UNIVERSITY OF KWAZULU NATAL

Forecasting electricity demand using univariate time series volatility forecasting models: A case study of Uganda and South Africa

by

Winnie Nakiyingi

A dissertation submitted in fulfilment for the degree of Master of Science in Statistics

in the

College of Agriculture, Engineering and Science
Department of Statistics

March 2016

Declaration of Authorship

The work described in this dissertation was carried out under the supervision of Mr. Oliver Bodhlyera and Prof. Henry Mwambi in the School of Mathematics, Statistics, & Computer Science, University of KwaZulu-Natal (Pietermaritzburg), from May 2014 to November 2015.

No portion of the work referred to in this dissertation has been submitted in support of an application for another degree or qualification of this or any other university or institution of learning. The dissertation is my original work except where due reference and credit is given.

Sign:  26/November/2015.
Winnie Nakiyingi Date

Sign: ....
Mr. Oliver Bodhlyera Date

Sign: ....
Prof. Henry Mwambi Date

"One of the statistics that always amazes me is the approval of the Chinese government, not elected, is over 80 percent. The approval of the U.S. government, fully elected, is 19 percent. Well, we elected these people and they didn't elect those people. Isn't it supposed to be different? Aren't we supposed to like the people that we elected?"

Bill Gates

UNIVERSITY OF KWAZULU NATAL

Abstract

College of Agriculture, Engineering and Science
Department of Statistics

Master of Science in Statistics

by Winnie Nakiyingi

Different sectors of economies are significantly affected by the supply of electricity. However, with the available limited resources, supply and demand of electricity in Africa are strongly correlated. In order to efficiently improve electricity supply, its demand has to be accurately predicted. In this research, we analyse electricity demand in two cases; peak monthly electricity demand in Uganda from January 2008 to December 2013, and daily electricity demand for South Africa from 1^{st} January 2004 to 30^{th} June 2008, using ARIMA and ARCH/GARCH models. We use this data to forecast future demand for both countries in order to help policy makers in the electricity sector make decisions for sustainable development of both countries. GARCH models are introduced to correct the volatility found in South Africa's daily demand data. Results from the study show that; for Uganda, a seasonal ARIMA(0,0,0)(1,1,1)[12] model describes the data better, with RMSE of 4.872027 and MAPE of 2.347028, and gives better forecasts which display a continued increase in electricity demand for months ahead. For South African data, a seasonal ARIMA(1,0,1)(0,1,0)[365] describes the data better but a standard GARCH(1,1) with normally distributed error terms accommodates volatility. Therefore, a combination of the two models produces better forecast accuracy.

Keywords: Seasonal ARIMA (SARIMA), ARCH/GARCH, peak electricity demand, Forecast accuracy.

Acknowledgements

I am extremely grateful to God for enabling me complete this research. I greatly thank my supervisors, Mr. Oliver Bodhlyera and Prof. Henry Mwambi, for their continued support and guidance while pursuing my Masters in Statistics. Their inspiring teaching and invaluable advice have been absolutely vital for the development of this research. I also thank Ms. Christel Barnard and Ms. Justine Nasejje for their unending support and contribution during my research period.

My special thanks also go to my best friend Mr. Joseph Malinzi for his genuine commitment, constructive suggestions and his enthusiastic support that enabled me to complete my research. He was originally there for me, encouraged me when I first started working on this project.

I also thank AIMS-NEI for the financial support which has made my research journey possible and easier than it would have been otherwise. I am truly grateful.

Last but not least, I would like to thank my family who have taught me the important things in life and have supported all my endeavours. My brothers and sister have given me great affection and support while patiently enduring my long absence from home.

Contents

List of Figures

List of Tables

Dedicated to Mr. Joseph Malinzi and the entire Kiwuuwa family.

Chapter 1

Introduction

Electricity is a form of energy that plays an essential role in modern life, bringing benefits and progress in various sectors; for example, transportation, manufacturing, mining and communication. Electricity is the backbone for an economy's prosperity and progress because it plays an important role in socio-economic development. It has the capability to make useful contributions to planning and future policy formulation of the energy sector. Uses of electricity are rapidly increasing day by day, leading to a tremendous advancement in human civilization. For this reason, demand for electricity is a vital topic to study since it is integrated with all aspects of development. This chapter covers the initial and introductory parts of this study. It gives the basis for definitions used in the study. Most of the topics and terminologies introduced here are used for further reference in the course of the study. This research is inspired by Sigauke and Chikobvu (2011), who did similar work with South Africa's daily electricity demand.

1.1 Forecasting

Forecasting is a statistical tool that helps to make predictions about the future, using past and present data (Bajpai, 2009; Mittra, 2002). It can also be defined as the process of making deductions about events whose actual outcomes have not yet been observed. The data used to carry out the forecasting exercise can be generated through different ways. The most common types of data are secondary and primary data. Secondary data is collected due to a focus on some study (other than the current study), but also happens to be useful to the current study. Such

data is usually saved in archives of official statistical and academic organizations (Hox and Boeije, 2005). On the other hand, primary data is generated specifically for an ongoing study. This data can be collected through; interviews, questionnaires, mailing and direct observations. Nowadays, data can also be captured from websites that have topics related to the study at hand.

The process of forecasting is mainly used to plan, make budgets and estimate future growth. Various departments in companies need forecasting for different reasons, for example, the accounting department uses forecasting to find out the cost and profit estimates during a period of interest, the finance department can find out how much cash flows in and out and how much funding is needed for a given investment, the human resource department forecasts an estimated number of people to hire, recruit or train, given the future of the company, and the marketing department can use forecasting to figure out how to price given products or services, where and when to give a promotion and the best strategy to use in order to maximize their profits or sales.

According to previous studies such as Hyndman and Athanasopoulos (2014), a forecast variable has never been 100% predictable. Therefore, when forecasting, one needs to keep in mind that there is no certainty about the occurrence of the event under study. Various studies, like, Zarnowitz (1984) indicate that group forecasts are better than individual forecasts, for example, one can forecast the average performance of the whole class using results from their midterm exams better than forecasting one student's performance using their previous grade. The accuracy of forecasts reduces with increase in the time horizon (Woodside and Martin, 2008). For that reason, some forecasting techniques, like, exponential smoothing attach a lower weight to observations in the further past than those in the recent past. There are two common approaches used when forecasting, depending on the problem at hand, but a combination of the two can also be applied to support their strength and reduce their individual weaknesses. These approaches are:

- *Judgemental Forecasting:* This is forecasting based on one's gut feeling about the event. It requires only one's intuition and experience. Human minds have the ability to make connections and understand situations in a way that no computerized system can (VSC website). However, there are limitations (like bias) that make analysis of large data complex. The most common

type of judgemental forecasting is the Delphi method. This method combines results (in form of questionnaires) from independent experts. Questionnaires are constructed about the topic under study and are sent out to experts in several rounds. These experts then anonymously comment on the topic and their responses are aggregated and discussed after every round. The process continues until a mutual agreement about the topic is reached. That agreement is considered as the correct response through consensus (Yousuf, 2007).

Judgemental forecasting works best under various cases, for example, when there is no historical data, when launching new products, when there is a new competitor in the market, when there are new growth plans, or during new and unique market conditions. The limitations of this method include; inconsistency (because it heavily depends on human intuition), unfair agendas (either personal or political) and anchoring (subsequent forecasts converging to an initial familiar reference point) (Hyndman and Athanasopoulos, 2014).

- *Quantitative Forecasting:* Here, numerical facts and prior experiences are used to analyze historical data and predict future events. This approach emphasizes developing numerical information about events. Quantitative forecasting techniques use mathematics for the systematic treatment of actual historical series of data to later identify and estimate functional relationships that can be used to make forecasts of such a series. There are deterministic techniques, for example, moving averages, exponential smoothing, trend analysis, and stochastic techniques, like, the ARIMA models (Chavez et al., 1999). Both deterministic and stochastic models can be used for any time series. The choice depends on the data available, its nature, and the degree of accuracy required for the analysis. For the strictly statistical criterion, methods producing lower forecast errors are of interest. Once one has numerical data, quantitative techniques can be used to carry out the forecasting exercise. There are two ways in which this can be done;

  - The time series method where by the researcher develops a model from identified repeated patterns in historical data, and uses that model to make forecasts. Usually in time series, the models developed use time as an independent variable. In other words, time is used as the main factor to help researchers understand all events that can not be measured but take place in time (Adhikari and Agrawal, 2013). The patterns

common in time series data are trend, seasonality, cycles, irregular and random variations. All these will be discussed further in the study. The disadvantages of using this method include; requirement of large data sets which might not be available in some cases, lack of variations in the data, autocorrelation and multicollinearity. However, it is also advantageous in that it provides cheap, fast and realistic forecasts. Regression analysis that does not consider lags fails to account for the relationship through time and also over estimates the relationship between the dependent and independent variables (Berger, 2003).

– The explanatory method which relates two or more variables. It assumes that there is a relationship between the variable of interest and other variables in the environment. For example, the number of mountain bikes that can be sold in an area depends almost entirely on the number of young people living in that same area. In this case, the mountains bikes sale is the predicted (dependent) variable and the number of young people is the predictor (independent) variable. Many statistical tools are available for determining such a relationship, but, liner regression analysis is the most preferred. In this method, the main objective is to obtain an equation of a straight line minimizing the sum of squared vertical deviations of data points from the line (Weisstein, 2002).

Some of the advantages of quantitative forecasting are (JRC Website); the ability to clearly examine the rates of change of events, which makes it more realistic to interpret data other than making theoretical conclusions. Also, results from this approach easily and efficiently communicate to people in case there is a lot of information being passed on. It allows employment of useful supplementary methods like accounting tools which require numerical data. Most importantly, it is less affected by bias. There are also some disadvantages, for example; limiting most people's contribution especially when it comes to using complex statistical techniques. Most of the experts employed when dealing with this approach are usually used to the same modeling technique, which hinders them from working with other experts. Various important social and political factors are usually neglected because they can not be expressed numerically yet they are very significant to studies. It also limits communication of results to people who can not easily interpret numerical expressions. In case of limited historical data or insufficiently

updated data, this approach may either be impossible to use, or result into incorrect inference.

The main steps to be followed while forecasting as described by Hyndman and Athanasopoulos (2014) involve:

- Defining the problem of the study. This is one of the hardest steps while carrying out a forecasting exercise. One needs to carefully define the problem according to the way they understand the requirements of the forecasts, that is, who needs the forecasts, how the forecasts will be used, how the determined model fits in the data at hand. This step can be simplified by having enough communication and discussion time with everyone involved in the study, like people responsible for collecting data, maintaining databases and using the forecasts for future planning. At this stage, it is important to note that any decision made basing on the results from the forecasts will affect the future of the organization.

- Time horizon: One needs to know how much time the forecasts should cover. Short-term forecasts usually cover a time period less than 1 year and they are mainly for scheduling and assigning purposes. Short-term forecasts in the electricity sector are useful in estimating load flows and making decisions on how to prevent load shedding. Medium-term forecasts range between 1 to 3 years and they are usually for determining and planning for future resource requirements. Lastly, long-term forecasts go for any period bigger than 3 years and are for strategic planning and development. Different time horizons develop different forecast accuracy. For example, it is very common to have more accurate load forecasts for the next day than forecasts for one year ahead. Accuracy of the forecasts keeps fading with increase in the time horizon.

- Data collection: Keeping in mind the event to be forecast helps in collecting valid and reliable historical data. As covered earlier in this chapter, data can either be primary or secondary. In most cases, it is hard to collect enough historical data for fitting the best model for the study. However, this is not always a problem if not much historical data is needed, especially in situation where very old data has become irrelevant due to changes in the event being studied. Good data should be reliable, accurate, relevant, consistent and timely.

- After collecting all the data necessary for the study, one needs to check the nature of that data. A time series plot is the best option for checking if there are any patterns, like trend, seasonality, cycles and their significance. Plotting also helps point out outliers if any exist and their meaning. Also the significance of the relationship between or among the variables studied can be identified through plotting the data. Data needs to be cleaned first, before it is used for any analysis in order to have clear quality and completeness.

- Choosing the forecasting approach to use. In case very little data exists, the product at hand is new in the market, or there is a new competitor, then the qualitative methods would be better. Otherwise, one needs to choose and fit models relevant to the data according to its nature. Choosing what models fit the data best depends on the availability and relevance of historical data. It is advisable to try and check different potential models before concluding about the best (Faraway and Chatfield, 1998).

- Lastly, the chosen model's parameters are estimated and the model is used to make forecasts. The accuracy of a model is determined either by waiting for actual data for the forecast period to become available then make comparisons, or by dividing the available data into two sets, one for training the model and estimating parameters and the other for evaluating the performance of the model. However, for crucial studies like electricity demand, policy makers might not have enough time to wait for the availability of actual data before taking decisions for the nation's development. Therefore, it is advisable to divide the available data in order to ease the decision making process.

Electricity demand is affected by different factors in different countries. For example, highly industrialised countries have a higher demand for electricity than low industrialised countries, countries with stable seasonal weather changes, like, winter and summer usually have an almost similar demand for electricity year after year, in the respective seasons. Other factors affecting electricity demand include; the supply of electricity, social factors and human activities. Since electricity demand keeps changing continuously in time, we consider it to be a time series set of data. Therefore, quantitative methods are preferred in making forecasts about it, specifically, time series techniques, depending on the nature of the available data.

This study will focus on forecasting electricity demand in two different scenarios; monthly peak demand for Uganda and daily demand for South Africa, using univariate time series volatility forecasting models. Univariate time series will be used because of various reasons;

- The available data is univariate because of the absence of sufficiently accurate data about other independent variables, such as weather changes. For that reason, we shall follow studies such as Saab et al. (2001) and Price and Sharp (1986) who found it more appropriate to use univariate forecasting techniques to reach the objectives of their studies.
- Researchers such as; Huss (1985) and Meese and Geweke (1984), show that univariate time series can give forecasts that are more accurate for a medium term period than a variety of other more complex forecasting data sets.
- In addition, since the main objective of the study is to find the best forecasting models for both short term (for South Africa) and medium term (for Uganda) electricity demand, we start with simpler (univariate) time series models, which can only be used with univariate time series data.

1.2 General discussion

Forecasting the demand of any commodity or service is very important for proper planning. Considering the electricity sector in this study, prediction of peak load demand is vital for decision making in this sector. For any economy, it is important to make accurate predictions in the electricity sector because most of the economic growth is affected by electricity. Accurate predictions can be made by taking some factors into consideration, for example, finding the maximum operation capacity of power plants, say in a month (Sigauke and Chikobvu, 2011). Predicting load demand is also a vital step to strategic planning for capacity expansion in a way that, it helps the government decide whether to build another power generation plant as a way of dealing with an increasing demand and catering for an unanticipated demand. It also helps identify strategies to reduce losses made in the electricity sector. Accurate prediction of peak load demand also helps determine consistent and reliable supply schedules during peak periods. It also enables effective load shifting between transmission substations, scheduling of start up times

of peak stations, load flow analysis and power system security studies (Sigauke and Chikobvu, 2010).

Due to changing economic situations in both Uganda and South Africa, and the unstable weather conditions, the demand of electricity keeps changing in short to medium intervals. For that reason, precise forecasting of electricity demand should be carried out for short periods. This helps in proper allocation of the available limited resources and planning for sustainable development. In this study, we shall consider both short and medium term forecasting for both data sets, which cover several days and months upto a year respectively. This is because our interest is in planning, maintenance and scheduling power supply in the future.

The available Ugandan data is in months and is analysed, modelled and used to forecast monthly peak demand up to a year ahead. Daily South African data is divided into two sets; where the training set is modelled and used to forecast daily demand up to a period compared with the test set. Since the electricity sector in these developing countries serves different groups of people for different purposes, for example, home use, commercial use and industrial use, the pattern of demand keeps changing for the different groups but stays almost constant for different people in the same group.

Currently around 80% of Uganda's power is generated by hydro power from 12 power stations with 3 main stations (Bujagali, Kiira and Nalubale with capacities of 250MW, 200MW and 180MW respectively) and 10 other stations that produce a total of 65.9MW (Baanabe, 2012). From all the available power stations, Uganda has a capacity of generating total electricity of about 695.9 MW. However, some of this electricity is lost through transmission and distribution. In 2009, the peak demand (evenings) was estimated to be about 380 MW and daytime demand about 260 MW (Saundry, 2009). If this was a constant demand, Uganda would not be suffering load shedding currently with all the increase in power supply through the new power plants. However, since it is a developing country, many economic changes take place and as a result, the demand of electricity keeps going up at an estimated average rate of 10% per year (New Vision Website). The growing demand for electricity and the lack of public and private investments in power infrastructure projects are major reasons affecting the electricity sector in Uganda. But also, droughts and increased discharges lower the water level and lead to significant power losses.

Most of the information about the current electricity situation in South Africa is from SAPower (2015). ESKOM is the South African electricity public utility, established in 1923 as the Electricity Supply Commission (ESCOM) by the government of South Africa in terms of the Electricity Act (1922). It was founded by a parliamentary act, namely the Electricity Act of 1922, which allowed the Electricity Control Board to appoint Hendrik Johannes van der Bijl as the Chairman of the Board (Conradie and Messerschmidt, 2000). The company was also known by its Afrikaans name Elektrisiteitsvoorsieningskommissie (EVKOM). The two acronyms were combined in 1986 and the company is now known as ESKOM, which represents South Africa in the Southern African Power Pool. ESKOM is; the largest producer of electricity in Africa, among the top seven utilities in the world in terms of generation capacity, and, among the top nine in terms of sales (Vedavalli, 2007). ESKOM operates a number of notable power stations, including Kendal Power Station, and Koeberg nuclear power station in the Cape Province, the only nuclear power plant in Africa. The company is divided into Generation, Transmission and Distribution divisions and together ESKOM generates approximately 95% of electricity used in South Africa.

Due to the South African government's attempted privatisation of ESKOM in the late 1990s, ESKOM's requests for budget to build new stations were denied. Former president Thabo Mbeki said in December 2007 that this was an error, and it is now adversely affecting the South African economy (Van Wyk, 2012). In January 2008 ESKOM introduced "load shedding", planned rolling blackouts based on a rotating schedule, in periods where short supply threatens the integrity of the grid. South Africa produces around 240,300 GW/hr (865,000 TJ) electricity annually (SAPower, 2015). Most of this electricity is consumed domestically, but around 12,000 GW/hr is annually exported to Swaziland, Botswana, Mozambique, Lesotho, Namibia, Zambia, Zimbabwe and other Southern African Development Community countries participating in the Southern African Power Pool. South Africa supplements its electricity supply by importing around 9,000 GW/hr per year from the Cahora Bassa hydroelectric generation station in Mozambique via the 1,920 MW Cahora Bassa high-voltage direct current transmission system (Africa., 2007). Most power stations in South Africa are owned and operated by ESKOM and these plants account for 95% of all the electricity produced in South Africa and 45% of all electricity produced on the African continent. In terms of share of GDP in 2012, South Africa was the 4th largest investor in renewable power in the world after Uruguay, Mauritius and Costa Rica (Martinot et al., 2005).

Often when forecasting, a key step is knowing whether it is possible to forecast a given event accurately or not. Good forecasts capture the genuine patterns and relationships which exist in the historical data, but do not replicate past events that will not occur again (Hyndman and Athanasopoulos, 2014). With the Ugandan data considered in this study, we use demand data from 2008 to 2013. This is because previous studies, like, Mawejje et al. (2013) show that from 1998, the average water level in Lake Victoria-which is the main reservoir for the Jinja Complex-dropped significantly and reached its lowest level in 2005/06, since 1951. This was because of the extended drought that occurred in Uganda, mainly the Eastern part, from 2003 to 2006. The water level conditions improved towards the end of 2006 which increased the supply and hence demand of electricity again. This is not a genuine pattern and we are not certain of it happening again. Therefore, we choose to ignore that data because it appears as a random fluctuation that does not need to be modelled and extrapolated.

The available data from UMEME of Uganda and ESKOM of South Africa shows monthly peak demand (maximum daily demand in 30-day period) of electricity in Uganda, measured in Gigawatts (GW), and daily demand (total hourly demand in 24-hour period) of electricity in South Africa, measured in GW/hr respectively. The data from Uganda is of limited size, covering a period of 6 years with corresponding months in each year, hence covering a total of 72 months (observations). However, South African data is rich in quantity, mainly because the levels of data management differ between these two countries. South African data runs from 1/01/2004 to 28/06/2008, covering a total of 1642 observations. This demand only includes households and companies willing and able to pay for electricity and actually have access to it. Data about households and companies willing and able to pay for electricity but do not have access to it is not included in this study. We are more worried about the peak demand than the off-peak demand because when dealing with electricity (its transmission and distribution), the capacity of any electricity generating plant has to be higher than the peak demand of all the customers it serves. If it can serve the high demand, it can serve a lower demand.

In this study, we forecast demand of electricity using time series volatility forecasting techniques. These techniques involve using historical data to construct a model that describes the nature of the data and can be used to make forecasts for the future. The study of medium-term demand forecasting of electricity began in the 1980's and since then, various techniques have been developed in the forecasting

discipline (Rallapalli and Ghosh, 2012). These include; exponential smoothing, ARIMA, and the recently developed ARCH and GARCH models. From these, we shall choose the best model that UMEME and ESKOM can use to predict electricity demand for a specified periods using the data available.

This study is organized in such a way that; Chapter 2 covers the literature studied in relation to the work done in this study. Some of the work done by other researchers and the current electricity situation in one of the countries - Uganda. This gives deeper insight about the significance of the study to developing countries. Chapter 3 looks at theoretical aspects of time series analysis in detail and all the relevant information needed to start a forecasting process. It also covers the preliminary analysis of the data at hand with the necessary adjustments. Chapter 4 studies all the possible linear models given the nature of both data sets. It specifically tackles ARIMA models that do not focus on the change in the variance of the data, and their properties. Chapter 5 covers modelling the available data using ARIMA models and using the developed models to forecast future values. Then Chapter 6 focuses on a change in variance of the data in detail. These are called volatility models, mainly the ARCH and GARCH models. Chapter 7 covers the application of volatility models to the residuals of the chosen ARIMA models. A conclusion about the whole study is covered in Chapter 8, with the appendices in the subsequent chapters. In this work, R-Studio software is used for all the necessary programming required, and the codes are displayed in the appendices.

Chapter 2

Literature review

Since the early 19^{th} century, electricity uses worldwide have expanded from lighting only to other uses like; cooking, washing, air conditioning, refrigeration, use of television and computers. This has gradually made electricity demand forecasting more complicated. Due to these changes, economic and natural factors have been employed to help make forecasting of electricity demand possible (Hong and Dickey, 2014). As a result of the oil crisis in the 1970's, policy makers desired to know the demand of energy even more. For that reason, many methodologies were introduced to help policy makers in the energy sectors make constructive decisions. From those methodologies a variety of models were developed which are now used to analyze and forecast energy demand (Bhattacharyya and Timilsina, 2009), electricity inclusive.

