to forecast the electricity maintenance cost in India. ARIMA models, also commonly known as Box-Jenkin's models were used in the study because they work best when forecasting single variables. The main reason for choosing ARIMA models for forecasting was because these model have the capabilities of making predictions using time series data with any type of pattern and the non zero autocorrelation between successive values of the time series data are taken into account (Dudek, 2010).

Data covering a period of 62 years of electricity maintenance cost was used to predict 5 years ahead. After modelling and analysing the data, an ARIMA(2,1,0) was chosen as the very best model explaining the patterns of the data perfectly. Attempts were made to forecast, as accurate as possible, the future electricity maintenance cost for a duration up to 5 years. Forecast results showed that the annual electricity maintenance cost would grow in year 2013, then take a sharp decline in 2014 and in years 2015 through 2017.

The study statistically tested that the successive residuals in the fitted ARIMA time series were not correlated, and the residuals seemed to be normally distributed with mean zero and constant variance. In conclusion, the selected ARIMA(2,1,0) was a

good predictive model for the electricity maintenance cost in India country. Similar to any other predictive models in forecasting, ARIMA also has limitations on accuracy of predictions, it is used widely for forecasting the future successive values in the time series (Kumar and Anand 2014).

In a case study of Dubai et al.,(2006) studied forecasting monthly peak load cost using time series models. In this study, an attempt to test and recommend reliable and accurate models of forecasting monthly peak load was carried out. Different time series models were developed to provide forecasts as accurate as possible. The univariate time series models used in the study involves a variety of complex methods, such as Box-Jenkins (BJ), exponential smoothing and dynamic regression.

The purpose was to yield short term monthly forecasts of one year ahead by analysing the behaviour of monthly peak loads. The study was carried out using Dubai data alone because other emirates refused to give timely data for reasons of secrecy and confidentiality. Data was used in two portions, for evaluation and validation of the performances of the models. Comparisons for how well the historical and forecast data for the holdout duration matched and correlated were also carried out. Such attempts reflected how the recommended models captured most of the characteristics of the data. In total, there were 267 cases available between 1985 and 2007. The data ranged from 296MW (January 1985) to 4113MW (August 2006) with a mean of 1395MW and a standard deviation of 862:3679MW. From the time series graphical representation of the data, there existed patterns of and trend. Cost was highest in July and August and lowest in January and February. A trend line equation was drawn, whose slope was estimated as 9:6643. This indicated a strong upward trend. The process used in the study followed seven steps, obtaining time series data, performing initial data screening to discover trend and seasonality, performing trend and seasonality analysis to identify data features, selecting time series models to use, analysing and obtaining results for each model with model performance statistics, performing out of sample diagnostics and validity tests, and lastly, recommending the final model. Through this process, different models were recommended, winters exponential smoothing and Box-Jenkins ARIMA model with root transform (Dubai et al., 2006).

The recommended models passed stringent diagnostic tests, including comparing outputs with selected holdout samples. A compregnation of the performance of the recommended models with those of electric authorities showed that the recommended model had better diagnostic results with the actual hold out sample. In conclusion, the developed model was recommended not only to the Dubai monthly

peak load data, but to other datasets showing seasonality and trends.

Table 1.1: Summary of the related literature on forecasting and modelling the electricity cost.

Authors	Data	Models
Saab et al. (2001)	01/1970-05/1999, Monthly electricity cost	AR, ARIMA models.
Kumar and Anand (2014)	Data for 62 years, Electricity maintenance cost	ARIMA models.
Dubai et al.(2006)	01/1985-08/2007, Monthly load cost	Box-Jenkins.

In conclusion, the developed models for each study were recommended not only to the electricity based data, but also to other data sets displaying seasonality and trends or any other kind of nature. These studies helped paint a clear picture of what the dissertation is going to be using which kind of models. The univariate time series models seem to be the best fit for the electricity maintenance cost study, but more accurately when multivariate models are used.

1.8 Research Layout

This study is organised in chapters that describe how each methodology considered is described and applied. A chapter by chapter outline of the study is given below.

Chapter 1 consists of initial and introductory parts of the whole study. It gives the definition of terms used in the study, an introduction to the research objectives and suggested statistical methods to be used. Chapter 1 also gives brief and simple definitions of univariate time series. Chapter 2 gives an explanatory description of the data and some statistical tools required to do time series analysis. In Chapter 3 and 4, the basic components of time series are described. Statistical tools that deal with uncertainty in the data are discussed. Non-linear models are required to define data that has changing variation over time. Chapter 5, here the study focuses on analysing the results using univariate ARIMA models, **R** software was used to plot the time series graph representing the data and came up with relevant results explaining the electricity data to determine the forecast and choose the accurate model to be used for our analysis. Chapter 6, focuses on volatility forecasting models for accurate forecasting. Chapter 7 applies ARCH and GARCH models to ARIMA model for forecasting electricity cost. Chapter 8 focuses on forecasting electricity using multivariate ARIMA, ARIMAX and VAR models. Chapter 9 introduces a new method Random Forest(RF) for forecasting electricity cost and chapter 10 is the discussion and conclusion.

Chapter 2

Exploratory Data Analysis(EDA)

Exploratory Data Analysis is a way used for analysis of data that allows a variety of techniques in most of the times it is graphical and tabular representation to extract important information, test underlying assumptions and help to develop parsimonious models. The EDA approach is a preliminary process about how analysis of data should be done. It was promoted by Rosenthal (1995) to encourage statisticians to visually examine the data sets at hand and to formulate hypotheses that could be tested on the datasets. EDA is a very crucial first step in analysis of any kind of data (Rosenthal, 1995). The main reasons why we use EDA are:

- To catch errors,
- To see patterns in the data,
- Checking of assumptions,
- To find violations of statistical assumptions,
- To generate hypotheses, and
- Defining relations between the explanatory variables.

2.1 Data description

In this research study, electricity cost data that has been composed over a duration of 6 years from 2012 to 2017 from eskom was used. The data is consist of 13 variables in total but I chose to focus only on the four relevant ones namely, transaction cost variable, travel cost variable, travel time variable and travel distance(km) variable. These four variables are the most relevant ones in the data since they give the better insights for forecasting the electricity data. The data had many observations such that it could not fit in R-studio, then I used (Structured Query Language)SQL in

order to cut and aggregated the data into fewer observations so that it can normally fit into **R**.

2.1.1 Data descriptive statistics

Statistics are suitable for defining the canonical components/attributes of the data. Descriptive statistics gives the basic sum-up of the data. As shown in the descriptive statistic tables below. A box plot gives a graphical summary of the distribution of a sample, it shows the central tendency, the shape and the variation of the electricity data. We use the box plot to investigate the stretch out of the electricity cost data and to distinguish any possible outliers and the box plot are very acuurate when the sample bulk is bigger than 20. With regards to each of the following plots, the upper part is the right spread of the data and the lower part is the left spread of the data.

1. Description of total transaction cost variable.

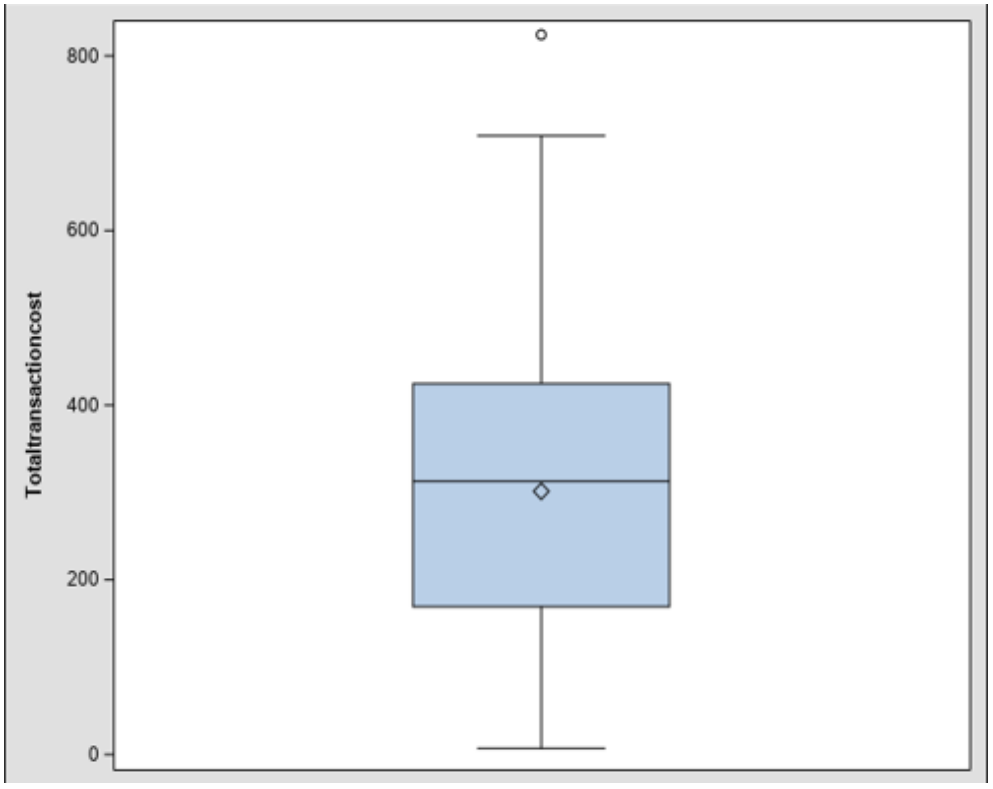

Figure 2.1: The box plot of transactional cost.

The box plot of transactional cost implies that the most of the data is left skewed and there is an outlier.

Table 2.1: The statistics of transactional cost.

Mean	Standard deviation	Standard error	Variance	Median	t-value	Min	Max
301.416	148.093	1.973	21931.583	124.933	152.732	7.031	824.527

Over-dispersion occurs when the variance of the data exceeds the mean which can lead to the occurrence of extreme values/outliers. On this variable there is over-dispersion based on the given values by the Table 2.1. However, it is clearly seen that the min and the max value is very far from the mean. Also, the standard deviation is very high. They represent that electricity cost of the areas/cities focused on are very different from each other. Since mean is greater than median, it shows right skewed distribution. Therefore, a transformation might be needed to satisfy the normality of the data. The difference between minimum and maximum values is also quite big.
2. Description of the travel cost variable.

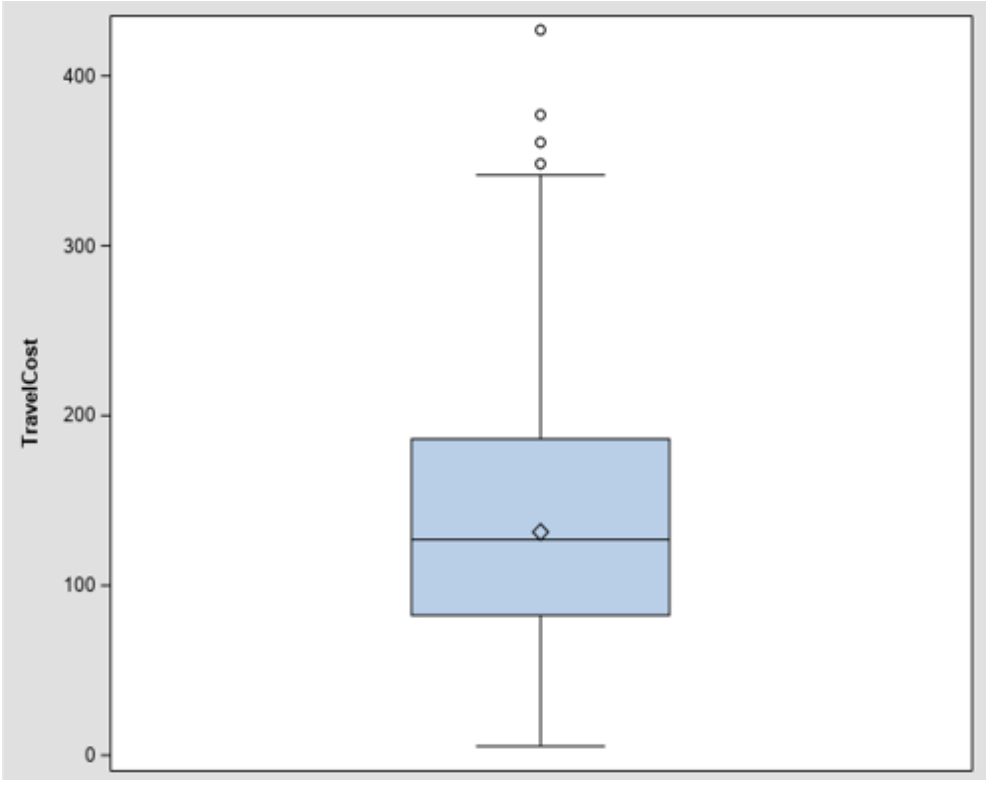

Figure 2.2: The box plot of travel cost.

The box plot of travel cost implies that the most of the data is right skewed and there are outliers.

Table 2.2: The statistics of travel cost.

Mean	Standard deviation	Standard error	Variance	Median	t-value	Min	Max
131.438	66.035	0.879	4360.622	126.937	149.352	5.210	427.011

Over-dispersion occurs when the variance of the data exceeds the mean which can lead to the occurrence of extreme values/outliers. On this variable there occurs over-dispersion based on the given values by the Table 2.2. However, it is clearly seen that the min and the max value is very far from the mean. They represent that electricity cost of the areas/cities focused on are also different from each other, even though the standard deviation is not too big. Since mean is greater than median, it shows right skewed distribution. Therefore, a transformation might be needed to satisfy the normality of the data. The difference between minimum and maximum values is also very big.

3. Description of travel time variable.

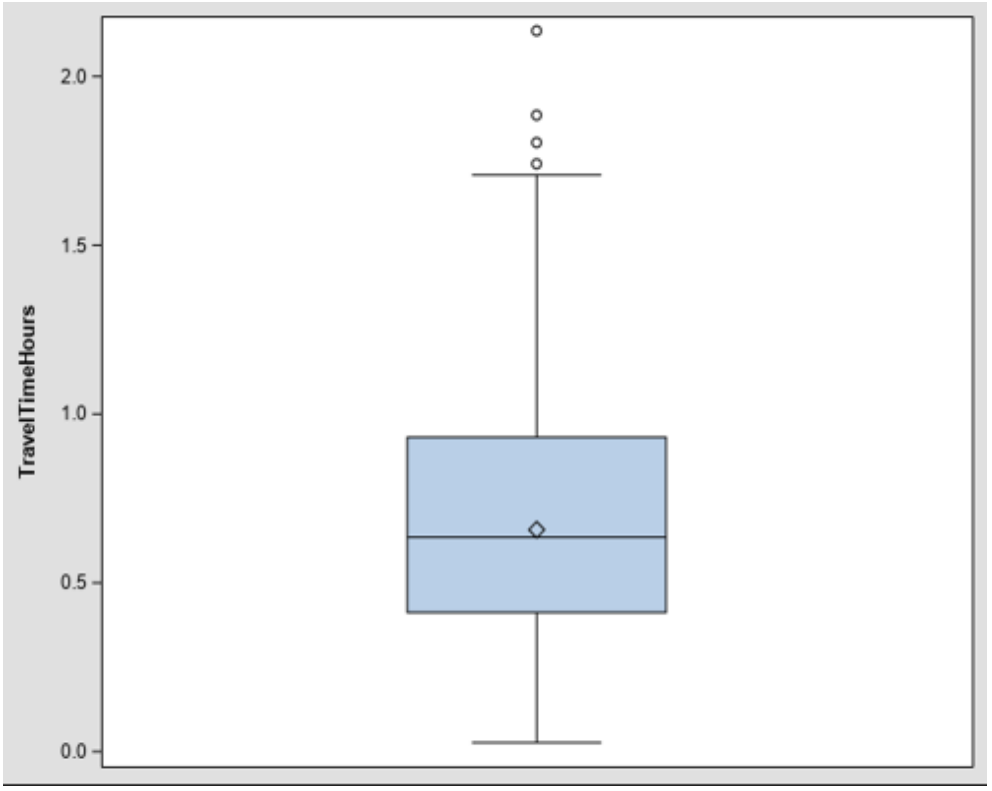

Figure 2.3: The box plot of travel time.

The box plot of travel time implies that the most of the data is right skewed and contains outliers.

Table 2.3: The statistics of travel time.

Mean	Standard deviation	Standard error	Variance	Median	t-value	Min	Max
0.657	0.330	0.004	0.109	0.633	50.254	0.033	2.147

Over-dispersion occurs when the variance of the data exceeds the mean which can lead to the occurrence of extreme values/outliers. On this variable there is no over-dispersion based on the given values by the Table 2.3. However, it is clearly seen that the min and the max value is very far from the mean. They represent that electricity cost of the areas/cities focused on are also different from each other. The standard deviation is quite good. Since mean is greater than median, it shows right skewed distribution. Therefore, a transformation might be needed to satisfy the normality of the data. The difference between minimum and maximum values is big.
4. Description of total travel distance(km) variable.

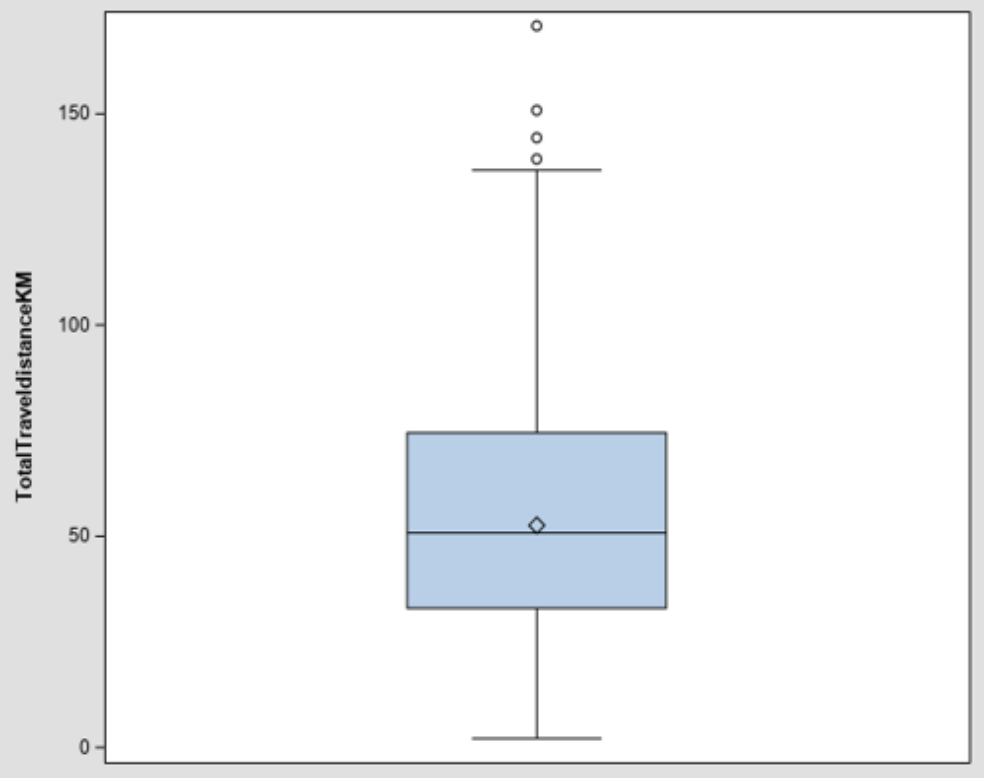

Figure 2.4: The box plot of total travel distance(km).

The box plot of total travel distance implies that the most of the data is right skewed and the outliers.

Table 2.4: The statistics of total travel distance.

Mean	Standard deviation	Standard error	Variance	Median	t-value	Min	Max
52.571	26.414	0.351	0.452	50.774	50.244	2.083	170.810

Over-dispersion occurs when the variance of the data exceeds the mean which can lead to the occurrence of extreme values/ outliers. On this variable there is no over-dispersion based on the given values by the Table 2.4. However, it is also clearly visible that the min and the max value is very far from the mean. They represent that electricity cost of the areas/cities focused on are also different from each other. The standard deviation is big. Since mean is greater than median, it shows right skewed distribution. Therefore, a transformation might be needed to satisfy the normality of the data. The difference between minimum and maximum values is big.

2.2 Testing for normality

Assumption of normality means that you have to make it a point that the data at hand fits a bell curve shape before running Shapiro-Wilk test or regression.

Here, we used travel cost, total travel distance(km), travel time and transaction cost variables in the electricity data to check for normality by using the Q-Q plots. To determine normality by using graphs, we can use the outcomes/results of a normal Q-Q plot. If the data is normally distributed, that is when data points are close to the diagonal line. If data points stray from the diagonal line in a non-linear manner, then it means the data is not normally distributed (Ahad et al., 2011).

We can see from the normal Q-Q plots in figures below, the data looks normally distributed for travel cost, total travel distance(km) and travel time. As for total transaction cost, the data points move stray away from the line which results into a non-linear manner.

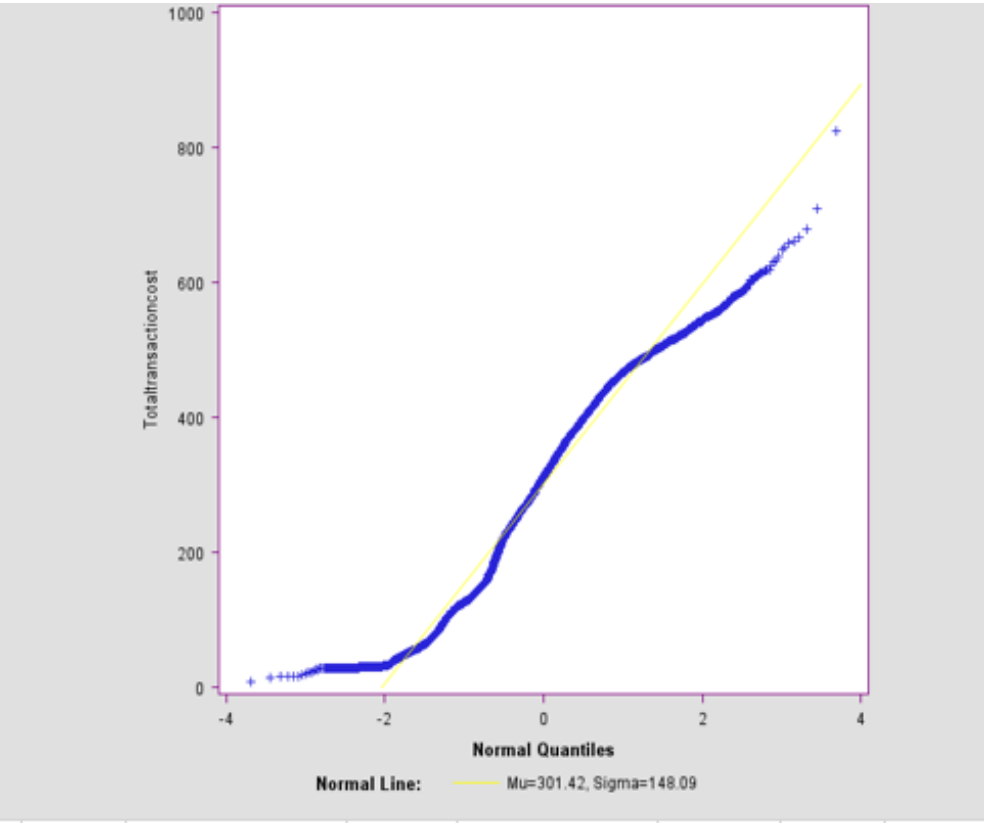

Figure 2.5: Q-Q graph for the transaction cost of electricity data.