After realizing the continuous increase in electricity demand, developed countries opted for deregulation which encourages using other power sources like solar and wind. 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 can not be predicted easily since it mainly depends on the weather.

The process of forecasting electricity demand 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 demand seems to be a world-wide problem, irrespective of the level of development. In this chapter, we look at various studies that have been carried out in order to forecast electricity demand with deviation from the actual demand as minimum as possible.

2.1 Empirical studies

Irrespective of the limitations in electricity forecasting, a couple of recent empirical studies show that forecasting electricity demand is actually possible using various methods. These studies include the following:

- Similar to the current study, Yasmeen and Sharif (2014) studied forecasting monthly electricity consumption (EC) for Pakistan. They analyzed the monthly EC of Pakistan using linear and non linear modelling techniques. Their emphasis was on ARIMA, Seasonal ARIMA (SARIMA) and ARCH/-GARCH models. Due to lack of appropriate EC data in a developing country like Pakistan, they approximated electricity production data for consumption in their study. A series of monthly EC data measured in GigaWatt hours (GWh) from January 1990 to December 2011 was used. The data from all economic sectors (industrial, residential, and commercial) of Pakistan was recorded by Department of Federal Bureau of Statistics Pakistan. Time series plots exhibited a significant trend in the data, so they took log transformations to stabilise the variance and mean. Seasonality was evident since plots showed higher EC in the months of May, June, July and August, due to the high temperature in those months.

  The data was divided into a training set January 1990-December 2006 (for training the models) and the test set January 2007-December 2011 (for comparing different models). The monthly behavior of forecast values depicted that EC was higher for summer season and was expected to increase in the future. The forecast model and forecast values revealed that EC was increasing with time. They evaluated the models by diagnostic tests and compared the forecast values to select the most appropriate model. The least out of sample forecast performance; Mean Absolute Percentage Error (MAPE) values and the minimum forecast standard deviation values showed that among

the four competing time series models, ARIMA(3,1,2) model was the most appropriate model to forecast EC in Pakistan.

- Sigauke and Chikobvu (2011) studied the prediction of daily peak electricity demand in South Africa using three volatility forecasting models; a seasonal auto regressive 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. They emphasized accurate load prediction for proper decision making in the South African electricity sector. Their study took into consideration the fact that for electricity as a good, its demand can not exceed its supply. This is because, electricity being a non storable good, there are no market forces influencing its prices. When dealing with the electricity market, the normal "buy-and-hold" theory for determining prices (as used for other goods) becomes irrelevant. Therefore, the number of power plants determine the supply which in turn determines the demand. If there is limited supply, say, few power plants available to produce electricity for current consumption and the unforeseen demand, the demand will automatically have to go down. If in any case demand ever exceeds supply, some areas in the country do not receive power and this might cause a system-wide blackout. In order to avoid this, last resort intervention policies are used like load shedding.

  Data used in the study was about the net energy sent out (NESO), which was defined as the rate at which electricity is delivered to customers and was measured in megawatts (MW). The data was in form of daily peak demand (the maximum hourly demand in a 24-hour period) from 1^{st} January 1996 to 14^{th} December 2009 with 5097 observations. They preferred daily peak demand data because their interest was in making shot-term forecasts. It excluded data on demand of those willing and able to pay for electricity but could not access it at the time of the study. After analysing the data, multiple seasonality was evident, corresponding to weekly and monthly periodicity. Since the data was short-term, frequent fluctuations led to non-zero mean and non-constant variance. Therefore, they applied volatility models, and GARCH was given preference. The developed models were also used for out of sample prediction of daily peak demand and concluded that the Reg-SARIMA-GARCH model produced better forecast accuracy with a mean absolute percent error of 1.42%.

- Another similar study was done by Ghosh (2008). They studied the monthly peak demand of electricity in the northern region of India, using univariate time series techniques. Two univariate models were applied in this study; Multiplicative SARIMA (MSARIMA) model and Holt-Winters Multiplicative Exponential Smoothing model (H/WMES). Originally, their focus was on studying non stationary homogeneous ARIMA models with seasonal variations and general multiplicative seasonal models. This was because most monthly demand data is faced by seasonal variations which cause non stationarity. For exponential smoothing, they preferred the H/WMES because it is more appropriate when it comes to data with seasonal variations.

  The data used in their study was collected from the Northern Regional Load Dispatch Center (NRLDC) website and was measured in Megawatts (MW). It covered a period between April 2000 and February 2007 and was used to make forecasts for the next 15 months. They also divided their data into 2 sets, the first one was from April 2000 to April 2006 and the second covered May 2006 to February 2007. Since they studied a very highly weather sensitive region, they used weather related variables to forecast the peak demand. After finding the error differences, root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) were calculated. Using the AIC and RMSE, the MSARIMA model gave better results than the H/WMES model. However, finding out which technique worked best was so general and they narrowed it down using the Box and Jenkins methodology. This helped them identify the most suitable ARIMA model. They found that SARIMA $(2,0,0)(0,1,1)^{12}$ was the best model to explain the monthly peak demand of electricity according to the data they had, and make appropriate forecasts. Results of the study were meant to help NRLDC make necessary arrangements to meet the future peak demand of electricity in Northern India.

- In their study, Kesavabhotla and Babu (2012) used Statistics, operations research and computer programming to discover and communicate meaningful patterns in data. This is commonly known as analytics. These techniques were used to forecast day-ahead electricity demand. The study investigated the application of ARMA and GARCH modeling techniques to fit the historical data and estimate the coefficients to predict the day-ahead electricity demand. R-Programming was used to fit the models. The data population of 375 observations from daily electricity demand of Andhra Pradesh

State, India between 2005 and 2006 was considered. To identify and fit the model, 365 daily electricity observations of 2005 were used for sampling and forecasting 10 days-ahead electricity demand. 10 observations of 2006 were considered for comparing the predicted electricity demand. The fitted model could be applied to all the other years for validation. The preliminary data analysis consisted of the average daily load in each year, year on year change in average daily load, minimum daily load in each year, and maximum daily load in each year.

The ARMA(1,1) model gave predictions of day-ahead electricity demand with $80-95\%$ confidence bounds for short duration. However the assumption of constant variance of residuals is not true in reality for various reasons. Therefore, the residuals of ARMA(1,1) were tested for autoregressive conditional heteroskedasticity (ARCH) effects using McLeod and Ljung-Box tests. The GARCH(1,1) model was identified, whose coefficients were estimated in order to use the model for prediction of conditional variances. From the results it was concluded that it is always a good practice to test the volatility of the errors after fitting linear models to the data. This helps to improve the accuracy of predictions. Non linear issues of errors can be handled appropriately through GARCH models which provide flexibility to coexist with other models. The combination of ARMA and GARCH models gave accurate forecasting in high volatility scenarios.

- An application of the linear models applied in this study was carried out by Kumar and Anand (2014). They used time series ARIMA forecasting models to predict sugarcane production in India. ARIMA models, also commonly known as Box-Jenkins' 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 to make predictions using time series data with any kind of pattern, and assume and take into account the non-zero autocorrelation between successive values of the time series data. Sugarcane was chosen because, apart from Brazil, India has the largest sugar production capacity in the whole world. Data covering a period of 62 years of sugarcane production was used to predict 5 years ahead. The data was taken from the Department of Agriculture and Cooperation (DAC) in India, from 1950 to 2012.

After modelling and analysing the data, an ARIMA(2,1,0) was chosen as the best model explaining the patterns of the data perfectly. Further, efforts were made to forecast, as accurate as possible, the future sugarcane production for a period upto 5 years. Forecast results showed that the annual sugarcane production would grow in 2013, then take a sharp dip in 2014 and in subsequent years 2015 through 2017. It would then continuously grow with an average growth rate of approximately 3% for the following years. The study statistically tested and validated that the successive residuals in the fitted ARIMA time series were not correlated, and the residuals seem to be normally distributed with mean zero and constant variance. In conclusion, the selected ARIMA(2,1,0) was an adequate predictive model for the sugarcane production in India. Although, like any other predictive models in forecasting, ARIMA also has limitations on accuracy of predictions yet it is used more widely for forecasting the future successive values in the time series.

- Saab et al. (2001) modelled and forecast electricity consumption in Lebanon using univariate approaches. Three univariate techniques were used to model and forecast EC; autoregressive (AR), ARIMA models and a combination of an AR(1) with a high pass filter (AR(1)/HPF). The main aim of their study was to investigate different univariate models and use them to forecast one month ahead electricity consumption in Lebanon. The interest was in identifying a forecasting method that would perform best on the unusual data 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 demand in order to help in the development of that sector and the country at large.

  Monthly average EC data was used, covering January 1970 until May 1999. A time series plot revealed 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 a non genuine 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 in the data, an adequate model was vital to carry out the forecasting exercise. A non linear deterministic model was used to represent the trend in the

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 39^{th} lag having a 0.057 standard deviation and maximum correlation of 20.155. For the AR(1)/HPF model, there were diverse correlations in all positive lags, with the 4^{th} lag having 0.09 standard deviation and 20.32 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. 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 resulted into better forecasts for the AR(1)/HPF model, as compared to both the AR and ARIMA models.

- In Rallapalli and Ghosh (2012), a study to forecast monthly peak demand of electricity was carried out. Similar to Ghosh (2008), this study was about India which is a developing country with a great need to accurately forecast demand in the energy sector using advanced forecasting methods. Due to the level of development in this country, there is a scarcity of resources. Therefore, India does not need to poorly invest in the electricity sector because it will affect the investments in other developmental activities, but still under estimation will cause electricity shortages. That is why proper forecasting is necessary for proper planning and sustainable development. Since all the forecasts made in India by the Central Electricity Authority (CEA) were usually over estimated because of poor techniques, this study tried to predict the same demand using MSARIMA model.

  An ARIMA model was considered, which was divided into the non stationary homogeneous models with seasonal variations and general multiplicative seasonal models. Data about the peak demand (measured in MW) of all the regions in India (north, west, east, south and south-east) for the period of April 2005 to March 2011 was collected from CEA. An adequate representation of the data through a model was important, that is why an ACF and PACF were used to determine the stationarity of the data and identify the

possible values of the regular part of the model. Point estimates of the coefficients of the model using the Maximum Likelihood Estimation method were then obtained in order to identify the seasonal part of the model. According to the standard errors of these coefficients, the most insignificant coefficients were dropped off from the model. The AIC information criterion and RMSE were used to identify the best model explaining the data, and an inspection of the model residuals for any remaining autocorrelation was carried out.

Model performance of both the MSARIMA model and the trend model used by CEA was evaluated, using RMSE, MAE and MAPE. Errors generated by the MSARIMA model were much smaller than the CEA trend model errors. For clarity, forecasts for the period April 2011 to July 2011 were calculated and compared with the CEA actual peak demand data at its time of publication. MSARIMA results were still doing better in all the five regions of India. CEA was advised to use some of the modern and more accurate forecasting techniques like MSARIMA, ARIMA-EGARCH, Exponential Smoothing, Vector Auto Regression and Neural Networks.

- In a case study of Dubai, Roken and Badri (2006) studied forecasting monthly peak load demand using time series models. In this study, an attempt to develop, 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 include a variety of complex techniques, such as exponential smoothing, Box-Jenkins (BJ) and dynamic regression. The objective was to produce 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 provide timely data for reasons of confidentiality and secrecy.

  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 period matched and correlated were also carried out. Such efforts reflected how the recommended models captured most of the characteristics of the data. Monthly electricity peak load data from January 1985 to March 2007 was provided by the Dubai Electricity and Water Authorities (DEWA). 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 plot of the data, there existed patterns of seasonality and trend. Demand 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 utilized in the study followed 7 main steps; Obtaining time series data, performing initial data screening to identify 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 (linear trend with multiplicative seasonality) and Box-Jenkins ARIMA model with root transform $[(1,1,1) * (0,1,1)]$.

The recommended models passed a sequence of stringent diagnostic tests, including comparing outputs with selected holdout samples. A comparison 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 also to other data sets displaying seasonality and trends. Given the similar climatic conditions in other regions of the country, the method and process used in the study can be reasonably generalized.

2.2 The electricity situation in Uganda

Uganda is a developing country in the eastern part of Africa, with agriculture as its main economic activity. It is an electricity deficient country and the electricity sector operates at bare capacity margin. Uganda fulfills its energy requirement through different sources, which include oil, coal, gas and firewood. In 1999 the government of Uganda embarked on the most extensive power sector reform program all over Africa, in order to deal with the power crisis (Mawejje et al., 2013;

Wamukonya, 2003). This reform and privatization policy resulted in the separation of the Uganda Electricity Board (UEB) into generation, transmission and distribution successor companies (Keating, 2006; Turkson and Wohlgemuth, 2001).

Electricity Regulatory Authority (ERA) is a statutory body established in 2000 in accordance with the Electricity Act 1999 to regulate the generation, transmission, distribution, sale, export and import of electrical energy in Uganda. Functions performed by this body as listed on the ERA website include; guiding the liberalization of the electricity industry, managing licensing, rates, safety and other matters concerning the electricity industry. ERA also supervises all licensed companies within the electricity sector to ensure they comply with the Electricity Act 1999 and Regulations thereto, and to safeguard all stakeholders' often competing interests. In performance of its functions, the authority ensures that electricity companies comply with the conditions of their licenses and protects the interests of electricity consumers in respect to prices, charges and other terms of supply of electricity and the quality, efficiency, continuity and reliability of the supply services (Mawejje et al., 2012).

Uganda Electricity Transmission Company Limited (UETCL) is a Public Limited Company which was incorporated on 26th March 2001. The company operates under policy guidance of the Ministry of Energy and Mineral Development. It is a public limited liability company owned by the Ministry of Finance, Planning and Economic Development. It has the operational mandate that is divided into the single buyer business and transmission system operator. It therefore undertakes bulk power purchases and sales, import and export of energy, operation of the high voltage transmission grid and plays the national system operator role (UETCL website).

Uganda Electricity Distribution Company Limited (UEDCL) is a limited liability company incorporated under the Companies Act and started operating on 1st April 2001. UEDCL is one of the successors of the Uganda electricity board and the owner of the electricity distribution network up to 33KV, as shown on the UEDCL website. The network was handed over to UMEME limited on the 1st march 2005, under a concession arrangement (Mawejje et al., 2012). The concession involved UMEME and other parties entering into a number of agreements, for example, the power sales agreement with UETCL and licenses for electricity distribution, supply, and embedded generation by ERA.

For Uganda to attain its development goals, the available resources need to be utilised sustainably. One of the most important sectors for a developing country like Uganda is the energy sector. Therefore, the ability to predict, plan and manage demand in this sector is very vital for the development of the country. Until July of 2012, the country was undergoing a deficiency in power supply. This was especially during the peak hours (evenings) where demand was 443MW yet supply was 330MW (New Vision Website). Peak demand in Uganda is in the evening because some medium scale industries usually extend their work till late evening hours whenever electricity is available. Likewise, during that time, people are free to do their leisure activities after a long day of work. The construction of the Bujagaali power plant increased the supply by 250MW and as a result, the demand of electricity increased aggressively (Baanabe, 2012). The increase in power supply was expected to enhance economic activities and reduce most of the expenses incurred while using generators.

From electricity demand forecasts made by UMEME, the future demand was supposed to be settled at least within the next 24 months by the surplus electricity that was produced, without any load shedding (Skyscrapercity). However, because of the escalated investments in industries whose activities heavily rely on electricity and increase in operation of industries that were originally under-producing, like China's Tiang Tang Steel factory, the forecasts made earlier have become irrelevant. The overall energy consumption in Uganda has increased exponentially in the previous years, as stated by New Vision Website, where the chief of ERA, Dr Benon Mutambi, was quoted saying *"although the electrification rate in Uganda is still low, the demand of electricity is currently increasing at a rate of* 10% *per year, compared to the previous years"*. Therefore, alternative power generating sources need to be established in the country to take care of the increasing demand. By the end of 2012, UETCL's statistics section found out that the peak demand had increased from the previous 443 MW to 487 MW in a very short time because of the economic activities that were increased (Skyscrapercity).

According to the Uganda Investment Authority (UIA), the favourable investment climate in Uganda has led to increased industrialization (Odenthal et al., 1999). Likewise, local commercial and agro-processing businesses are on a rise and in order for them to produce goods of standard quality and value, they will need electricity. This means a high increase in electricity demand, yet there are no profound strategies for significantly increasing supply. Therefore, to create an

even more favourable environment for these investments, adequate, affordable and reliable electricity should be available. The National Development Plan (NDP) identified low electricity generation transmission and distribution capacity as the main limitations faced by the electricity sector and suggested that construction of larger power plants would be the first intervention strategy (New Vision Website).

Uganda's manufacturers and traders in the industrial sector were relieved from the problem of load shedding and using generators after the construction of the Bujagali hydro power dam. By the end of 2013, ERA recorded statistics of around 500MW of peak demand, yet the total capacity of the available power plants was 682MW (Emma Onyango). This proved the fact that peak demand kept increasing at a $10 - 12\%$ rate annually. Due to the threatening increase in the demand of electricity, the Government of Uganda, through the Ministry of Energy and Mineral Development signed a contract with a Chinese company called Sinohydro to construct a 600MW dam at Karuma falls towards the end of 2013 and it is anticipated to be completed after a period 60 months (Mawejje et al., 2013). For the same cause, another 183MW power plant is under construction at Isimba falls in Kamuli District. This contract is also with a Chinese company called China International Water and Electric Corporation (CWE) and it is expected to take 40 months.

The former Prime Minister of Uganda Mr. J. P. Amama Mbabzi also launched a nuclear power generating program which he said would help in sustainable development (Uganda). This contract is with the International Atomic Energy Agency (IAEA) and was signed towards the end of 2013. Mr. Mbabazi's reason for the launch of the program was to attract more investment opportunities that the country needs, which he said would not be possible without proper infrastructure, most especially electricity. In an article by Emma Onyango, Dr Mutambi was also quoted talking about two heavy fuel oil-based plants of 100MW reserved in Tororo and Namanve that are able to meet the growing demand for some time. He however said that these two renewable energy sources can only be used as last resort because they incur a high maintenance cost. Dr. Mutambi talked about the policies currently being undertaken by ERA like the Global Energy Transfer for Feed-in-Tariffs (GETFiT) scheme, which acts as an incentive for investors to embark on renewable energy projects to produce energy.

The first time ERA sent out bids for the GETFiT, a capacity of 83.7MW from eight renewable energy projects was realized. These projects were meant to start in

2014 and go for either two or three year. Bids equivalent to a capacity of 67MW were established towards the end of the year 2013 from which more renewable energy projects are expected to qualify for the GETFiT premium. ERA has also taken a step in the direction of solar energy as a remedy to solve the increasing electricity demand with limited supply. During the first half of the financial year 2013/14, five permits were given to different companies to study the anticipated development of solar photovoltaic power of 99MW which would supply different parts of the country.

More projects of 50MW from solar photovoltaic are expected to be commissioned by mid-2015. Five other permits were issued for prospected development of 33.7MW of small hydro power plants, ERA also extended the duration of three permits for the prospected development of 36.2MW from renewable energy. All these power sources are aimed at increasing the power generation capacity hence increasing power supply. However, not knowing the demand for which supply is increasing is another problem. Therefore, the first step to these strategies is knowing how much electricity is demanded and all the relevant information necessary to affect the supply of electricity. This means having accurate electricity demand forecasts made if the electricity sector is to meet and adequately supply the demand of the country.

Given that Uganda is a developing country, the process of forecasting its future electricity demand is more complicated because of various factors like; poor performance of the energy sector, poor infrastructure and denial to transform from traditional to modern energy sources. In addition to those factors, it is hard to estimate demand for the whole country since it is divided into different economical and social classes. There is a fast growing urban sector co-existing with a rural "dormant" sector. Rural areas are characterized by informal economic activities, unemployment or semi-employment, more itemized than monetized transactions and either low payments or payment in kind as rewards of work. This division leads to a non uniform level of industrialization in the country, which results into non uniform changes in the economic structure of the country as a whole.

In their study, Bhattacharyya and Timilsina (2009) note that due to the existence of the rural sector and the use of traditional energy sources, developing countries have "incomplete markets" whose prices are hard to determine, hence complicating output and income distribution, irrespective of the existing supply and demand. This makes it hard for policy makers to find solutions to the problems existing

in such countries, since the common neoclassical paradigm can not be applied (Herr et al., 2009). Changes in social factors like technology and fuel consumption in developing countries are important dynamics for determining future energy demand, since they have an effect on the environment and the sustainability of the economy as a whole. It is always wise to incorporate such dynamics when modeling the transition of these countries.

Due to poor policies and wrong investment decisions, developing countries suffer from electricity supply shortages. This means that not everyone who demands for electricity is supplied. Therefore, the recorded consumption data does not exactly represent actual demand because there is a portion of demand that is not supplied. This means that market forces of supply and demand are distorted, so the market does not freely clear up. The big difference in income level also leads to inequality in the consumption of electricity. This discourages the put up of social policies which in the end results into losses in the energy sector. The costs of distributing electricity in all regions of the country are the same but the recovery from rural regions is usually low. This leads to low profit inflows into UMEME, which results into low financial performance, leading to a reduction in capital for more investments hence reducing the capacity to supply electricity even more.

In reality, it seems "impossible" to forecast electricity demand in a developing country given all constraints discussed in this section. That is why, this study concentrates on finding a model that will help make forecasts as accurate as possible, given the available data.

Chapter 3

Theoretical aspects of time series analysis

Time series analysis is used to either model randomness in a given data series or forecast future values basing on observed historical data. Time series data can be from any field, but is more often collected when monitoring industrial and corporate business processes (Chukwukelue et al., 2013). This chapter gives a brief 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, transformations, differencing, autocovarinace, autocorrelation functions (ACF) and Partial Autocorrelation Functions (PACF) in detail. It also covers preliminary analysis of electricity demand data from both Uganda and South Africa. Concepts in this chapter are useful for reference purposes in the following chapters.

3.1 General description of time series data

A time series is a sequence of data ordered in uniform time intervals. An example of time series data is monthly electricity demand observed over many years. It is also called a historical or chronological series (Chavez et al., 1999). Univariate time series analysis involves using data about a single variable to build a model that describes the behaviour of the variable in the past. Basing on the built model, satisfactory forecasts for the future are ably made. Since the analysis of time series data depends on what is observed in the past about a specific variable, it becomes

more difficult to analyse this type of data as compared to other data types. 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; $X_1, X_2, \cdots, X_N$ and want to forecast its value at time $N+h$. Denote the forecast as $\hat{X}_N(h)$, where;

- $\hat{X}$ is 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.