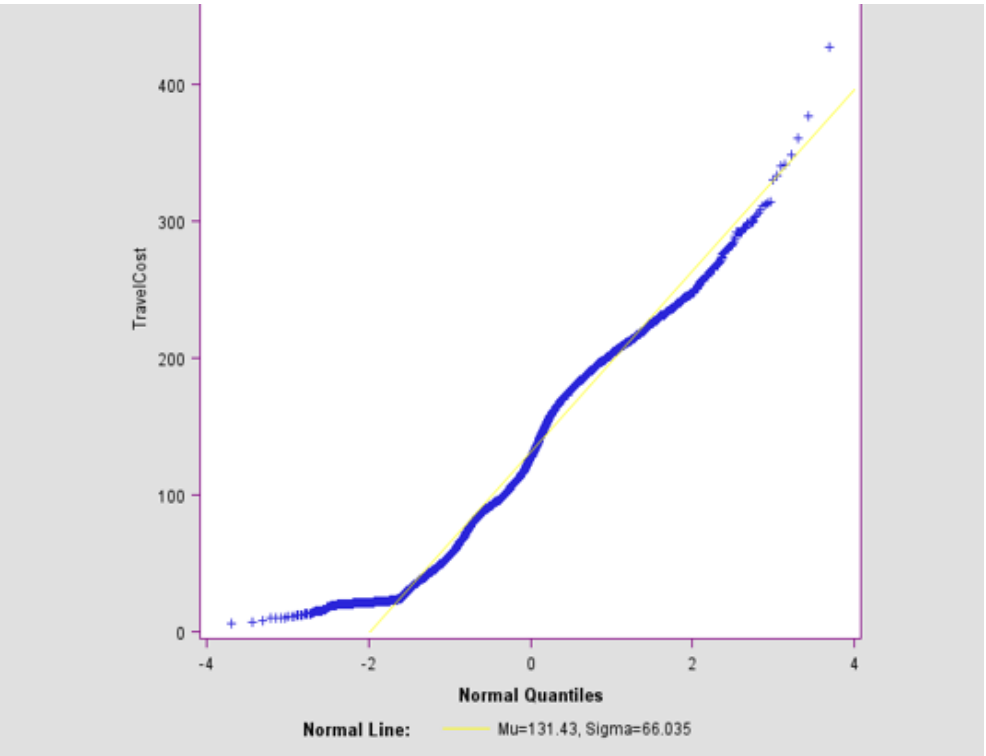

Figure 2.6: Q-Q graph for travel cost of electricity data.

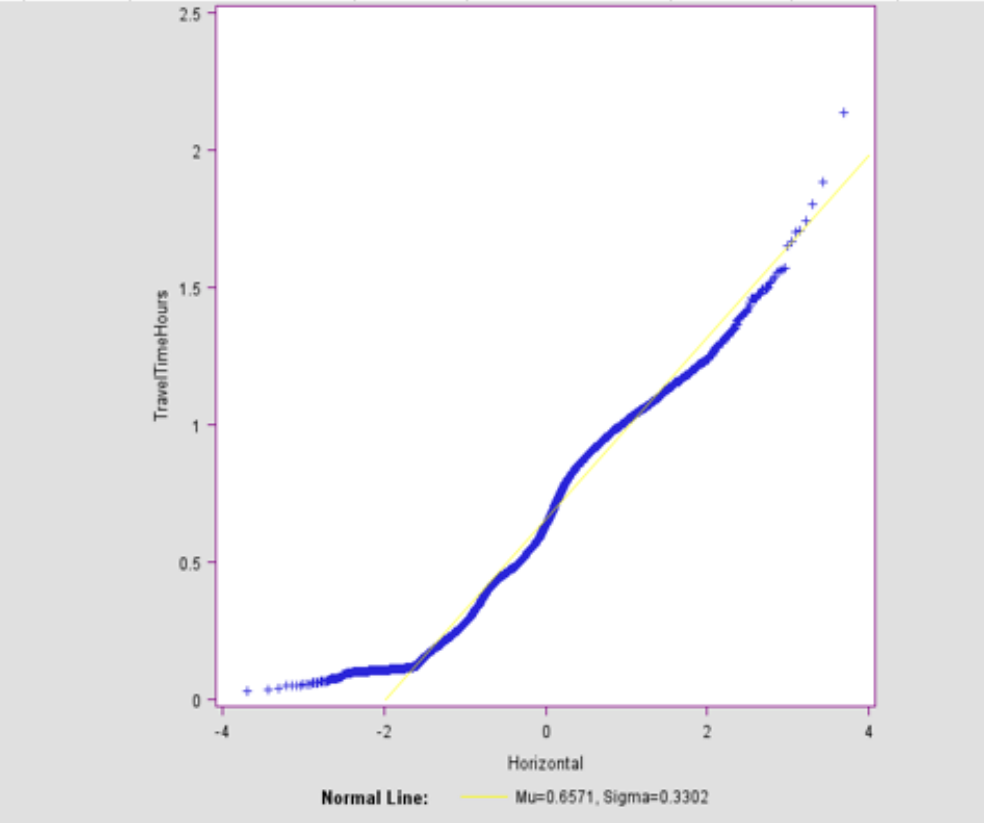

Figure 2.7: Q-Q graphical representation for travel time of electricity data.

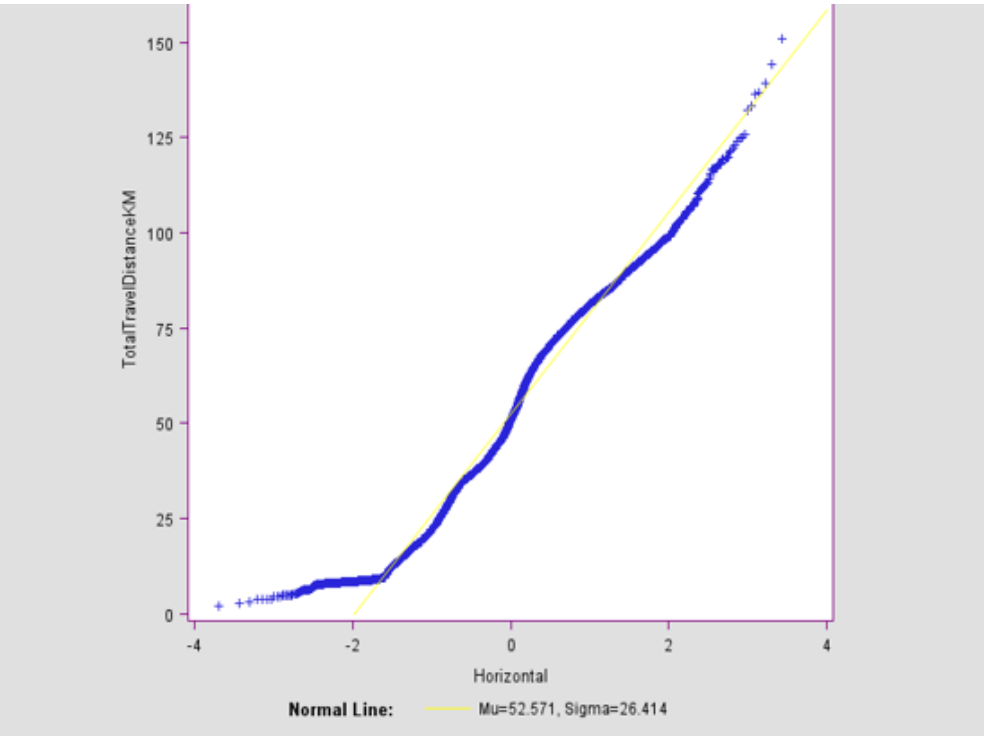

Figure 2.8: Q-Q graph for total distance(km) of electricity data.

2.2.1 Shapiro-Wilk Test

For testing the normality of the data, we use the Shapiro-Wilk test. After running this test in **R**, the used function returns a p-value of 0.00025, which is far less than 0.1. Hence, we reject H_0 and deduce that the data does not follow a normal distribution. The Q-Q plots in figures above also shows non normality of data because most of the data points move away from the main diagonally plotted line.

2.3 Conclusion

Fitting the normal distribution approximates the standard deviation and the mean(average) from the sample. By looking at the plots above, they are all of a poor fit since the tallness of the bars do not pursue the exact shape of the line. The data that well fit the normal distribution have the bars that are close enough to the line.

Chapter 3

The basic theoretical aspects of time series resolution

A time series is an order of observations on a particular variable. Time series modelling is an exploration area that has drawn attentions of investigators community for the past years. Usually the observations are taken at periodical intervals (days, months, years). The purpose of time series modelling is to congregate and study the elapsed observations of a time series to develop an accurate model which illustrate the structure of the series (Adhikari and Agrawal, 2013).

Time series analysis is used to either model randomness in a given data series or forecast future values basing on observed historical data (Brockwell et al., 2002). This chapter present an overview of some of the basic tools and concepts used to model and analyse time series data. Areas covered include, describing different features and patterns of time series data, the data being stationary, the process of differencing, Autocovariance, Autocorrelation functions (ACF) and partial Autocorrelation Functions (PACF) in detail.

A time series analysis consists of two steps:

- Construction of a model.
- Using the model to foretell (forecast) forthcoming values.

3.1 The general description of time series data

Univariate time series analysis involves using data about a single variable to build a model that illustrate the action of the variable in the past and the multivariate time series analysis involves using data about many variables to develop a model that

define the behaviour of the variable in the past (Brockwell et al., 2002).

The element of correlation within the data has to be taken into consideration. Suppose we have a series of N observations for a variable X observed over time, and want to forecast its value at time N+h. Denote the foretell as $\hat{X}_N(h)$, where,

- $\hat{X}$ stands for the forecast of X.
- N is the base time at which forecasting is done.
- h is the time horizon which shows how far ahead the forecast covers.

Time series analysis covers two types of quantitative forecasting, namely, univariate (analysing historical data of a single series) and multivariate (analysing historical data of more than one variable). Before carrying out a forecasting exercise, it is very essential to know the features of the data available to be able to choose the right model. The easiest method to go about doing this is to have a time series plot with observations against time. Using the time series plot, features like trend, seasonality, outliers, transformation in structure, turning points and sudden discontinuities are easily observed (Yamin et al., 2002).

3.2 Constituents of time series

The main use of time ordered data is to discover trends and other patterns that take place over time. When the pattern is abstracted from the data, all that should be remain is random, a stable process, called a stationary process.

Time series is consists of 4 constituents, namely trend constituent, seasonal constituent, cyclical constituent and random or irregular constituent (Chatfield, 2002).

3.2.1 Trend constituent

Chatfield (2002) defined that the trend pattern may be linear or non-linear, linear trend pattern is the most common trend pattern. This is when the series has a steady up or down motion. This is brought about by long term factors that affect the variable being observed.

In economic variables we might have the steady economic growth giving a steady upward trend to economic indicators. This movement can either be linear or non-linear. If there is no increase or decrease pattern then the series is said to be stationary

in the mean. Global warming will produce an upward trend in monthly temperatures over the years.

3.2.2 Seasonal constituent

Time series data may include a seasonal component. Regular, relatively short-term repetitive up and down fluctuation of the variable Y.

This type of component generally repeats itself at fixed intervals within a period of time, for example, daily, weekly, monthly or quarterly. Similar patterns of behaviour are observed during these intervals, there is repeatition. Seasonality occurs when a series is influenced by seasonal element and is usually predictable. It usually happens during a fixed and known time interval (Chatfield, 2002). Here we would somehow expect the ice cream sales to be high in the summer and very low in winter. The variation in ice cream sales is seasonal fluctuation.

3.2.3 Cyclical constituent

The cyclical constituent modifies the trend constituent. A gradual, up and down potentially irregular swings of the variable Y. Business and economic data go through the cycles of expansion and contraction which span over more than one year (Bhar and Sharma, 2005; Jebb et al., 2015).

There has been periods where economies have been observed to be growing followed by years of economic declines. Drought have been observed to occur after a certain period of few years. If the period between the occurrence of these events is constant then the series has cycles or a cyclic component. This pattern takes place when the data series exhibits ups and downs that are not of fixed periods. The period of these fluctuation takes about at least two years. The main difference between cycles and seasons is that, if the changes are not of fixed duration then it means they are cyclic. Otherwise, if the duration is constant then the pattern is said to be seasonal (Hyndsight Website, 2015).

3.2.4 Irregular or Random constituent

An increase or decrease of the variable Y for a particular time period. This describes the fluctuation in the series that is due to irregular or non-recurring factors like

strikes or earthquakes.

These factors present a randomness in time series that can not be forecasted. It is the variation left in a data series after removing all systematic effects, like, trend, seasonality and cycles. In other words, they can not be forecasted. During a forecasting exercise, the objective is to model all the constituents until the only unexplained one is an irregular fluctuation (Hyndman and Athanasopoulos, 2014).

3.2.5 Stationarity

A time series is stationary if there is no change in the mean and variance. The attributes of the data are more uniform throughout all sections of one series. In simple terms, a stationary time series will have no predictable shape in the long run. Therefore, a series with trend and seasonality components is not stationary because these components influence the value of the series at different intervals of observation (Hyndman and Athanasopoulos, 2014).

A time series is strictly stationary if the joint distribution of $X_{t1}, X_{t2}, \cdots, X_{tn}$ is the similar as that of $X_{t1-k}, X_{t2-k}, \cdots, X_{tn-k}$ for all time periods t and all time lags k (Shumway and Stofer, 2010, Washington et al., 2010). Shifting the time commencement by an amount k has no influence on the joint distributions, which must therefore depend only on the intervals between $t_1, t_2, \cdots, t_n$. This means that for a strictly stationary procedure, the mean, $E(X_t) = E(X_{t-k}) = \mu$ and variance, $var(X_t) = var(X_{t-k}) = \sigma^2 = \gamma(0)$ are constant throughout time (McCleary et al., 1980).

The definition of a weakly stationary process is that,

- The mean value is constant.
- The covariance function is time-invariant.
- The variance is constant,

Where the strictly stationary process is a process whose probability distribution does not change over time.

Likewise, the covariance between any 2 observations rely only on the time lag between them, $(t, t-k)$ rely on amount k only. A series is second order stationary if both the 1st and 2nd order instants do not depend on time and both the covariance

and correlation are functions of the time lag only. Second order stationarity is also called weak stationarity (McCleary et al., 1980).

Majority of time series analysis methods can be applied to stationary time series. Therefore, it is very essential to know whether the data at hand is stationary or non-stationary before doing any further analysis (McCleary et al., 1980). If not stationary, it's very much important to use appropriate transformations to achieve stationarity. Testing for stationarity helps to find out if there is any correlation that needs to be dealt with and determining which model best suits the data. Different methods can be used to test for stationarity, for example, software like the Augmented Dickey-Fuller (ADF) test or unit root test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS), plotting an ACF, and a time series plot.

A non-stationary series can be transfigured to become stationary through different ways, for example, de-trending (using regression to fit the trend), taking logs (stabilize the variance), differencing (to stabilize the mean by eliminating trend and seasonality) and using moments (Franses, 1998; Heij et al., 2004).

To perform forecasting process, most methods demand the stationarity conditions to be satisfied:

- 1st order stationary : A time series is a 1st order stationary if expected value of X_t remains similar for all t. For example, in economic time series, a process is 1st order stationary when we abstract any kinds of trend by some mechanisms such as differencing (Montgomery et al., 1990).

- 2nd order stationary : A time series is a 2nd order stationary if it is 1st order stationary and covariance between X_t and X_s is function of length (t-s). In economic time series, a procedure is 2nd order stationary when we stabilize its variance by some translations, such as taking the square root (Montgomery et al., 1990).

Stationarizing a time series through differencing is a very significant part of the process of fitting an ARIMA model.

One other reason for stationarizing a time series is to be able to find meaningful statistics such as means, variances, and correlations with other variables. Such statistics are beneficial as descriptors of future behaviour only if the series is stationary.

For instance, if the series is congruously increasing over time, the sample mean and variance will advance with the size of the sample, and they will always underestimate the mean and variance in forthcoming time. If the mean and variance of a series are not well illustrated, then neither are its correlations with other variables. For this reason, you should be cautious about trying to extrapolate regression models fitted to non-stationary data.

In simple terms, a stationary time series will have no predictable shape/pattern in the long run. Therefore, a series with trend and seasonality elements is not stationary because these elements affect the value of the series at different periods of observation. Shifting the time origin by an amount k has no influence on the joint distributions, which must therefore depend only on the intervals between $t_1, t_2, \cdots, t_N$. This means that for a weakly stationary process, the mean,

$$\begin{aligned} E(X_t) &= E(X_{t-k}) \\ &= \mu \end{aligned} \tag{3.1}$$

and variance,

$$\begin{aligned} var(X_t) &= var(X_{t-k}) \\ &= \sigma^2 \\ &= \gamma(0) \end{aligned} \tag{3.2}$$

are constant throughout time. Likewise, the covariance between any 2 observations depends only on the time lag between them, (t, t-k) depends on amount k only. A series is 2nd order stationary if both the 1st and 2nd order moments do not depend on time and both the covariance and correlation are functions of the time lag only. Testing for stationary helps to find out if there is any correlation that needs to be dealt with and determining which model best suits the data (Montgomery et al., 1990).

Different methods can be used to test for stationarity, for example, tests like the Augmented Dickey-Fuller (ADF) test, Kwiatkowski Phillips Schmidt Shin (KPSS), plotting an ACF, and a time series plot as well. Stationarity is primarily violated when the mean of a series changes. Differencing the series accedes a stationary stochastic process. Time series with a stochastic incline/trend have forecast intervals that grow over time and their shock effects are permanent. Unit root tests work well when assessing the presence of a stochastic trend in any observed series. A non stationary series can be transformed to become stationary through different ways, for example,

de-trending (using regression to the trend), taking logs (stabilize the variance of the series), differencing (to stabilize the mean of the series by eliminating trend and seasonality) and using moments (Clements et al., 2001).

3.2.6 Differencing

Differencing is a distinctive type of filtering used to remove trend from time series data until stationarity is achieved. Suppose we have a stochastic procedure X_t. The 1st difference of X_t is designated as, equation $\nabla X_t = X_t - X_{t-1}$.

This means the second difference $\nabla^2 X_t$ is defined as,

$$\begin{aligned} \nabla^2 X_t &= \nabla(\nabla X_t) \\ &= \nabla(X_t - X_{t-1}) \\ &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= X_t - 2X_{t-1} + X_{t-2} \end{aligned} \tag{3.3}$$

In general, the d^{th} difference process $\nabla^d X_t$ is defined as,

$$\begin{aligned} \nabla^d X_t &= \nabla^{d-1}(\nabla(X_t)) \\ &= \nabla^{d-1} X_t - \nabla^{d-1} X_{t-1}^d \end{aligned} \tag{3.4}$$

A proper type of filtering used for removing trend is differencing a granted time series until it is stationary. The type of filtering is mainly used in Box-Jenkins process. The 1st difference of a time series is the series of transformations from one duration to another. If Y_t denotes the value of the time series Y at duration t, then the 1st difference of Y at period t equals to $Y_t - Y_{t-1}$. If the 1st difference of Y is stationary and also entirely random (not auto-correlated), then Y is described by a random walk model (Shumway and Stoffer, 2010).

The random walk theory propose that stock price changes have the same distribution and are autonomous of each other, so the past trend of a stock price can not be used to forecast its future movement. This is the idea that stocks take a random and unpredictable path if the first difference of Y is stationary but not totally random. Differencing is a type of transformation that accomplishes several things, making a time series stationary. Stabilizing the mean of the time series. Stationarity is a very useful statistical property, it is important to understand why. It means that the ef-

fect of time is removed, and now you can reason about the statistical distribution as you would with a standard probability distribution function. Differencing makes the times series stationary which is required if we want to forecast for time periods outside the range of data set (Chatfield, 2002).

3.2.7 The Autocovariance and Autocorrelation functions

To define a proper model for a given time series data, it is needful to carry out the ACF and PACF analysis. These statistical measures contemplate how the observations in a time series are connected to each other. For modelling and forecasting aim, it is frequently advantageous to plot the ACF and PACF against consecutive time lags. These plots assist in defining the order of AR and MA terms. Beneath we bestow their mathematical definitions: For a time series $X_t(t) = 0, 1, 2 \cdots$, the Autocovariance at lag k is specified as:

$$\begin{aligned} \gamma(k) &= Cov(X_t, X_{t+k}) \\ &= E[(X_{t-\mu})(X_{t+k} - \mu)] \end{aligned} \tag{3.5}$$

Covariance is an expectation of the product of two random variables minus the product of their mean. Autocovariance is the covariance between a stochastic process at different times. Autocorrelation is preferred to autocovariance when interpreting results because auto-covariance lies on the units of measurement of the variable under study (Schefer and Young, 2009). Autocorrelation is usually measured on a scale of -1 to +1.

In time series we are more concerned in how the current and past values can be used to estimate future values. To be able to do this we need to understand the relationship between the present value and the values in the past. The sample autocorrelation coefficients are the correlations of sample data at different lags apart. The ACF and PACF can be used to define whether the data is stationary or not, and to discover the best model to fit to the data. When testing for stationarity , the ACF plays a vital role. If a series is stationary, the ACF drops to zero relatively quicker than that of a non-stationary series, which shows a slower decay and longer tails (Vivanco, 2008). For model identification, the PACF is used to discover an autoregressive(AR) process, and the ACF is used to identify MA process.

If plotting a given data set shows a sharp cut of in the PACF and a slower decay in the ACF, then we can deduce that the series is more of an AR process(vice versa). The lag at which the PACF cuts off denotes the order of the AR process. On the other

hand, if the ACF of a differenced series shows a serrated cut off and the autocorrelation value at the first lag is non negative, then we can conclude that the series has a moving average (MA) element in it. The lag at which the ACF cuts off indicates the number of MA provisions to be considered when building the model (Wang and Jain, 2003; Wei, 1994).

For example, for AR(1) process, the ACF rejects in geometric progression from its highest value at lag 1, while the PACF cuts off after lag 1. The contradictory pattern applies to an MA(1) process, where the ACF elude off after lag 1 and the PACF drop in geometric progression from its highest value at lag (Reza, 1961),

$$\begin{aligned} \rho(k) &= Corr(X_t, X_{t-k}) \\ &= \frac{\rho(k)}{\rho(0)} \end{aligned} \tag{3.6}$$

For a stationary process,

$$Var(X_t) = Var(X_{t-k}). \tag{3.7}$$

$\rho(k) = \rho(-k), \rho(0) = 1.$
Another measure, known as the partial autocorrelation function (PACF) is used to measure the correlation between an observation k period ago and the present observation, after controlling for observations at intermediate lags (i.e. at lags k less than 1). At lag 1, PACF(1) is similar as ACF(1).

Normally, the stochastic procedure governing a time series is not clear enough and so it is not possible to define the actual or theoretical ACF and PACF values. These values are to be figured from the training data, i.e. the known time series at hand. The forecasted ACF and PACF values from the training data are respectively termed as sample ACF and PACF. A PACF is another indicator of correlation. It measures the relationship between observations X_t and X_{t-k} after removing the influences of the other time lags. This means the first value of the PACF is identical to the first value of the ACF because there is no lag whose effect should be removed. The ACF and PACF can be used to define whether the data is stationary or not, and to distinguish the best model to the data. When testing for stationarity, the ACF plays a vital role (Monsen and Van Horn, 2007). If a series is stationary, the ACF drops to zero relatively quicker than that of a non-stationary series, which shows a slower decay and longer tails.

For model identification, the PACF is used to identify an AutoRegressive (AR) process. If plotting a given data set shows a sharp cut off in the PACF and a slower decay in the ACF, then we can conclude that the series is more of an AR process. The lag at which the PACF slice off indicates the order of the AR process. On the other hand, if the ACF of a differenced series shows a sharp cut off and the autocorrelation value at the first lag is non negative, then we can conclude that the series has a moving average (MA) element in it. The lag at which the ACF cut off indicates the number of MA terms to be considered when building the model (Farooque, 2002).

For example, for an AR(1) process, the ACF drops in geometric progression from its highest value at lag 1, while the PACF cuts off abruptly after lag 1 (Chramcov, 2011). The coefficient of correlation between two values in a time series is designated as the auto-correlation function (ACF), for example, the ACF for a time series Y_t is given by, $Corr(Y_t, Y_{t-k})$.

A lag 1 autocorrelation (i.e., k = 1) is the correlation between values that are one time duration apart. More broadly, a lag k autocorrelation is the correlation between values that are k time periods apart. The ACF is used to measure the linear relationship between an observation at time t and the observations at prior times.

In a plot of ACF versus the lag, if there is a large ACF values and a non-random pattern, then likely the values are correlated. In a plot of PACF versus the lag, the pattern will commonly appear random, but large PACF values at a given lag indicate this value as a potential choice for the order of an autoregressive model. It is considerable that the choice of the order makes sense (McDonald, 2009). For example, assume you have blood pressure readings for every day over the past 2 years. You may detect that an AR(1) or AR(2) model is appropriate for modelling blood pressure.

However, the PACF may show a large partial autocorrelation value at a lag of 17, but such a large order for an autoregressive model likely does not make sense. As illustrated by Box and Jenkins, the sample ACF plot is profitable in determining the type of model to a time series of length N. Since ACF is symmetrical about lag zero, it is only demanded to plot the sample ACF for positive lags, from lag one onwards to a maximum lag of about N/4. The sample PACF plot aids in distinguishing the maximum order of an AR process. The procedure for calculating ACF and PACF for ARMA models (Reinert, 2002).

Table 3.1: Chramcov (2011) related the PACF and ACF functions to the MA and AR models.

	AR(p)	MA(q)
ACF	Tails off	Cuts off after lag q
PACF	Cuts off after lag p	Tails off

3.3 Analysis of the available cost electricity data

The time series plot in Figure 3.1 shows the daily electricity cost for South Africa from 04 January 2012 to 03 January 2017. The data is collected from a confidential source who prefers the data to remain private. This data originally contains no missing values which makes it not hard to fit ARIMA models directly without correcting the missing values first.

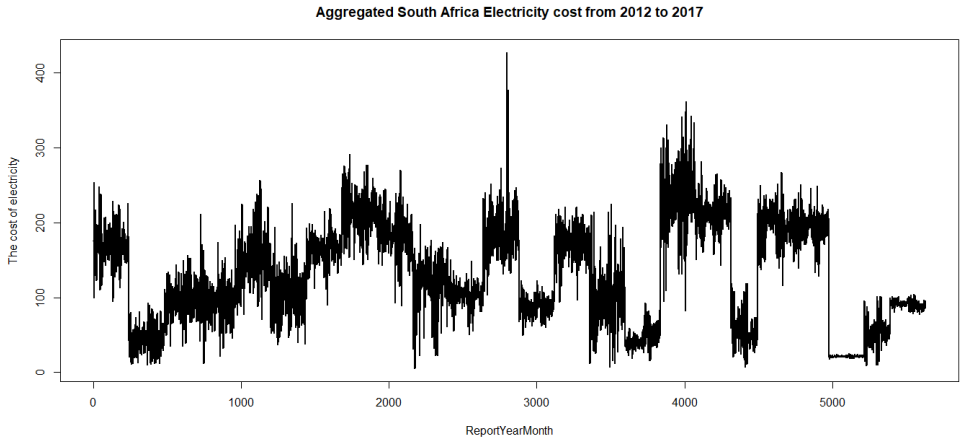

Figure 3.1: The original time series plot of four variables data set before cleaning and taking the seasonal difference.

Considering the time series plot in Figure 3.1, here we used **R** software to plot the time series original graph for South Africa's daily electricity cost data using transaction cost, travel cost, travel time and travel distance variables. We managed to verify and identify a seasonal pattern in the data at hand, concluding from the time series plot. The data shows an unstable variance, simply meaning that the data is not stationary and can not carry on with forecasting process. Hence it is very essential and vital to come up with some statistical methods and adjust it accordingly to achieve stationarity.

Chapter 4

Introduction to forecasting with ARIMA models

The ARIMA process analyses and forecasts uniformly spaced univariate time series data, transfer function data, and intervention data by using the AutoRegressive Integrated Moving Average (ARIMA) or AutoRegressive Moving Average (ARMA) model. An ARIMA model forecasts a value in a response time series as a linear combination of its own past values, past errors, current and past values of other time series.

4.1 ARIMA models

The word ARIMA stands for AutoRegressive(AR) Integrated(I) Moving Average(MA). In this study, we shall tackle each element of this model individually as we build up to its general purpose. ARIMA models are mainly used for forecasting data that is originally non-stationary but it can be made stationary by differencing (Karamouz et al., 2012; Nason, 2006).

4.2 The AR model

This kind of a model is in the similar form as the well known simple linear regression model in which r_t is the dependent variable and r_{t-1} is the explanatory variable. In time series literature, the model is assigned to as an AutoRegressive(AR) model of order 1 or an AR(1) model. This uncompounded model is also broadely used in stochastic volatility modelling when r_t is reinstated by its log volatility (Nogales et

al., 2003).

The AR(1) model has diverse properties similar to those of the isolated linear regression model. However, there are some important dissimilarities between the models. Here it suffices to note that an AR(1) model denotes that, conditional on the past return r_{t-1}, we have ductile model. A series X_t is an AutoRegressive process of order p, denoted by AR(p), if it can be expressed in the form,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-k} + e_t \tag{4.1}$$

In back shift operator notation, the model can be written as,

$$X_t(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) = e_t \tag{4.2}$$

Where,

- B is back shift operator.
- $\phi_1, \cdots, \phi_p$ are parameters of the model at hand.
- e_t is normally distributed with mean 0 and a constant variance σ_e^2 . This term is said to be independent of all previous process values $X_{t-1}, X_{t-2}, \cdots$

An AutoRegressive model equation is similar to a multiple regression model with the value of X at time t linearly depending on a union of its weighted p past values. The term AutoRegressive means that its a regression of the variable of interest against its past values plus an error term e_t at time t. AR models are normally qualified to stationary data (Hyndman and Athanasopoulos, 2014).

That is why it is always needful to check for stationarity of the data way before fitting such models. In this model, it is assumed that the mean is $E(X_t) = 0$. However, a non zero mean could be added to the model by reinstating X_t with $X_{t-\mu}$, for all t (MAHIEUA et al., 2007). This would not impact the attributes of the model. After applying the back shift notation operator, Equation 4.2 can be used to yield the AR(p) characteristic equation as,

$$\begin{aligned} \phi(x) &= 1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p \\ &= 0 \end{aligned} \tag{4.3}$$

It is important to note that an AR(p) process is stationary if the p roots of $\phi(x)$ each

surpass 1 in absolute value (Cryer and Kellet, 2006). The autocovariance and autocorrelation functions of an AR process can be derived using the Yule-Walker equations (Eshel, 2010). If we assume a stationary AR(p) process with zero means, multiplying both sides by X_{t-k} yields,

$$X_t X_{t-k} = \phi_1 X_{t-1} X_{t-k} + \phi_2 X_{t-2} X_{t-k} + \cdots + \phi_p X_{t-p} X_{t-k} + e_t X_{t-k} \tag{4.4}$$

Since we assumed zero means, it means the autocovariance of the process at lag k is given by,

$$\begin{aligned} \rho(k) &= Var(X_t X_{t-k}) \\ &= E(X_t X_{t-k}) - E(X_t)E(X_{t-k}) \\ &= E(X_t X_{t-k}) \end{aligned} \tag{4.5}$$

Taking expectations of Equation 4.4 gives,

$$E(X_t X_{t-k}) = E(\phi_1 X_{t-1} X_{t-k} + \phi_2 X_{t-2} X_{t-k} + \cdots + \phi_p X_{t-p} X_{t-k} + e_t X_{t-k}) \tag{4.6}$$

$$\gamma(k) = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p} \tag{4.7}$$

Dividing through by the process variance $\gamma(0)$, we get,

$$\rho(k) = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p} \tag{4.8}$$

Equation 4.8 gives a set of Yule-Walker equations, for $k > 0$. For the very known values of $\phi_1, \phi_2, \cdots, \phi_p$, we can calculate the first lag p autocorrelations $\rho_1, \rho_2, \cdots, \rho_p$ (Plasmans, 2006). Values of ρ_k, for $k > p$, can be prevailed by using the recursive relation (Von Storch and Zwiers, 2001).

4.3 The MA model

In time series analysis, the Moving Average (MA) model is a familiar approach for modelling univariate time series. The moving average model indicate that the output variable depends on the current and numerous past values of a stochastic (im-

perfectly predictable) term, together with the AutoRegressive (AR) model, the moving average model is a distinctive case and key element of the general ARMA and ARIMA models of time series, which have a more intricated stochastic organization (GHASEMI and SHAYEGHI, 2013).

Opposed to the AR model, the restricted MA model is always stationary. A series with white noise process of mean 0 and variance σ_e^2 is a moving average process of order q, noted as MA(q), if it can be expressed as a weighted linear sum of the past forecast errors.

$$X_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \tag{4.9}$$

In back shift operator notation, the model can be written as,

$$X_t = e_t(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) \tag{4.10}$$

Where,

- B is said to be the back shift operator notation.
- $\theta_0, \theta_1, \cdots, \theta_q$ are the coefficients of the lagged error terms. θ_0 is usually equated to 1 (Broersen, 2006; Hipel and McLeod, 1994).
- e_t is normally distributed white noise with mean 0 and variance.

We can write the MA(1) (moving average of order one) and MA(q) (moving average of order q) as:

$$X_t = eX_{t-1} + e_t \tag{4.11}$$

Equation 4.11 is named as Moving Average (MA) model. Some authors note the parameters of an MA process as negatives, in order to have attributes operators of the same signs for AR and MA processes. However, this has no expressive change to the interpretation of the model (Chatfield, 2002). The autocovariance functions of an MA(q) model is expressed as $\gamma(k) = Cov(X_t, X_{t-k})$, where it becomes the variance of the process if k = 0. Therefore, from the definition of the model,

$$\gamma(0) = \sigma_e^2(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2) \tag{4.12}$$

Generally, an MA(q) process is invertible if all roots of the MA(q) characteristic polynomial, $\theta_x = 1 + \theta_1 X + \theta_2 X^2 + \cdots + \theta_q X^q$, exceed 1 in absolute values, or lie outside of the unit circle.

4.4 The ARMA model

Aggregating both the AR(p) and MA(q) models yields rise to an AutoRegressive Moving Average model (ARMA(p,q)) which is expressed as,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \quad (4.13)$$

Re-arranging the model gives,

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \quad (4.14)$$

Using the back shift operator,

$$X_t(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) = e_t(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) \quad (4.15)$$

This can be simplified to,

$$\phi(B) X_t = \theta(B) e_t \quad (4.16)$$

where,

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p); \quad (4.17)$$

and,

$$\theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) \quad (4.18)$$

ARMA model supplies one of the canonical instruments in time series modelling. A significant parametric family of stationary time series, the AutoRegressive moving average or ARMA. For a big class of autocovariance functions, it is likely to get an ARMA process X_t. Especially, for any positive integer K, there is an ARMA process X_t such that $\gamma_x(h) = \gamma(h)$ for $h = 0, 1, \cdots, K$. The family of ARMA processes plays a big role in the modelling of time series data. The linear structure of ARMA processes also direct to a substantial simple ARMA(p,q) process if,

- there is stationarity.
- It (the deviations $X_t - E(X_t)$) satisfies the linear difference equation in regression form.

Both the AR(p) and MA(q) are special cases for the ARMA model. An ARMA(p,0) process is similar to an AR(p) process and an ARMA(0,q) process is similar to an MA(q) process. If the data at hand is stationary, it is better modelled using an ARMA(p,q) model rather than AR(p) or MA(q) models individually (Chin and Fan, 2005).

This is mainly because an ARMA(p,q) in such a case uses rare parameter than the individual models and bestow a better representation of the data. This is called the Principle of Parsimony (Singh, 2002; Woodward et al.,2011). For ARMA(p,q) process to be stationary, an absolute value of the roots of all the AR(p) characteristic polynomials should be greater than 1.

For invertibility, the absolute values of the roots of all the MA(q) characteristic polynomials should be greater than 1. For example, given a model in Equation 4.19,

$$X_t + 0.2X_{t-1} - 0.48X_{t-2} = Z_t \tag{4.19}$$

Since we can express the procedure as an AR process, it is invertible. For stationarity, the polynomial,

$$1 + 0.2B - 0.48B^2 = 0 \tag{4.20}$$

must have the roots out of the unit circle. By using the quadratic equation we obtain B=1.667 and B=-0.125, the two solutions are both out of the unit circle hence the process is stationary.

4.5 The ARIMA model

The ARIMA process analyses and foretells adequately spaced univariate time series data, transfer function data, and intervention data by using the AutoRegressive Integrated Moving Average (ARIMA) or AutoRegressive Moving Average (ARMA) model (Chatfield, 1991).

One can neither apply the AR, MA, nor ARMA models straightly. The most suitable method of getting stationarity when dealing with ARIMA models is through differencing (Espinola, 2005).

Generally, the data for time series can be differenced many times(d) times, $d = 1, 2, 3, \cdots$ until it becomes stationary. The first difference $X_t - X_{t-1}$ can also be expressed using a back shift notation as $(1 - B)X_t$. If the primary data series is differenced d times before fitting ARMA(p,q) model, then the model for original undifferenced series is an AutoRegressive Integrated Moving Average model (ARIMA(p,d,q)), where d potray the number of times the data has been differenced (Janacek and Swift, 1995). Taking first differences normally removes a linear deterministic trend (Hendry, 1995).

There are three stages of ARIMA modelling. The analysis executed by PROC ARIMA

is divided into three stages, coinciding with the stages described by Box and Jenkins (1976).

- Identification stage: Identify statement is used to specify the response series and discover candidate ARIMA models for it. The identify statement reads time series that are used in the later statements, perhaps differencing them, and calculate autocorrelations, inverse autocorrelations, partial autocorrelations, and cross correlations (TIAO, 2015).

  Stationarity tests can be done to define if differencing is essential. The analysis of the identify statement output insinuate one or more ARIMA models that could be fit.

- Estimation and Diagnostic Checking stage: Estimate statement is used to specify the ARIMA model to fit to the variable indicated in the previous identify statement and to estimate the parameters of that model (WANG and CHE, 2010). The estimate statement also can exhibit diagnostic statistics to assist you judge the adequacy of the model.

  Significance tests for parameter estimate show whether some terms in the model might be unnecessary. Goodness-of-fit statistics helps in assimilating this model to other models. The Outlier statement supplies another profitable tool to check whether the currently estimated model accounts for all the variation in the series. If the diagnostic tests show any problematic areas with the model, try another model and then repeat the estimation and diagnostic checking stage (Systematics, 1994).

- Forecasting stage: Forecast statement is used to forecast future values of the time series and to propagate confidence intervals for these forecasts from the ARIMA model given by the preceding estimate statement. In Box-Jenkins approach to ARIMA modelling, the sample autocorrelation function, inverse autocorrelation function, and partial autocorrelation function are assimilated with the theoretical correlation functions anticipated from different kinds of ARMA models (Khandakar and Hyndman, 2008).

  The matching of theoretical autocorrelation functions of different ARMA models to the sample autocorrelation functions calculated from the response series is the main center of the identification stage of Box-Jenkins modelling. Most

textbooks on time series analysis, such as Chatfield (2000), do discuss the theoretical autocorrelation functions for different types of ARMA models.

Differencing deals with the observed values at different times, not the error terms. Therefore, in an ARMA model, adding a differencing term changes only the AR side, not the MA side. Differencing d times changes and results to,

$$X_t(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)(1-B)^d = e_t(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q) \quad (4.21)$$

and can be simplified to,

$$\phi(B)(1-B)^d X_t = \theta(B)e_t \quad (4.22)$$

Equation 4.22 is called an ARIMA model with the term ϕ_B corresponding to the AR characteristic polynomial of order p, $(1-B)^d$ for the integrated part of order d, and θ_B for the MA characteristic polynomial of order q. All models discussed in section above are a special type of ARIMA models. For example, the white noise ARIMA(0,0,0), the random walk ARIMA(0,1,0), an autoregression ARIMA(p,0,0), moving average ARIMA(0,0,q) and an autoregressive moving average ARIMA(p,0,q).

4.6 The SARIMA model

ARIMA models can also be used to model seasonal data. Box and Jenkins have generalized this model to deal with seasonality. ARIMA models that incorporate seasonal patterns occurring over time are called Seasonal Autoregressive Integrated Moving Average models (SARIMA). In this model seasonal differencing of appropriate order is used to remove non stationarity from the series. A first order seasonal difference is the difference between an observation and the corresponding observation from the previous year and is calculated as $Z_t = Y_t - Y_{t-s}$ (Ghysels et al., 2006).

Basically, for monthly time series s=12 and for quarterly time series s=4. Seasonality in a time series is a regular pattern of changes that repeats over S time periods, where S defines the number of time periods until the pattern repeats again. ARIMA models can also be used to model seasonal data. ARIMA models that incorporate seasonal patterns occurring over time are called Seasonal Autoregressive Integrated Moving Average models (SARIMA). With seasonal data, dependence with the past occurs most prominently at multiples of an underlying seasonal lag, denoted by s (Ghysels

etal., 2006). SARIMA models include an additional seasonal term as indicated,

$$ARIMA(p,d,q)(P,D,Q)_s \tag{4.23}$$

where s denotes the number of periods per season. The upper case notation in equation above is for the seasonal part and the lower case notation for the non seasonal parts of the model. The seasonal part of the model consists of terms that are very similar to the non seasonal components of the model, but they involve back shifts of the seasonal period. For example, if a seasonal component is added to equation above, the resultant model will be:

$$\phi_B(B^s)(1-B)^d(1-B^s)^D X_t = \theta_B(B^s)e_t \tag{4.24}$$

According to Chatfield (2002), the most common SARIMA model for monthly data is the $SARIMA(0,1,1)(0,1,1)_{12}$ it is defined as,

$$(1-B)(1-B^{12})X_t = (1+\theta_B)(1+\theta_B^{12})e_t \tag{4.25}$$

4.7 Model specification

Statistical model structure has three main stages, namely:

- Model identification - The goal is to detect seasonality, if it exists, and to identify the order for the seasonal autoregressive and seasonal moving average terms. For example, for monthly data we would include either a seasonal AR 12 term or a seasonal MA 12 term.
- Model fitting - choosing the statistical model that predict values as close as possible to the ones observed in the whole eskom data. While doing a statistical analysis, it is very important to make sure about the goodness of fit of the model used.
- Model verification - is the task of confirming that the outputs of a statistical model are acceptable with respect to the real data generating process

Bad selection of orders p, d, and q results to bad models, which lead to bad forecasts of future values (Okyere and Nanga, 2014). It is therefore essential to make sure that the choices made are accordant with the structure of the observed data.

For any series of the data, a clear indicant of non-stationarity is that the ACF demonstrate a slow deterioration across lags. This usually occurs because in a non-stationary

process, the series hangs together and exhibit trends. If the data is non-stationary seasonally, then the ACF displays clusters of either positive and/or negative autocorrelation (Ord and Fildes, 2012). However, there are other common methods of determining non-stationarity, for example, the ADF and/or KPSS test and using a time series plot (Ord and Fildes, 2012).

When there is a definite linear trend in the data and ACF for the series decline very gradually, it is usually advisable to take first differences (Ord and Fildes, 2012). If the ACF of first differenced data represents that of a stationary ARMA process (declines quickly), then the value d in ARIMA(p,d,q) is taken to be 1. The ACF and PACF of first differenced data can then be used to discover plausible p and q values. Otherwise, second differences are taken and d = 2 is used instead (WANG JUN et al., 2010). Then the plotted ACF and PACF at that point is used to discover plausible values of p and q. The order of differencing can take on any value until stationarity is attained.

Once the model order has been identified, then all parameters in the model can be approximated (SHEIKH-EL-ESLAMI, 2015). This can be done using different software like R-Studio which estimates the ARIMA model by using MLE (Hevia, 2008). This method finds the values of parameters which maximize the probability of getting the data that has been observed.

Chapter 5

Applying univariate ARIMA models to electricity cost data

Here in this chapter, we fit an ARIMA model to the available time series data.

We want to fit an ARIMA model to South Africa's daily electricity data. We were able to identify a seasonal pattern, with no trend, from the time series graphical representation in Figure 3.1. The data shows an unstable variance over time. It is necessary to adjust it accordingly to gain stationarity. We plot the ACF and PACF to check for stationarity because visual inspection of the time series plot is some times misleading. From Figure 5.1 and Figure 5.2, the data is not stationary. The ACF decays off at a very slow rate.

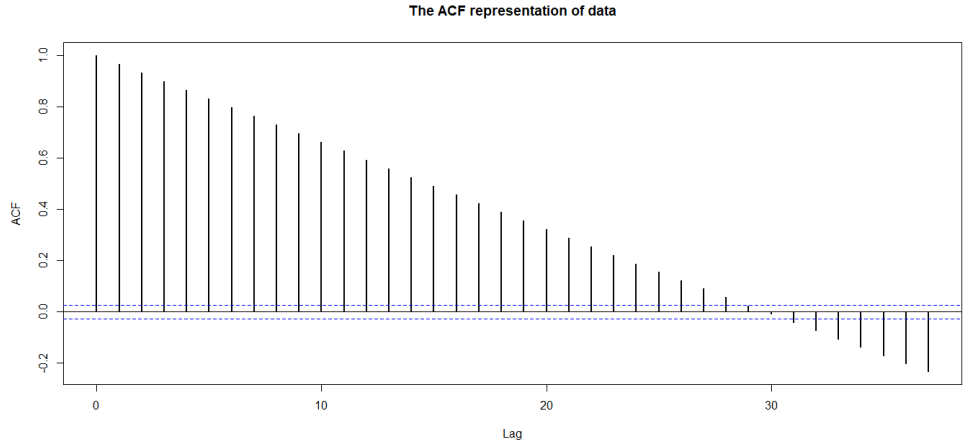

Figure 5.1: The ACF of a non stationary data set.

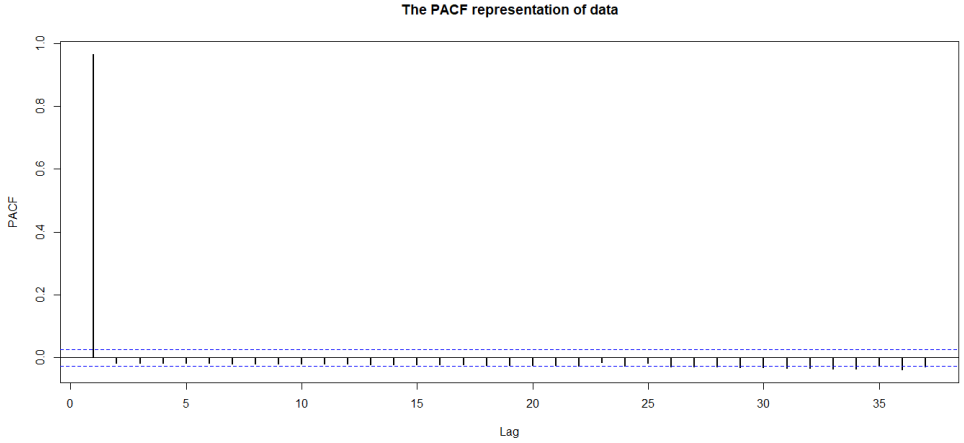

Figure 5.2: The PACF of a non stationary data set.

5.1 Testing for stationarity

5.1.1 Test for stationarity in the data using KPSS test

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test figures out if time series is stationary or non-stationary around the mean or linear trend. A stationary time series is the one where statistical attributes like mean and variance are constant over time (Shin and Schmidt, 1992).

- Null hypothesis: The data is stationary.
- Alternative hypothesis: The data is not stationary.

The KPSS test is based on linear regression, with regression equation:

$$X_t = r_t + \beta_t + \epsilon_t \tag{5.1}$$

We use a significance level value of 5% to make decisions. That means a p-value that is less than 0.05 defines that an advance differencing is needed. A disadvantage of KPSS test is that it has got a high rate of type I error(where it tends to reject the null hypothesis more frequently). If there are attempts made to manage these errors(by having high p-values), then that impact negatively the power of the test. To deal with potential for high type I error you can aggregate KPSS with ADF test. If the results from both tests propose that the time series is stationary, then it is stationary (Kwiatkowski et al., 1992).