If the forecast $\hat{X}_N(h)$ is a future value calculated using a model developed from all observations up to period N, then it is called an out-of-sample forecast. The problem with out-of-sample forecasts is that their accuracy can not be evaluated until real observed data for the initially forecast time horizon is available. Alternatively, if the used model is developed from all the available data and it is used to forecast a value within the available data, then the resultant forecast is called an in-sample forecast. The accuracy of an in-sample forecast can be evaluated but it is usually not genuine because the data used to develop the model is the same data used to test it (Chatfield, 2002). A better way of dealing with time series data for forecasting is to split the series into two parts. The first part is called the *estimation/training sample* and it is used to estimate the starting values, smoothing parameters and also train the model. It usually contains $(75-80)\%$ of the observations, depending on the size of the series. The remaining $(20-25)\%$ makes the *test sample*, which is used to check the performance and accuracy of the forecasting model (Hyndman and Athanasopoulos, 2014).

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, one needs to know the features of the data available in order to choose the right model to fit to the data. The easiest way to do this is to make a time series plot with observations against time. Using the time series plot, features like trend, seasonality, outliers, changes in structure, turning points and sudden discontinuities are easily observed.

3.2 Components of a time series

Time series processes generally contain two different types of variation, namely, the systematic variation (trends, seasonal, cyclic, that we would like to capture and model), and the random variation (inherent background noise in the process).

3.2.1 Seasonal component

This type of variation generally repeats itself at fixed intervals within a year, for example, weekly, monthly or quarterly. During these intervals similar patterns of behaviour are observed. Seasonality exists when a series is influenced by seasonal factors and is usually predictable. It always happens during a fixed and known time interval. However, if a time series is measured only once per year, detecting seasonality might be complicated (Chatfield, 2002).

3.2.2 Cyclic component

This pattern exists when the data series exhibits rises and falls that are not of fixed periods (Bhar and Sharma, 2005; Jebb et al., 2015). The duration of these fluctuations is usually of at least 2 years. Cyclic variations are regular in nature and often occur in periods of more than one year. Cyclic patterns are common in economic and business data where declines or growths can happen over a period of time, say, five years but the duration is not known beforehand. The main difference between cycles and seasons is that; if the changes are not of fixed period then they are cyclic. Otherwise, if the period is constant and associated with some aspect of the calendar, then the pattern is seasonal (Hyndsight Website).

3.2.3 Trend component

A trend exists when a series exhibits steady upward or downward movement over a long period of time. This movement can either be linear or non-linear. It is defined by Chatfield (2002) as the long-term change in the mean level per unit time. If a time series does not show an increasing or decreasing pattern then the series is stationary in the mean. A trend is usually caused by long term factors

affecting the variable under study, for example, population growth, price inflation and general economic changes.

3.2.4 Random component

This is also called the irregular fluctuation. It is the variation left in a data series after removing all systematic effects, like, trend, seasonality and cycles. Random effects are changes in data caused by non-recurring factors, for example, tsunamis, earthquakes. These effects are completely random and unpredictable. In other wards, they can not be forecast. During a forecasting exercise, the main objective is to model all the systematic components until the only unexplained component is the irregular fluctuation.

Time series plots showing some of the different time series patterns are shown in Figure 3.1 (Hyndman and Athanasopoulos, 2014).

Time series data features

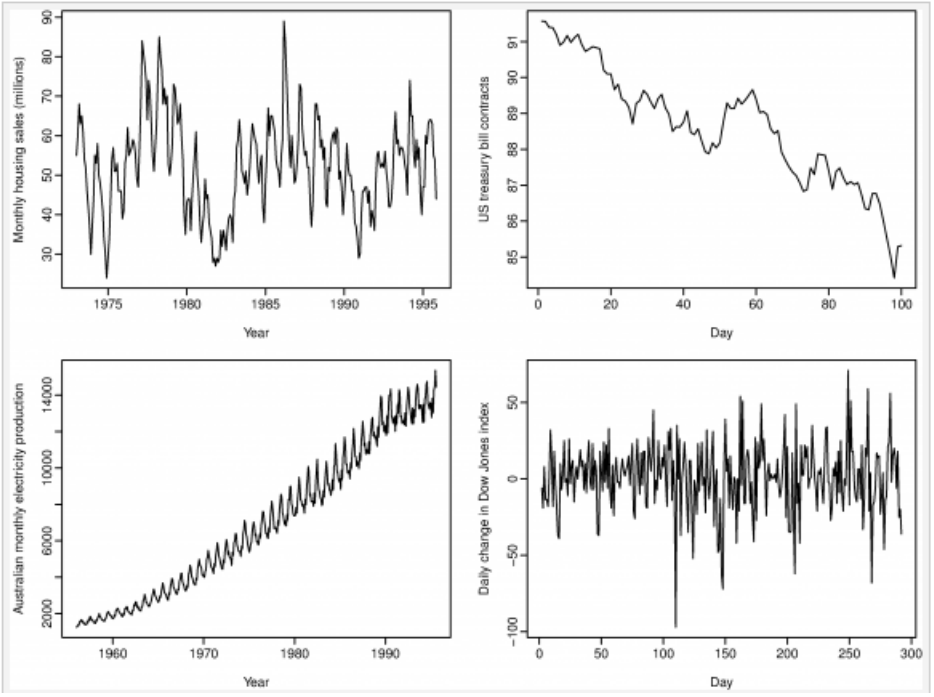

FIGURE 3.1: Top-left corner shows seasonality, top-right corner shows a downward trend, bottom-left corner is an upward trend and bottom-right corner is a random fluctuation.

In this study, we focus more on the trend component because of the nature of the data. Originally, trend was statistically defined as (Chatfield, 2002);

$$T_t = \alpha + \beta t, \tag{3.1}$$

where; T_t is the trend at time t, α is the intercept, and β is the slope.

Equation 3.1 is called a simple linear global trend model which shows that trend does not depend on time. However, recent studies show that trend is better modeled with an effect of time, for example, studies by Chatfield (1996) and University of South Carolina include local α and β which evolve through time. This results into a trend model

$$\mu_t = \alpha_t + \beta_t t, \tag{3.2}$$

where; μ_t is the local mean level at time t, α_t is the local intercept, and β_t is the local slope.

Equation 3.2 is called a simple local linear trend equation which considers the fact that parameters α and β change over time. This type of equation is preferred to the classic global liner trend equation because models developed from it (equation 3.2) produce more realistic results when applied to real data (Chatfield, 2002). There are other types of trend which will not be discussed in this study, for example, quadratic and logarithmic.

3.3 Stationarity

A time series is said to be stationary if there is no systematic change in mean and variance. In other words, the properties of the data are much more uniform throughout all sections of one series. In simple terms, a stationary time series will have no predictable patterns in the long run. Therefore, a series with trend and seasonality components is not stationary because these components affect the value of the series at different times of observation (Hyndman and Athanasopoulos, 2014).

A time series is said to be strictly stationary if the joint distribution of $X_{t_1}, X_{t_2}, ..., X_{t_n}$ is the same as that of $X_{t_1-k}, X_{t_2-k}, ..., X_{t_n-k}$ for all time periods t and all time lags

k (Shumway and Stoffer, 2010; Washington et al., 2010). Shifting the time origin by an amount k has no effect on the joint distributions, which must therefore depend only on the intervals between $t_1, t_2, ..., t_n$. This means that for a strictly stationary process, the mean, $E(X_t) = E(X_{t-k}) = \mu$ and variance, $\text{var}(X_t) = \text{var}(X_{t-k}) = \sigma^2 = \gamma(0)$ are constant throughout time. Likewise, the covariance between any two observations depends only on the time lag between them; $\gamma(t, t-k)$ depends on k only. A series is said to be second order stationary if both the first and second order moments do not depend on time and both the covarinace and correlation are functions of the time lag only (Reinert, 2002). Second order stationarity is also called weak stationarity.

Majority of time series analysis methods can easily be applied to stationary time series. Therefore, it is always important to know whether the data at hand is stationary or non-stationary before any further analysis. If not stationary, one is required 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 staionarity, for example, software like the unit root test or 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, especially in a trendy manner (Maradiaga et al., 2013). There are two popular approaches for non stationary series with a trending mean;

- Trend stationary: In this case, the mean trend is deterministic. Once the trend is estimated and removed from the data (using regression), the residual series is a stationary stochastic process (Clements et al., 2001). Effects of shocks in series with a deterministic trend are always eliminated in the long run. As a result, forecast intervals from this approach have constant width.
- Difference stationary: Here the mean trend is stochastic. Differencing the series yields a stationary stochastic process. Time series with a stochastic trend have forecast intervals that grow over time and their shock effects are permanent (Franses, 1998; Heij et al., 2004). 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, detrending (using regression to fit 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.

3.4 Differencing

Differencing is a special type of filtering used to remove trend from time series data until stationarity is achieved. Suppose we have a stochastic process $\{X_t\}$. The first difference ∇X_t is defined as;

$$\nabla X_t = X_t - X_{t-1}. \tag{3.3}$$

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

$$\begin{aligned} \nabla^2 X_t &= \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}), && (3.4)\\ &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}), && (3.5)\\ &= X_t - 2X_{t-1} + X_{t-2}. && (3.6) \end{aligned}$$

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

$$\begin{aligned} \nabla^d X_t &= \nabla^{d-1}(\nabla X_t), && (3.7)\\ &= \nabla^{d-1} X_t - \nabla^{d-1} X_{t-1} \; ; \quad d = 1, 2, \cdots && (3.8) \end{aligned}$$

In most cases, when dealing with real data, instead of fitting a stochastic model to a non stationary series to remove the linear trend, preference is given to taking the first difference and examining the results for stationarity. One advantage of differencing over detrending is that no parameters are estimated in taking differences. However, differencing does not provide an estimate of the error process. Therefore, if an estimate of the error process is crucial, detrending may be more appropriate. Otherwise, if the goal is only to make the data stationary, then differencing may be preferred (Shumway and Stoffer, 2010).

Taking first differences removes a linear deterministic trend. However, if the data exhibits a quadratic trend, then it is necessary to take the second difference in

order to remove that trend. It is not so common to take third or higher order differences when dealing with real data (Chatfield, 2002).

3.5 The autocovariance and autocorrelation functions

Assume moments are taken on a series $\{X_t\}$;

$$M_N = E[X - (E(X))^N]. \tag{3.9}$$

In studies like Reza (1961) and Press et al. (1992), it is indicated that the first order moment about the origin is the mean $E(X_t)$ of the data and the second order moment about the mean is the variance $\text{Var}(X_t)$ which is mathematically the same as the covariance $\text{Cov}(X_t, X_t)$. Generally, the second order moment is the covariance between variables X_t and X_{t-k}, which becomes the variance when $k = 0$. Such a type of covariance in a series between current and lagged values is called autocovariance, and it is usually represented by γ (Wang and Jain, 2003; Wei, 1994). The autocovariance function at lag k, γ_k is expressed as;

$$\gamma(k) = \text{Cov}(X_t, X_{t-k}), \tag{3.10}$$

$$= E[(X_t - E(X_t))(X_{t-k} - E(X_t))]. \tag{3.11}$$

In order to understand how a variable of interest is related to its own past observed values, one can use the autocorrelation (or serial correlation ρ). In simple terms, autocorrelation explains the correlations between values of a random process at two different times, as a function of those two time points (Haag, 2005). Autocorrelation is preferred to autocovariance when interpreting results because autocovariance depends on the units of measurement of the variable under study (Scheaffer and Young, 2009). Autocorrelation is usually measured on a scale of -1 to 1. The autocorrelation function (ACF) at lag k is expressed as;

$$\rho(k) = \text{Corr}(X_t, X_{t-k}) = \frac{\text{Cov}(X_t, X_{t-k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t-k})}} \Longrightarrow \frac{\gamma_k}{\gamma_0}. \tag{3.12}$$

For a stationary process, $\sqrt{\text{Var}(X_t)\text{Var}(X_{t-k})} = \sigma^2$ because $\text{var}(X_t) = \text{var}X_{t-k}$, $\rho(k) = \rho(-k)$, $\rho(0) = 1$, and $|\rho(k)| \leqslant 1$.

A partial autocorrelation function (PACF) is another indicator of correlation. It measures the relationship between observations X_t and X_{t-k} after removing the effects of the other time lags; $1, 2, 3, \cdots, k-1$ (Vivanco, 2008). 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 PACF, which is also called the conditional correlation function is expressed as;

$$\phi_k = \text{Corr}(X_t, X_{t-k}|X_{t-1,\cdots,X_{t-k+1}}). \tag{3.13}$$

The ACF and PACF can be used to determine whether the data is stationary or not, and to identify 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. 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 cuts 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) component in it. The lag at which the ACF cuts off indicates the number of MA terms to be considered when building the model. For example, for an AR(1) process, the ACF declines in geometric progression from its highest value at lag 1, while the PACF cuts off abruptly after lag 1 (Chramcov, 2011). The opposite pattern applies to an MA(1) process, where, the ACF cuts off abruptly after lag 1 and the PACF declines in geometric progression from its highest value at lag 1. Table 3.1 can be of good importance when choosing appropriate values of p and q using the PACF and ACF.

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

TABLE 3.1: Relating the ACF and PACF to the AR and MA processes.

3.6 Transformations

Transforming data means performing the same mathematical operation on each piece of the original data (McDonald, 2009). Many common methods of analysis assume that the amount of variability in a time series is constant across time, which is not always true. When dealing with non stationary time series, it is important to transform the data first before any further analysis. For example, if there is clear evidence of non constant variance over time, then a suitable transformation to the data might remove, or reduce the impact of, the non constant variance pattern.

A situation in which the variability of a time series is unequal across time is called heteroscedasticity (de Carvalho et al., 2014; Rubliková and Hill, 2006). In case the time series plot of the data displays a change in variance, then a transformation is necessary prior to the main analysis. Transforming the data helps to stabilise the variance thereby making the series homoscedastic. A transformation may also be helpful in making the data more normally distributed, especially if the observations appear to be skewed to some direction (Monsen and Van Horn, 2007). Likewise, if the seasonal effect in the data appears to be multiplicative, it would be desirable to make the effect additive and this can be done through a transformation (Farooque, 2002). This is because linear effects are generally easier to handle.

Data transformations are an important tool for the proper statistical analysis of any type of data. For presentation purposes, it is essential that one is able to defend their choice of data transformations. There are many types of transformations, but it is better to use a transformation that other researchers within the same field of study commonly use, such as the square-root transformation for count data or the log transformation for size data (O'hara and Kotze, 2010; Osborne, 2005). In this study, we consider a transformation commonly used by researchers in the electricity demand forecasting specialisation, and this is either differencing or log transformation. Furthermore, with a limited data set, we may not be able to see much effect of the transformations on the normality and homoscedasticity.

Let $X_1, \cdots, X_t$ denote the original observations and $W_1, \cdots, W_t$, the transformed observations. Then, a logarithmic transformation is given as $W_t = \log(X_t)$. Logarithms are commonly preferred because they do not require complex interpretation. Changes in a log value are relative or percentage changes on the original scale (Keene, 1995). If log base 10 is used, then an increase of 1 on the log scale

corresponds to a multiplication of 10 on the original scale. Log transformations also constrain the forecasts to stay positive on the original scale (Hyndman and Athanasopoulos, 2014). Log transformation can lead to substantial reductions in forecast mean squared error (MSE), if taking logs leads to a more stable variance of the data. Otherwise, if the transformation is applied but does not make the variance more homogeneous, it can be damaging to the forecast precision (Lütkepohl and Xu, 2012).

Other transformations are also used, for example, square and cube roots. These are called power transformations because they can be written in the form $W_t = X_t^p$. A useful family of transformations that includes logarithms and power transformations is the Box-Cox transformations developed by Box and Cox (1964). These transformations depend on the parameter λ and are defined as;

$$W_t = \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & ; \ \lambda \neq 0 \\ \log X_t & ; \ \lambda = 0. \end{cases} \tag{3.14}$$

This class of transformations is preferable for theoretical analysis because it takes into account the discontinuity at $\lambda = 0$ (Sakia, 1992). The definition in equation 3.14 is only valid for positive values of X, $(X_t > 0)$. Therefore, modifications had to be made for negative observations. Box and Cox proposed the shifted power transformation of the form;

$$W_t = \begin{cases} \frac{(X_t + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & ; \ \lambda_1 \neq 0 \\ \log(X_t + \lambda_2) & ; \ \lambda_1 = 0. \end{cases} \tag{3.15}$$

Where λ_1 is a parameter defining a particular transformation and λ_2 is chosen in such a way that $X_t > -\lambda_2$ (Li, 2005). The aim of the Box-Cox transformations is to ensure that the usual assumptions for linear models hold. Most common transformations reduce positive skewness but may worsen negative skewness unless the variable is reflected prior to transformation. The Box-Cox transformation eliminates the need to reflect variables (Osborne, 2010).

The main objective in the analysis of Box-Cox transformation model is to estimate the transformation parameter λ. Most, but not all, of the modern Statistical packages have implemented ways of estimating λ. For example, SAS has a convenient and well done implementation of Box-Cox within *proc transreg* that iteratively tests a variety of λ and identifies the best options (Osborne, 2010). Otherwise, λ

can be estimated by hand. Box and Cox considered two approaches; the Bayesian method which requires one to first ensure that the model is fully identifiable, and the Maximum Likelihood method (MLE). MLE is commonly used because of its easy concepts and an easy-to-compute profile likelihood function. It is also easy to obtain an approximate confidence interval for λ because of the asymptotic property of MLE (Li, 2005).

Another proposed procedure for manual calculation of λ is given by Osborne (2010), where they used an example to verify their proposal.

- They divided their data into at least 10 regions which turned out evenly distributed.
- They selected each part and calculated their respective means and standard deviations.
- Then took $\log_{10}$ of each mean and standard deviation (sd) and plotted the resultant data as $\log(sd)$ against $\log(mean)$ for all parts.
- They estimated the slope (b) of the plot, and used $1-b$ as the initial estimate of λ.

In case the plot is in form of a curve, it is better to estimate the slope for each segment of the line and calculate the average of all. This produces an average slope (b) which can then be used to calculate λ. The resultant value of λ from the proposed procedure was interestingly very close to the empirically derived value.

The Box-Cox class of transformation incorporates many traditional transformations (see table 3.2):

λ	Effect
1.00	Produces results identical to original data
0.50	square root transformation
0.33	cube root transformation
0.25	fourth root transformation
0.00	natural log transformation
-0.50	reciprocal square root transformation
-1.00	reciprocal (inverse) transformation

TABLE 3.2: Box-Cox Transformations

3.7 Preliminary analysis of the available data

The first step in any time series analysis is to make a time series plot (Reinert, 2002). The plot should show all important features of the data such as trend, seasonality, cycles and randomness. A time series plot is very important for data description and in helping to formulate a sensible model for both analysis and forecasting (Adhikari and Agrawal, 2013; Chatfield, 2000).

3.7.1 Uganda monthly electricity demand data

The time series plot in Figure 3.2 shows the monthly peak demand of electricity in Uganda for 72 months. We faced a big challenge of limited data, given that Uganda is a developing country, characterised by limited data and poor data storage systems. The series shows a positive linear trend along with random fluctuations about this trend. This long term increase is mainly due to economic development of the country over the years. With the increase in Lake Victoria water levels since the end of 2006, and the entry into service of additional plants, there has been a continuous increase in the supply and demand of electricity (New Vision Website).

The series starts in January 2008 and increases with random fluctuations until the middle of the year, from where it keeps increasing gradually. At the beginning of 2009, which is the 13^{th} month, it drops back but at a level higher than the beginning of the previous year. This trend continues, until the big decline at the end of 2009 (24^{th} month), before surging back at the beginning of the next year. The trend keeps on for the rest of the year until the beginning of the next year where it drops again but still at a level higher than the beginning of the yeah before. This pattern continues with the demand dropping at the beginning of every year but to a level higher than the beginning of the year before.

In most countries, electricity demand is expected to be influenced by seasonal changes. However, in Uganda, there are no extreme weather changes that can affect the demand of electricity (Safari Website). Daily mean temperature in major towns of Uganda measured in degrees Celsius as reported on the Kabiza Website is shown in Table 3.3

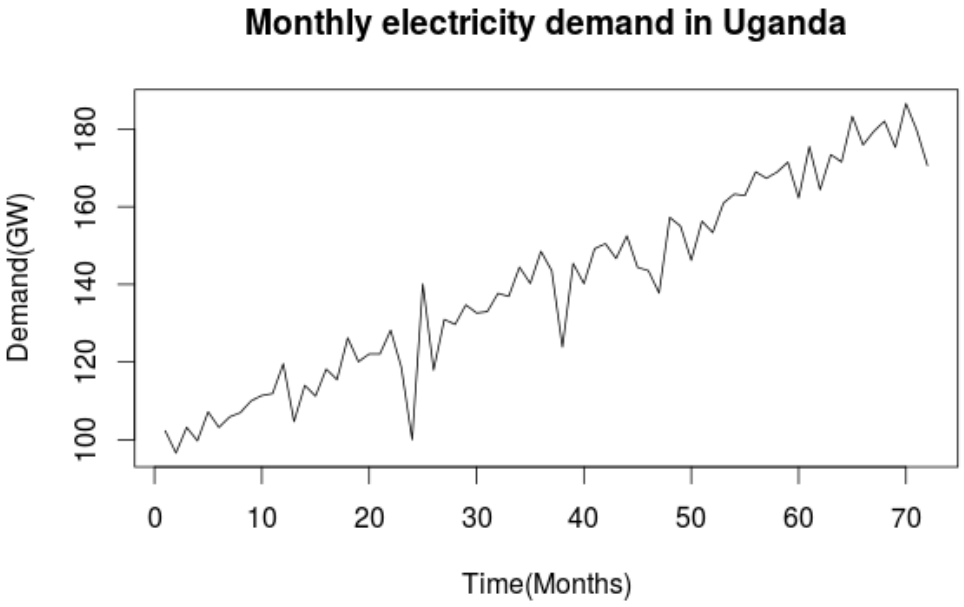

FIGURE 3.2: Time series plot for monthly peak demand (in Gigawatts) for the period 01/01/2008 to 31/12/2013.

Region	High range	Low range
Kampala	27-30	16 - 18
Entebbe	26	16 - 18
Fort Portal	25	12 - 14
Kabale	23	9 - 11
Gulu	29-33	16 - 18
Masindi	28	12 - 13
Jinja	28	14 - 15

TABLE 3.3: Mean daily temperature in Uganda.

From Table 3.3, we see that there are no regions with extreme temperatures that would require use of electricity for either cooling or heating in Uganda. There are two rainy seasons in Uganda; March to May and September to November. This means during those rainy months electricity demand is low because people do not use their fans for cooling. Which then means demand should be higher during other months, for example, January, February, June,. However, this contradicts the pattern seen from the time series plot where demand falls every January. Therefore, any fluctuations seen in the data might be due to changes other than temperature. Likewise, the standard of living in Uganda is not high enough for people to afford cooling and heating appliances in their houses during extreme weathers (if any).

The other main preliminary task is to make a careful assessment of the data to see if there are any obvious errors, outliers, missing observations, smooth changes in structure and turning points or sudden discontinuities. It is of no purpose to focus on forecasting if the data is originally of poor quality. Therefore, it is essential to clean the data before proceeding with the analysis, in case there are unusual components identified (Elliott et al., 2006). In this study, the time series plot shows a big fall in the 24^{th} month. We can not visually conclude that it is an outlier, therefore, we make a box plot in R to help in identifying whether this value is truly an outlier or not. However, this can easily return a biased result because the data is not stationary in the mean. A box plot using first differences of the data is inspected (Perron and Rodríguez, 2003). We consider first differences because stationarity is attained after taking the first difference. Figure 3.3 shows no evidence of outliers in the stationary data.

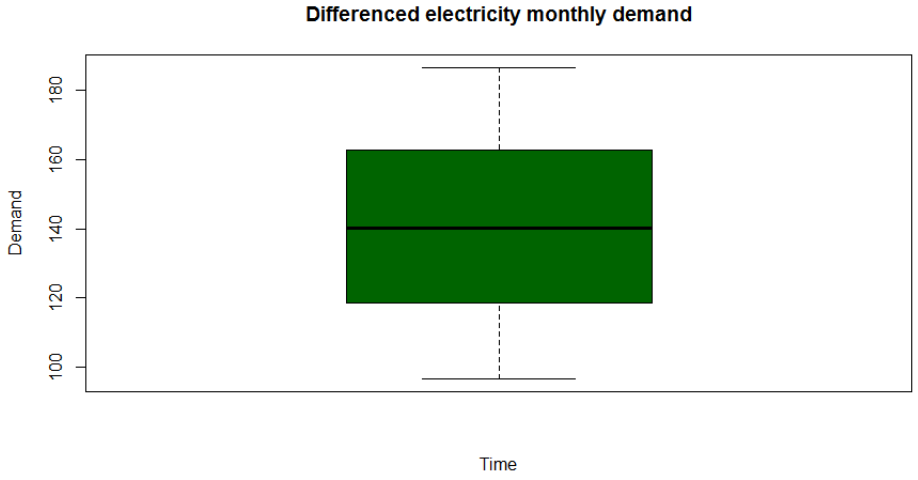

FIGURE 3.3: Box plot showing no outliers in stationary data.

In 2009, the Government of Uganda, through ERA, announced that power tariffs reduce by an average of 8% for all tariff categories. This was attributed to increased energy generation by new entrants and the strengthening of the Ugandan Shilling against the US Dollar (EPRC). The reduction changed the tariff for domestic consumers from USh.426.1/kWh to USh.385.6/kWh. Therefore, there was an increase in demand. However, as the year progressed, there was a decline in the water levels of lake Victoria where Nalubaale and Kiira power stations, the biggest power station in Uganda then, have their reservoirs. Hence affecting the supply of electricity generally, and the demand in return. Lake Victoria water levels hit

their lowest (in 2009) towards the end of the year, as seen in figure 3.4 (Mawejje et al., 2013).

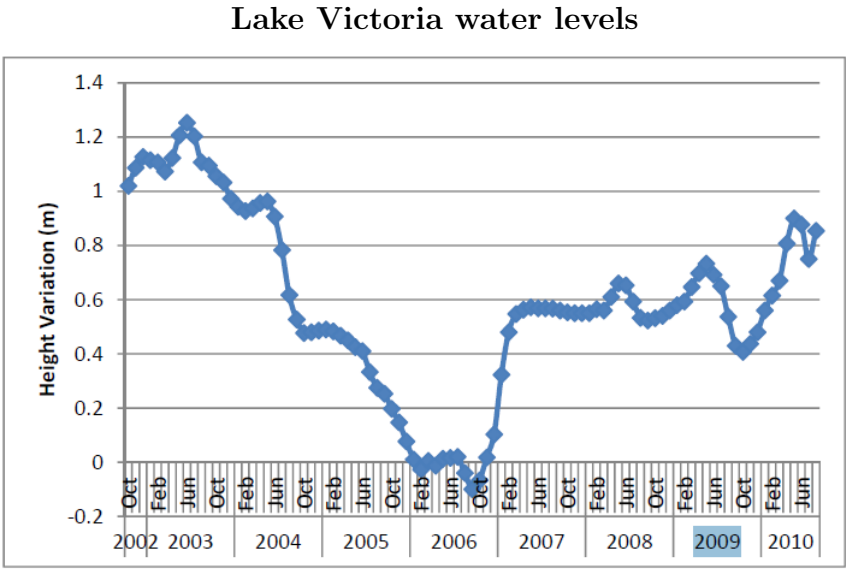

FIGURE 3.4: Water levels of Lake Victoria.

Hence, dropping electricity supply to its lowest in that year, and the demand as well. This is because the other power generating plants could not supply enough electricity to meet the had-been-increasing demand. Since electricity is a non-storable good, supply influences how much is demanded (Ghosh, 2008). Therefore, irrespective of the incentives to demand, supply hindered this demand from being as high as it would have been in case there was no decline in water levels.

The scale on the vertical axis of Figure 3.4 shows the height variations in water levels. The figure displays negative levels for the whole of 2006. This does not indicate that the water levels became negative. It rather shows that the difference in water levels between a current period and one preceding it, was negative. Meaning the water kept decreasing month after month and hit the lowest levels towards the end of the year. It however started rising again between November and December.

Since there is no evidence of outliers in the data, the unusual value at time 24 is replaced with the arithmetic average of its 2 preceding and 2 succeeding values. This is because it looks like an outlier and more likely to distort the general pattern of the data. The new series is shown in Figure 3.5 and it is what we use for all the following analyses in this study.

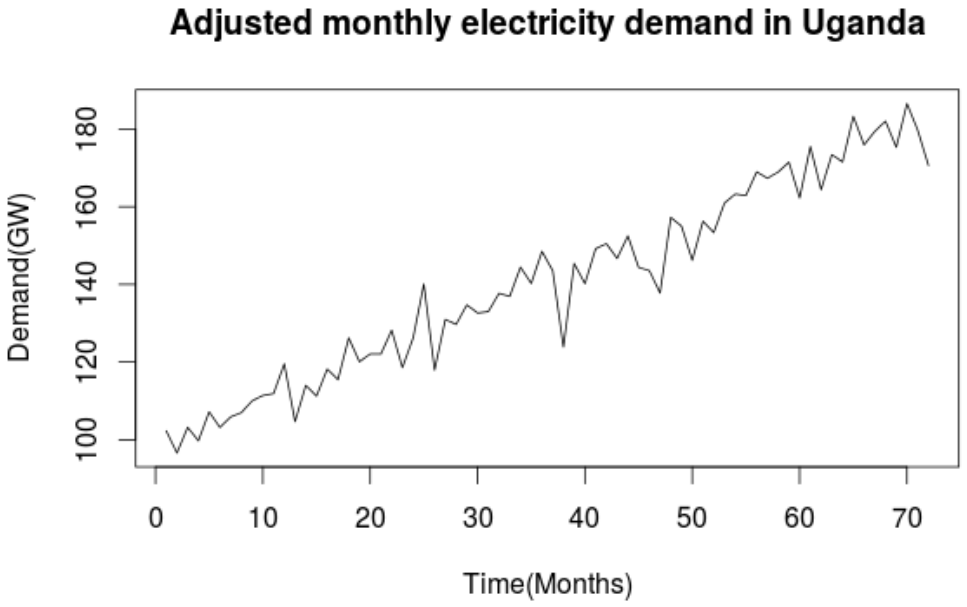

FIGURE 3.5: New series plotted after adjusting the values at time 24.

3.7.2 South Africa daily electricity demand data

The time series plot in Figure 3.7 shows the daily electricity demand for South Africa from 1^{st}-January-2004 to 28^{th}-June-2008. The data is collected from a confidential source who prefers the province and sector where the data is from remain classified. This data originally contains missing values which makes it hard to fit ARIMA models directly without correcting the missing values first. Therefore, certain approaches have to be employed, depending on the size and nature of both the missing values and the overall sample.

3.7.2.1 Missing Data Mechanisms

Most of the discussions in this section are found in Moritz et al. (2015) who study imputation of missing values in univariate time series data.

Most of the commonly used statistical methods depend on complete data. Having missing values in a data set can be a very big hindrance to statistical analysis. Therefore, it is vital to account for the missing values before carrying out further analysis. There are different approaches used to analyse data with missing values, but, commonly analysts apply established statistical methods to impute these

missing values. This approach is called imputation of missing values (Honaker and King, 2010).

Understanding the distribution of the gaps caused by missing values in a given data set is important when dealing with any type of data. It is vital because it can be used as defence for choosing an appropriate imputation method. Categorising imputation methods depends on the approach used and the output got as a result of applying the approach (Little and Rubin, 2014). Some approaches impute a single value and use it to replace the missing value. This has a limitation of neglecting the uncertainty that the new imputed value introduces to the data (Bögl et al., 2015). Other approaches use repeated re-sampling, which makes it possible to calculate the standard error from the variability of estimates. Approaches that use multiples imputation techniques, like, Monte Carlo based simulations, make the computation of estimates and confidence intervals possible, which can be used to communicate the uncertainty of the imputation (Schafer, 1999). Missing data mechanisms are divided into three categories;

- Missing completely at random (MCAR): In this case, there is no systematic mechanism on the way the data is missing. Missing data points occur entirely at random. Since there are no other variables existent for univariate time series data other than time, MCAR in this case means the probability for a certain observation being missing is independent of the point of time of this observation in the series. For other data types, MCAR means; the probability for a certain observation being missing is independent from the values of other variables, and the probability for an observation being missing is also independent of the value of the observation itself. For example, an administrative error causing several test results to be misplaced prior to data entry.

- Missing at random (MAR): This is similar to MCAR, but the information is not missing entirely at random. Here, the propensity for missing data is correlated with other variables related to the study. In MAR, the probability for an observation being missing is also independent of the value of the observation itself, but is dependent on other variables. For example, consider a study for assessing the relationship between drug use and self esteem in high school students. Drug abuse may be associated with frequent absenteeism, leading to a higher probability of missing data on the self esteem measure.

This happens mainly because students using drugs are always absent on the days researchers carry out the self esteem survey. For univariate time series data, since other than time no other variables are given, the probability for an observation being missing is dependent of the point in time of this observation in the series.

- Not missing at random (NMAR): In this case, observations are not missing in a random manner. The data is neither MCAR nor MAR. That means, the probability for an observation missing depends on the value of the observation itself. Furthermore the probability can be dependent on other variables (for other data types) or point of time (for univariate time series). For example, temperature sensor gives no values for temperatures over 100 degrees.

In practice, testing for MCAR can be done using different tests like, the t-test, and Little's test developed by Little (1988). Software like R also have packages like *MissMech*, developed by Jamshidian et al. (2014), whose functions can be used to test if missing data is MCAR. However, for data either MAR or NMAR, checking requires manual analysis of the patterns in the data. Most of the missing data approaches consider either MCAR or MAR because the missing data mechanism is considered ignorable for both cases. (Rubin, 1976).

For univariate time series dealing with missing data is different, this is because there is one variable of the data. Therefore, time has to be treated as another variable to make imputations possible. The other difference is that algorithms used to impute univariate time series data can use characteristics of the series instead of covariates for missing value estimates only, like for other data types.

The missing values in the data available for this study show evidence of MAR. This is because according to the definition of MAR in univariate time series data, the probability that an observation will be missing depends on the point in time of this observation in the series. Most of the variables missing happen to be clustered at different point. That means all the points close to a missing value, in time, have a high probability of missing as well. The missing values are shown by breaking patterns in the time series plot of South Africa electricity demand data seen in figure 3.6, for example, between the 700-800th days, there exists breaks.

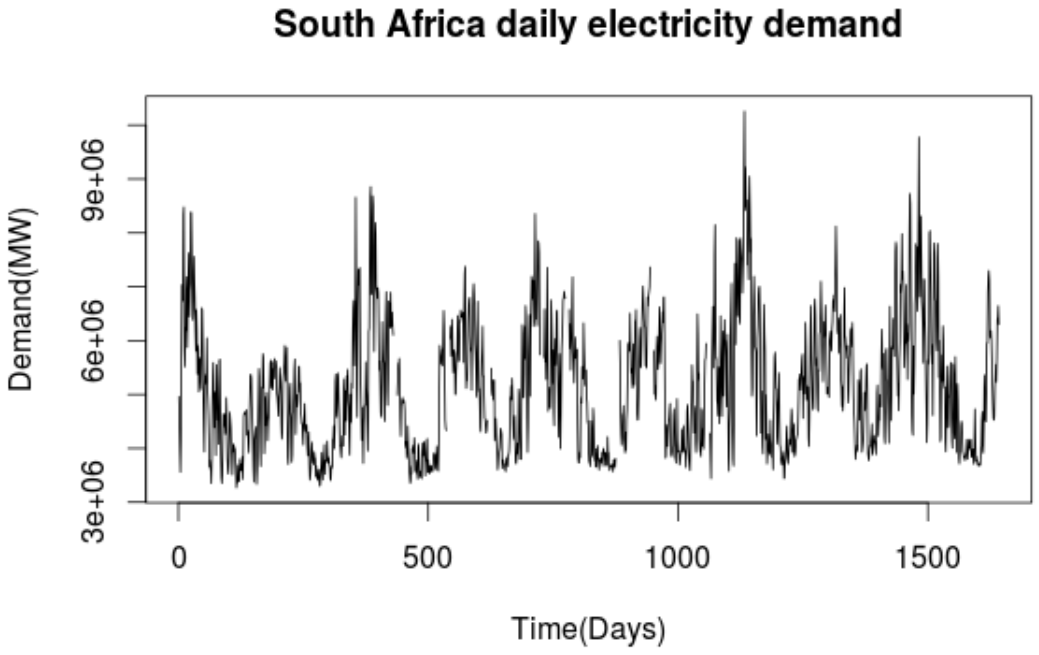

FIGURE 3.6: A time series plot of South Africa's original data with missing values

.

3.7.2.2 Univariate time series imputation

Instead of using covariates like in multivariate datasets, univariate time series use time dependencies to perform an effective imputation. Techniques capable of doing imputation for univariate series can be roughly divided into three:

- Univariate algorithms. These are algorithms employed when imputing missing data in a univariate data set. They however do not consider the time series characteristics of the data. Therefore, can not be used to impute the missing data in this study.
- Univariate time series algorithms. Unlike the univariate algorithms, univariate time series algorithms take into consideration the time series aspect of the data. Examples of such algorithms include; locf (last observation carried forward), nocb (next observation carried backward), arithmetic smoothing and linear interpolation. The more advanced algorithms are based on structural time series models and can handle seasonality.

- Multivariate algorithms on lagged data. These algorithms are not normally used when dealing with univariate data. However, for univariate time series, time is implicitly considered as another variable whose characteristics can then be used as covariates. This can be done using lags (variables that take the value of another variable in the previous time) and leads (variables that take the value of another variable in the next time). This then makes it possible to use multivariate imputation algorithms.

In this study we use linear interpolation to impute missing values. This is because the available data is a time series of univariate nature with no evidence of trend. For the same reason, we do not use multiple imputation, which is a better way of imputing missing data, but works best with multivariate data. Likewise, if the optimal reason for the project is statistical inference, then multiple imputation is greatly advised. Otherwise, for optimal point forecasting (which is the aim of this study), researchers such as Rubin (1996) advise against using multiple imputation.

We use the *na.interp* algorithm in R which was developed by Hyndman et al. (2015). This algorithm uses linear interpolation to replace the missing values for non seasonal data. It can also be applied to seasonal data, where its periodic *stl* decomposition works on the seasonal component of the data. We use a more general function called *tsclean*; a combination of *na.interp* and *tsoutliers* which handles both missing values and outliers (in case there exists any). It returns a cleaned version of a time series with outliers and missing values replaced by estimated values. The new dataset produced is plotted against time in figure 3.7

The new plotted dataset shows no evidence of trend, but a pattern of seasonality. Mainly fluctuations are happening at different points in time. There is one point in time when the demand is slightly higher than all other points but that does not confirm that it is an outlier. Likewise, from the time series plot, 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 deeper analysis of this data will be studied in the next chapters to confirm the stationarity of the data and how to attain it in case the data is not actually stationary.

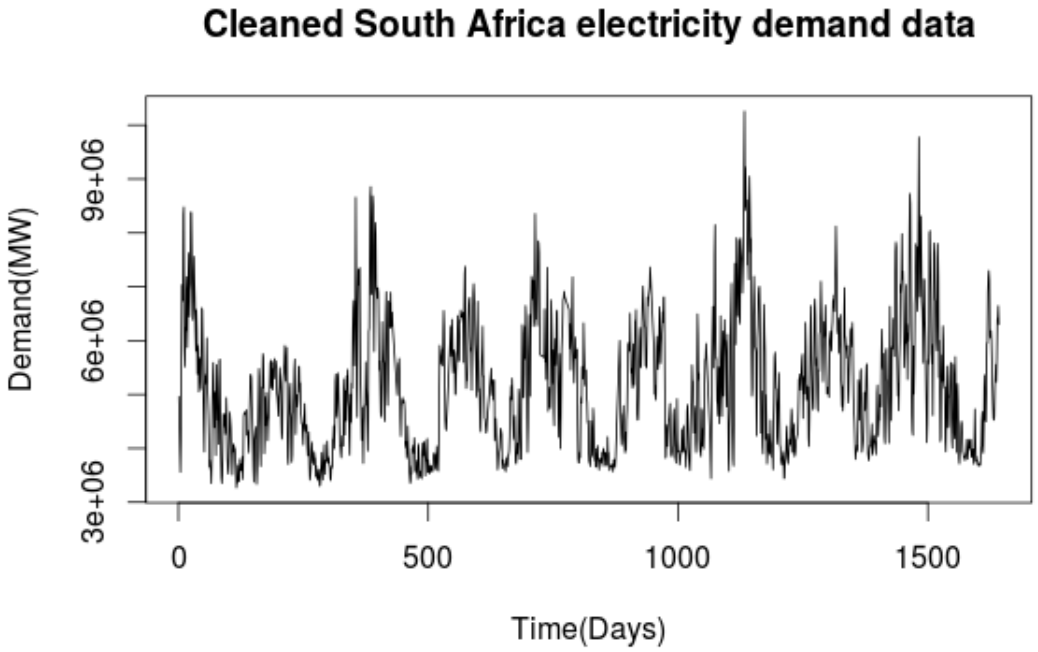

FIGURE 3.7: A time series plot of South Africa's daily electricity demand data

.

Chapter 4

Time series analysis using ARIMA models

A forecasting model is a statistical description of the data generating process from which a forecast may be derived (Ord and Fildes, 2012). It is very important to study different time series models before choosing a method that best suits the forecasting exercise at hand. Scientifically, a forecasting method is an efficiently established systematic process of generating a forecast with certain levels of accuracy, using a series of steps (Hyndman and Athanasopoulos, 2014). In this study, we shall focus on univariate time series models because of the available data. A univariate time series model describes the distribution of a single random variable (X) at time t, in terms of its relationship with past values (Adhikari and Agrawal, 2013).

4.1 ARIMA models

The word ARIMA stands for AutoRegressive(AR) Intergrated(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 can be made stationary by differencing (Karamouz et al., 2012; Nason, 2006). These models are also called Box-Jenkins models because they were developed by Box and Jenkins (1976).

4.1.1 AR models

A series X_t is said to be an autoregressive process of order p, denoted by $AR(p)$, if it can be expressed in the form;

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

In back shift operator notation, we can write the model as;

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

Where;

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

Equation 4.2 is called an **Autoregressive model**. It is similar to a multiple regression model with the value of X at time t linearly depending on a combination of its weighted p past values. The term autoregressive means that it is a regression of the variable of interest against its past values plus an error term e_t at time t. AR models are normally restricted to stationary data (Hyndman and Athanasopoulos, 2014). That is why it is always necessary to check for stationarity of the data before fitting such models. In this model, it is assumed that $E(X_t) = 0$. However, a non zero mean could be added to the model by replacing X_t with $X_t - \mu$, for all t. This would not affect the properties of the model.

After applying the back shift notation operator, equation 4.2 can be used to yield the AR(p) characteristic equation as;

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

It is important to note that an AR(p) process is stationary if and only if the p roots of $\phi(x)$ each exceed 1 in absolute value (Cryer and Kellet).

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

$$\begin{aligned} \gamma_k &= \text{Var}(X_t X_{t-k}) = E(X_t X_{t-k}) - E(X_t)E(X_{t-k}), \\ &= E(X_t X_{t-k}). \end{aligned}$$

Taking expectations of equation 4.4 gives;

$$\begin{aligned} E(X_t X_{t-k}) &= E(\phi_1 X_{t-1} X_{t-k} + \phi_2 X_{t-2} X_{t-k} + \cdots + \phi_p X_{t-p} X_{t-k} + e_t X_{t-k}), \\ \gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}. \end{aligned} \tag{4.5}$$

Dividing through by the process variance γ_0, we get;

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}. \tag{4.6}$$

Equation 4.6 gives a set of Yule-Walker equations, for $k > 0$. For known values of $\phi_1, \phi_2, \cdots, \phi_p$, we can compute the first lag p autocorrelations $\rho_1, \rho_2, \cdots, \rho_p$ (Plasmans, 2006). Values of ρ_k, for $k > p$, can be obtained by using the recursive relation in equation 4.6 (Von Storch and Zwiers, 2001).

4.1.1.1 MLE estimation of AR(1) parameters

The most important step to study the Maximum Likelihood Estimation method is to evaluate the sample joint distributions. These joint distributions form likelihood functions if they are treated as a function of the parameters given by the data. In the case of identically and independently distributed samples, the likelihood function is the product of marginal densities of individual samples (Reid, 2010). However, with time series analysis, the dependence structure of observation is specified. Therefore, it is not advisable to use the product of marginal densities to evaluate the likelihood function. It is important to note that while we assume

that $e_t \sim i.i.dN(0, \sigma^2)$, it does not imply that the data X_t is $i.i.d$ as well. To evaluate the sample likelihood, the conditional density is needed.