5.1.2 Test for stationarity in the data using ADF

The original Dickey Fuller test, developed by Dickey and Fuller (1979), is used to test whether a unit root is present in an autoregressive model. The condition for stationarity states that for an AR model to be stationary, $\phi < 1$. The case where $\phi = 1$ corresponds to the random walk which is not stationary. In this test, the null hypothesis of the variable containing a unit root is tested against the alternative that the variable was generated by a stationary process. The general idea is to set up an AR model for the observations X_t and test if $\alpha = 1$.

Consider the AR(1) model,

$$X_t = \alpha X_{t-1} + e_t. \tag{5.2}$$

Augmented Dickey Fuller (ADF) test, tests big and more complicated sets of time series models by getting rid of all the autocorrelation in the time series. The unit root or ADF null hypothesis against the stationary alternative corresponds to:

- Null hypothesis: $H_0 : \phi \geq 1$

- Alternative hypothesis: $H_1 : \phi < 1$

Again here we use a significance level of 5% to make a decision. That means a p-value that is less than 0.05 defines that differencing is needed. The disadvantage of using the ADF test is that the normal test significance level value (usually 5%) is not reliable/trustworthy when the error terms ϵ_t are autocorrelated. The bigger the autocorrelation of ϵ_t, the more distorted the significance of the test becomes. The main usual assumption in lot of time series data is that the data is always stationary (Jürgen et al., 2011).

In order to gain stationarity, different transformations can be used. A log transformation is commonly used in electricity studies to correct for an increasing variance, the log transformation alone is not effective enough to attain stationarity. Due to the seasonality component exhibited, a seasonal difference is a better way of making the data stationary. The tbats function in **R** shows us the presence of annual seasonality in the data, as shown in Figure 5.3.

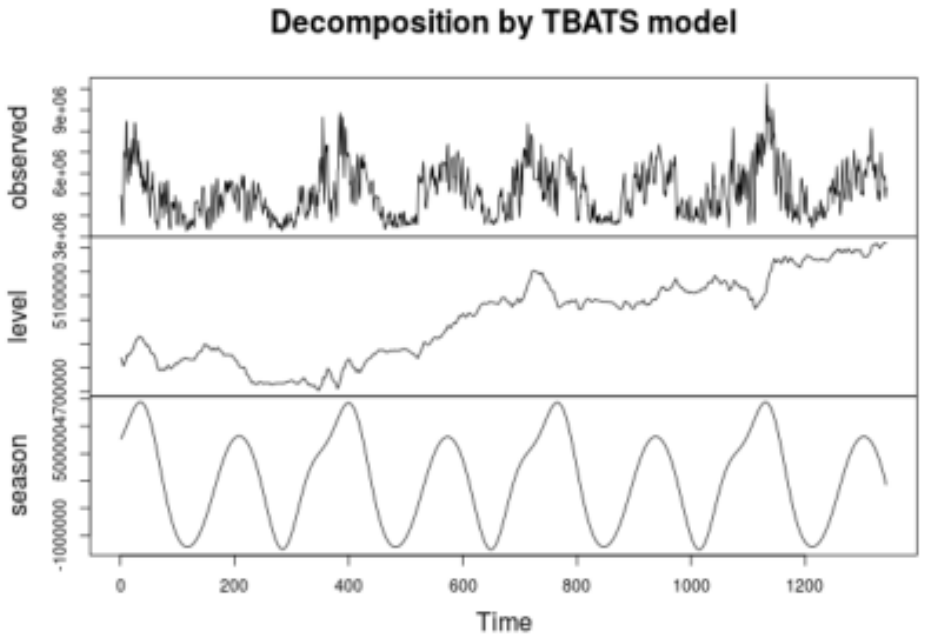

Figure 5.3: Annual seasonality in South Africa data.

After identifying that the data at hand is not stationary, a seasonal difference is applied to the data in **R** software and produced the results as shown in Figure 5.4.

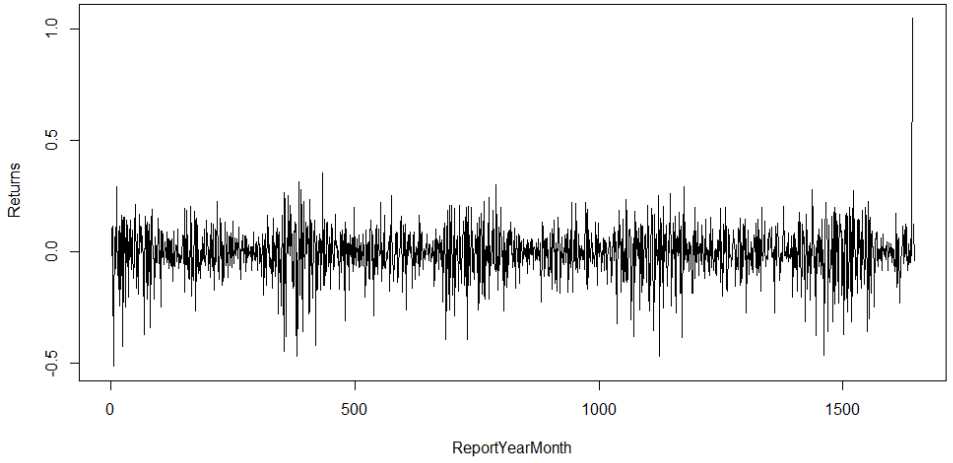

Figure 5.4: Time series graphical representation of the electricity data set after cleaning and the seasonal differencing.

At the beginning, Figure 5.4 shows stationarity in the data set used. After that we run a KPSS test, which returns a p-value of 0.3. Since the returned p-value is greater than significance level value of 0.05, the conclusion is that we fail to reject the null

hypothesis and deduce that the data is now stationary, which allows us to proceed with the forecasting process.

Even though the new plotted time series dataset shows a bit of pattern of seasonality around the end, it is not clear whether the variance of the data is stationary or non stationary through time. At this point, there is no certainty to perform transformations. However, a strong analysis of this data will be studied in the next chapters to prove the stationarity of the data and how to attain it in case the data is not actually stationary.

5.2 Model order selection

Since the data is seasonally differenced and is now stationary, we plot the graphs of ACF and PACF with seasonal differenced data in order to select the the correct p and q values to use when constructing the model. However, since the data shows evidence of seasonality, we can use the auto.arima function in R-studio, together with different values of D (the seasonal difference), to develop different possible models from which the best is chosen using accuracy measurements.

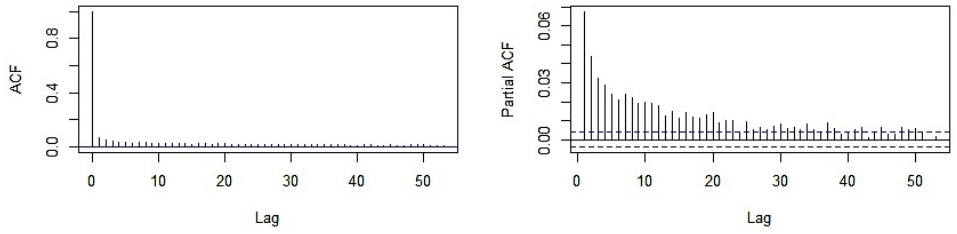

Figure 5.5: ACF and PACF showing stationarity of the electricity data set.

Here, we do not choose the best model depending on its AIC, AICc, or BIC because doing so requires that all models have same orders(d) of differencing, which is not the case for the seasonal difference used (Hyndman and Athanasopoulos, 2014). Here, we compare 2 models with D = 1,2 because, it is not common to difference data more than twice before stationarity is achieved. The models are shown in Table 5.1:

Table 5.1: Models from the auto.arima function using the electricity data.

	ARIMA(1,0,1)(0,1,0)[365]	ARIMA(1,0,1)(0,2,0)[365]
RMSE	635400.5	8845422
MAPE	6.78943	6.94378

Then by using the two main accuracy measures (mainly RMSE and MAPE) as shown in Table 5.1, the very best model for the available data is said to be ARIMA(1,0,1)(0,1,0). We chose it because it has the most lowest RMSE and MAPE. Then the coefficients of the model are shown in Table 5.2:

Table 5.2: The Coefficients of the ARIMA(1,0,1)(0,1,0)[365] model.

	ar1	ma1
Parameters	0.7301	0.6639
Standard error	0.0173	0.3278
P-values	0.01	0.01

Hence, since the p-values are less than the significance level(0.05), the conclusion is that the parameter estimates are statistically significant.

5.2.1 Forecasting data with seasonal ARIMA(1,0,1)(0,1,0)[365] model

After figuring out the most accurate model to use for forecasting, we then use the it to make the process. Using the forecast package in **R**, we develop both point and interval forecasts for 95% confidence intervals.

Table 5.3: Observation forecasting.

Obs	Forecast	(Lower Confidence Interval) - (Higher Confidence Interval)
1440	96.2948	(-367.6249) - (560.2145)
1441	96.3488	(-357.8031) - (550.5007)
1442	96.4027	(-347.7666) - (540.5720)
1443	96.4567	(-337.5005) - (530.4139)
1444	96.5107	(-326.9882) - (520.0095)
1445	96.5646	(-316.2110) - (509.3402)
1446	96.6186	(-305.1476) - (498.3849)
1447	96.6726	(-293.7740) - (487.1192)
1448	96.7266	(-282.0623) - (475.5154)
1449	96.7805	(-269.9802) - (463.5413)
1450	96.8345	(-257.4900) - (451.1590)
1451	96.8885	(-244.5472) - (438.3241)
1452	96.9424	(-231.0983) - (424.9832)
1453	96.9964	(-217.0787) - (411.0715)
1454	97.0504	(-202.4085) - (396.5093)
1455	97.1044	(-186.9873) - (381.1960)
1456	97.1583	(-170.6859) - (365.0025)
1457	97.2123	(-153.3330) - (347.7576)
1458	97.2663	(-134.6936) - (329.2261)
1459	97.3202	(-114.4292) - (309.0696)
1460	97.3742	(-92.0202) - (286.7686)
1461	97.4282	(-66.5922) - (261.4486)
1462	97.4821	(-36.4399) - (231.4042)
1463	97.5361	(2.8389) - (192.2333)

By looking at the forecast Table 5.3 for randomly chosen observations, we expect an increase in electricity cost in South Africa for the next few coming years, even Figure 5.6 implies the increase in electricity cost, hence great measures should be taken.

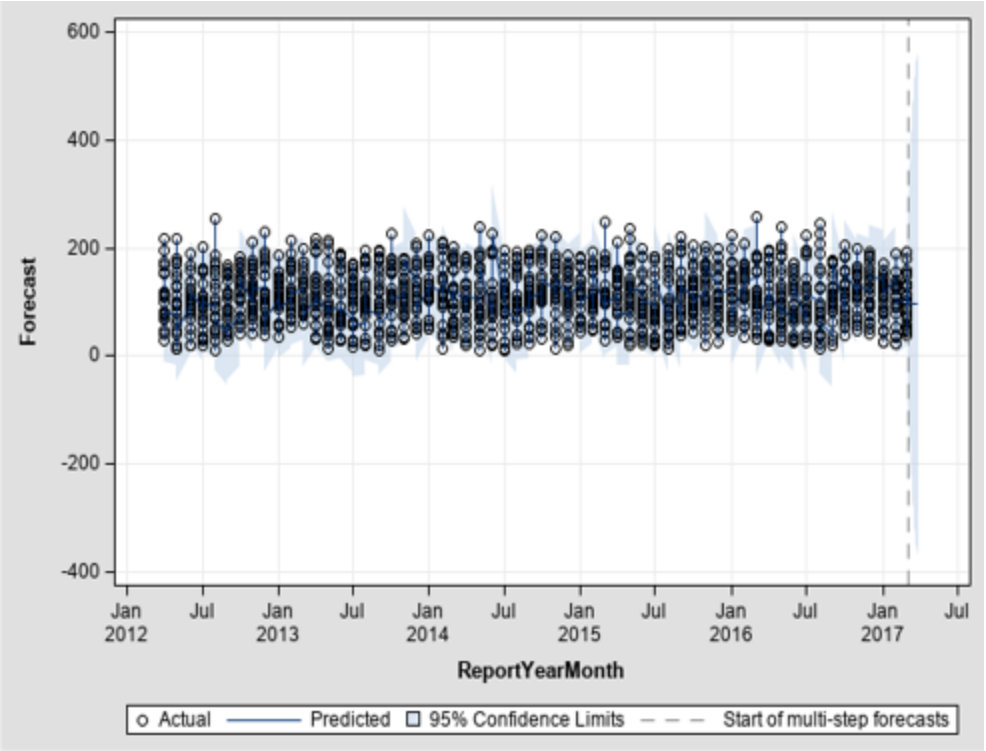

Figure 5.6: Point forecasts shown by the extending blue line.

5.3 Diagnosis of seasonal ARIMA(1,0,1)(0,1,0)[365]

It is advisable to investigate whether the forecast errors for an ARIMA model are homoskedastic, normally distributed, and if there are any correlations between the successive errors. This can be done by plotting down the ACF of the residuals, and doing a portmanteau test for the residuals using the Ljung-Box test. If residuals do not look like white noise, it means that the model can be modified to improve the forecasts. Once the residuals look exactly like the white noise, the model can then be considered effective and used for forecasting.

5.3.1 Autocorrelation

Ljung Box test developed by Ljung and Box (1978) can be used to check for any evidence of autocorrelation. The test is usually applied to the residuals of a time series after fitting ARIMA model, not the original data, and it tests all autocorrelations of the residuals (Arranz, 2005). In this test, the null hypothesis (H_0) of zero autocorrelation is tested against an alternative (H_1) of autocorrelation.

A conclusion can also be drawn using the p-value, and this is the considered option in this research. We can use the Ljung Box Statistic test in **R** to test the same hypotheses as theoretically suggested. The null hypothesis of randomness or no autocorrelation is tested against the alternative hypothesis of non randomness or autocorrelation. We test lags from 6 to 24 in intervals of 6. According to the test, all lag returns a p-value that is less than 0.05 as shown in Table 5.4. Meaningly we reject the null hypothesis and deduce that, the residuals are not random, meaning there is autocorrelation in the residuals of the chosen ARIMA model.

Table 5.4: Ljung Box results for ARIMA(1,0,1)(0,1,0)[365] model residuals.

Lags	Test Statistics(Chi-Square)	P-Values	Autocorrelation
Lag 6	306.06	0.0086	0.015
Lag 12	553.16	0.0031	0.306
Lag 18	693.34	0.0049	-0.021
Lag 24	1371.31	0.0020	-0.378

Graphically, Figure 5.11 displays the ACF of the residuals, with significant spikes in most lags. Hence autocorrelated residuals of an ARIMA(1,0,1)(0,1,0)[365] model.

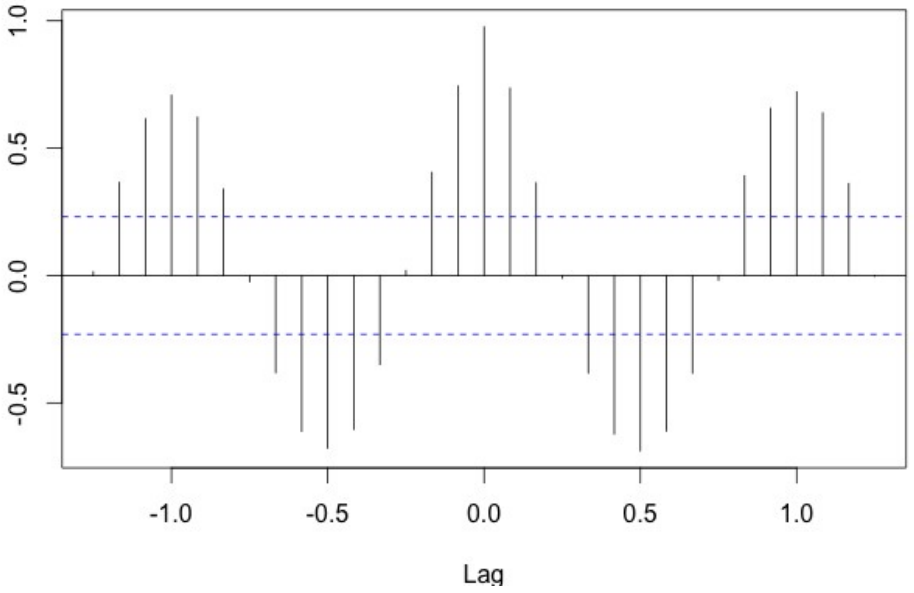

Figure 5.7: The ACF residuals.

Graphically, for each of the four variables we display the figures showing the ACF of the residuals, with significant spikes in most lags. Hence autocorrelated residuals of the ARIMA(1,0,1)(0,1,0)[365] model.

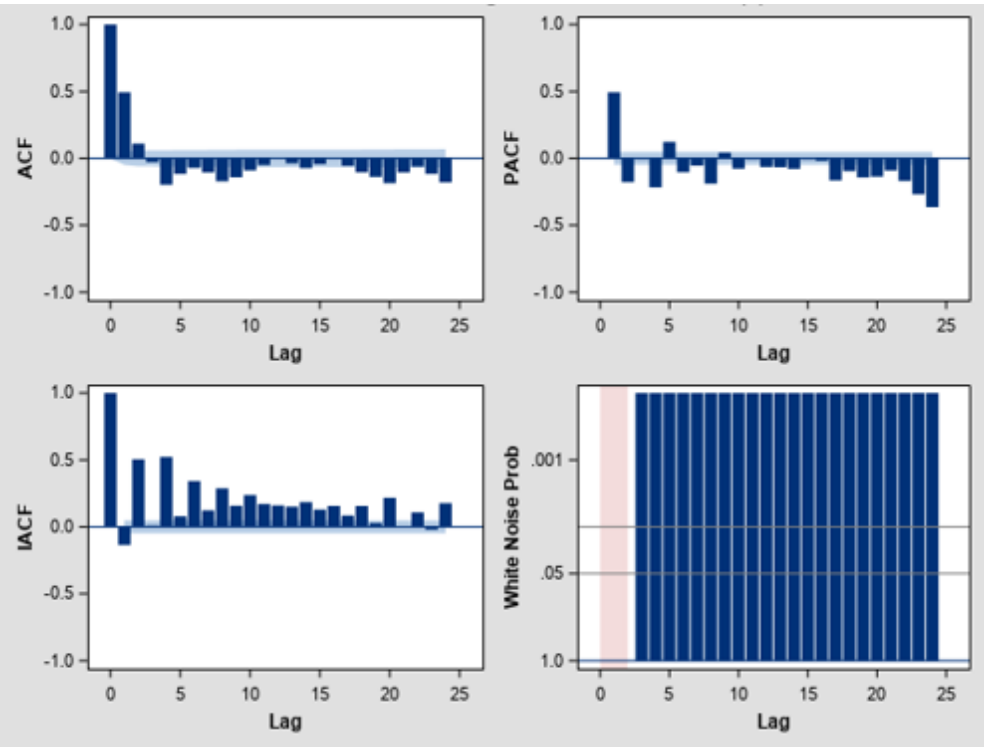

Figure 5.8: ACF residuals for transactional cost.

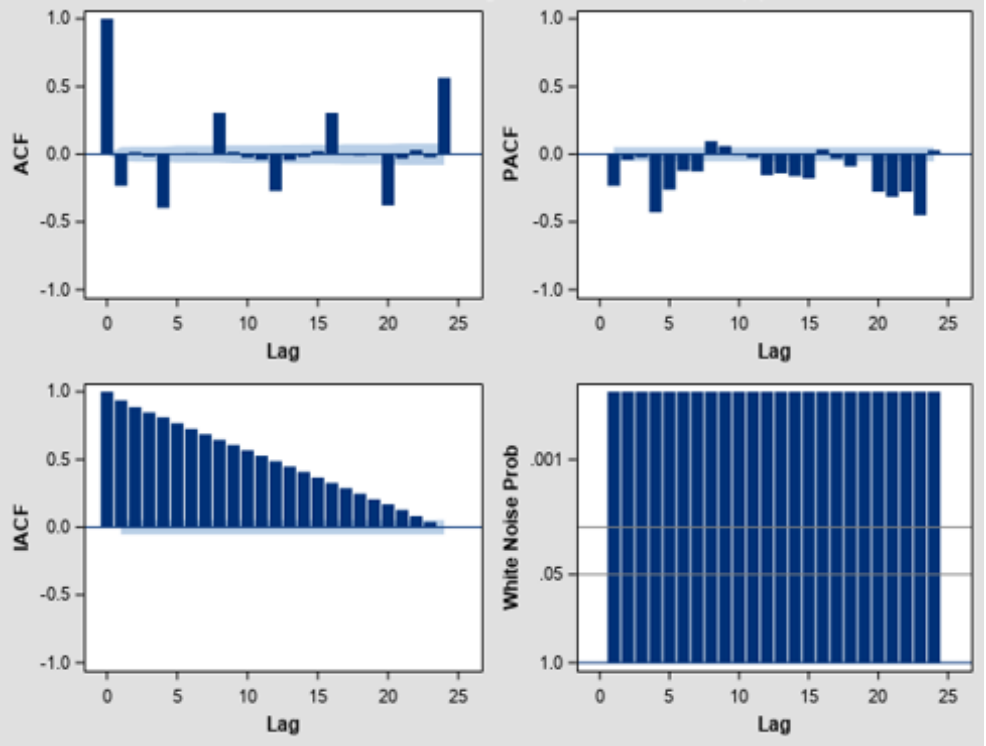

Figure 5.9: ACF residuals for travel cost.

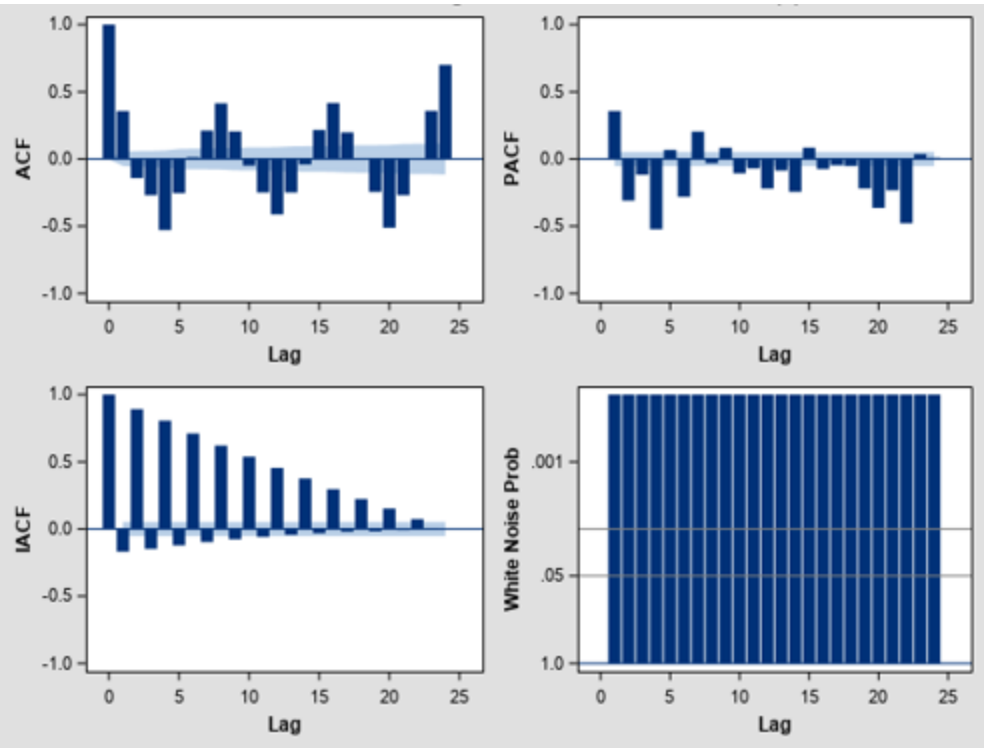

Figure 5.10: ACF residuals for travel time.