A stationary normally distributed AR(1) process takes the form

$$X_t = c + \phi X_{t-1} + e_t. \tag{4.7}$$

We know from the conditions of stationarity of an AR process, that, assuming stationarity means that $|\phi| < 1$. The parameters to be estimated in this model are; $\Theta = (c, \phi, \sigma^2)'$. Consider the probability density function of the first observation in the sample, X_1. This is a random variable with mean and variance respectively,

$$E(X_1) = \frac{c}{1-\phi}. \tag{4.8}$$

$$E(X_1 X_1) = \frac{\sigma^2}{1-\phi^2}. \tag{4.9}$$

Since the error term is assumed to be normally distributed, the data is also normal. That means $X_1 \sim N(\frac{c}{1-\phi}, \frac{\sigma^2}{1-\phi^2})$. Therefore,

$$fX_1(x_1; \Theta) = fX_1(x_1; c, \phi, \sigma^2), \tag{4.10}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma^2}{1-\phi^2}}} \exp\left\{ -\frac{1}{2} \frac{\left(x_1 - \frac{c}{1-\phi}\right)^2}{\frac{\sigma^2}{1-\phi^2}} \right\}. \tag{4.11}$$

Next, we consider the distribution of the second observation X_2 conditional on the first observation $X_1 = x_1$. This means we are treating the random variable X_1 as a deterministic constant x_1. Relating to equation 4.7,

$$X_2 = c + \phi X_1 + e_2. \tag{4.12}$$

From the conditional perspective, equation 4.12 can also be defined as the constant $(c + \phi x_1)$, plus the normally distributed error term e_t. Therefore,

$$(X_2 | X_1 = x_1) \sim N((c + \phi x_1), \sigma^2). \tag{4.13}$$

This means that the probability density function (pdf) of the second observation is;

$$fX_2|X_1(x_2|x_1;\Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x_2 - c - x_1)^2}{\sigma^2}\right\}. \tag{4.14}$$

The joint density of observations X_1 and X_2 is then given by;

$$fX_2, X_1(x_2, x_1;\Theta) = fX_2|X_1(x_2|x_1;\Theta)fX_1(x_1;\Theta). \tag{4.15}$$

Generally, the values of $X_1, X_2, \cdots, X_{t-1}$ matter for X_t only through the value of X_{t-1} (Songsiri et al., 2009). The pdf of the t^{th} observation conditional on the preceding $t-1$ observations is given by;

$$fX_t|X_{t-1}, \cdots, X_1(x_t|x_{t-1}, \cdots, x_1;\Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x_t - c - x_{t-1})^2}{\sigma^2}\right\}. \tag{4.16}$$

The likelihood of the general sample can then be calculated as;

$$fX_T, X_{T-1}, \cdots, X_1(x_T, x_{T-1}, \cdots, x_1;\Theta) = fX_1(x_1;\Theta)\prod_{t=2}^{T} fX_t|X_{t-1}(x_t|x_{t-1};\Theta). \tag{4.17}$$

The product in equation 4.17 starts counting from the second period ($t = 2$) because the first observation is not included in the general multiplication. In equation 4.11, we saw the pdf of the first observation which has a different formulation from the rest of the observations' pdfs. Therefore the generalisation in the overall pdf excludes it. Taking logs of equation 4.17 gives the log likelihood function as;

$$L(\Theta) = \log fX_1(x_1;\Theta) + \sum_{t=2}^{T} fX_t|X_{t-1}(x_t|x_{t-1};\Theta). \tag{4.18}$$

Using the pdfs for the respective observations we get;

$$L(\Theta) = -\frac{1}{2}\left[\log(2\pi) + \log\left(\frac{\sigma^2}{1-\phi^2}\right) + \frac{\left(x_1 - \frac{c}{1-\phi}\right)^2}{\frac{\sigma^2}{1-\phi^2}}\right] - \frac{T-1}{2}\left[\log(2\pi) + \log(\sigma^2)\right] - \sum_{t=2}^{T}\frac{e_t^2}{2\sigma^2}. \tag{4.19}$$

where $e_t = x_t - c - x_{t-1}$ Using equation 4.19, we can either calculate the exact or conditional MLEs. For conditional MLEs, we let the value of the first observation become deterministic ($fX_1(x_1) = 1$) and maximize the likelihood conditioned on the first observation (Kirk and Stumpf, 2009). Therefore, our objective changes to maximising

$$L(\Theta) = -\frac{T-1}{2}\left[\log(2\pi) + \log(\sigma^2)\right] - \sum_{t=2}^{T}\frac{e_t^2}{2\sigma^2}. \tag{4.20}$$

Maximising equation 4.20 with respect to c and ϕ is equivalent to minimising the autoregression vector (Zivot and Wang, 2007)

$$(y - X\beta)'(y - X\beta) = \sum_{t=2}^{T}(x_t - c - x_{t-1})^2, \tag{4.21}$$

which is achieved by an ordinary least square (OLS) regression of x_t on a constant and its own lagged value (Eden, 2008), where

$$y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_T \end{bmatrix}, \quad \beta = \begin{bmatrix} c \\ \phi \end{bmatrix}. \tag{4.22}$$

The conditional maximum likelihood estimates of c and ϕ are therefore given by;

$$\begin{bmatrix} \hat{c} \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} T-1 & \sum_{t=2}^{T} x_{t-1} \\ \sum_{t=2}^{T} x_{t-1} & \sum_{t=2}^{T} x_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=2}^{T} x_{t-1} \\ \sum_{t=2}^{T} x_{t-1}x_t \end{bmatrix}. \tag{4.23}$$

The conditional maximum likelihood estimator of σ^2 is found by setting

$$\frac{\partial L(\Theta)}{\partial \sigma^2} = -\frac{T-1}{2\sigma^2} + \sum_{t=2}^{T} \frac{(x_t - c - x_{t-1})^2}{2\sigma^4} = 0. \tag{4.24}$$

4.1.2 MA model

A series with a white noise process of mean 0 and variance σ_e^2 is said to be a moving average process of order q, denoted as $MA(q)$, if it can be expressed as a weighted linear sum of the past forecast errors.

$$X_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}. \tag{4.25}$$

In back shift operator notation, we can write the model as;

$$X_t = e_t(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q). \tag{4.26}$$

Where;

- B is the back shift operator notation.
- $\theta_0, \theta_1, \cdots, \theta_q$ are coefficients of the lagged error terms. θ_0 is usually equated to 1 (Broersen, 2006; Hipel and McLeod, 1994).
- e_t is a normally distributed white noise with mean 0 and variance σ_e^2.

Equation 4.26 is called a **Moving Average model**. Some authors note the parameters of an MA process as negatives, in order to have characteristic operators of the same signs for both AR and MA processes. However, this has no significant change to the interpretation of the model (Chatfield, 2002).

The autocovariance functions of an MA(q) model is defined as $\gamma_k = \text{Cov}(X_t, X_{t-k})$, where it becomes the variance of the process if $k = 0$. Therefore, from the definition of the model,

$$\gamma_0 = \sigma_e^2(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2). \tag{4.27}$$

The general autocovarinace function of the model is given by;

$$\gamma_k = \begin{cases} \sigma_e^2 \sum_{i=0}^{q} \theta_i^2 & \text{for } k = 0 \\ \sigma_e^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k} & \text{for } k = 1, 2, \cdots, q \\ 0 & \text{for } k > q \end{cases}. \tag{4.28}$$

Dividing equation 4.28 by γ_0, we obtain the autocorrelation function (Hipel and McLeod, 1994);

$$\gamma_k = \begin{cases} 1 & \text{for } k = 0 \\ \dfrac{\sigma_e^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{\sigma_e^2 \sum_{i=0}^{q} \theta_i^2} & \text{for } k = 1, 2, \cdots, q \\ 0 & \text{for } k > q \end{cases}. \tag{4.29}$$

An MA model is said to be invertible if it is algebraically equivalent to a converging infinite order AR model (Rao, 2008). Convergence means that the coefficients of the AR model decrease to 0 as we move back in time. Invertibility is a restriction programmed into time series software used to estimate the coefficients of models with MA terms. It is some what unusual to check for invertibility while carrying out a data analysis exercise (Oduro-Gyimah, 2011). From the definition, we see that stationary AR models are automatically invertible (since they contain no MA terms). However, the condition is not obvious for MA models, not all stationary MA models are invertible (Bartlett). Taking an example of an MA model with $q = 1$, equation 4.26 becomes;

$$X_t = (1 + \theta B) e_t, \tag{4.30}$$

which can also be written as;

$$e_t = (1 + \theta B)^{-1} X_t. \tag{4.31}$$

Obtaining the binomial expansion of $(1+\theta B)^{-1}$ gives

$$e_t = (1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \cdots)X_t, \tag{4.32}$$

$$= \sum_{i=0}^{\infty} (-\theta)^i X_{t-i}. \tag{4.33}$$

Equation 4.33 shows that e_t can be expressed as a causal function of X_t. This shows that the MA(1) model has been expressed as an infinite-order AR model $AR(\infty)$, which converges if $|\theta| < 1$. Therefore, we can say that this $MA(1)$ process is invertible if and only if $|\theta| < 1$.

Generally, an MA(q) process is inverible if and only 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 the unit circle.

4.1.2.1 MLE estimation of MA(1) parameters

Assume a normally distributed invertible ($|\theta| < 1$) MA(1) model

$$X_t = c + e_t + \theta e_{t-1}, \tag{4.34}$$

where $e_t \sim i.i.dN(0, \sigma^2)$. The aims is to estimate parameters $\Theta = (c, \theta, \sigma^2)'$ using MLE. We consider the pdf of the first observation of the sample

$$X_1 = c + e_1 + \theta e_0, \tag{4.35}$$

whose mean and variance are defined as

$$E(X_1) = c, \tag{4.36}$$

$$E(X_1 X_1) = \sigma^2(1 + \theta^2). \tag{4.37}$$

The normality definition of e_t, where t takes on all possible values, makes us define X_1 as normal too. Therefore, $X_1 \sim N(c, \sigma^2(1+\theta^2))$. The pdf of the first

observation then changes to

$$fX_1(x_1;\Theta) = fX_1(x_1;c,\theta,\sigma^2), \tag{4.38}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2(1+\theta^2)}} \exp\left\{-\frac{1}{2}\cdot\frac{(x_1-c)^2}{\sigma^2(1+\theta^2}\right\}, \tag{4.39}$$

The next observation X_2 has a distribution conditional on the first observation $X_1 = x_1$.

$$X_2 = c + e_2 + \theta e_1. \tag{4.40}$$

Following the method of calculating the joint density of the complete sample in part 4.1.1.1, conditional on $X_1 = x_1$ means treating the random variable X_1 as the deterministic constant x_1. Therefore, equation 4.40 defines X_2 as a constant $(c+\theta e_1)$ plus the normally distributed error term e_2. However, since we are dealing with an MA model which focuses on error terms not observations, this method of calculation might not apply directly. This is because observing $X_1 = x_1$ gives no information on the realization of e_1 because we can not distinguish e_1 from e_0 even after the first observation on x_1 (C. Lee, 2006).

In order to make the information of the observation on $X_1 = x_1$ useful, C. Lee (2006) suggests imposing an additional assumption such that we know with certainty that $e_0 = 0$. This means

$$(X_1|e_0 = 0) \sim N(c,\sigma^2). \tag{4.41}$$

This means the pdf of the first observation given e_0 is;

$$fX_1|e_0 = 0(x_1|e_0 = 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\cdot\frac{(x_1-c)^2}{\sigma^2}\right\}, \tag{4.42}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{e_1^2}{2\sigma^2}\right\}. \tag{4.43}$$

Given the observation x_1, the value of e_1 is then known with certainty as well. That means

$$(X_2|X_1 = x_1, e_0 = 0) \sim N(c+\theta e_1\sigma^2). \tag{4.44}$$

Hence;

$$fX_2|X_1, e_0 = 0(x_2|x_1, e_0 = 0; \Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \cdot \frac{(x_2 - c - \theta e_1)^2}{\sigma^2}\right\}, \quad (4.45)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{e_2^2}{2\sigma^2}\right\}. \quad (4.46)$$

Continuing in this same manner, it implies that given condition that $e_0 = 0$, the full sequence $\{e_1, e_2, \cdots, e_T\}$ can be calculated from $\{x_1, x_2, \cdots, x_T\}$ by iterating on

$$e_t = x_t - c - \theta e_{t-1} \text{ for } \ t = 1, 2, \cdots, T$$

. The general conditional pdf can then be calculated as;

$$fX_t|X_{t-1}, \cdots, X_1, e_0 = 0(x_t|x_{t-1}, \cdots, x_1, e_0 = 0; \Theta) = fX_t|e_{t-1}(x_t|e_{t-1}; \Theta), \quad (4.47)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{e_t^2}{2\sigma^2}\right\}. \quad (4.48)$$

The likelihood function of the complete sample conditional on $e_0 = 0$ can thus be calculated as the product of these individual densities (Hurlin, 2013)

$$fX_T, \cdots, X_1|e_0 = 0(x_T, \cdots, x_1|e_0 = 0; \Theta) \quad (4.49)$$

$$= fX_1|e_0 = 0(x_1|e_0 = 0; \Theta) \cdot \prod_{t=2}^{T} fX_t|X_{t-1}, \cdots, X_1, e_0 = 0(x_t|x_{t-1}, \cdots, x_1; \Theta).$$

The conditional log likelihood function is therefore defined by

$$\ln L(\Theta) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=2}^{T} \frac{e_t^2}{2\sigma^2}. \quad (4.50)$$

The log likelihood function is a fairly complicated non linear function of c and θ. Therefore, an analytical expression for the MLE of c and θ is not readily calculated. Hence even the conditional MLE for an MA(1) process must be found by numerical optimization (C. Lee, 2006). For MA models, estimating the process by conditional MLE requires the use of all the T observations in the sample since the estimation is conditional on $e_0 = 0$ and not on the first observation X_1. More studies about

approximations and conditions made when estimating MA(1) parameters have been covered, for example, in Shephard (1993) and Davis and Dunsmuir (1996).

4.1.3 ARMA model

Combining both the AR(p) and MA(q) models gives rise to an **Autoregressive Moving Average model** (ARMA(p,q)) which is expressed as;

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}. \tag{4.51}$$

Re-arranging the model gives

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}. \tag{4.52}$$

Using the back shift operator,

$$X_t(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) = e_t(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q). \tag{4.53}$$

This can be simplified to;

$$\phi(B) X_t = \theta(B) e_t, \tag{4.54}$$

where;

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p), \tag{4.55}$$

$$\theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q). \tag{4.56}$$

Both the AR(p) and MA(q) are special cases of the ARMA model. An ARMA(p,0) process is the same as an AR(p) process and an ARMA(0,q) process is the same as an MA(q) process. If the data available is stationary, it is better modeled using an ARMA(p,q) model than AR(p) or MA(q) models individually (Chin and Fan, 2005). This is because an ARMA(p,q) in such a case uses fewer parameter than

the individual models and gives a better representation of the data. This is called the principle of parsimony (Singh, 2002; Woodward et al., 2011).

For an ARMA(p,q) process to be stationary, the absolute value of the roots of all the AR(p) characteristic polynomials should be greater than 1. For invertibility, absolute values of the roots of all the MA(q) characteristic polynomials should be greater than 1. For example, given a model below;

$$\begin{aligned} X_t &= 0.5X_{t-1} + e_t - 0.3e_{t-1} + 1.2e_{t-2} \\ X_t &= 0.5BX_t + e_t - 0.3Be_t + 1.2B^2e_t \\ X_t(1-0.5B) &= e_t(1-0.3B+1.2B^2) \end{aligned}$$

From the left hand side, the AR characteristic polynomial is $\phi(x) = 1-0.5x$, with root $x = 2$ which is greater than 1. Therefore, the model is stationary. The right hand side gives the MA characteristic polynomial, $\theta(x) = 1 - 0.3x + 1.2x^2$ which gives complex roots $|x| = 0.125 \pm 0.9.4i \implies |x| = \sqrt{(0.125)^2 + (0.9.4)^2} = 0.913$ which is less than 1, hence the process is not invertible. The invertibility and stationarity conditions of any proces are independent of each other. For example, a pure, finite order AR process is always invertible but not necessarily stationary, while a pure, finite order MA process is always stationary, even if it is not invertible (Chatfield, 2002).

4.1.3.1 MLE estimation of ARMA(p,q) parameters

Let us assume a stationary and invertible normal ARMA(p,q) process with an error term $e_t \sim N(0, \sigma^2)$. The parameters to be estimated in this case are $\Theta = (c, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma^2)'$. The approximation to the likelihood function for an AR model is conditional on initial values of the $x's$. The approximation to the likelihood function for an MA model is conditional on initial values of the error terms $e's$. Therefore, a common approximation to the likelihood function of an ARMA process is conditional on initial values of both the process and error terms (Hamilton, 1994).

The $(p+1)^{th}$ observation of an ARMA(p,q) process can be given by;

$$X_{p+1} = c + \phi_1 X_p + \phi_2 X_{p-1} + \cdots + \phi_p X_1 + e_{p+1} + \theta_1 e_p + \cdots + \theta_q e_{p-q+1}. \quad (4.57)$$

There are different schools of thought when dealing with this type of parameter estimation. Among all, Box et al. (2011) suggested that all random variables $X's$ equal to their actual observations and error terms equal to their individual expectation, which is 0. Therefore, the distribution of the observation X_{p+1} becomes;

$$X_{p+1} \sim N((c + \phi_1 X_p + \phi_2 X_{p-1} + \cdots + \phi_p X_1), \sigma^2). \tag{4.58}$$

Then the conditional log likelihood function calculated from $t = (p+1), \cdots, T$ is

$$\log L(\Theta) = \log f(x_T, x_{T-1}, \cdots, x_{p+1} | x_p, \cdots, x_1, e_p = e_{p-1} = \cdots = e_{p-q+1} = 0; \Theta), \tag{4.59}$$

$$= -\frac{T-p}{2}\log(2\pi) - \frac{T-p}{2}\log(\sigma^2) - \sum_{t=p+1}^{T} \frac{e_t^2}{2\sigma^2}. \tag{4.60}$$

The sequence $\{e_{p+1}, e_{p+2}, \cdots, e_T\}$ can be calculated from $\{x_1, x_2, \cdots, x_T\}$ by iterating on (Hamilton, 1994)

$$e_t = X_t - c - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}. \tag{4.61}$$

Solutions to the conditional log likelihood function in equation 4.60 can be calculated using the method of conditional sum of squares estimation as studied by Robinson (2006).

4.1.4 ARIMA model

In reality, most time series data is not stationary because of seasonality and trend. Therefore, one can neither apply the AR, MA, nor ARMA models directly. The most suitable method of obtaining stationarity when dealing with ARIMA models is differencing (Systematics, 1994). Generally, time series data can be differenced 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 $(1 - B)X_t$. Therefore, by convention, differencing d times can be written as; $(1-B)^d X_t$. If the original data series is differenced d times before fitting an ARMA(p,q) model, then the model for the original undifferenced series is said to be an **Autoregressive Integrated Moving Average model** (ARIMA(p,d,q)), where d represents the number of

times the data has been differenced. Taking first differences normally removes a linear deterministic trend (Hendry, 1995).

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 equation 4.53 to;

$$X_t(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)(1-B)^d = e_t(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q), \tag{4.62}$$

which can be simplified to;

$$\phi(B)(1-B)^d X_t = \theta(B)e_t. \tag{4.63}$$

Equation 4.63 is called an **ARIMA model** with the term $\phi(B)$ corresponding to the AR characteristic polynomial of order p, $(1-B)^d$ for the integrated part of order d, and $\theta(B)$ for the MA characteristic polynomial of order q. All the models discussed in section 4.1 are a special type of ARIMA models. For example, white noise-ARIMA(0,0,0), random walk-ARIMA(0,1,0) autoregression-ARIMA(p,0,0), moving average-ARIMA(0,0,q) and autoregressive moving average-ARIMA(p,0,q).

4.1.5 SARIMA model

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 et al., 2006). SARIMA models include an additional seasonal term as indicated;

$$ARIMA(p,d,q)(P,D,Q)_s, \tag{4.64}$$

where s denotes the number of periods per season. The uppercase notation in equation 4.64 is for the seasonal part and the lowercase 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 backshifts of the seasonal period. For example, if a seasonal component is added

to equation 4.63, the resultant model will be;

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^s)e_t. \tag{4.65}$$

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

$$(1-B)(1-B^{12})X_t = (1+\theta B)(1+\Theta B^{12})e_t. \tag{4.66}$$

4.2 Model specification

Statistical model building usually has three main stages, namely: Model specification (or model identification), model fitting (or model estimation) and model checking (or model verification). Bad choices of orders p, d, and q lead to bad models, which in turn lead to bad forecasts of future values (Okyere and Nanga, 2014). It is therefore essential to make sure that the choices made are consistent with the underlying structure of the observed data.

For any series of data, a clear indicator of non stationarity is that the ACF exhibits a very slow decay across lags. This occurs because in a non stationary process, the series tends to "hang together" and displays trends. If the data is non stationary seasonally, then the ACF displays clusters of either positive and/or negative autocorrelation. However, there are other common methods of determining non stationarity, for example, the ADF and/or KPSS test and using a time series plot. When there is a clear linear trend in the data and the ACF for the series decays very slowly, it is usually advisable to take first differences (Ord and Fildes, 2012).

If the ACF of first differenced data resembles that of a stationary ARMA process (decays quickly), then d in ARIMA(p,d,q) is taken to be 1. The ACF and PACF of first differenced data can then be used to identify plausible values of p and q. Otherwise, second differences are taken and $d = 2$ is used instead. Then the plotted ACF and PACF at that point is used to identify plausible values of p and q. The order of differencing can take on any value until stationarity is attained. However, in real data analysis, there is rarely a need to consider values of $d > 2$ (Chatfield, 2002).

For MA(q) models, the ACF is usually non zero for lags $k < q$ and is equal to 0 for lags $k > q$. That means, the ACF for an MA(q) process "drops off" to zero after lag q. Therefore, the ACF provides a considerable amount of information about the order when the process is truly a moving average. On the other hand, if the process is AR(p), then the ACF may not tell us much about the order. It is therefore advisable to use the PACF to determine the order of an AR(p) process.

Once the model order has been identified, then the parameters in the model can be estimated. This can be done using different software like R-Studio which estimates the ARIMA model using MLE (Hevia, 2008). This technique finds the values of the parameters which maximize the probability of obtaining the data that has been observed. For ARIMA models, MLE is very similar to the least square estimates that would be obtained by minimising squared errors. In practice, R reports the value of the log likelihood of the data, which is the logarithm of the probability of the observed data coming from the estimated model. For given values of p, d and q, R tries to maximize the log likelihood when finding parameter estimates (Hyndman and Athanasopoulos, 2014).