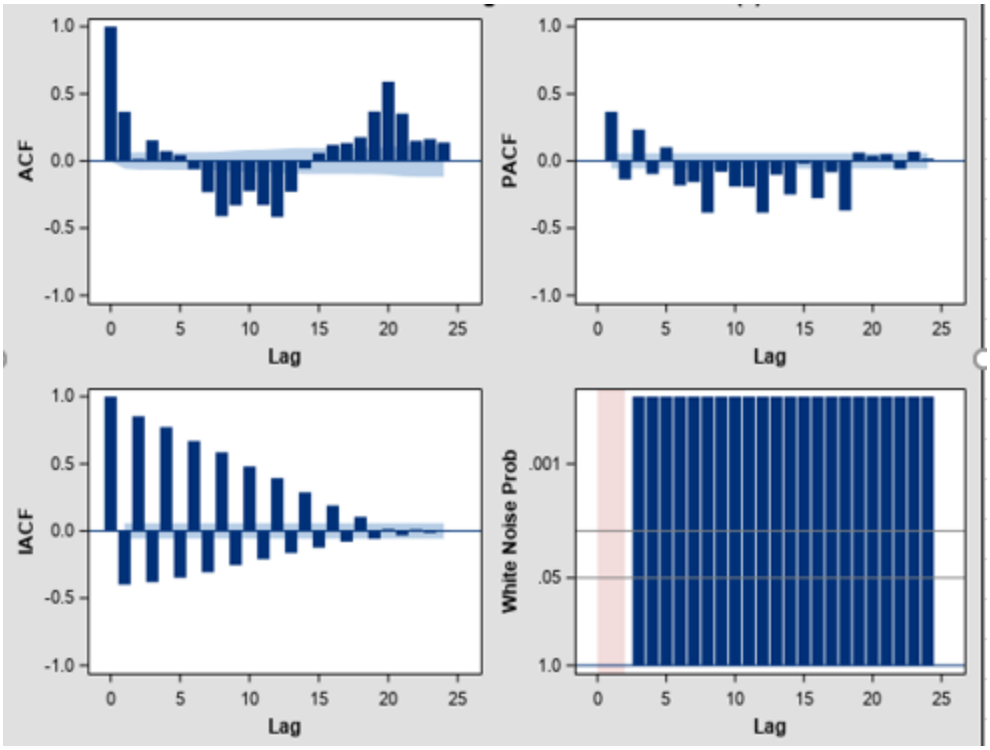

Figure 5.11: ACF residuals for total travel distance(km).

5.3.2 Normality data test

By using the Jarque-Bera (JB) test in **R**, a p-value of 0.0013 is calculated, which is less than 0.05 significance level. Hence we reject H_O , and conclude that the residuals have a non normal distribution.

The histograms and Q-Q plots for four variables in figures below shows the residuals from out-of-sample forecasts. The histograms also shows that the residuals have flatter tails and a higher kurtosis than a normal data set. Therefore, we confirm that residuals are not normally distributed. They neither have a constant variance nor 0 mean.

The normal Q-Q plot in figures also confirms the non normality of residuals.

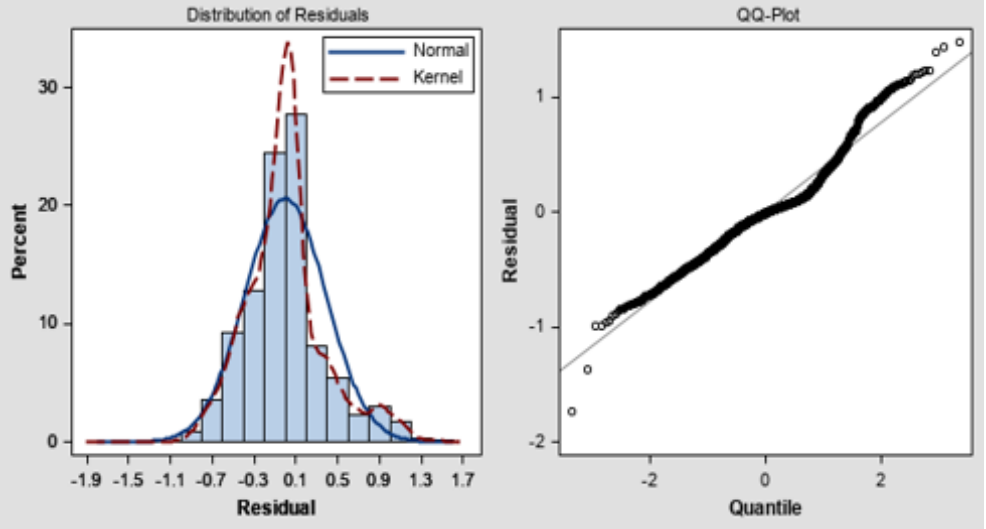

Figure 5.12: Histogram and Q-Q graphical representation of residuals for transactional cost variable.

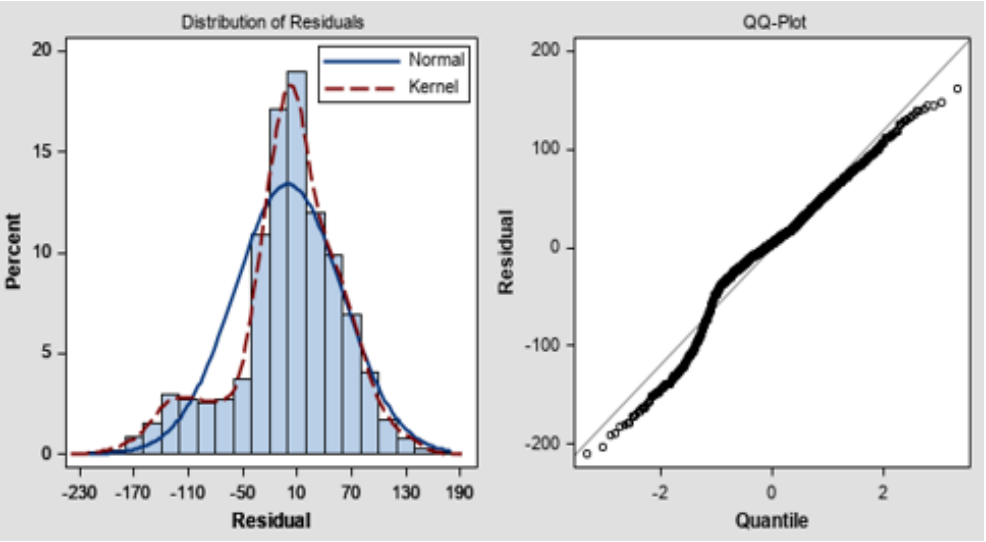

Figure 5.13: Histogram and Q-Q graphical representation of residuals for travel cost variable.

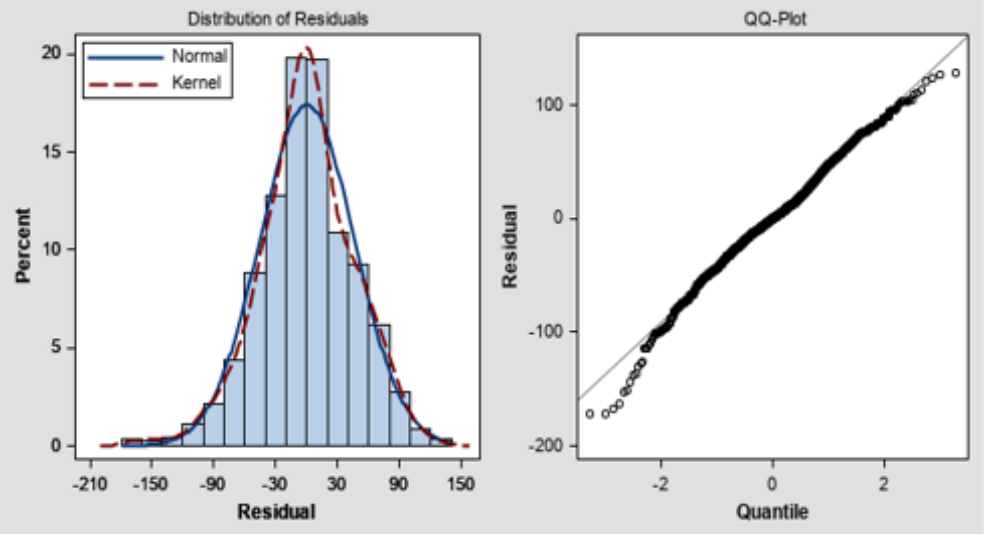

Figure 5.14: Histogram and Q-Q graphical representation of residuals for travel time variable.

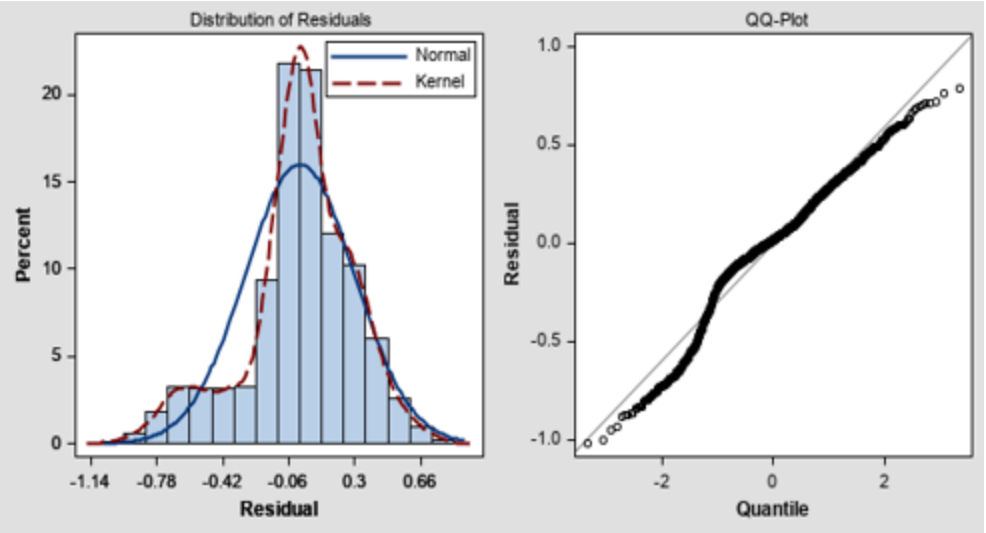

Figure 5.15: Histogram and Q-Q graphical representation of residuals for total travel distance(km) variable.

5.3.3 Conclusion

ARIMA models are capable of making predictions using time series data with any form of a pattern and with autocorrelations, that is why we are going to use them. We have statistically tested and validated that residuals in the fitted ARIMA model are correlated, and are not normally distributed. Therefore, ARIMA(1,0,1)(0,1,0)[365] was chosen and can be improved to produce better forecasts. Similar to other forecasting models, ARIMA models are limited somehow on accuracy of predictions. Taking into consideration the change in variance of the data at hand, the forecasts provided by an ARIMA model can be improved in various ways. This is called volatility testing and various models have been made over the past years to take volatility into consideration while forecasting.

The univariate ARIMA model is good but not adequate as this is a multivariate electricity data. Hence, multivariate approach will be used for further analysis.

Chapter 6

Modelling variations in cost of attending to electrical faults using volatility forecasting models

Here we study some of the statistical methods for analysing and modelling volatility in any given data set, with specific emphasis on electricity cost data. The models to be studied are called conditional heteroscedastic models. Since our emphasis is on multivariate models, we shall study the Multivariate AutoRegressive Conditional Heteroscedastic (ARCH) model developed by Engle (1982) and the Generalized ARCH (GARCH) model developed by Bollerslev (1986) but first starting with the univariate part for our analysis.

6.1 The importance of the chapter

The purpose of this part of the study is to study some of the volatility forecasting models that are used in the analysis of multivariate time series data but we start with univariate and to use these processes to model volatility in the residuals of electricity cost data for South Africa. This helps improve the level of accuracy of the forecasts.

The specific objectives are to:

- Discover the best fitting model for the electricity cost data available.
- Assess contribution of these models to understanding of volatility in electricity cost.

- Examine and compare the ARCH model and its extensions with ARIMA models, both theoretically and practically.

To achieve these objectives, this chapter will be organized as follows, meaning and more understanding of volatility, development, then extend to multivariate ARCH and GARCH models, model specification, application to electricity cost data and lastly the conclusion.

6.2 The definition of volatility

Volatility is basically a statistical measure of the dispersion of returns for a given security. Volatility can be measured by using the standard deviation or variance between returns from that same security. Ordinarily, the higher the volatility, the riskier the security. Volatility is actually the amount of uncertainty or risk about the size of changes in the value of a security. It is also defined as level of uncertainty about changes in the value of a given variable (Islam et al., 2013).

A higher volatility means that a security's value can possibly be spread out over a larger range of values. A lower volatility defines that a security's value does not oscillate dramatically, but changes in value at a steady pace over a duration of time. Modelling volatility in electricity cost is very vital because many factors affecting the cost of electricity change in very short time intervals (Patton and Engle, 2001).

Volatility modelling improves the accuracy of forecasts by giving better variance estimates which can be used to compute more reliable prediction intervals (Tsay,2005). It also improves the efficiency in parameter estimation, especially when we deal with time series data. With electricity cost, the higher the volatility, the more complicated it gets to forecast the cost accurately because the values are widely spread out. Therefore, complex forecasting techniques are employed in such cases in order to accommodate all the values (Dralle, 2011).

6.3 The ARCH model

In 1982, Robert Engle created the AutoRegressive Conditional Heteroskedasticity (ARCH) models to model the time changing volatility often observed in economical time series data. These models are useful when the aim of the study is to analyse

and forecast volatility.

This part of the study gives the motivation behind the simplest GARCH model and demonstrate its helpfulness in examining portfolio risk. ARCH models assume that the variance of the current error term to be a function of the actual sizes of the previous time periods error terms, often the variance is related to the squares of the previous innovations (Diebold and Nerlove, 1989). This was the first and simplest model to provide a framework for volatility modelling. The acronym ARCH stands for AutoRegressive Conditional Heteroscedasticity. The AR part comes from the fact that this model is a type of autoregressive model.

Heteroscedasticity means non constant variance. However, with an ARCH model, it's not the variance that changes with time, rather, the conditional variance (Hassan and Malik, 2007). It represents the uncertainty about the next periods observation given all the information currently available. ARCH models are usually employed to data that assumes an unstable variance in the error term at any given point in the series (Engle, 1992). ARCH models assume that the variance of the current error term is a function of the previous time periods error terms (Perrelli, 2001). Eberly College Website recommend the possibility of using an ARCH model for any series that has changing variance, for example, residuals after fitting the ARIMA model to the data.

Y_t follows an ARCH process if,

$$Y_t = \sigma_t \epsilon_t \tag{6.1}$$

where σ_t is the the local conditional standard deviation of the process and is not directly observable (Tsay, 2005). It can be calculated from the conditional variance σ_t^2 which is connected to squares of the previous error terms, depending on the order of the process.

6.3.1 The ARCH(1)

An ARCH(1) is the simplest version of ARCH models. The number 1 in the brackets shows that it is of order 1. In ARCH(1) model, the conditional variance σ_t^2 is calculated as,

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 \tag{6.2}$$

where α_0 and α_1 are parameters, carefully chosen in order to avoid a negative conditional variance. That is, for positive variance, the conditions that $\alpha_0 > 0$ and $\alpha_1 > 0$

are assumed, and $\alpha_1 < 1$ is assumed for stationarity (Chatfield, 2002). It is clear from Equation 6.2 that variance at time t is connected to the value of the series at time t-1. Therefore, a big past residual implies a big conditional variance which in turn gives a large current residual Y_t, in absolute terms.

That is why it is common to expect large residuals to be followed by other large residuals and the same applies to smaller residuals (Talke, 2003). Due to the dependence of the conditional variance on past series values, the process Y_t is not independent. Substituting the Equation 6.2 gives an ARCH(1) model, which is represented as,

$$Y_t = \epsilon_t \sqrt{\alpha_0 + \alpha_1 y_{t-1}^2} \tag{6.3}$$

where Y_{t-1} defines the observed value of the derived series at time t-1.

6.4 Testing for ARCH effect

The history of ARCH models is indeed a very short one. For it was introduced by Robert Engle years ago. Ever since the development of Autoregressive Conditional Heteroscedasticity (ARCH) model Engle (1982), testing for the presence of ARCH has become a routine diagnostic.

To test for ARCH effects in ARIMA residuals, one can use the McLeod-Li test. This test was developed by McLeod and Li (1983) who proposed a formal test for ARCH effect based on the Ljung-Box test. The test looks at the autocorrelation function of the squares of the residuals and tests whether the first chosen, say L, autocorrelations for the squared residuals are collectively small in magnitude. The Ljung-Box Q-statistic of McLeod-Li test is given by:

$$Q = T(T+2) \sum_{k=1}^{L} (T-k)^{-1} r_k^2, \tag{6.4}$$

where in this case r_k is the sample autocorrelation of squared residual series at lag k. The statistic Q is used to test the null hypothesis of no ARCH effect in the data against the alternative hypothesis of the presence of ARCH effect.

The test statistic is asymptotically $\chi^2(L)$ distributed with L degrees of freedom (Janacek et al, 2000; Wei, 2007). However, we are not sure if the residuals are identically independently distributed through time. We use visual inspection of the time series plot of residuals and find tendency of large (small) absolute values of the residual process being followed by other large (small) absolute values, which is a common

behaviour of ARCH processes. This is evidence of the absence of ARCH effects.

In their work, Wang et al. (2005) suggested the use of the ACF of squared residuals in identifying dependency in the series. Therefore, the next step is to plot the ACF of squared residuals. If there are no significant spikes all through the lags that are considered, we conclude that there is no evidence of dependency in the residuals. This means that the variance of residual series is not conditional on its history. Therefore, the residual series does not exhibit an ARCH effect.

6.4.1 Forecasting electricity cost with ARCH(1) model

The ARCH(1) model can be extended to include many parameters. This means the conditional variance will depend on observations from q previous times, hence the term ARCH(q). In this case,

$$\begin{aligned} \sigma_t^2 &= Var(Y_t, Y_{t-1}, \cdots, Y_{t-p}) \\ &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \cdots + \alpha_q Y_{t-1}^2 \end{aligned} \tag{6.5}$$

where the restrictions $\alpha_0 > 0$ and $\alpha_i >= 0$, $i = 1, 2, \cdots, q$ for positive variance still hold like in ARCH(1). The properties of an ARCH(q) process are similar to those of an ARCH(1) process. The mean is still 0 and the variance takes into consideration the other parameters introduced in the model. ARCH models are suitably used when the change in variance takes short intervals. They can also be used for gradual changes over time, but, gradual increasing variance connected to a moderately increasing mean can be handled best when using transformation methods (Eberly College Website, 2015). Some disadvantages of using ARCH Models include (Tsay, 2005):

- The model assumes the same effect on volatility from both positive and negative errors, since it uses squares of previous errors. However, this is not correct, for instance, from financial point of view, reality shows that the price of a financial asset accord differently to positive and negative shocks.
- ARCH models do not provide fresh ideas for understanding the source of variations of any given time series. They only help us understand the behaviour of the conditional variance (Kearney and And Patton, 2000).

- ARCH models are credible to over predict the volatility since they respond slowly to large isolated errors to the new developed series.

When we forecast with volatility models, we mostly consider the variance of the data. Assume a series $Y_T = y_1, \cdots, y_T$, the forecast $y_T(l)$ is the minimum square error predictor and it minimises the expression $E(y_{T+l} - f(y))^2$ among all functions of observations [y, f(y)] (Talke, 2003). When dealing with time series data, $y_T(l)$ is calculated depending on observed data as,

$$\begin{aligned} y_T(l) &= E(y_{T+l}|Y_T) \\ &= E(\sigma_{T+l}\epsilon_{T+l}|Y_T) \\ &= \sigma_{T+l}E(\epsilon_{T+l}|Y_T) \\ &= 0 \end{aligned} \tag{6.6}$$

where,

$$y_t = \sigma_t \epsilon_t \tag{6.7}$$

Shephard (1996) suggested the use of squares of the series to make more meaningful forecasts for an ARCH model. They calculated $y_T(l)$ using,

$$\begin{aligned} y_T^2(l) &= E(y_{T+l}^2|Y_T) \\ &= E(\sigma_{T+l}^2\epsilon_{T+l}^2|Y_T) \\ &= \sigma_{T+l}^2 \\ &= \hat{\alpha_0} + \hat{\alpha_1}E(y_T^2) \end{aligned} \tag{6.8}$$

where,

$$E[\epsilon_{T+l}^2] \tag{6.9}$$

At time T, y_T is already observed, therefore its expectation takes the real observed value. The parameter $\hat{\alpha_0}$ and $\hat{\alpha_1}$ are the conditional maximum likelihood estimates.

$$\begin{aligned} y_T^2(1) &= \hat{\alpha_0} + \hat{\alpha_1} y_T^2 \\ &= \sigma_T^2(l) \\ &= E(\sigma_{T+l}^2 | y_T) \end{aligned} \tag{6.10}$$

Say l = 2, then the forecast is given as,

$$\begin{aligned} y_T^2(l) &= E(y_{T+2}^2 | Y_T) \\ &= E(\hat{\alpha_0} + \hat{\alpha_1} y_{T+1}^2 | Y_T) \\ &= \hat{\alpha_0} + \hat{\alpha_1} E(y_{T+1}^2 | Y_T) \\ &= \hat{\alpha_0} + \hat{\alpha_1}(\hat{\alpha_0} + \hat{\alpha_1} y_T^2) \\ &= \hat{\alpha_0} + \hat{\alpha_1} \sigma_T^2(1) \end{aligned} \tag{6.11}$$

6.5 The GARCH model

Due to the restrictions presented by the ARCH models, a better model was proposed by Bollerslev in 1986 (Bollerslev, 1986) to solve the problem of requiring many parameter to adequately describe any given data while using an ARCH model. It is said to be called the Generalised Auto Regressive Conditional Heteroskedasticity (GARCH) model. GARCH models allows the conditional variance σ_t^2 to depend on both previous conditional variances σ_{t-1}^2 and previous squared values of the series Y_{t-1}^2 (Dralle, 2011).

Using GARCH models to control the problem of heteroskedasticity helps to obtain valid standard errors, which can be used to evaluate the chosen model and also construct forecasts with correct prediction intervals (Bollerslev, 1996). GARCH models fit any data as well as any high order ARCH model, but are more advantageous because they hold the condition of parsimony. The idea behind a GARCH model is same to that behind ARMA model (Gaston, 2016). A high order AR or MA model may frequently be approximated by a mixed ARMA model, with fewer parameters (Chatfield, 2000). Just like an ARCH process, a GARCH process is where ϵ_t is still assumed to be a pattern of iid random variables with mean 0 and variance 1 (Bollerslev, 1996). σ_t^2 is a function of previous conditional variances and previous observed values of the series. However, it specifically depends on the order of the model.

6.5.1 GARCH(1,1)

A GARCH(1,1) procedure is simply a prolongation of an ARCH(1) process. In this specification, the current conditional variance σ_t^2 is expected to be an average of a past derived series and a past conditional variance, plus a constant,

$$\sigma_t^2 t = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{6.12}$$

The assumptions for stationarity and positive variance still hold like for an ARCH process, with the inclusion of the coefficient of the past conditional variance, β_1 (Gao, 2007). This can be explained by expansion of the model in Equation 6.12,

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1(\alpha_0 + \alpha_1 y_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \\ &= \alpha_0 + \beta_1\alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1\alpha_1 y_{t-2}^2 + \beta_1^2 \sigma_{t-2}^2. \end{aligned} \tag{6.13}$$

The expansion for the conditional variance can go on until infinity. That is a not so desirable situation, especially when applying the models to practical data.

6.5.2 GARCH(p,q)

A GARCH model of order (p,q) assumes the conditional variance rely on the squares of the last p-values of the series and on the last q-values of the conditional variance. The properties and applications of this model are not different from those of a GARCH(1,1) model, however, it is very rare to require the use of a GARCH model of order higher than (1,1) (Engle and Sheppard, 2001). If such a model is fitted to data and the stationarity condition is not satisfied, squared observations can be made stationary after taking the first differences.

This results into the Integrated GARCH (IGARCH) model. Other prolongations of the GARCH model involve Quadratic GARCH (QGARCH), which allows for negative shocks to have more influence on the conditional variance than positive shocks, and exponential GARCH (EGARCH) which allows an asymmetric response by modelling log σ_t^2, rather than σ_t^2 (Bauwens et al., 2006).

6.5.3 Model specification

Time series graphical representation of the data set is the very best identification tool that may be used. It is commomly easy to spot periods of increased variation throughout the series. It is also helpful to study the ACF and PACF of both Y_t and

Y_t^2. For example, if Y_t appears to be white noise and the PACF of the Y_t^2 suggests AR(1), then ARCH(1) model for the variance is proposed (Shephard, 1996). In practice, it is advisable to exemplify with various ARCH and GARCH structures after realising the importance in the time series plot of the series (Eberly College Website, 2015).

Discovering a correct ARCH or GARCH model is not as easy as dealing with linear models, which partially describes why many analysts assume GARCH(1,1) to be the standard model (Chatfield, 2002). A series with GARCH(1,1) variances may look like uncorrelated white noise if 2nd order attributes alone are examined, and so non-linearity has to be evaluated by pondering the properties of higher order moments (as for other non-linear models). If Y_t is GARCH(1,1), then it can be shown that Y_t^2 has the same autocorrelation structure as an ARMA(1,1) process. In their study, Garcia et al. (2005) propose a general scheme for obtaining a desired and appropriate GARCH model as follows :

- Models are formulated assuming some certain hypotheses. In this step, a general GARCH formulation is appointed to model the available data. This selection is carried out by inspection of the main characteristics of the series. For example, in most of the competitive electricity markets, the data usually exhibits high frequency, non constant mean and variance, and multiple seasonality. These factors are among the main ones applied when selecting the GARCH model.

- A model is discovered for the observed data. A trial model must be distinguished for the available data, as seen in the first step. In the very first trial, the observation of the ACF and PACF graphs of the data can help to make this selection. In successive trials, the same observation of the residuals obtained can polish the structure of the functions in the model.

- The model parameters are calculated. After the functions of the model have been specified, the parameters of these functions must be figured. Good estimators of the parameters may be found by maximizing the likelihood with respect to the parameters. Any Statistical software system can be used to estimate the parameters of the model in the previous step.

- The model can now be used to forecast future values of the data.

6.5.4 Remarks

In this chapter we studied statistical methods for forecasting, analysing and modeling volatility in a given data set, specifically on electricity cost data. These models are called conditional heteroscedastic models. Looking at the theory for forecasting electricity cost with ARCH and GARCH models. The next chapter is all about application of ARCH and GARCH effect models to Arima residuals.

Chapter 7

Applying ARCH and GARCH models to ARIMA residual for forecasting electricity cost

Here, we use the residuals from an ARIMA(1,0,1)(0,1,0)[365] model for South Africa respectively. We test for the ARCH influence in the data set. We need to test whether these residuals display a change in variance before applying volatility models.

7.1 ARCH effect in ARIMA(1,0,1)(0,1,0)[365] model's residuals

To test for ARCH effects in ARIMA residuals, one can use the McLeod-Li test. This test was developed by McLeod and Li (1983) who suggested a formal test for ARCH effect based on the Ljung-Box test. The test checks at the autocorrelation function of the squares of the residuals and tests whether the first chosen, say L, autocorrelations for the squared residuals are small in magnitude.

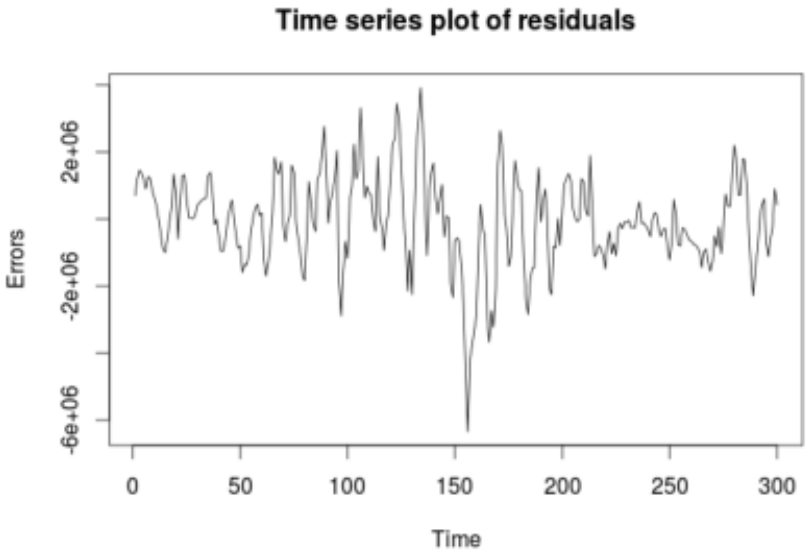

Figure 7.1: Time series graphical representation of residuals.

By visual examination of the time series graph of residuals in the Figure 7.1, there exists a tendency of large and small absolute values of the residual process being followed by other large and small absolute values. This is evidence of an ARCH processes. To certainly confirm presence of ARCH effects, we plot the ACF of squared residuals. There are many significant spikes, indicating the existence of dependency in the residuals. This means the variance of residuals is conditional on it's own history.

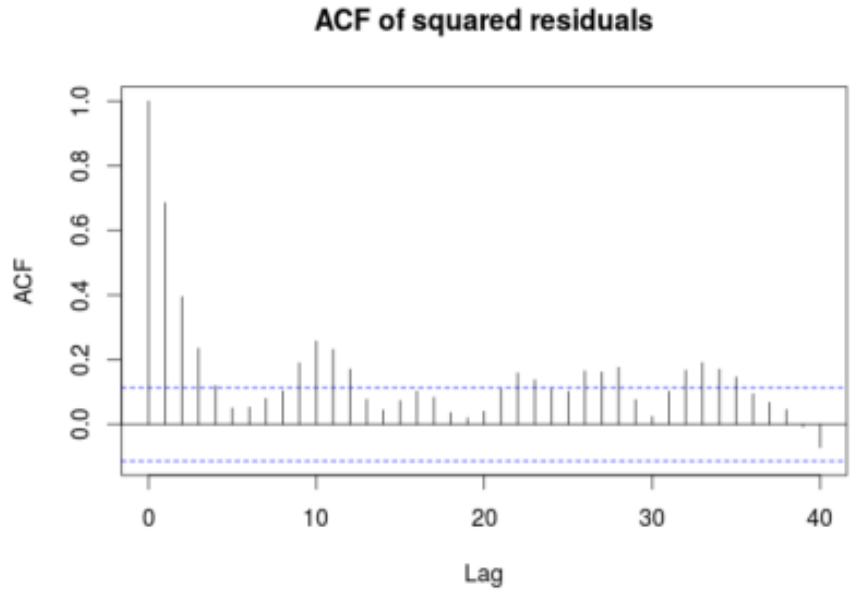

Figure 7.2: ACF of squared residuals.

Using the formal ARCH test in **R**, a p-value of 0.00014 is returned which is less than 0.05. Therefore, we reject H_0 and deduce that there is an ARCH effect in the squared residuals of the ARIMA(1,0,1)(0,1,0)[365] model. Thus, the heteroskedasticity of errors needs to be further analysed using a volatility model such as GARCH where the variances are modelled as an AR(p) model. Since we observe both positive and negative changes in the daily electricity load, this is one other factor for considering a non-linear model to analyse the errors.

7.1.1 Model order selection

We plot the PACF of squared residuals in order to select the correct q-values to use when constructing the appropriate model. This is shown in Figure 7.3.

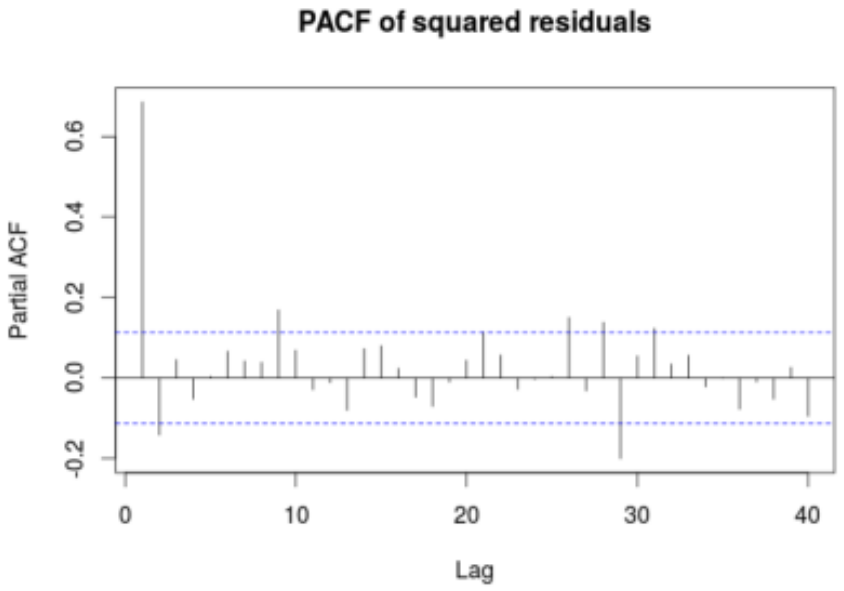

Figure 7.3: PACF of squared residuals.

From the PACF in Figure 7.3, we read off q = 1,2 because there are two significant spikes. Using the the rugarch function in **R**, we try different plausible models and compare their information criteria to choose the best. We also fit a standard GARCH(1,1) to the data, whose errors are assumed to follow a normal distribution.

Table 7.1 shows the Akaike(A), Bayes(B), Shibata(S), and Hannan-Quinn(H-Q) results for all the possible models,

Table 7.1: The criteria for different models.

Model	A	B	C	D
ARCH(1)	-0.7524	-0.6639	-0.6351	-0.7532
ARCH(2)	-0.6354	-0.4542	-0.5436	-0.6554
GARCH(1,1)	-0.6789	-0.4378	-0.6849	-0.5465
RGARCH(1,1)	-0.2574	-0.3447	-0.3574	-0.8136
STANDARD GARCH(1,1)	-2.5547	-2.4977	-1.4359	-1.3465

By looking at the Table 7.1, we can deduce that the best model from the analysis is the Standard GARCH(1,1) assimilated to all the other suggested models. It has the lowest values for all information criteria. Then we look at the coefficients of the model as in the Table 7.2:

Table 7.2: The coefficients for the Standard GARCH(1,1) model.

	μ	w	α_1	β_1
Parameter	-0.07264	0.0662	0.7676	0.2572
Standard Error	-0.03914	0.0045	0.0253	0.0964
P-values	0.001	0.001	0.001	0.001

In conclusion, the chosen best model is,

$$\sigma_t^2 = 0.0662 + 0.7676\epsilon_{t-1}^2 + 0.2572\sigma_{t-1}^2. \tag{7.1}$$

Therefore, since the p-values are less than the significance level, the conclusion is that the parameter estimates are statistically significant.

7.1.2 Forecasting volatility with Standard GARCH(1,1) model

The identified model can then be used to make volatility forecasts. The chosen Standard GARCH(1,1) model forecasts the volatility in residuals for the cost and fault of electricity.

7.1.3 Dependency of Standard GARCH(1,1) residuals

Another set of information returned from fitting the standard GARCH(1,1) model to the residuals is the test for dependency on the model's standardised squared residuals. Different lags are tested and none of them returns a p-value below 0.05. Therefore we fail to reject the null and deduce that there is no dependency in the squared residuals. Table 7.3 shows a sample of the tested lags.

Table 7.3: Dependency test for different lags.

Lag	Statistic	P-Value
6	0.00724	0.8845
12	3.40914	0.6987
18	6.23433	0.7456

7.1.4 Remarks

A Standard GARCH(1,1) model has been applied to residuals from a seasonal ARIMA model. Residuals of the Standard GARCH(1,1) model show a big improvement as compared to residuals from the linear seasonal ARIMA(1,0,1)(0,1,0)[365] model. This means, a combination of these two models gives better forecasts as compared

to a simple ARIMA model.

More sophisticated models have not been considered in this study for various reasons. For example;

- The purpose of this part of the research was to find a model that fits the data and can be used by policy and decision makers in the electricity sector to make the most accurate forecasts possible. If the purpose was to study various volatility models then more complicated models would have been considered.
- Extended versions of the GARCH model (IGARCH, EGARCH, TGACRH, to mention a few), work best when dealing with financial data.

Chapter 8

Time series analysis using multivariate models in forecasting electricity cost.

In this chapter, we tackle each element of this model individually as we build up to its general purpose. ARIMA models are mainly used for forecasting data that is originally non-stationary but can be made stationary by the process of differencing (Karamouz et al., 2012; Nason, 2006).

8.1 The multivariate AR model

Given a univariate time series, its orderly measurements hold information about the process that generated it. An attempt at illustrating this underlying order can be attained by modelling the current value of the variable as a weighted linear sum of it's previous values. This is an AutoRegressive(AR) process and is a simple, yet powerful, approach to time series characterisation (Chatfield, 1996).

The order of this model is the number of preceding observations that are used, and the weights characterise the time series. Multivariate autoregressive models lengthen this approach to multiple time series so that the vector of current values of all variables is modelled as a linear sum of previous activities. A MAR(m) model foretell the next value in a d dimensional time series, Y_n as a linear combination of the m previous vector values.

$$Y_n = \sum_{i=1}^{m} y_{n-i} A(i) + e_n \tag{8.1}$$

where $y_n = [y_n(1), y_n(2), \cdots, y_n(d)]$ is the n^{th} sample of a d dimensional time series, and $e_n = [e_n(1), e_n(2), \cdots, e_n(d)]$ is additive Gaussian noise with zero mean and covariance. We assume that the data mean has been removed from the time series. The model can be in a standard form of a multivariate linear regression model as follows,

$$y_n = x_n w + e_n \tag{8.2}$$

where $x_n = [y_{n-1}, y_{n-1}, \cdots, y_{n-m}]$ are the m previous multivariate time series samples and **W** is a (m-d)-by-d matrix of MAR coefficients. If the n^{th} rows of **Y** , **X** and **E** are y_n, x_n and e_n respectively and there are $n = 1, \cdots, N$ samples then we can write,

$$Y = XW + E \tag{8.3}$$

where **Y** is an (N - m)-by-d matrix, **X** is an (N - m)-by-(md) matrix and **E** is an (N - m)-by-d matrix. The number of rows N-m arises as samples at time points before m do not have sufficient preceding samples to allow forecasting.

8.2 The multivariate MA model

A Vector Moving Average(VMA) model of order q, or VMA(q), is written as,

$$r_t = \theta_0 - \alpha_1 a_{t-1} - \cdots - \alpha_q a_{t-q} \tag{8.4}$$

or

$$r_t = \theta_0 + \alpha(B) a_t \tag{8.5}$$

where θ_0 is a k dimensional vector, α_i are k(k) matrices, and $\alpha(B) = I - \alpha_1 B - \cdots - \alpha_q B^q$ is the MA matrix polynomial in the back shift operator **B** (Nyblom and Harvey, 2000).
The mean of VMA(q) is

$$\mu = E(r_t) = \theta_0 \tag{8.6}$$

8.3 The multivariate ARMA model

A VARMA(p, q) model is written as follows,

$$\phi(B)r_t = \phi_0 + \alpha(B)a_t \tag{8.7}$$

where $\phi(B)$ and $\alpha(B)$ are two matrix polynomials. For $v > 0$, the (i, j) the components of the coefficient matrices ϕ_v and α_v measure the linear dependence of r_{1t} on $r_{j,t-v}$ and $a_{j,t-v}$. The very important and adequate condition of weak stationarity for rt is similar as that for the VAR(p) model with matrix polynomial $\phi(B)$ (Enders, 2004).

8.4 Multivariate modelling for electricity cost using ARIMAX model

ARIMAX is an abbreviation for AutoRegressive Integrated Moving Average with exogenous variables. It is a logical prolongation of ARIMA modelling that incorporates independent variables which sum up explanatory value. When AR and MA terms in a pure ARIMA model are not adequate to supply an acceptably some measure of a model's overall explanatory power, it is very good to look for other driving phenomena whose inflames over time is not sufficiently implanted in the historical values of the dependent time series (Karlaftis and Vlahogianni, 2011).

Structuring an ARIMAX model calls for the combination of the predictive value of both the trailing time series values themselves (y_t) and the trailing model errors (ϵ_t) with the predictive value of exogenous variables. For example, if a set of exogenous variables serving as independent variables in a multiple regression were all highly significant, did not show significant cross correlation and yielded a high R^2 with the time series of residuals approximating white noise, then there would be no need for ARIMAX modelling (Karlaftis and Vlahogianni, 2011). However, if that same multivariate regression equation produced residuals that yielded significant serial correlation, then pure ARIMA modelling of the residuals would be required to remove the serial correlation so that t-statistics could be properly calculated and the significance of the independent variables could be judged properly.

Simply put, an ARIMAX model can be seen as a multiple regression model with one or more AutoRegressive (AR) terms, one or more Moving Average (MA) terms. AutoRegressive terms for a dependent variable are lagged values of that dependent

variable that have a statistically significant relationship with its most recent values. Moving average terms are nothing more than residuals resulting from previously made estimates (Karlaftis and Vlahogianni, 2011).

For instance, an unknown time series dependent variable y_t might be well estimated by a weighted combination of the following 4 right hand side (RHS) variables.

- x_t = is the value of independent variable x at time t.
- y_{t-1} = the immediately preceding value of the contigent variable y_t at time t-1.
- y_{t-2} = the immediately preceding value of the contigent variable y_t at time t-2.
- ϵ_{t-4} = is the estimation error exhibited by the model at time t-4.

This single independent variable, multiple regression like model for estimation of the dependent variable y_t relies on the predictive value of the independent variable x, the dependent variable (lagged by 1), the dependent variable again (lagged by 2) and a previously exhibited error term (lagged by 4) (Beal and Dennis, 2007). That is,

$$\hat{y_t} = \hat{\beta}_1 x_t + \hat{\phi}_1 y_{t-1} + \hat{\phi}_2 y_{t-2} + \hat{\theta}_1 \epsilon_{t-4} \tag{8.8}$$

where $\hat{\beta}_1$, $\hat{\phi}_1$, $\hat{\phi}_2$ and $\hat{\theta}_1$ are estimated coefficients.

There are statistical assumptions that must be tested for ensuring that the produced ARIMAX model is valid at each level of its evolution. The first two of these assumptions pertain to the residuals produced by the model, and the other four left relate to the exogenous variables that comprise the model.

- Assumption one: ARIMAX model building may not be initiated until the time series is stationary. The degree of stationarity of the residuals may be statistically assessed by using the Augmented Dickey-Fuller(ADF) test. With pure ARIMA model building, the p-values for the augmented Dickey-Fuller test for a single mean must be acceptably small for ensuring stationarity (Brocklebank et al., 2003).
- Assumption two: The residual series must not show significant serial correlation. The Ljung-Box test may be used to statistically assess the degree to which the residuals are correlated. If significant serial correlation is there among the residuals, it may be reduced by adding an appropriate combination of significant AR and/or MA terms distinguished from the PACF and ACF (Brocklebank et al., 2003).

- Assumption three: The coefficient estimated for an exogenous variable must be significantly different from 0, as judged by it's t-statistic. The calculation of the significance levels of t-statistics (p-values) for regression coefficients assumes that regression residuals are white noise. If Assumption two is disobeyed, and the residuals are not white noise, then their serial correlation must be abstracted with ARIMA modelling. This bids for pure ARIMA modelling process (Brocklebank et al., 2003).

- Assumption four: An exogenous variable mustn't exhibit evidence of receiving feedback from the dependent variable. That is, an attractive exogenous variable petitioner should show a significant relationship with the dependent variable without the dependent variable showing a causal relationship with it. The directions of significant causality between an exogenous variable and the dependent variable can be tested using the Granger Causality test. If the reverse causality is discovered, the exogenous variable must be removed from the independent variable candidates. This test must be executed on the dependent and independent variable in their current form (transformed or untransformed) (Brocklebank et al., 2003).

- Assumption five: An indication of the coefficient for each significant exogenous variable must be reasonable. The expected (reasonable) indicator can be defined prior to model construction by examining the signs of exogenous variable correlation coefficients that show a significant correlation with the dependent variable. If the dependent variable required a transformation to attain stationarity, that similar transformation would also be applied to the independent variable candidates, and the bivariate correlation analysis would then focus on the transformed variables (Brocklebank et al., 2003).

- Assumption six: The surviving exogenous variables composing the final model must not show a significant degree of multicollinearity. To meet this condition, each of the surviving exogenous variables must be individually tested for significant multicollinearity by using the variance inflation factor (VIF = $1/[1\text{-}R^2]$) for ensuring they are all linearly independent. A VIF of 10 or less is considered to show an acceptable level of correlation among the exogenous variables (Brocklebank et al., 2003).

8.5 Forecasting electricity cost for all four variables using multivariate ARIMAX model

These 4 very small p-values from the Ljung-Box test support the rejection of the null hypothesis that there is absolutely no autocorrelation in the residuals. This provides an indication that AR or MA terms must be added into the model to get rid of the serial correlation.

Table 8.1: SAS software output autocorrelation check of residuals.

Lag	chi square	p values	Autocorrelation
6	67.15	0.0001	-0.457
12	78.25	0.0001	-0.567
18	77.97	0.0001	0.032
24	80.56	0.0001	0.084

Same as in the case of structuring a pure ARIMA model, the process of adding AR and/or MA terms into the model is driven by the significance of the ACF and PACF spikes in the residual time series. Testing of both ACF and PACF displays that the most significant spike is in the ACF at lag 4, showing that θ_4 term should be involved in the model. The PACF and ACF of the time series of revised residuals shows that there are equally significant spikes at lag 1 in both the ACF and the PACF. As shown in rows 2 and 3, there is an introduction of MA1 term into the model and AR1 term. The correlation between the θ_1 and θ_4 terms is 0.323, and the correlation between ϕ_1 and θ_4 is 0.162. This defines that the impact that including an θ_1 term has on the t-statistic of the θ_4 term, decreasing it from 10.34 to 3.24. In opposition, an introduction of ϕ_1 term instead of θ_1 term only reduce the θ_4 t-statistic from 10.34 to 9.36.

Table 8.2: SAS output ARIMAX model building results.

Model	coefficient	t-statistic	p values
MA4	$\hat{\theta_4} = 0.7456$	$\hat{\theta_4} = 10.34$	0.0001
AR1, MA4	$\hat{\phi_1} = -0.4673, \hat{\theta_4} = 0.7566$	$\hat{\phi_1} = -3.56, \hat{\theta_4} = 9.36$	0.0001
MA1, MA4	$\hat{\theta_1} = 0.4234, \hat{\theta_2} = 0.7345$	$\hat{\theta_1} = 1.28, \hat{\theta_2} = 3.24$	0.0001

With pure ARIMA model, it is important to ensure that the residuals of the ARIMAX model satisfy the two conditions of normality and homoscedasticity. The Kolmogorov Smirnov test on the residuals from the ARIMAX model produce a statistic of 0.071 with a p-value of 0.180, which doesn't support the rejection of the null hypothesis of normally distributed residuals.