When fitting SARIMA models, one must choose suitable values for the two orders of differencing, both seasonal (D) and non-seasonal(d) first, so as to make the series stationary and remove (most of) the seasonality. Then an ARMA-type model is fitted to the differenced series with the added complication that there may be AR and MA terms at lags which are a multiple of the season length s.

Chapter 5

Applying ARIMA models to electricity demand data

In this chapter, we follow a general procedure to help us fit an ARIMA model to the available time series data.

5.1 Uganda monthly electricity demand data

It is clear from the time series plot in Figure 3.2 that there are variations in the amount of electricity demanded per month. There also is evidence of a reduction in demand towards the end of every year, and a rise every beginning of a new year. This time series can be described using an additive model because the random fluctuations seem to be roughly constant in size over time (Hyndman and Athanasopoulos, 2014). That means that the data has a stable variance, thus no need to transform it.

We plot the ACF of the data for visual inspection of stationarity. The ACF for a non stationary series shows large autocorrelations that diminish only very slowly at large lags (Montgomery et al., 2015). From Figure 5.1 the ACF decays off at a very slow rate, hence indicating non stationarity of the data. This means there is need to take the first difference of the data in order to make it stationary.

Figure 3.2 also shows some evidence of seasonality, because of the repeated drop in demand at the end of every year and increase at the beginning of the following year. However, the ACF in figure 5.1 displays neither positive nor negative

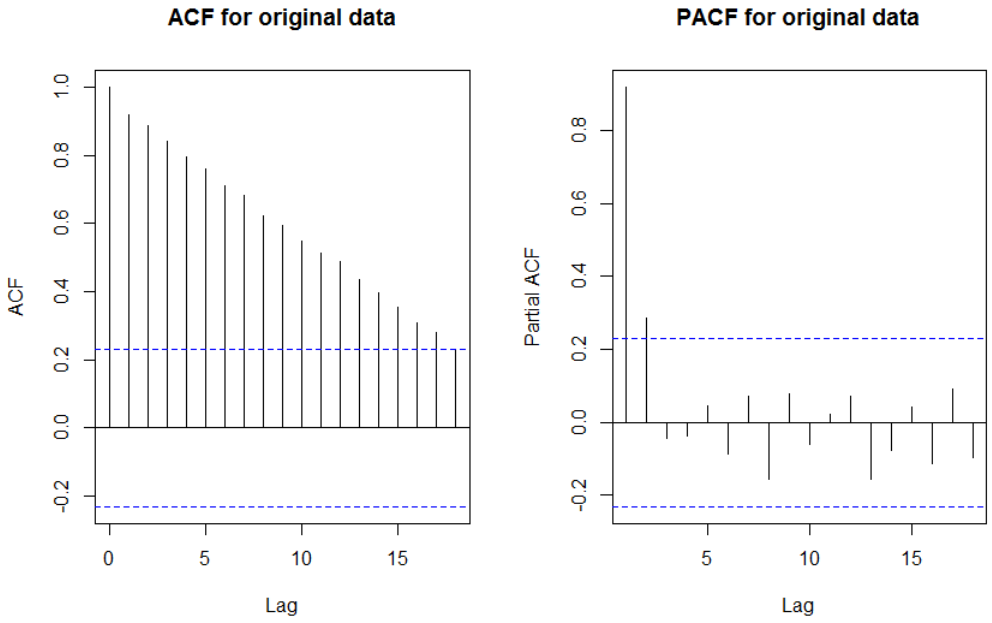

FIGURE 5.1: ACF and PACF of non stationary data.

clusters of autocorrelations. Therefore, seasonality might be present but not significant enough to be considered while modelling. For certainty, we run a monthly seasonality check in R using the *tbats* function, which returns positive monthly seasonality in the data. This is shown in Figure 5.2

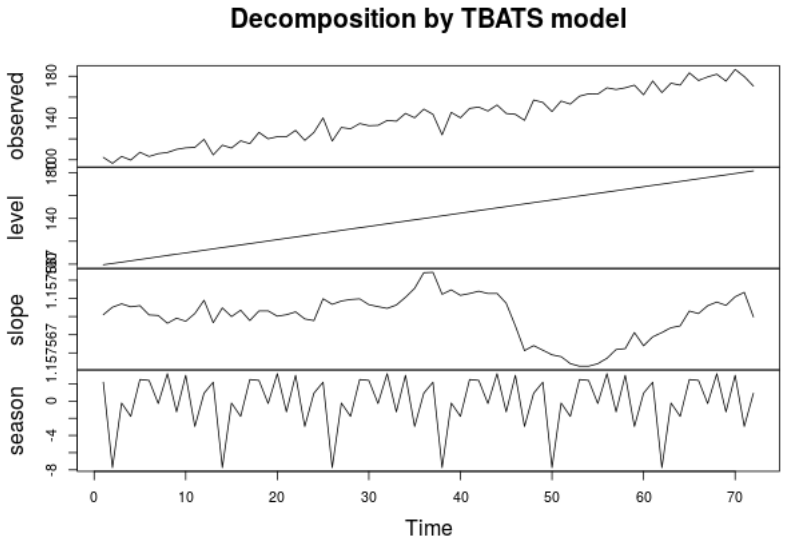

FIGURE 5.2: Monthly seasonality in Ugandan data

With a positive confirmation of monthly seasonality in the data, we apply seasonal differencing in order to make the data stationary. At first glance of figure 5.3,

the data now looks stationary because it displays a constant mean over time. However, we run a unit-root test for stationarity to be sure that the data is now stationary. Different tests can be used to test for stationarity of data, for example, the Kwiatkowski Phillips Schmidt Shin (KPSS) test, Augmented Dickey Fuller (ADF) test.

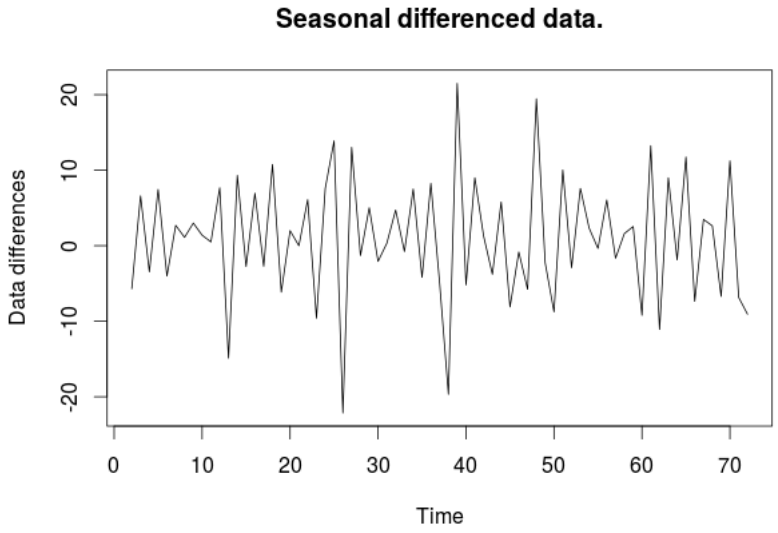

FIGURE 5.3: Time series plot of the data after taking the seasonal difference.

5.1.1 Testing for stationarity using the ADF test

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 as studied in subsection 4.1.1 shows 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 $\phi = 1$.

Consider the AR(1) model

$$X_t = \phi X_{t-1} + e_t. \tag{5.1}$$

The unit root null hypothesis against the stationary alternative corresponds to

$$H_0 : \phi \geqslant 1$$
$$H_1 : \phi < 1.$$

Alternatively, the model can be formulated as (Nielsen, 2006);

$$\Delta X_t = (\phi - 1)X_{t-1} + e_t, \tag{5.2}$$
$$= \pi X_{t-1} + e_t. \tag{5.3}$$

where $\pi = \phi - 1 = \phi(1)$. The unit root hypotheses therefore change to: $H_0 : \pi = 0$ against $H_1 : \pi < 0$. The DF test is simply the standard t-statistic calculated as;

$$\hat{t}_n = \frac{1 - \hat{\phi}}{\sqrt{\hat{\sigma}^2 \left(\sum_{t=2}^{n} X_{t-1}^2 \right)^{-1}}}, \tag{5.4}$$

where $\hat{\phi}$ and $\hat{\sigma}^2$ are estimators for ϕ and the variance σ^2 of e_t respectively. For increasing n the statistic calculated in equation 5.4 does not converge to a standard normal distribution but instead to the distribution of a function following a Wiener process (da Silva Lopes, 2006). All critical values considered to draw conclusions in the DF distribution are one sided (Nielsen, 2006). The null hypothesis H_0 of a unit root is rejected when the test statistic in equation 5.4 is smaller than the critical value.

The disadvantage of using the DF test is that the normal test significance level (usually 5%) is not reliable when the error terms e_t in 5.3 are autocorrelated. The larger the autocorrelation of e_t, the more distorted the significance of the test becomes (Jürgen et al., 2011). These autocorrelations can not be ignored because they might affect the decision of the test. H_0 can be rejected at a low significance level say, 5%, when in reality the significance level lies at, for example, 30%. To solve the problem of negatives related to their test, Dickey and Fuller suggest another test which contains lagged differences.

This is called the Augmented Dickey Fuller (ADF) Test, which tests larger and more complicated sets of time series models by removing all the autocorrelation in the time series, then using the same procedure as the DF (Dwyer et al., 2012).

The ADF statistic used in the test is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence. The rejection criterion for this test is similar to that of a simple DF test. Jürgen et al. (2011) simulated a process in which they correlated errors e_t through a relationship

$$e_t = \beta\varepsilon_{t-1} + \varepsilon_t, \tag{5.5}$$

where ε_t are i.i.d. $(0, \sigma^2)$. This is the same as an MA model of order 1, whose variance, covariance and ACF can be calculated from subsection 4.1.2. Detailed examples with simulations about the theoretical application and calculation of the ADF can be found in both Nielsen (2006) and Jürgen et al. (2011).

5.1.2 Testing for stationarity using the KPSS test

The KPSS test is used to test a null hypothesis that an observable time series is stationary around a deterministic trend. Such models were proposed in 1982 by Bhargava (1986), where several sample tests for unit roots were developed. Later, Kwiatkowski et al. (1992) proposed a test of the null hypothesis that an observable series is stationary around a deterministic trend. The series is expressed as the sum of deterministic trend, random walk, and stationary error. The test is a Lagrange Multiplier (LM) test of the hypothesis that the random walk has zero variance. By testing both the unit root hypothesis and the stationarity hypothesis, one can distinguish series that appear to be stationary, series that appear to have a unit root, and series for which the data are not sufficiently informative to be sure whether they are stationary or integrated (Nusair et al., 2003).

Results from the KPSS test are intended to be complementary to those derived from unit root tests. However, the KPSS test uses a null hypothesis of stationarity and an alternative of a unit root. The test may be conducted under the null of either trend stationarity or level stationarity. For trend stationarity, the regression model with a time trend takes the form

$$X_t = c + \mu t + k\sum_{i=1}^{t}\varepsilon_i + \eta_t, \tag{5.6}$$

where η_t are assumed to be stationary and $\varepsilon_t \sim i.i.dN(0,1)$. If $k = 0$, equation 5.6 becomes trend stationary, otherwise ($k \neq 0$), the process is integrated (Shin and Schmidt, 1992). Therefore, the null and alternative hypotheses are defined as;

$$H_0 : k = 0$$
$$H_1 : k \neq 0.$$

Using an estimation method, for example least square estimation, it is possible to obtain residuals $\hat{\eta}_t$ from equation 5.6. These residuals can then be used to calculate a partial sum

$$S_t = \sum_{i=1}^{t} \hat{\eta}_i,$$

which increases linearly with t. The KPSS test statistic is then calculated as;

$$KPSS_T = \frac{\sum_{t=1}^{n} S_t^2}{n^2 \hat{\omega}_T^2}, \tag{5.7}$$

where (Jürgen et al., 2011)

$$\hat{\omega}_T^2 = \hat{\sigma}_\eta^2 + 2 \sum_{\tau=1}^{T} \left(1 - \frac{\tau}{T-1}\right) \hat{\gamma}_\tau, \tag{5.8}$$

is an estimator of the spectral density at a frequency of zero, with $\hat{\sigma}_\eta^2$ as the variance estimator of η_t and

$$\hat{\gamma}_\tau = \frac{1}{n} \sum_{t=\tau+1}^{n} \hat{\eta}_t \hat{\eta}_{t-\tau}$$

is the covariance estimator.

The main problem with a KPSS test is to determine the reference point T. T must be chosen with a very fragile procedure because, if T is very small, the test will become biased especially if there is evidence of autocorrelation. Otherwise, a very large T also makes the test lose its power (Cappuccio and Lubian, 2010).

If we assume there is no trend, the regression model formed takes the form;

$$X_t = \varepsilon_t + e_t, \tag{5.9}$$

where e_t is assumed to be stationary and ε_t is a random walk of the form

$$\varepsilon_t = \varepsilon_{t-1} + v_t; \quad v_t \sim i.i.d(0, \sigma_v^2). \tag{5.10}$$

If the variance σ_v^2 is zero, then $\varepsilon_t = \varepsilon_0$ for all t and X_t is stationary. Therefore, in this case, the null and alternative hypotheses will respectively be;

$$H_0 : \sigma_v^2 = 0$$
$$H_1 : \sigma_v^2 > 0.$$

The KPSS test statistic is given by

$$KPSS = \frac{1}{T^2} \cdot \frac{\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_\infty^2}, \tag{5.11}$$

where $\hat{\sigma}_\infty^2$ is a heteroscedasticity and autocorrelation consistent (HAC) estimator of the variance of $\hat{e}_t$, and S_t is the partial sum defined by

$$S_t = \sum_{s=1}^{t} \hat{e}_s. \tag{5.12}$$

The estimate $\hat{e}_t$ can be calculated using a simple regression, for example (Nielsen, 2006),

$$X_t = \hat{\mu} + \hat{e}_t, \tag{5.13}$$

assuming the null that $\hat{e}_t$ is stationary.

In this study, we only use the KPSS function in R to test for stationarity of the data after taking the first seasonal difference. We use a significance level of 5% to make a decision. That means a p-value less than 0.05 suggests that further differencing is required. In this case, the KPSS p-value is 0.1, which is greater than 0.05. Therefore, we accept the null hypothesis of stationarity.

5.1.3 Testing for normality

We check for normality of the data because the methods we are applying in the study, for example, ARIMA models, assume normality of the data. In their study, Ahad et al. (2011) tested the sensitivity of normality tests when given non normal data. Their results showed that the Shapiro-Wilk test is the best normality test because it rejects the null hypothesis of normality at the smallest sample size compared to the other tests, for all levels of skewness and kurtosis of these distributions. Therefore, we use the same test to check if the available data follows a normal distribution, given how small the sample size is.

The Shapiro-Wilk test was developed by Samuel Sanford Shapiro and Martin Wilk in Shapiro and Wilk. This test uses the null hypothesis to check whether a given sample came from a normally distributed population. It has a test statistic $W \in (0, 1])$ calculated as

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_i\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \tag{5.14}$$

where the constants a_i $(a_1, \ldots, a_n)$ are given by

$$a_i = \frac{m' V^{-1}}{(m' V^{-1} V^{-1} m)^{1/2}} \tag{5.15}$$

and

$$m = (m_1, \cdots, m_n)'. \tag{5.16}$$

$m_1, \cdots, m_n$ are the expected values of the order statistics of i.i.d random variables sampled from the standard normal distribution, and V is the covariance matrix of those order statistics. Under the null hypothesis H_0, the numerator is an estimator for $(n-1)\sigma^2$, and the denominator is also an estimator for $(n-1)\sigma^2$ (Castro, 2013). Hence, under $H_0, W \approx 1$. Under H_1, the numerator tends to be smaller. Therefore, we reject the null hypothesis for small values of W.

Otherwise, one can also use the p-value from the test to draw conclusions. If the p-value is less than the chosen significance level, then H_0 is rejected and there is

evidence that the data tested is not from a normally distributed population. In our study, we use the Shapiro-Wilk test in R to check for normality. This test was developed by Royston (1995) who suggested that an approximate p-value for the test is said to be adequate for p-value < 0.1. With the available data, the function returns a p-value of 0.836, which is greater than 0.1. Therefore, we fail to reject the H_0 and conclude that the data is from a normally distributed population. For verification purposes, we also use a Q-Q plot in addition to the test.

In order to determine normality graphically, we can use the output of a normal Q-Q Plot. If the data is normally distributed, data points are close to the diagonal line. If data points stray from the line in an obvious non-linear manner, then the data is not normally distributed. As we can see from the normal Q-Q plot in figure 5.4 below, the data looks normally distributed.

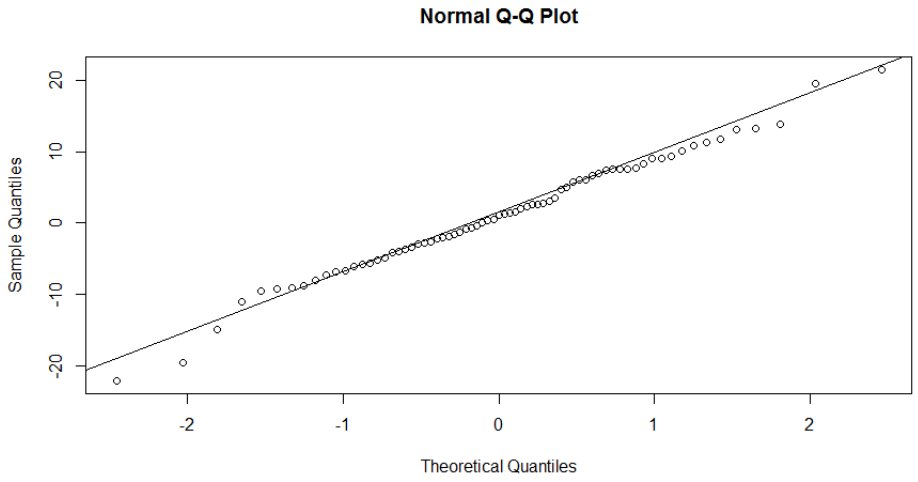

FIGURE 5.4: Q-Q plot of data.

5.1.4 Model order selection

We plot the ACF and PACF again, but this time with differenced data in order to examine which p and q values would give an appropriate model, as discussed in section 3.5. This is shown in figure 5.5.

We are able to read different plausible values of p and q by visual inspection of the ACF and PACF. From the ACF, we read off an MA(1) model because only one lag exceeds the significance bounds. From the PACF we read off an AR(2) model because only the first 2 lags exceed the significance bounds, and combining

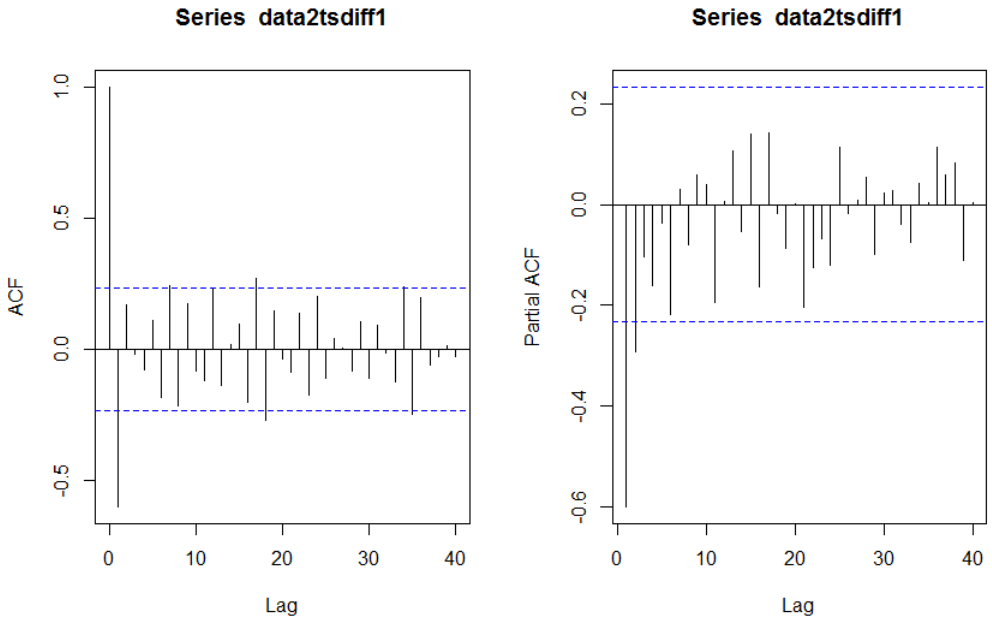

FIGURE 5.5: ACF and PACF plots for stationary data.

the ACF and PACF results, we get an ARMA(2,1) model. Since stationarity is attained after the taking the first difference, $d = 1$. This results into 3 possible models, that is, ARIMA(0,1,1), ARIMA(2,1,0) and ARIMA(2,1,1). Using the law of parsimony, the best model is ARIMA(0,1,1) because it has the least number of parameters. However, we choose to try out all possible models and select the best using different criteria. In this study, we compare models using the AIC, AICC and BIC criteria.

The oldest and most popular criterion is the Akaike information criterion (AIC) developed by Akaike (1973). This criterion is a measure of the comparative quality of a statistical model for a given set of data (Kiche et al., 2014; Vandekerckhove et al., 2015). It estimates the quality of each model as compared to other proposed models. Theoratically, the AIC is defined as

$$AIC = -2\log(L) + 2(p + q + k + 1), \tag{5.17}$$

where L is the likelihood of the data, $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$. The last term in brackets is the number of independent parameters estimated in the model, including the variance of the residuals, σ^2 (Chatfield, 2002).

A study by Hurvich and Tsai (1989) was done to improve the effectiveness of the AIC and correct it of bias. The corrected method is called the AICC, often

written as AICc. The correction is of particular use when the sample size is small, or when the number of fitted parameters is either a moderate or large fraction of the sample size. The corrected method is asymptotically efficient if the true model is infinite dimensional. Furthermore, when the true model is of finite dimension, AICC is found to provide better model order choices than any other asymptotically efficient method. For ARIMA models, AICC can be written as (Hyndman and Athanasopoulos, 2014);

$$AICC = AIC + \frac{2(p+q+k+1)(p+q+k+2)}{T-p-q-k-2}. \tag{5.18}$$

The Bayesian information criterion (BIC) is another famous criterion developed by Schwarz et al. (1978). It is partially based on the likelihood function and closely related to the AIC. The BIC can be written as (Wang and Liu, 2006)

$$BIC = -2\log(L) + (p+q+k+1)\log(T) \tag{5.19}$$

$$= AIC + (\log(T) - 2)(p+q+k+1). \tag{5.20}$$

All rejection criteria choose the model with the best fit, as measured by the likelihood function, subject to a penalty term that increases with the number of parameters fitted in the model. The best model is obtained by minimizing the AIC, AICC and BIC as seen in table 5.1. The "arima()" function in R, applied to all the possible models gives the following AIC, AICC and BIC results;

Model	AIC	AICc	BIC
ARIMA(0,1,1)	477.08	477.25	481.60
ARIMA(2,1,0)	473.57	473.93	480.36
ARIMA(2,1,1)	475.56	476.16	484.61

TABLE 5.1: Information criteria for the three models

We see that the BIC gives ARIMA(1,1,0) as the best model but the AIC and AICC give ARIMA(2,1,0). The best model from the analysis is ARIMA(2,1,0) relative to all other models. It does not have a very high BIC and has the lowest AIC and AICC.