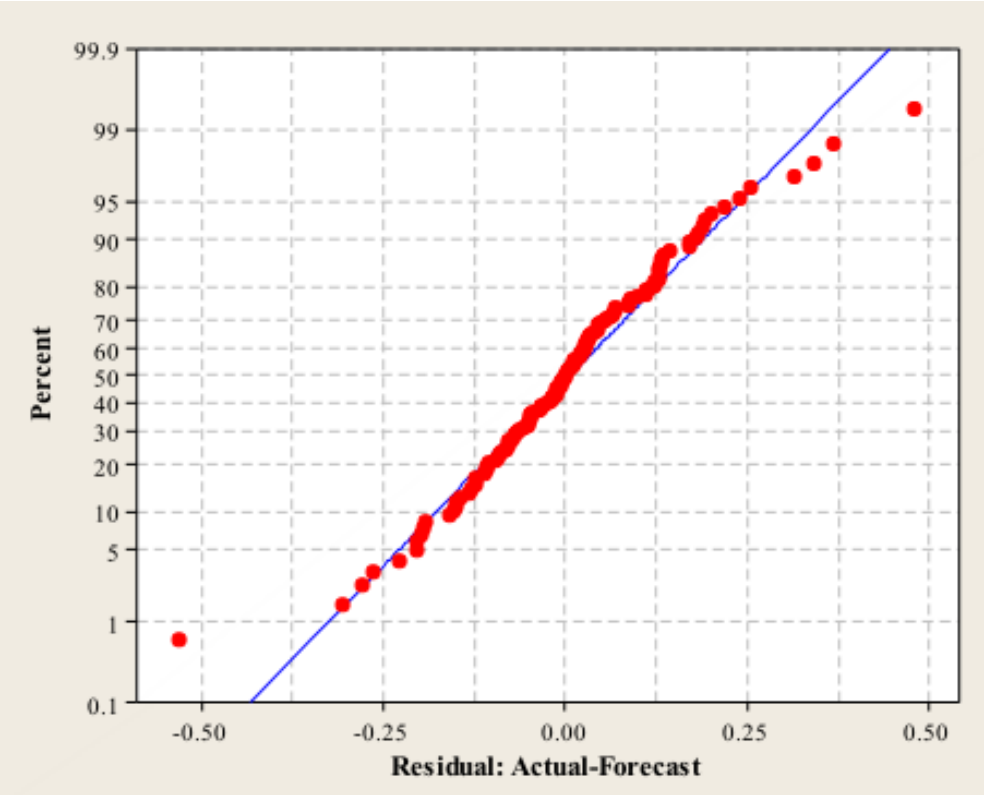

Figure 8.1: Normal probability plot of residuals.

The Schwarz Bayesian criterion(-110.996) of the ARIMAX model is attractive than the value of the pure ARIMA model (-94.537). It is advantageous to test the goodness-of-fit measures such as the RMSE and MAPE for the new developed ARIMAX model and to compare them to those of the pure ARIMA model to assess the improved precision of the ARIMAX model.

The White's test, with a p-value of 0.3021, shows that the null hypothesis of homoscedasticity can not be rejected.

When comparing the values for MAPE, mean and standard deviation in Table 8.3, the ARIMAX model is preferable to the ARIMA model in every manner, with it general lower MAPE, means and standard deviations. Therefore, ARIMAX could be the better model for forecasting electricity.

Table 8.3: Performance comparison of ARIMA and ARIMAX models.

Model	Mean	Std. Dev.	MAPE
ARIMA	3.94	2.70	3.25
ARIMAX	3.12	2.31	2.79

The p-values for the coefficients of AR1, MA4 and the independent variable were all valued 0.0001, and the p-value for AR2 term is 0.0064. After the significant (p-values less than 0.05) AR1, AR2 and MA4 terms are entered, it is very important to ensure that the exogenous variable(X) in the model stay significant.

Table 8.4: ARIMAX model.

Parameter	estimate	t value	P-value
$\phi_2 = AR1, 2$	-0.34577	-3.25	0.01
$\phi_1 = AR1, 1$	-0.73422	-4.79	0.01
ϕ_0	-0.000356	-6.44	0.01
$\theta_1 = MA1, 1$	0.84563	7.36	0.01

The composition of ARIMAX model was built using the SAS routines. After differencing to remove seasonality, the AR1, MA4 model exhibited a mean error value of 0.00272 and a standard error value of 0.0260. . The best fitting ARIMAX model, that is,

$$\hat{y_t} = \phi_0 - \phi_1 y_{t-1} - \phi_2 y_{t-2} + \theta_1 \epsilon_{t-1}, \tag{8.9}$$

which results to,

$$\hat{y_t} = -0.000356 - 0.73422 y_{t-1} - 0.34577 y_{t-2} + 0.84563 \epsilon_{t-1} \tag{8.10}$$

8.6 Vector AutoRegressive model for tackling electricity cost crisis

Vector AutoRegression (VAR) model is a prolongation of univariate autoregression model to multivariate time series data. VAR model is said to be a multi equation system whereby all the variables are handled as endogenous (dependent) variables. In it's reduced form, the right hand side of each equation involves lagged values of all dependent variables in the system, there is no contemporaneous variables (Karlaftis and Vlahogianni, 2011).

In multiple time series context, vector autoregressive (VAR) models are perhaps the common and broadely used models able to account for linear relationships among different time series. Unlike univariate, VAR is a multivariate modelling technique that considers multiple equation system or a multiple time series generalisation of AR models (Karlaftis and Vlahogianni, 2011).

In VAR models, each variable is a linear function of past lags of itself and of other variables taking into account the interdependence among variables included in the model. The vector autoregressive model of the order p, defined as VAR(p), is as follows:

$$x_i = \phi_1 x_{i-1} + \cdots + \phi_p x_{i-p} + \epsilon_i \tag{8.11}$$

where x_i is a multivariate random variable, ϕ_j $(j = 1, \cdots, p)$ are coefficient matrices. As in case of AR, parameters can be estimated by OLS or ML methods. In case of stationary series, VAR is fitted directly to the data otherwise differentiation are made before fitting a VAR model. In general, two choices to be made in prior using a VAR model to forecast. The first one corresponds to the number of variables, say j, whereas the second is the number of lags, say p, to be involved in the system. Thus, the number of coefficients to be estimated in a VAR model is equal to $j + j^2 p$. Generally, cross validation techniques and different information criteria are commonly used for the selection of number of lags (Yang, 2013). Apart from the fact that VAR models gives a systematic way to capture rich dynamics of the given multiple time series, they become difficult to estimate when the number of variables get higher.

8.6.1 Cost forecasting using VAR model

The forecasting ability of each model was evaluated by different prediction accuracy statistics. The result concerning the day ahead maintenance cost forecasting are listed in Table 8.5 and the reported results indicate that multivariate models produce lower prediction error compared to univariate models in general. However, the differences are small compared to error sizes.

Table 8.5: Electricity cost prediction accuracy statistics.

Model	MAPE	MAE	Median
AR	5.02	12.58	4.02
NPAR	5.57	14.87	3.96
FAR	4.32	11.83	3.42
VAR	5.63	12.42	3.53

These other models are AutoRegressive (AR), Nonparametric AutoRegressive (NPAR), Functional AutoRegressive (FAR), Vector AutoRegressive (VAR).

Table 8.6: Electricity cost day specific mean absolute percentage errors(DS-MAPE).

Models	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
AR	7.83	6.21	5.43	5.34	4.78	6.35	6.92
NPAR	8.63	5.83	4.95	6.01	5.42	7.64	7.64
FAR	6.86	5.35	3.85	4.66	4.32	5.32	4.87
VAR	6.71	5.75	4.40	4.87	4.62	5.49	5.87

From Table 8.6 we can see that the day specific mean absolute percentage errors(DS-MAPE) values are relatively higher on Monday, Saturday and Sunday and smaller on other weekdays. The effect of parametric and non-parametric approach is also evident in this table as parametric approach produces lower errors compared to other non-parametric approach.

8.7 Conclusion

Modelling and forecasting electricity cost gained an increasing attention in competitive electricity markets. This chapter considered these issues by using different multivariate approaches. For the residual component, different univariate and multivariate models have been considered with advancing level of being complex. Linear parametric and non-linear non-parametric models, have been estimated and compared in a one day ahead out of sample forecast. By looking at the results, the analyses propose that the multivariate approach leads to better results than the univariate one and that, within multivariate framework, functional models are the most accurate ones, with VAR being a very competitive model.

Chapter 9

A Random Forest for forecasting the electricity cost

The topic of real-time electricity cost forecasting has been discussed thoroughly in the past few years. Too much attempt has been put into this from electricity sellers and buyers for the aim of obtaining the very best bidding strategy. The previously proposed tools have several bottlenecks.

Simple cost prediction is not that much of help compared with the cost probability distribution, that can help the sellers or buyers estimate the risk of their bidding decisions (Breiman, 2001). In cost probability distribution, the probability for a specific electricity cost can be known. The previously used forecasting models are not updatable. The market and climate are changing because of this technology taking over, which simply implies that we need a model that can adjust to the latest market and climatic condition automatically.

A Random Forest adaptive forecasting model is proposed in this chapter. By using its bootstrap distribution, it can produce an accurate prediction (Breiman, 2001). Furthermore, the random forest model adjust to the latest forecasting condition, for example, the latest climatic, seasonal and market conditions, by updating random forest parameters with the new observations. This kind of adaptability of random model avoids the model failure in a climatic or economic condition different form the training set.

Random Forests (RFs) are an ensemble learning process for both classification and regression problematic areas. In simple terms, random forest is a collection of decision trees that grow in randomly selected subspaces of the feature space. Random Forest is robust and easier to train compared to others (Olshen et al., 1984). The

power of random forest is to aggregate a set of binary decision trees (Breiman's CART – Classification And Regression Trees), each is constructed using bootstrap sample from the learning sample and a subset of features (input variables or predictors) that are randomly chosen at each node. In contrast to the CART model construction strategy, one tree in random forest is built on a subset of learning points and on subsets of features considered at each node to split on (Olshen et al., 1984). More over, trees in the forest are grown to maximum size and the pruning step is not done.

After a number of trees in ensemble are fitted using bootstrap samples, the final decision is attained by combining over the ensemble, for example, by averaging the output for regression or by means of voting for classification. This method of bagging improves stability and accuracy of the model, lessens the variance and helps to avoid over-fitting (Friedman et al., 2009). The bias of bagged trees is similar as that of the individual trees, but the variance is reduced by lessening the correlation between the trees.

9.1 The attributes of trees

- They handle large datasets.
- They can maintain mixed predictors quantitative and qualitative.
- They can easily disown redundant variables.
- Can handle missing data very well.
- Small trees are the easiest to understand and interpret (Friedman et al., 2009).

9.2 Bagging or Bootstrap Aggregating process

9.2.1 Why bagging?

- It reduce over fitting.
- It reduce bias.
- It break the bias variance trade-off.

Through this process, it is possible to form an ensemble/forest of trees where multiple training sets are generated with substitution, meaning data instances. Bootstrap is a statistical re-sample method (Breiman, 1996).

9.2.2 Bootstrapping algorithm

- Used in statistic when you want to estimate a statistic of a random sample (mean, variance, mode, etc...).
- Using bootstrap we diverge from traditional parametric statistic, we don't assume a distribution of the random sample.
- What we do is sample our only data set (random sample) with replacement. We take up to the number of observations in our original data.
- We again do step 3 for a large number of times, B times. Once done we have B number of bootstrap random samples (Aggarwal, 2009).
- We then take the statistic of each bootstrap random sample and average it.

The main disadvantages of bootstrapping could be the Bagging algorithm using CART. CART uses Gini-Index, a greedy algorithm to find the best split, so we end up with trees that are structurally similar to each other (Angeralides and Cronie, 2006). The trees are highly correlated among the predictions, but random forest address this.

9.2.3 Problem Random Forest is trying to solve

With bagging process, we have an ensemble of structurally similar trees. This causes highly correlated trees and somehow the results could not be what we want, but with random forest, we can create trees that have no correlation or weak correlation (Bunn, 2004).

9.3 Random Forest algorithm

- Take a random sample of size N with substitution from the data. This is just bootstrapping on our data as mentioned how it is done above (Ziegler et al., 2017).
- Take a random sample without substitution of predictors. Predictor sampling / bootstrapping, this is bootstrapping our predictors and it's without replacement and this is random forest solution to highly correlated trees that arises from bagging algorithm (Ziegler et al., 2017)..
- Construction of the first CART partition of the data. We partition our first bootstrap and use Gini-index on our bootstrapped predictors sample in step 2 to decided the split (Ziegler et al., 2017)..

- Again repeat step 2 for each split until the tree is as huge as wanted. Don't prune (Ziegler et al., 2017)..
- Repeat steps 1 to 4 a large number of times (e.g. 500). Steps 1 to 4 is to build one tree. We repeat step 1-4 a large number of times in order to build a forest. There's no magic number for large number, you can build 101, 201, 501, 1001, etc.. (Ziegler et al., 2017)..

9.3.1 An Out-Of-Bag(OOB) error estimation

There is a straight forward way to calculate the test error of a bagged model, without any need to execute the validation test set approach. The key to bagging is that the trees are repeatedly fit to bootstrapped subsets of the observations. One can show that on average, each bagged tree use around two-third of observations (Breiman, 1999). The remaining one-third of the observations that is not used to fit a given bagged tree are called out-of-bag(OOB) observations. We can also foretell the response for the i^{th} observation using each of the trees in which that observation was OOB. This will result approximately to [B/3] predictions for the i^{th} observation (Breiman, 1998).

To get a single prediction for the i^{th} observation, we can average the predicted responses. This yields a single OOB prediction for the i^{th} observation. An OOB prediction can possibly be found in this approach for each of the n observations, from which the overall OOB MSE or a classification error can be computed (Breiman, 1998). Bagging results in improved accuracy over prediction using just a single tree. Thus the resulting model can be somehow hard to interpret. One advantage of decision trees is the ease of interpretation. When we bag a very big number of trees, we can no longer represent the resulting statistical learning process using a single tree and it is somehow not clear which variables are the most significant than others. Bagging does improve the prediction accuracy at the cost of interpretability (Weron, 2006).

There are only 2 main reasons for using bagging process. The first one is that the use of bagging process enhances accuracy when random features are used. The second is that bagging process can be used to give the ongoing estimates of the generalization error of the aggregated ensemble of trees, as well as the estimates for the strength and correlation (Dietterich, 1998).
Breiman displayed that random forests do not over-fit as more trees are added, but exhibit a limiting value of the generalization error. The Random Forest generalization error is estimated by an out-of-bag (OOB) error, for example, the error for

training points which are not in the bootstrap training sets (Tibshirani, 1999). The big advantage of random forest is that they can be fit in one sequence, with cross-validation being the performer.

The algorithm for Random Forest:
1. k = 1 to K:
1.1. The first step is to draw a bootstrap sample L of size N from the training data.
1.2. Grow a random forest tree T_k to the bootstrapped dataset, by repeating the following steps for each node of the tree, until the minimum node size m is achieved.
1.2.1. Select F variables randomly from n variables.
1.2.2. Pick the very best variable among the F.
1.2.3. Split the node into 2 nodes.
2. Then output the ensemble of trees $T_k, k = 1, 2, \cdots, K$

For making a prediction at point x:

$$f(x) = 1/k \sum_{K=1}^{K} T_k(x) \tag{9.1}$$

9.3.2 Random Forest assumptions

1. At each stage of constructing individual tree, we find the best split of data.
2. While building a tree, we do not use the whole dataset, but we bootstrap sample.
3. We combine the individual tree outputs by averaging.

The predecessor of Random Forest, Classification and Regression Tree (CART), was proposed by Breman in 1984. Breiman present another important method for random forest, called Bagging, in 1996. Random Forest is an ensemble learning technique for classification process. It is based on 2 methods, the CART and Bagging. The CART is a tree built classification model that maps observations about an item to deduce about the item's class (Breman, 1984).

Figure 9.1 also gives a hint to the CART growing process, splitting each node into 2 sub nodes by finding the best split variable along with the best split value till achieving minimum node size. CART's benefit is that it can be fitted into data perfectly. When conducting prediction, CART's accuracy is not that good. CART got low bias but suffers from the high variance. In order to solve this issue, random forest extends CART by presenting Bagging technique. This implies that random forest fits a large number of CARTs into bootstrap sets resampled from the origin data set and random forest forecasts through the mode of the predictions generated by the fitted

CARTs. The Bagging process will lessen the variance of CART while keeping the bias low (Girish et al., 2013).

Adding to the low bias and low variance, random forest has other features, summarized below:

- Random Forest needes only three parameters. Following recommended values are very easy to tune.
- Random Forest can produce an out-of-bag error, a good estimate of the generalization error, in its growing process, while other models only need multiple training procedures like cross-validation to generate such estimates.
- Random Forest can produce variable significant indices in its growing process and they turn out to be good estimates of variable relevancies.
- Random Forest is robust against irrelevant components and outliers in training data.
- Structured as tree, random forest is easy to spread itself to fit more data by growing more branches. This leads to the random forest online learning algorithm and has made random forest a very good adaptive machine learning model (Ghosh and Kanjilal, 2014).

Random Forest parameters are, B(the number of trees), m(the number of candidate variables in each split), n_{min}(the minimum node size), α(the confidence level).

Table 9.1: Predictors for random forest (Breiman, 1998).

Predictor	Description
P(c-3)	The real-time cost 3 hours ago.
P(c-24)	The real-time cost 1 day ago.
P(c-168)	The real-time cost 1 week ago.
P(c-720)	The real-time cost 1 month ago.
L(c)	The current load.
W(c)	The current temperature.
D(c)	The day indicator.

The accuracy is presented in Table 9.2. By looking at the results in Table 9.2, it is displayed that random forest model outperforms both ANN and ARIMA models by having the smallest MAPE value.

Table 9.2: Predictors for Random Forest.

Model	random forest	ANN	ARIMA
MAPE	11.32%	12.94%	13.58%

Here, we assimilate the random forest model with STLF(Short Term Load Forecasting) models such as ARIMA, exponential smoothing and artificial neural network(ANN). The time series is preprocessed for the ANN model in a similar approach as for random forest. To get the best ARIMA and ES models automated procedures implemented in the forecast package for the **R** system were used.

MAPE (Mean Absolute Percentage Error) is used here to examine the performance of the forecasting models. The results of forecasts (MAPE for the test samples MAPE test and the interquartile range (IQR) of MAPE test) in Table 9.3 are shown. In this table, the results from using the NAIVE technique are also displayed. The forecast principle in this case states that the forecasted daily cycle is similar as seven days ago. Note that the lowest errors were gotten by ANN and RF. Mean errors for random forest and ANN are statistically indistinguishable (Wilcoxon signed-rank test was used) by using SAS software.

Table 9.3: Results of Forecasting.

Model	$MAPE_{test}$	IQR	MEAN($MAPE_{test}$,IQR)
RF	1.45	1.40	(1.15, 1.17)
CART	1.72	1.54	(1.41, 1.40)
Fuzzy CART	1.65	1.49	(1.38, 1.35)
ARIMA	2.68	2.39	(1.80, 1.66)
ES	2.38	1.78	(1.79, 1.57)
ANN	1.37	1.34	(1.13, 1.18)
NAIVE	5.34	4.65	(3.77, 3.87)

The error (PE) histograms in Figure 9.3 are also displayed. The most favourable error distributions are observed for random forest and ANN models. The distributions for ARIMA and ES are flattened and asymmetrical.

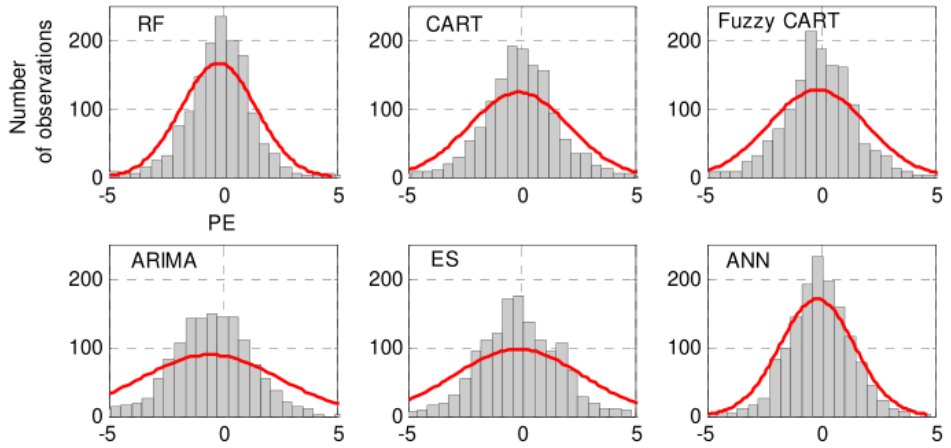

Figure 9.1: Histograms of errors.

9.4 Conclusions

The suggested approach allows us to forecast time series with multiple seasonal variations. It is due to the data pre-processing and defining the sequences of the seasonal cycles on which the model produce an effect. This elucidate the forecasting issue and leads to the better accuracy. The random forest forecasting model is attributed by being simple. The number of parameters to be estimated is small, which means a simple procedure of model optimization. In application, the random forest model produced an accurate forecasts as ANN and out performed the crisp and fuzzy CART, ARIMA and ES models. The random forest model is more simpler to train and tune than the above mentioned other models, random forest does not over fit and it reduces variance due to averaging the outputs of many simple regression trees over ensemble.

Chapter 10

Discussion and Conclusion

In this dissertation, we have researched models for forecasting electricity cost in South Africa. For this project, daily data from 4 January 2012 to 3 June 2017 was collected from ESKOM. We faced a problem of constrained details about the data. The purpose of the study was to search for the very best model together with the volatility forecasting model to deal with the data at hand and how better volatility models forecast electricity cost. However, it is hard to find a single model that outperforms all others in every situation. This research project addressed the issue of modelling and forecasting electricity cost following different approaches.

10.1 Comparison of models for forecasting electricity cost

Table 10.1: Comparison of models for forecasting.

Model	VAR	ARIMA	ARIMAX	Random Forest
MAPE	9.64%	9.72%	10.29%	15.74%

Based on Table 10.1 of MAPE values for different models, we can conclude that ARIMA, ARIMAX and VAR are the best models for the electricity data at hand for forcasting electricity cost, given that the models has the small and close values. The analyses propose that multivariate approach leads to good results than the univariate approach and that, within the multivariate time series models are, in general, the most accurate ones. This is confirmed by the results in Table 10.1.

For this South African dataset, a seasonal ARIMA(1,0,1)(0,1,0)[365] model was found to be the most suitable model. However, it returned autocorrelated non normal residuals which also tested positive for ARCH effects. Therefore, we were able to im-

prove the forecasts by modelling volatility. Different ARCH models were tried based on the PACF of residuals and the standard GARCH(1,1) model, with assumed normal residuals was chosen because of its lowest information criteria values. The Standard GARCH(1,1) parameter coefficients were estimated in order to fit the model to the residuals and use it to forecast the conditional variances.

The non-linear issues of variances were handled appropriately through the fitted standard GARCH models. Therefore, we conclude that it is normally the best way to test the volatility with variances and standard deviations after fitting linear models in order to improve the accuracy of the forecasts. These models gives flexibility to co-exist with other models.

The aggregation of the two Seasonal ARIMA and GARCH gives more accurate forecasts than just a model on it's own. R-programming is the best tool for modelling and forecasting electricity cost. This research is close to being similar to one carried out by Yasmeen and Sharif (2014) and Nkiyingi Winnie (2016), where monthly electricity consumption (EC) for Pakistan was studied and a model developed to forecast four years ahead.

Another study was carried out by Sigauke and Chikobvu (2011), who studied the prediction of daily peak electricity demand using three volatility forecasting models, a Seasonal AutoRegressive Integrated Moving Average (SARIMA) model, a SARIMA with Generalized AutoRegressive Conditional Heteroskedastic errors (SARIMA-GARCH) model and a Regression-SARIMA-GARCH (Reg-SARIMA-GARCH) model. Emphasis was given to both linear and non-linear models, ARIMA, Seasonal ARIMA (SARIMA), ARCH and GARCH models. Unlike results in our study, the ARIMA(3,1,2) model was the most appropriate model to forecast monthly electricity cost in Pakistan.

Similar to our results, the non-linear volatility model out performed linear models when dealing with daily cost data. This indicates that when policy makers want to make forecasts for medium term periods, the best models to consider are linear models. This is because monthly, quarterly, annual or any other longer period historical data does not experience high volatility. The residuals of such series are homoskedastic. Therefore, it would be both time and resource wastage to apply volatility models to such data.