We also consider possible seasonal models because the data tests true for monthly seasonality. These models are developed using the *auto.arima()* function in R,

with different levels of differencing. The best model is selected depending on accuracy measures, rather than information criteria. This is because, comparing models using information criteria requires that all models have the same orders of differencing, which is not true in this case (Hyndman and Athanasopoulos, 2014). In this study, we mainly focus on the mean absolute percent error (MAPE) and root mean square error (RMSE). Theoretically, MAPE and RMSE are respectively calculated as;

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\frac{|A_i - F_i|}{A_i},$$
$$\text{RMSE} = \frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2,$$

where A_i and F_i represent the actual observation and forecast respectively, at a given time i and n represents the number of observations. In R, accuracy measures are calculated using the *accuracy* function which in this case returns in-sample accuracy measures because of the absence of actual data for comparisons.

The returned accuracy measures are shown in Table 5.2

ARIMA	RMSE	MAPE
(2,1,0)	6.036516	3.080902
(0,0,0)(2,2,0)[12]	5.926322	2.426974
(0,0,0)(1,1,1)[12]	4.872027	2.347028

TABLE 5.2: Accuracy measures for the different models

From Table 5.2, the best model is SARIMA(0,0,0)(1,1,1)[12] with coefficients:

	sar1	sma1	drift
Parameters	-0.0996	-0.7421	1.1580
s.e.	0.2135	0.3634	0.0289

TABLE 5.3: Coefficients for the SARIMA(0,0,0)(1,1,1)[12] model

5.1.5 Forecasting with seasonal ARIMA(0,0,0)(1,1,1)[12]

After identifying the best model, we use it to make forecasts for future values of the series. Considering a time horizon of 12 months ahead, we apply the *forecast.Arima()* function in the *forecast* R package to equation 7.2. The point and interval forecasts developed using this function for both 80% and 95% confidence intervals are shown in Table 5.4

Point Forecast	Lo-80	Hi-80	Lo-95	Hi-95
185.5849	178.6963	192.4736	175.0497	196.1202
175.0248	168.1361	181.9134	164.4895	185.5601
185.4329	178.5443	192.3216	174.8976	195.9682
183.7596	176.8709	190.6482	173.2243	194.2949
190.5924	183.7038	197.4811	180.0571	201.1277
190.7364	183.8477	197.6250	180.2011	201.2717
190.1033	183.2146	196.9919	179.5680	200.6385
194.2202	187.3316	201.1089	183.6849	204.7555
191.3214	184.4328	198.2101	180.7861	201.8567
195.6508	188.7622	202.5395	185.1156	206.1861
192.1151	185.2265	199.0038	181.5799	202.6504
195.0531	188.1645	201.9418	184.5178	205.5884

TABLE 5.4: Forecasts for 12 months ahead.

The SARIMA(0,0,0)(1,1,1)[12] model chosen gives the forecast demand of electricity ranging from 175GW to 195GW as displayed in Figure 5.6. There is an expected increase in demand of electricity in Uganda during the following 12 months. We see that as we go further in time, the interval of the forecasts increases because the level of uncertainty also increases. A longer time horizon is likely to have more inaccurate forecasts than a short time horizon.

5.1.6 Diagnosis of seasonal ARIMA(0,0,0)(1,1,1)[12]

It is advisable to investigate whether the forecast errors of an ARIMA model are homoskedastic, normally distributed, and whether there are correlations between successive forecast errors. This can be done by plotting the ACF of the residuals, and doing a portmanteau test of the residuals using the Ljung-Box test. If residuals do not look like white noise, it means the model can be modified to improve

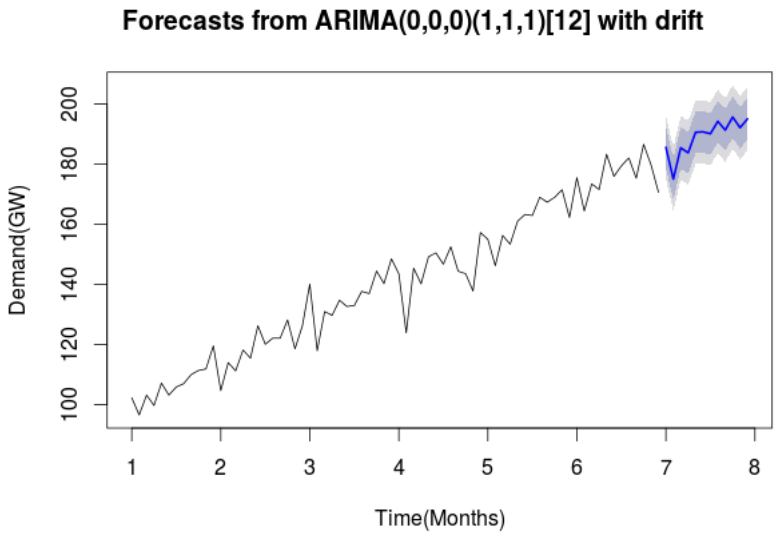

FIGURE 5.6: Point forecasts shown by the extending blue line.

the forecasts. Once the residuals look like white noise, the model can then be considered effective and used for forecasting.

5.1.6.1 Autocorrelation

We use the Ljung Box test developed by Ljung and Box (1978) to check for any evidence of autocorrelation. The test is usually applied to the residuals of a time series after fitting an ARIMA model, not the original data, and it examines all autocorrelations of the residuals (Arranz, 2005). In this test, the null hypothesis (H_0) of zero autocorrelation is tested against the alternative (H_1) of autocorrelation. For any given series, say, Y_t of length T, the test statistic Q is calculated as;

$$Q = T(T+2)\sum_{k=1}^{m}(T-k)^{-1}r_k^2, \tag{5.21}$$

where; T is the length of the time series, r_k is the k^{th} autocorrelation coefficient of the residuals, and m is the number of lags to be tested. The null hypothesis is rejected if

$$Q > \chi^2_{1-\alpha,h},$$

where α is the level of significance and h are the degrees of freedom. The degrees of freedom need to be adjusted to account for the estimated model parameters.

Therefore, $h = m - p - q$, where p and q are the parameters from the chosen ARIMA model (Maynard, 2003). In their study, Hyndman and Athanasopoulos (2014) recommended using $h = 10$ for non-seasonal data and $h = 2m$ for seasonal data, where m is the period of seasonality. They based their recommendations on power considerations, because they wanted to ensure that h was large enough to capture any meaningful and troublesome correlations.

A conclusion can also be drawn using the p-value, and this is the considered option in this study. We 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 of non randomness or autocorrelation. We test lags from 5 to 50 in intervals of 5. According to the test, no lag returns a p-value less than 0.05 as seen in Table 5.5.

Lag	test stat	p-value
5	2.858157	0.7218427
10	8.310156	0.5985696
15	13.489736	0.5645262
20	19.528189	0.4877694
25	23.668310	0.5385987
30	25.257068	0.7124451
35	38.768240	0.3035306
40	39.500716	0.4925622
45	44.573411	0.4899108
50	46.564569	0.6120341

TABLE 5.5: Absence of autocorrelation in SARIMA(0,0,0)(1,1,1)[12] residuals.

Therefore, we fail to reject the null hypothesis and conclude that, the data is random. That means, there is no evidence of autocorrelation in the residuals of the SARIMA(0,0,0)(1,1,1)[12] model chosen.

An alternative test is the Breusch-Godfrey serial correlation Lagrange multiplier test, where the test is equivalent to one based on the idea of Lagrange multiplier testing.

Graphically, we also make a plot to test for the autocorrelation effect in the residuals of the SARIMA(0,0,0)(1,1,1)[12] model as shown in Figure 5.7. None of the lags shows significant spikes, indicating no evidence of autocorrelation in the residuals of the SARIMA(0,0,0)(1,1,1)[12] model.

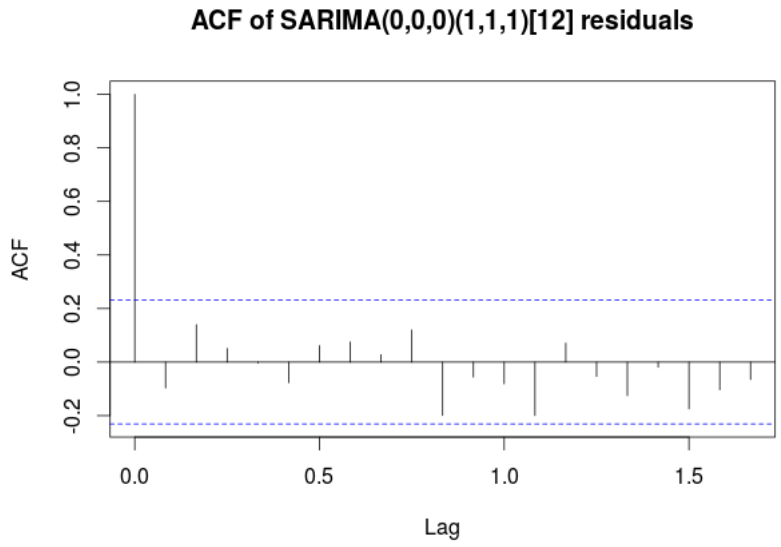

FIGURE 5.7: ACF of in-sample residuals.

5.1.6.2 Normality

We use the Jarque-Bera (JB) test to investigate whether the skewness and kurtosis of the residuals are the same as for normally distributed data. The null hypothesis is a joint hypothesis of the skewness being 0 and the excess kurtosis being 0. The JB test is defined by the following procedure; consider testing the null hypothesis H_0 : normal distribution (both skewness and excess kurtosis are zero), against the alternative hypothesis H_1 : non-normal distribution. The JB test statistic is calculated as:

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(C-3)^2\right), \tag{5.22}$$

where; n is the number of observations, also called the degrees of freedom, S is the sample skewness calculated by:

$$S = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{\frac{3}{2}}}, \tag{5.23}$$

and C is the sample kurtosis. We use $C-3$ in this expression because for normally distributed data, the kurtosis is expected to be 3. This means any excess kurtosis

should be 0 ($C - 3$). The sample kurtosis is calculated by:

$$C = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}. \tag{5.24}$$

The calculated test statistics in equation 5.22 can be compared with a χ^2 (chi-square) value with 2 degrees of freedom and a chosen level of significance α (Thadewald and Büning, 2007). The null hypothesis of normality is rejected if the calculated test statistic is greater than the tabulated value. Otherwise, a conclusion can also be drawn using the p-value. In our study, we choose this option because of the software used (R). The same hypotheses are tested. A p-value of 0.002698 is calculated, which is less than 0.05. Hence leading to the rejection of the null hypothesis of normality and concluding that the residuals show evidence of a non normal distribution.

As the definition of JB shows, any deviation from skewness and excess kurtosis of 0 increases the JB statistic. With the available data, the calculated JB statistic is 11.831, which is not a high value. This means there is not much deviation from skewness and excess kurtosis of 0, according to the available data. For small samples the chi-squared approximation is overly sensitive, often rejecting the null hypothesis when it is in fact true (Tao et al., 2014). This might be the reason behind these JB test results with a small sample of 72 observations. The sample size of 72 is too small for the underlying asymptotic approximations for the JB test to work (Wuertz and Katzgraber, 2005). Therefore, the reported p-value is neither uniform nor sufficient enough for use in hypothesis testing.

To further investigate the non normality of the residuals, we make a time series plot and a histogram (with overlaid normal curve) of residuals. We use residuals from in-sample forecasts because observed data for the forecast is not readily available. The histogram in figure 5.8 shows that the distribution of residuals from in-sample forecasts is not centred on zero and has flatter tails compared to a normal distribution.

The same conclusion is drawn from the Q-Q plot of residuals plotted in Figure5.9. The residuals have heavier tails on both the right and left ends. Therefore indicating non normality. However, in their book, Hyndman and Athanasopoulos (2014) point out that it is useful, but not necessary for the residuals to also have a constant variance and be normally distributed. These two properties only make the

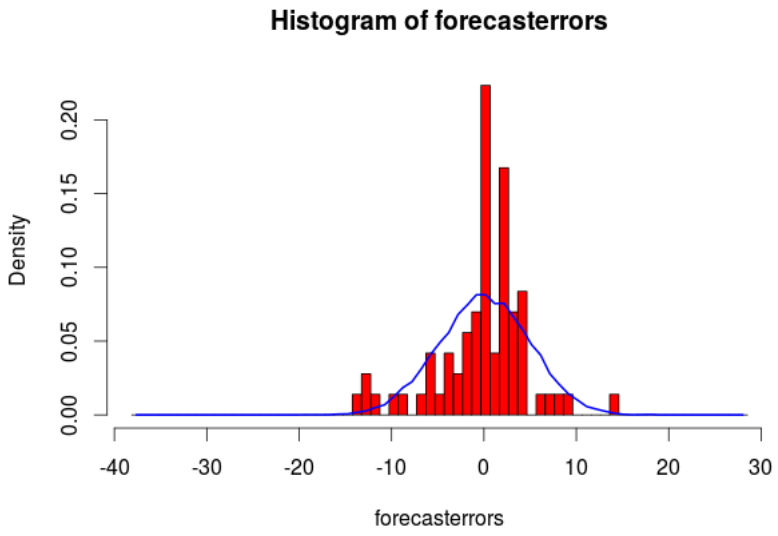

FIGURE 5.8: Histogram of in-sample residuals.

calculation of prediction intervals easier. If a forecasting model produces residuals that are homoskedastic and show no evidence of autocorrelation, then little can be done to achieve normality from the model's residuals. That models can be considered sufficient enough for forecasting if the main aim for building it was to carry out forecasting exercises.

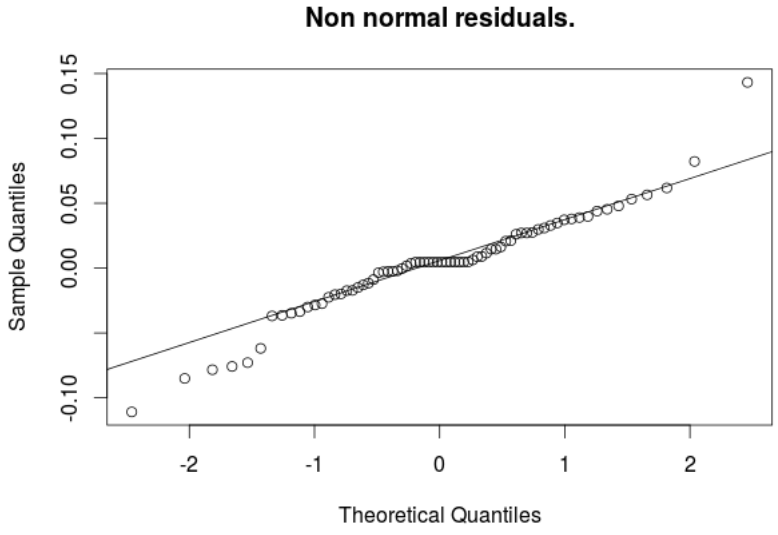

FIGURE 5.9: Q-Q plot showing non normality of residuals.

The time series plot in figure 5.10 shows that the variance of the forecast errors is not constant over time. However, more formal tests for homoskedasticity are

carried out in Chapter 7.

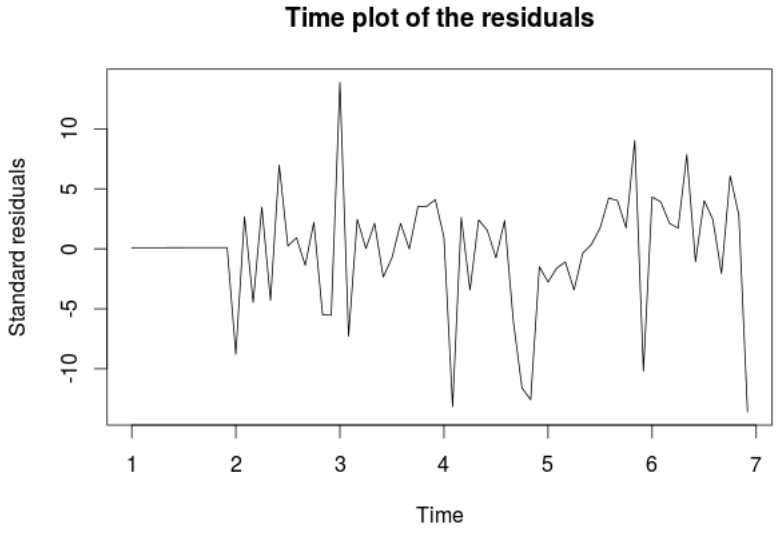

FIGURE 5.10: Time plot of in-sample residuals.

The main objective of this study is to predict electricity demand for Uganda in order to help electricity generating bodies and policy makers plan for the future of Uganda sustainably and adequately. Since successive forecast errors do not show evidence of auto correlation, we concluded that the SARIMA(0,0,0)(1,1,1)[12] model is an adequate predictive model for the peak monthly demand of electricity in Uganda.

5.2 South Africa daily electricity demand data

In this section, we fit an ARIMA model to South Africa's daily electricity demand data. We are able to identify a seasonal pattern, with no trend, from the time series plot in Figure 3.7. From the time series plot, the data shows an unstable variance over time. It is necessary to adjust it accordingly to gain stationarity. We divide the data into two sets; the first set has the first 1342 observations used to train the model, and the second set has the last 300 observations used to test the efficiency of the ARIMA model.

We plot the ACF of the training set in order to check for stationarity because visual inspection of the time series plot is some times misleading. From Figure 5.11 the data is not stationary. The ACF decays off at a very slow rate.

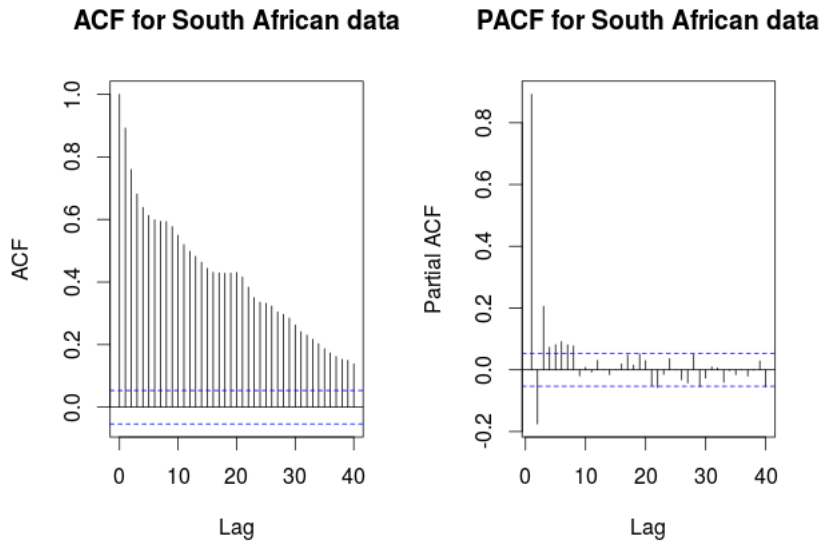

FIGURE 5.11: ACF and PACF of non stationary data.

5.2.1 Testing for stationarity

In order to gain stationarity, different transformations can be used. A log transformation is commonly used in electricity studies but for the available data, the log transformation alone is not effective enough to attain stationarity. Due to the seasonality component exhibited by the data, a seasonal difference is a better way of making the data stationary. The *tbats* function in R reveals the presence of annual seasonality in the data, as shown in Figure 5.12.

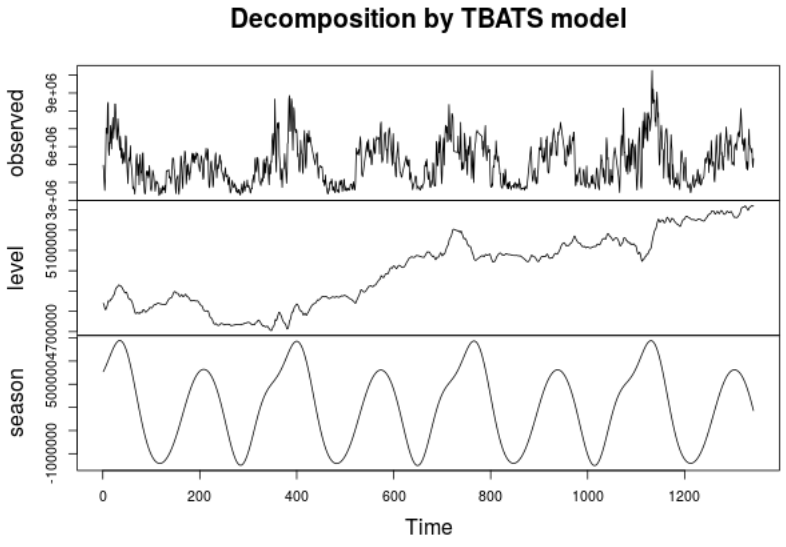

FIGURE 5.12: Annual seasonality in South Africa data.

After a log transformation, a seasonal difference is applied to the data too and the results are shown in Figure 5.13. At initial glance, Figure 5.13 shows stataionarity in the data. However, we run a KPSS test for certainty, which returns a p-value of 0.1. Since the returned p-value is greater than 0.05, we fail to reject the null hypothesis and conclude that the data is now stationary.