This study also proposes a multivariate ARIMAX, VAR and random forest based adaptive model for electricity cost forecasting. For Random Forest, the main con-

tribution lies in these two aspects as follows, firstly, the model gives a confidence interval with the prediction. Secondly, the model can adjust to the latest forecasting areas by updating itself with new observations. Random forests is a powerful and conducive technique in prediction. Because of the law of big numbers they do not over fit.

On the other side, weekly, daily, hourly or any other shorter period historical data requires application of volatility models. This is because during short durations there is a lot of variation in data points. In order to save time, non-linear models should be taken into consideration initially when dealing with such data.

The technique used in constructing VAR and VARMA models is quite hard to apply, but on small time horizons, the forecasts based on these models are better than others for variables not affected by structural shocks. This conclusion has also been reached by other researchers, (Athanasopoulos and Hyndman, 2014). In future, it would be interesting to estimate the VAR and VARMA models using the state space form for which the technique based on Kalman filter are used for optimization. Other areas for further study would include,

- Fitting models which would accommodate non normality in case the available time series data isn't normal. For example, the Gamma function and student t-distribution.
- Introducing hybrid models in the forecasting process. This can be done through various ways for example, using Artificial Neural Networks (ANNs) or combining different models to develop one model that suits the data perfectly.
- Introduction of more complex time series models such as VARMA, VARMAX, EGARCH, ICA-GARCH and Support Vector Machine (SVM).

References

Winnie Nakiyingi. Forecasting electricity demand using univariate time series volatility forecasting models: A case study of uganda and south africa. Master of Science in Statistics, 2016.

C Sigauke and D Chikobvu. Prediction of daily peak electricity demand in south africa using volatility forecasting models. Economics, 33(5):882–888, 2011.

Samer Saab, Elie Badr, and George Nasr. Univariate modeling and forecasting of energy consumption: the case of electricity in lebanon. Energy, 26(1):1–14, 2001.

Winnie Nakiyingi. Forecasting electricity demand using univariate time series volatility forecasting models: A case study of uganda and south africa. Master of Science in Statistics, 2016.

George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M. Ljung. Time series analysis: forecasting and control. Economic Modelling, John Wiley, 2015.

Rob J Hyndman and George Athanasopoulos. Forecasting: principles and practice. 2014.

Rob J Hyndman and George Athanasopoulos. Forecasting: principles and practice. 2014.

Tao Hong and David A Dickey. Electric load forecasting: fundamentals and best practices. 2014.

Farah Yasmeen and Muhammad Sharif. Forecasting electricity consumption for pakistan. IJETAE, 2014.

Subhes C. Bhattacharyya and Govinda R. Timilsina. Energy demand models for policy formulation. The World Bank, Policy Research Working Paper 4866, 2009.

AnagnostakisE.MKodogiannisV.S. Softcomputingbasedtechniquesforshort- term load forecasting. Fuzzy Sets and Systems, 2002.

G. Dudek. Similarity-based approaches to short-term load forecasting. Forecasting Models: Methods and Applications. iConcept Press, 2010.

DHR Price and JA Sharp. A comparison of the performance of different uni- variate forecasting methods in a model of capacity acquisition in uk electricity supply. International Journal of Forecasting, 1986.

Hamilton and James Douglas. Time series analysis. Princeton university press Princeton, 1994.

George EP Box and Gwilym M Jenkins. Time series analysis: forecasting and con-

trol, revised ed. Holden-Day, 1976.

Dudek G. Short-Term Load Forecasting Using Fuzzy Regression Trees. 2010.

Dudek G. Short-Term Load Forecasting Using Fuzzy Regression Trees. 2010.

Samer Saab, Elie Badr, and George Nasr. Univariate modeling and forecasting of energy consumption: the case of electricity in lebanon. Energy, 26(1):1–14, 2001.

Freund, Y. and Schapire, R. [1996] Experiments with a new boosting algorithm, Machine Learning: Proceedings of the Thirteenth International Conference, pp. 148-156

Farah Yasmeen and Muhammad Sharif. Forecasting electricity consumption for pakistan. IJETAE, 2014.

Farah Yasmeen and Muhammad Sharif. Forecasting electricity consumption for pakistan. IJETAE, 2014.

Manoj Kumar and Madhu Anand. An application of time series arima forecasting model for predicting sugarcane production in india. Studies in Business and Economics, 9(1):81–94, 2014.

Farah Yasmeen and Muhammad Sharif. Forecasting electricity consumption for pakistan. IJETAE, 2014.

Manoj Kumar and Madhu Anand. An application of time series arima forecasting model for predicting sugarcane production in india. Studies in Business and Economics, 9(1):81–94, 2014.

Rashid Mohammed Roken and Masood A Badri. Time series models for forecasting monthly electricity peak load for dubai. Chancellor's Undergraduate Research Award, 2006.

Rashid Mohammed Roken and Masood A Badri. Time series models for forecasting monthly electricity peak load for dubai. Chancellor's Undergraduate Research Award, 2006.

Eric Ghysels, Denise R Osborn, and Paulo MM Rodrigues. Forecasting seasonal time series. Handbook of Economic Forecasting, 1:659–711, 2006.

Caston Sigauke and Delson Chikobvu. Daily peak electricity load forecasting in south africa using a multivariate non-parametric regression approach. ORiON: The Journal of ORSSA, 26(2), 2010.

Ratnadip Adhikari and RK Agrawal. An introductory study on time series modeling and forecasting. 2013.

Peter J Brockwell, Richard A Davis, and Matthew V. Calder. Introduction to time series and forecasting. 2002.

Peter J Brockwell, Richard A Davis, and Matthew V. Calder. Introduction to time series and forecasting. 2002.

H. Yamin M. Shahidehpour and Z. Li. Market operations in electric power systems. Forecasting, Scheduling and Risk Management, 2002.

Chris Chatfield. Time-series forecasting. CRC Press, 2002.

Chris Chatfield. Time-series forecasting. CRC Press, 2002.

Chris Chatfield. Time-series forecasting. CRC Press, 2002.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the Econo- metric Society, 1982.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3):307–327, 1986.

LM Bhar and VK Sharma. Time series analysis. Indian Agricultural Statistics Research Institute, New Delhi, pages 1–15, 2005.

Andrew T Jebb, Louis Tay, Wei Wang, and Qiming Huang. Time series analysis for psychological research: examining and forecasting change. Name: Frontiers in Psychology, 6:727, 2015.

Hyndsight Website. Cyclical and seasonal time series. http://robjhyndman.com/hyndsight/cyclicts/, Accessed May 2015.

Rob J Hyndman and George Athanasopoulos. Forecasting: principles and practice. 2014.

Rob J Hyndman and George Athanasopoulos. Forecasting: principles and practice. 2014.

Richard McCleary, Richard A Hay, Erroll E Meidinger, and David. McDowall. Applied time series analysis for the social sciences. 1980.

Richard McCleary, Richard A Hay, Erroll E Meidinger, and David. McDowall. Applied time series analysis for the social sciences. 1980.

Richard McCleary, Richard A Hay, Erroll E Meidinger, and David. McDowall. Applied time series analysis for the social sciences. 1980.

Philip Hans Franses. Time series models for business and economic forecasting. Cambridge university press, 1998.

Christiaan Heij, Paul De Boer, Philip Hans Franses, Teun Kloek, Herman K Van Dijk, et al. Econometric methods with applications in business and economics. OUP Oxford, 2004.

Douglas C Montgomery, Lynwood A Johnson, and John S. Gardiner. Forecasting and time series analysis. 1990.

Douglas C Montgomery, Lynwood A Johnson, and John S. Gardiner. Forecasting and time series analysis. 1990.

Douglas C Montgomery, Lynwood A Johnson, and John S. Gardiner. Forecasting and time series analysis. 1990.

Michael P Clements, David Hendry, et al. Forecasting with difference stationary and trend stationary models. The Econometrics Journal, 4(1):1–19, 2001.

Robert H Shumway and David S Stoffer. Time series analysis and its applications: with R examples. Springer Science and Business Media, 2010.

Chris Chatfield. Time-series forecasting. CRC Press, 2002.

Richard Scheaffer and Linda Young. Introduction to Probability and its Applications. Cengage Learning, 2009.

Daniel A Vivanco. Towards Emulation of Large-scale IP Networks Using End-to-end Packet Delay Characteristics. ProQuest, 2008.

G.C.S. Wang and C.L. Jain. Regression Analysis: Modeling and Forecasting. Graceway Pub., 2003. ISBN 9780932126504. URL https://books.google.co.za/ books?id=dRQAkwHHmtwC.

William Wu-Shyong Wei. Time series analysis. Addison-Wesley publ, 1994.

Fazlollah M Reza. An introduction to information theory. Courier Corporation, 1961.

Elaine R Monsen and Linda Van Horn. Research: successful approaches. American Dietetic Associati, 2007.

David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association,74(366a):427–431, 1979.

W Wang, PHAJM Van Gelder, JK Vrijling, and J Ma. Testing and modelling autoregressive conditional heteroskedasticity of streamflow processes. Non linear Processes in Geophysics, 12 (1), 2005, 2005.

Golam M Farooque. Effects of transformation choice on seasonal adjustment diagnostics and forecast errors. Report the esult of research and analysis undertaken by US Cencus Bureau, Washington DC, 2002.

John H McDonald. Handbook of biological statistics, volume 2. Sparky House Publishing Baltimore, MD, 2009.

G. Reinert. Lecture notes on time series, January-March 2002.

Bronislav Chramcov. Identification of time series model of heat demand using mathematica environment. ACMOS, 11:346–351, 2011.

Mohammad Karamouz, Sara Nazif, and Mahdis Falahi. Hydrology and hydrocli-

matology: principles and applications. CRC Press, 2012.

GP Nason. Stationary and non-stationary times series. Statistics in Volcanology. Special Publications of IAVCEI, 1:000–000, 2006.

F. Nogales J. Contreras, R. Espinola and A. Conejo. Arima models to predict next day electricity prices. IEEE Transactions on Power Systems, vol. 18, 2003.

Rob J Hyndman and George Athanasopoulos. Forecasting: principles and practice. 2014.

Chris Chatfield. Time-series forecasting. CRC Press, 2002.

Jonathan D Cryer and Natalie Kellet. Time series analysis.2006.

RONALD MAHIEUA. RONALD HUISMANA, CHRISTIAN HUURMANA. Hourly electricity prices in day ahead markets. Energy Economics, 2007.

Gidon Eshel. The yule walker equations for the ar coefficients. Internet resource, 2010.

Joseph Plasmans. Modern linear and nonlinear econometrics, volume 9. Springer Science and Business Media, 2006.

Hans Von Storch and Francis W Zwiers. Statistical analysis in climate research. Cambridge university press, 2001.

A. GHASEMI. H. SHAYEGHI. Day-ahead electricity prices forecasting by a modified cgsa technique and hybrid wt in lssvm based scheme. Energy Conver- sion and Management, 2013.

Breiman. L. [1998b] Randomizing Outputs To Increase Prediction Accuracy. Technical Report 518, May 1, 1998, Statistics Department, UCB (in press Machine Learning).

Petrus MT Broersen. Automatic autocorrelation and spectral analysis. Springer Science and Business Media, 2006.

Keith W Hipel and A Ian McLeod. Time series modelling of water resources and environmental systems. Elsevier, 1994.

Chris Chatfield. Time-series forecasting. CRC Press, 2000.

Lawrence Chin and Gang-Zhi Fan. Autoregressive analysis of singapore's private residential prices. Property Management, 23(4):257–270, 2005.

Vijay P Singh. Mathematical models of small watershed hydrology and applications. Water Resources Publication, 2002.

Wayne A Woodward, Henry L Gray, and Alan C Elliott. Applied time series analysis. CRC Press, 2011.

E. Chatfield. The analysis of time series: An introduction. Chapman and Hall Ltd, 1991.

R. Espinola A. Conejo, M. Plazas and A. Molina. Day ahead electricity price forecasting using the wavelet transform and arima models. IEEE Transactions on Power Systems, 2005.

G. Janacek and L. Swift. Time Series Forecasting, Simulation. 1995.

David F Hendry. Dynamic econometrics. Oxford University Press, 1995.

JIANZHOU WANG. JINXING CHE. Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving av- erage modeling. Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling, 2010.

Cambridge Systematics. A guidebook for forecasting freight transportation demand. Number 388. Transportation Research Board, 1994.

Douglas M Patterson and Richard A Ashley. A Nonlinear Time Series Workshop: A Toolkit for Detecting and Indentifying Nonlinear Serial Dependence, volume 2. Springer Science and Business Media, 2000.

Khandakar Y. Hyndman, R.J. Automatic time series forecasting: The forecast package for r. Journal of Statistical Software, 2008.

Chris Chatfield. Time-series forecasting. CRC Press, 2000.

Francis Okyere and Salifu Nanga. Forecasting weekly auction of the 91-day treasury bill rates of ghana. 2014.

Keith Ord and Robert Fildes. Principles of business forecasting. Cengage Learning, 2012.

Keith Ord and Robert Fildes. Principles of business forecasting. Cengage Learning, 2012.

JIANHUI WANG JUN XU. ZHONGFU TAN, JINLIANG ZHANG. Day-ahead electricity price forecasting using wavelet transform combined with arima and garch models. Original Research Article. Applied Energy, 2010.

Allan I McLeod and William K Li. Diagnostic checking arma time series models using squared-residual autocorrelations. Journal of Time Series Analysis, 4(4):269–273, 1983.

M.K.SHEIKH-EL-ESLAMI TARBIAT MODARES. M.SHAFIE-KHAH, M.PARSA MOGHADDAM. Price forecasting of day-ahead electricity markets using a hybrid forecast method. Energy Conversion and Management, 2015.

Constantino Hevia. Maximum likelihood estimation of an arma (p, q) model. The World Bank, DECRG, 2008.

Yongcheol Shin and Peter Schmidt. The kpss stationarity test as a unit root test. Economics Letters, 38(4):387–392, 1992.

Franke Jürgen, Wolfgang ardle, and Christian Hafner. Statistics of Financial Markets: An Introduction. Springer, 2011.

Nor Aishah Ahad, Teh Sin Yin, Abdul Rahman Othman, and Che Rohani Yaacob. Sensitivity of normality tests to non-normal data. Sains Malaysiana, 40(6):637–641, 2011.

Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. Biometrika, 65(2):297–303, 1978.

Miguel A Arranz. Portmanteau test statistics in time series. Time Orientated Lan-

guage, pages 1–8, 2005.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the Econometric Society, 1982.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3):307–327, 1986.

Faridul Islam, Muhammad Shahbaz, Ashraf U Ahmed, and Md Mahmudul. Alam. Financial development and energy consumption nexus in malaysia: a multivariate time series analysis. Economic Modelling, Elsevier, 2013.

Patton A. J. et al. Engle, R. F. What good is a volatility model? Quantitative Finance, 2001.

B. (2011). Dralle. Modelling volatility in financial time series. Master's thesis, Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, 2011.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the Econometric Society, 1982.

Syed Aun Hassan and Farooq. Malik. Multivariate GARCH modelling of sector volatility transmission. 2007.

R.F. Engle. Autoregressive conditional heteroskedasticity with estimates of variance of United Kingdom inflation' Econometrica. 1992.

TA Roberto Perrelli. Introduction to arch and garch models. 2001.

Chris Chatfield. Time-series forecasting. CRC Press, 2002.

Ismael Suleman Talke. Modelling volatility in Time series data. Master's thesis, University of KwaZulu Natal, South Africa, 2003.

Eberly College Website. Applied time series analysis. https://onlinecourses. science.psu.edu/stat510/node/78, Accessed April 2015.

Ruey S Tsay. Analysis of financial time series, volume 543. John Wiley and Sons, 2005.

ColmKearneyandAndrewJ.Patton. Multivariategarchmodellingofexchange rate volatility transmission in the european monetary system. Financial Review, Wiley Online Library, 2000.

Ismael Suleman Talke. Modelling volatility in Time series data. Master's thesis, University of KwaZulu Natal, South Africa, 2003.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3):307–327, 1986.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3):307–327, 1986.

B. (2011). Dralle. Modelling volatility in financial time series. Master's thesis, Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, 2011.

T. Bollerslev. 'generalized autoregressive conditional heteroskedasticity',. Journal of Econometrics, 1996.

Mr Rugiranka Tony Gaston. Modelling of volatility in the south african mining sector: Application of arch and garch models. School of Mathematics, Statistics and Computer Science Pietermaritzburg Campus, 2016.

Chris Chatfield. Time-series forecasting. CRC Press, 2000.

T. Bollerslev. 'generalized autoregressive conditional heteroskedasticity',. Journal of Econometrics, 1996.

R. Napoli C. Gao, E. Bompard and H. Cheng. Price forecast in the competitive electricity market by support vector machine. IEEE Power Engineering Society General Meeting, vol. 1, 2007.

Robert F Engle and Kevin Sheppard. Theoretical and empirical properties of dynamic conditional correlation multivariate garch. National Bureau of Economic Re-

search, 2001.

Luc Bauwens, Sébastien Laurent, and Jeroen VK. Rombouts. Multivariate garch models: a survey. Journal of applied econometrics, Wiley Online Library, 2006.

Neil Shephard. Statistical aspects of arch and stochastic volatility. Monographs on Statistics and Applied Probability, 65:1–68, 1996.

Eberly College Website. Applied time series analysis. https://onlinecourses. science.psu.edu/stat510/node/78, Accessed April 2015.

Chris Chatfield. Time-series forecasting. CRC Press, 2000.

Reinaldo C Garcia, Javier Contreras, Marco Van Akkeren, and Jo˜ ao Batista C Garcia. A garch forecasting model to predict day-ahead electricity prices. Power Systems, IEEE Transactions on, 20(2):867–874, 2005.

Allan I McLeod and William K Li. Diagnostic checking arma time series models using squared-residual autocorrelations. Journal of Time Series Analysis, 4(4): 269–273, 1983.

L. Breiman. Random forests. Machine learning, 2001.

L. Breiman. Random forests. Machine learning, 2001.

Olshen R.A. Stone C.J. Breiman L., Friedman J.H. Classification and Regression Trees. 1984.

Olshen R.A. Stone C.J. Breiman L., Friedman J.H. Classification and Regression Trees. 1984.

Friedman J. Hastie T., Tibshirani R. The elements of statistical learning. data mining. Inference, and Prediction. Springer, 2009.

Friedman J. Hastie T., Tibshirani R. The elements of statistical learning. data mining. Inference, and Prediction. Springer, 2009.

Breiman, L. (1996). Bagging predictors. Machine Learning, 24, 123–140.

Aggarwal, S. K., Saini, L. M., and Kumar, A. (2009). Electricity price forecasting in deregulated markets: A review and evaluation. International Journal of Electrical Power and Energy Systems, 31.

Alain Angeralides and Ottmar Cronie. Modelling electricity prices with orn- stein uhlenbeck processes. Masters Thesis, Dept. of Mathematical Statistics, Chalmers University of Technology, 2006.

Bunn, D. W. (Ed.) (2004). Modelling prices in competitive electricity markets. Chichester: Wiley.

Ziegler A. ranger Wright, M. N. A fast implementation of random forests for high dimensional data in C++ and R. 2017.

Ziegler A. ranger Wright, M. N. A fast implementation of random forests for high dimensional data in C++ and R. 2017.

Ziegler A. ranger Wright, M. N. A fast implementation of random forests for high dimensional data in C++ and R. 2017.

Ziegler A. ranger Wright, M. N. A fast implementation of random forests for high dimensional data in C++ and R. 2017.

Ziegler A. ranger Wright, M. N. A fast implementation of random forests for high dimensional data in C++ and R. 2017.

L. Breiman. Using adaptive bagging to debias regressions. 1999.

Breiman. L. Randomizing outputs to increase prediction accuracy. Technical Report 518, May 1, 1998, Statistics Department, UCB, 1998.

L. Breiman. Out-of-bag estimation. 1998.

Weron R. Modeling and Forecasting Electricity Loads and Prices. 2006.

T. Dietterich. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting and Randomization, Machine Learning. 1998.

R. Tibshirani. Bias, variance, and prediction error for classification rules. Technical Report, Statistics Department, University of Toronto, 1999.

L. Breiman. Out-of-bag estimation. 1984.

Girish, G. P., Vijayalakshmi, S., Ajaya, K., and Badri, R. 2013. Forecasting Electricity Prices in Deregulated Wholesale Spot Electricity Market: A Review. International Journal Of Energy Economics And Policy 4(1): 32-42.

Ghosh, S., and Kanjilal, K. (2014). Modeling and Forecasting day ahead electricity price in Indian Energy Exchange: Evidence from MSARIMA-EGARCH model. Int. J. Indian Culture and Business Management, 8 (3), 413-423.

C Sigauke and D Chikobvu. Prediction of daily peak electricity demand in south africa using volatility forecasting models. Economics, 33(5):882–888, 2011.

Rob J Hyndman and George Athanasopoulos. Forecasting: principles and practice. 2014.

Chris Chatfield. Time-series forecasting. CRC Press, 1996.

Nyblom, J. and A. C. Harvey (2000). Tests of common stochastic trends. Econometric Theory 16, 176-99.

Enders, W., (2004), Applied Econometric Time Series, 2nd Edition, John Wiley and Sons, Hoboken, NJ.

M. G. Karlaftis and E. Vlahogianni, Statistical methods versus neural networks in transportation research: differences, similarities and some insights, Transportation Research Part C: Emerging Technologies,2011.

M. G. Karlaftis and E. Vlahogianni, Statistical methods versus neural networks in transportation research: differences, similarities and some insights, Transportation Research Part C: Emerging Technologies,2011.

M. G. Karlaftis and E. Vlahogianni, Statistical methods versus neural networks in transportation research: differences, similarities and some insights, Transportation Research Part C: Emerging Technologies,2011.

Beal, Dennis J. "Information Criteria Methods in SAS for Multiple Linear Regression Models."SA05, presented at SouthEast SAS Users Group (SESUG) Conference, Hilton Head, SC, November 4–6, 2007. http://analytics.ncsu.edu/sesug/2007/SA05.pdf.

Brocklebank, John C., and David A. Dickey. SAS for Forecasting Time Series. 2 nd edition. Cary, NC: SAS Institute Inc. and Wiley, 2003.

Brocklebank, John C., and David A. Dickey. SAS for Forecasting Time Series. 2 nd edition. Cary, NC: SAS Institute Inc. and Wiley, 2003.

Brocklebank, John C., and David A. Dickey. SAS for Forecasting Time Series. 2 nd edition. Cary, NC: SAS Institute Inc. and Wiley, 2003.

Brocklebank, John C., and David A. Dickey. SAS for Forecasting Time Series. 2 nd edition. Cary, NC: SAS Institute Inc. and Wiley, 2003.

Brocklebank, John C., and David A. Dickey. SAS for Forecasting Time Series. 2 nd edition. Cary, NC: SAS Institute Inc. and Wiley, 2003.

Brocklebank, John C., and David A. Dickey. SAS for Forecasting Time Series. 2 nd edition. Cary, NC: SAS Institute Inc. and Wiley, 2003.

M. G. Karlaftis and E. Vlahogianni, Statistical methods versus neural networks in transportation research: differences, similarities and some insights, Transportation Research Part C: Emerging Technologies,2011.

M. G. Karlaftis and E. Vlahogianni, Statistical methods versus neural networks in transportation research: differences, similarities and some insights, Transportation Research Part C: Emerging Technologies,2011.

Y. Yang, J. Wu, Y. Chen and C. Li,A New Strategy for Short-Term Load Forecasting, Hindawi Publishing Corporation, Abstract and Applied Analysis,2013.

P Karamouz, J. Wu, Y. Chen. Introduction to Time Series Analysis and Forecasting. Hoboken, NJ: John Wiley and Sons, Inc., 2012.

Nason. SAS for Forecasting Time Series. 2 nd edition. Cary, NC: SAS Institute Inc. and Wiley, 2006.