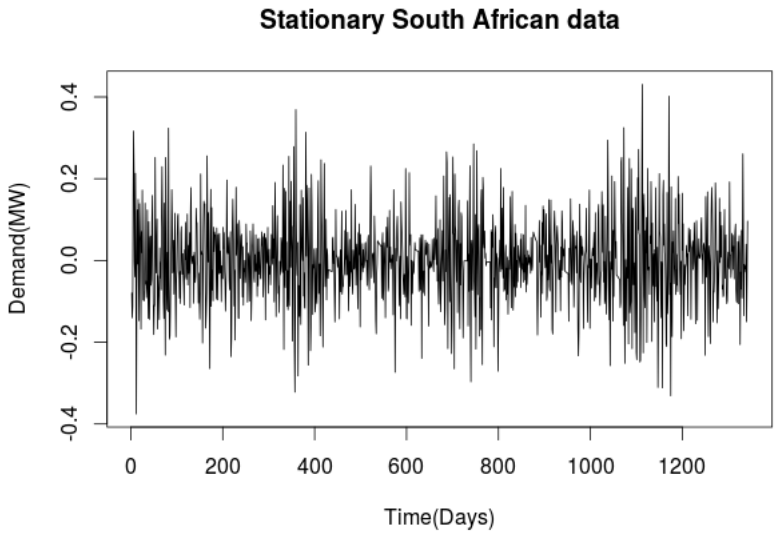

FIGURE 5.13: Time series plot of the data after taking the seasonal difference.

5.2.2 Testing for normality

We use the Shapiro-Wilk test for normality. The function returns a p-value of $4.781e-09$, which is less than 0.1. Therefore, we reject H_0 and conclude that the data does not follow a normal distribution. The Q-Q plot in Figure 5.14 also shows non normal data because most of the data points stray away from the diagonal line.

5.2.3 Model order selection

We plot the ACF and PACF with seasonal differenced data in order to select the appropriate p and q values to use when constructing the model. This is shown in figure 5.15.

However, since the data shows evidence of seasonality, we use the *auto.arima* function in R, together with different values of D (the seasonal difference), to

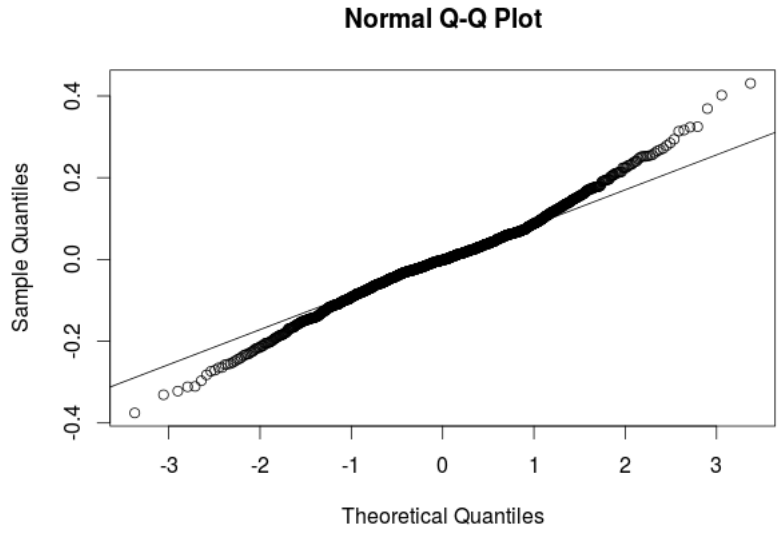

FIGURE 5.14: Q-Q plot of South Africa data.

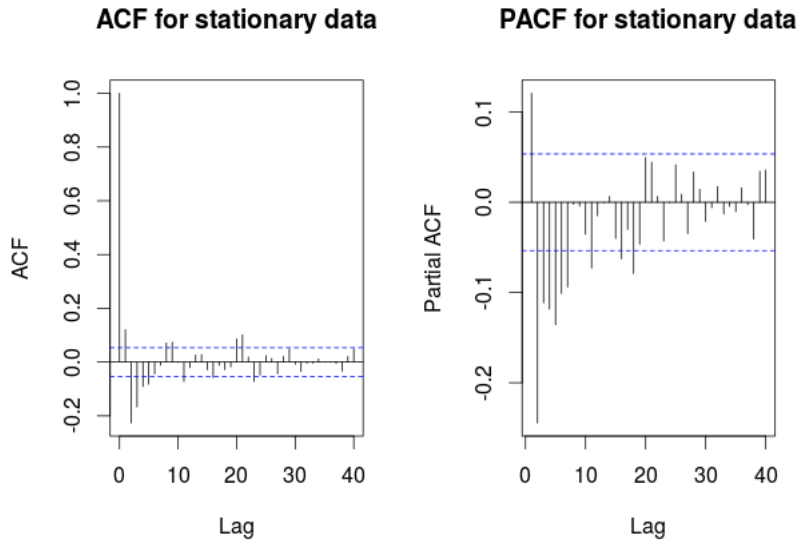

FIGURE 5.15: ACF and PACF of seasonal differenced data.

develop different possible 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 the same orders of differencing, which is not the case for the seasonal difference used (Hyndman and Athanasopoulos, 2014). These models are shown in Table 5.6;

We compare 2 models with $D = 1, 2$ because in reality, it is not common to difference data more than twice before stationarity is achieved (Chatfield, 2000).

	ARIMA(1,0,1)(0,1,0)[365] with drift	ARIMA(1,0,1)(0,2,0)[365]
ME	1553.408	-17639.39
RMSE	581600.1	775767
MAE	373029.5	394677.6
MPE	-0.6057378	-0.688367
MAPE	7.098341	7.549908
MASE	0.4243369	0.4489625
ACF1	-0.001359556	-0.003033473

TABLE 5.6: Plausible models from the *auto.arima* function

Using the accuracy measures (mainly RMSE and MAPE) as shown in Table 5.6, the best model for the available data is ARIMA(1,0,1)(0,1,0)[365] with a drift. It has the lowest RMSE and MAPE. The coefficients of the model are shown in Table 5.7:

	ar1	ma1	drift
Parameters	0.6518	0.3511	531.1491
s.e.	0.0296	0.0364	231.3203

TABLE 5.7: Coefficients for the ARIMA(1,0,1)(0,1,0)[365] model

5.2.4 Forecasting with seasonal ARIMA(1,0,1)(0,1,0)[365]

We then use the identified model to make forecasts. The desired time horizon is 300 days ahead, in order to compare the forecasts with the test set of actual data. Using the *forecast* package in R, we develop both point and interval forecasts for 80% and 95% confidence intervals, covering a period of 300 days, as shown in Figure 5.16.

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

Here we investigate the normality of the errors and autocorrelation between successive forecast errors. To get the forecast errors, we calculate the difference between the observed test set and the forecasts from the chosen model.

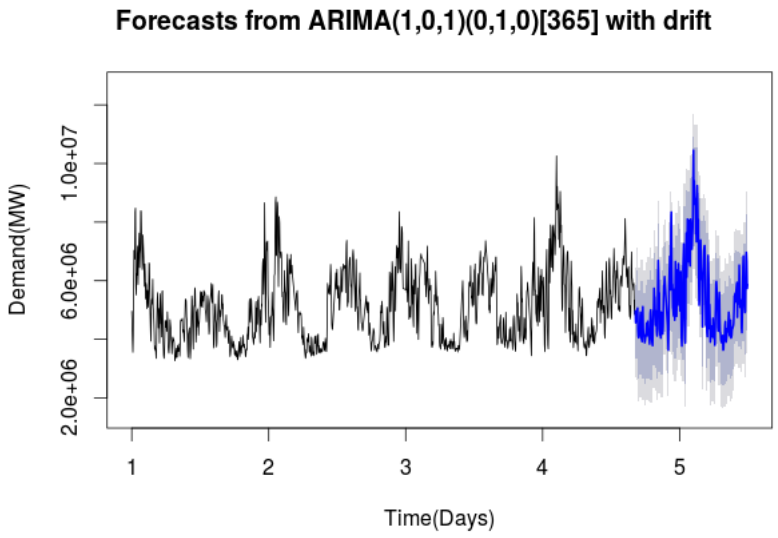

FIGURE 5.16: Point forecasts shown by the extending blue line.

5.2.5.1 Autocorrelation

We use the Ljung Box Statistic test with the null hypothesis of zero autocorrelation against the alternative of autocorrelation. According to the Ljung Box Statistic test used in the study, all lags return p-values less than 0.05. This is shown in Table 5.8. Therefore, we reject the null hypothesis and conclude that, the residuals are not random, meaning there is evidence of autocorrelation in the residuals of the chosen ARIMA model.

Lags	Test stat	P-value
Lag 5	266.1336	1.893091e-55
Lag 10	293.0550	4.563708e-57
Lag 15	302.5744	1.630482e-55
Lag 20	315.3051	5.987907e-55
Lag 25	346.9746	1.996977e-58
Lag 30	367.9272	8.046908e-60
Lag 35	377.7595	4.309539e-59
Lag 40	388.6129	1.133349e-58
Lag 45	431.3102	1.535623e-64
Lag 50	438.3353	1.789481e-63

TABLE 5.8: Ljung Box results for ARIMA(1,0,1)(0,1,0)[365] model residuals.

Graphically, Figure 5.17 shows 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.

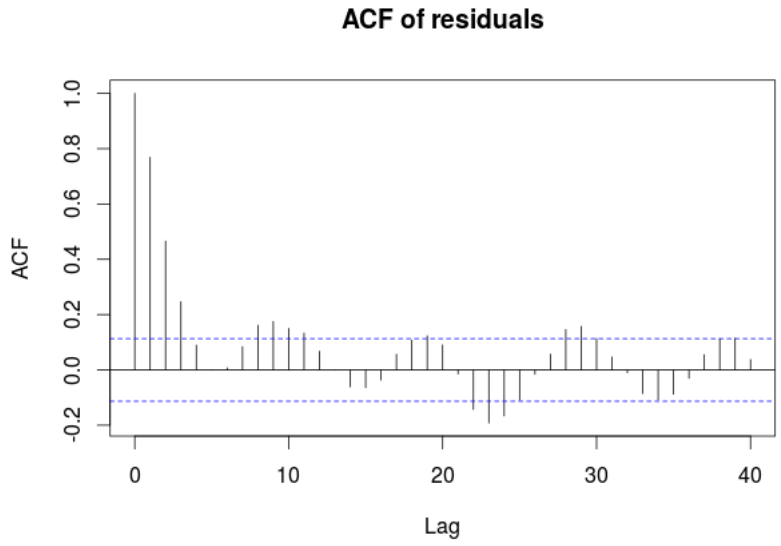

FIGURE 5.17: ACF of out-of-sample residuals.

5.2.5.2 Normality

Using the JB test in R, a p-value of $1.376e-12$ is calculated, which is less than 0.05. Hence rejecting H_O, and concluding that the residuals have a non normal distribution.

Graphically, we plot a time series plot and a histogram with an overlaid normal curve of residuals. The time series plot in Figure 5.18 shows that the mean of the forecast errors fluctuates about zero but the variance still varies roughly. We conclude that the data does not have a constant variance.

The histogram in Figure 5.19 shows that the residuals from out-of-sample forecasts are skewed to the left. The calculated mean of the residuals is -55714.73, which highly varies from 0. The histogram also shows that the residuals have flatter tails and a higher kurtosis than a normal data set. Therefore, we confirm that the residuals are not normally distributed. They neither have a constant variance nor 0 mean.

The normal Q-Q plot in figure 5.20 also confirms the non normality of the residuals.

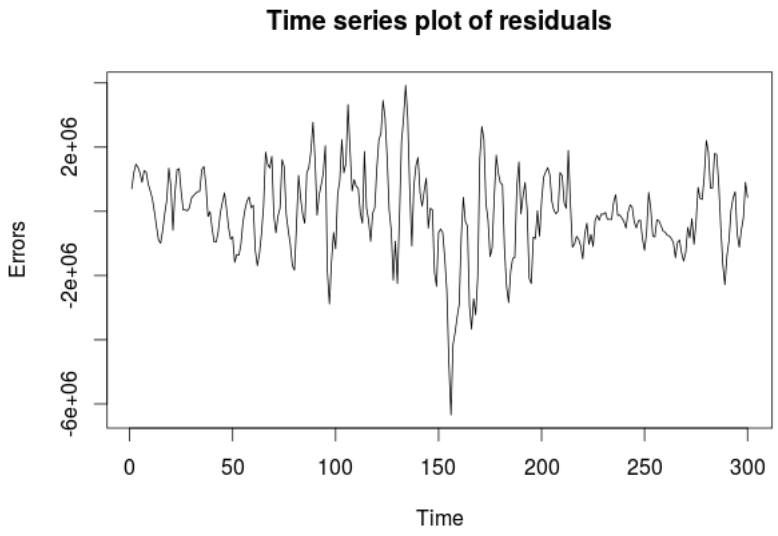

FIGURE 5.18: Time plot of out-of-sample residuals.

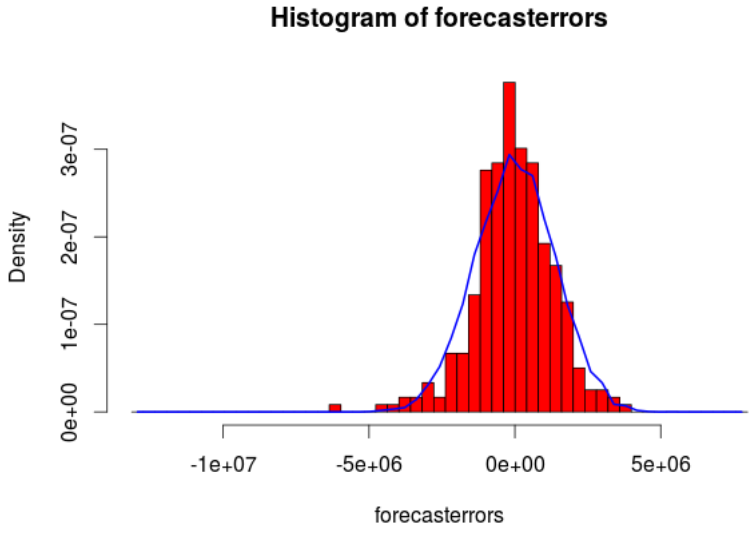

FIGURE 5.19: Histogram of out-of-sample residuals.

5.2.6 Conclusion

ARIMA models are used because of their capabilities to make predictions using time series data with any kind of pattern and with autocorrelations between the successive values in the series (Kumar and Anand, 2014). We have statistically tested and validated that the residuals in the fitted ARIMA model are correlated, and are not normally distributed. Therefore, the chosen ARIMA(1,0,1)(0,1,0)[365] can be improved to produce better forecasts.

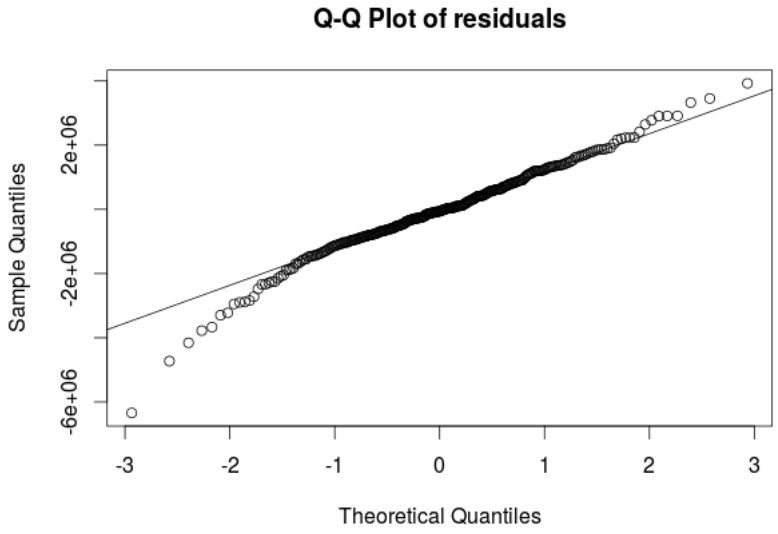

FIGURE 5.20: Q-Q plot of out-of-sample residuals.

Like other forecasting models, ARIMA models have limitations on accuracy of predictions. The forecasts provided by an ARIMA model can be improved in various ways, for example, taking into consideration the change in variance of the data. This is called volatility testing and various models have been developed over the years to take volatility into consideration while forecasting.

Chapter 6

Volatility forecasting models

In this chapter we study some of the statistical methods for analysing and modeling volatility in any given data set, with specific emphasis on electricity demand data. The models to be studied are called conditional heteroscedastic models. Since our emphasis is on univariate models, we shall study the AutoRegressive Conditional Heteroscedastic (ARCH) model developed by Engle (1982) and the generalized ARCH (GARCH) model developed by Bollerslev (1986).

Linear methods often work well and may provide an adequate approximation for the task at hand. These methods can also be used as a basis for comparison with the results from more complicated alternative analyses. However, in reality, life processes can not be restricted to linearity only. Therefore, it is necessary to apply the use of non linear models as well for time series analysis. Data that looks stationary in the mean but is non-stationary in variance can not be explained by a linear model, and so non-linear models are needed to describe such data because of the change in variance through time.

In our study, we have analysed and forecast the available data using ARIMA models. However, we need to check the residuals for normality, autocorrelation and homoscedasticiy. If any of these tests returns a negative result, then the model chosen is not adequate enough and forecasts can be improved with further analysis of extra data features. For that reason, volatility models are important to accommodate the volatility in the residuals of the data after applying ARIMA models.

6.1 Significance of this chapter

The main aim of this chapter is to study some of the volatility forecasting models used in the analysis of univariate time series data and to use these processes to model volatility in the residuals of electricity demand data for both Uganda and South Africa. This helps improve the level of accuracy of the forecasts. The specific objectives are to;

- identify the best fitting model for the electricity demand data available.
- assess the contribution of these models to understanding of volatility in electricity demand.
- examine and compare the ARCH model and its extensions with ARIMA models, both theoretically and practically.

To attain these objectives, this chapter will be organised as follows; Meaning and more understanding of volatility, development and properties of an ARCH model, extension to the GARCH model and its properties, model specification, application to electricity demand data and lastly the conclusion.

6.2 Definition of Volatility

Volatility is the rate at which the variance of a given variable under study changes over time. It can also be defined as the level of uncertainty about changes in the value of a given variable. Modelling volatility in electricity demand is very vital because many factors affecting the demand of electricity change in very short time intervals. 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 dealing with time series data. High volatility means that the value of a variable can potentially be spread out over a larger range of values (Investopidia-Volatility). With electricity demand, the higher the volatility, the more complicated it gets to forecast demand accurately because the values are widely spread out. Therefore, complex forecasting techniques are employed in such cases in order to accommodate all the values.

Volatility can either be measured using the standard deviation or variance. Authors like Talke (2003) prefer using the standard deviation because it is measured in the same units as the original data. In this study, we shall also use the standard deviation symbol, σ, to denote volatility. According to Tsay (2005), the most common characteristics of volatility include; no direct observation especially when dealing with univariate time series data, existence of volatility clusters when the data is plotted, continuous progress over time, variation in a fixed range-meaning divergence does not reach infinity.

6.3 The ARCH model

This was the first and simplest model 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. This variance is conditional on the available data. It represents the uncertainty about the next period's 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. In particular, 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 suggests the possibility of using an ARCH model for any series that has changing variance, for example, residuals after fitting an ARIMA model to the data. Therefore, in this study we use residuals from our ARIMA analysis as data to which volatility models are applied. We have chosen to use the notation Y_t instead of X_t, to emphasize that these volatility models are not applied directly to the observed data. The derived series Y_t is from the residuals of the direct observations. Assume y_t represents residuals and ε_t is a normal white noise process where, $\varepsilon_t \sim iid(0,1)$, and let Y_{t-1} be a set of information available at time $t-1$. That means, at a current time, $Y_{t-1} = \{y_1, y_2, \cdots, y_{t-1}\}$. Then, y_t is said to follow an ARCH process if;

$$y_t = \sigma_t \varepsilon_t, \tag{6.1}$$

where σ_t is 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 related to squares of the previous error terms, depending on the order of the process.

6.3.1 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 this 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 $\alpha_1 \geqslant 0$ are assumed, and $\alpha_1 < 1$ is assumed for stationarity (Chatfield, 2002).

It is clear from equation 6.2 that the variance at time t is connected to the value of the series at time $t-1$. Therefore, a large past residual implies a large 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 equation 6.2 into equation 7 gives an ARCH(1) model, which is represented as;

$$y_t = \varepsilon_t\left(\sqrt{\alpha_0 + \alpha_1 y_{t-1}^2}\right), \tag{6.3}$$

where y_{t-1} denotes the observed value of the derived series at time $t-1$. Considering the assumption of normality of ε_t, the ARCH(1) process can be written conditional on Y_{t-1} as;

$$y_t|Y_{t-1} = y_t|y_{t-1} \sim N(0, \sigma_t^2). \tag{6.4}$$

6.3.1.1 Properties of ARCH(1) process

- Mean: This is calculated conditional on all past information and it is given by;

$$\begin{aligned} E(y_t|Y_{t-1}) &= E(\sigma_t\varepsilon_t|Y_{t-1}), && (6.5)\\ &= \sigma_t E(\varepsilon_t|Y_{t-1}), && (6.6)\\ &= 0. \qquad \text{(From the normality assumption of } \varepsilon_t.) && (6.7) \end{aligned}$$

  This means the series y_t is a martingale difference sequence (MDS) because its expectation with respect to the past observations is zero.

- Variance: If it is assumed that y_t is second order stationary, then the variance is constant at all times. That means $\text{Var}(y_t) = E(y_t^2) = E(y_{t-1}^2) = \text{Var}(y_{t-1})$. Therefore;

$$\begin{aligned} \text{Var}(y_t) &= E(y_t^2) - (E(y_t))^2 && (6.8)\\ &= E(\sigma_t^2\varepsilon_t^2) - 0 && (6.9)\\ &= E[\sigma_t^2(E(\varepsilon_t^2|Y_{t-1}))] && (6.10)\\ &= E[\sigma_t^2(1)] \qquad \text{(From the normality assumption of } \varepsilon_t.) && (6.11)\\ &= E[\alpha_0 + \alpha_1 y_{t-1}^2] && (6.12)\\ &= \alpha_0 + \alpha_1 E(y_{t-1}^2) && (6.13)\\ &= \alpha_0 + \alpha_1 \text{Var}(y_t) && (6.14)\\ &= \frac{\alpha_0}{1-\alpha_1}. && (6.15) \end{aligned}$$

- Other properties of the ARCH(1) model are its skewness and kurtosis that have not been discussed in this study.

6.3.1.2 Parameter estimation in ARCH(1)

Parameters α_0 and α_1 in equation 6.3 have to be estimated for the model to be considered for analysis. Different methods can be used but in this study we shall use the MLE method because we assumed normality of the process, with 0 mean and constant variance (see equation 6.4). If we assume a series $y_1, y_2, ..., y_T$ to be a realization of the ARCH(1) process, we begin by defining the distribution function

of the process as;

$$f(y_T|y_{T-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{y_t^2}{2\sigma_t^2}\right). \tag{6.16}$$

We then derive the likelihood function containing the unknown parameters of the model as a product of all conditionals.

$$L(y_1, ..., y_T|\beta) = f(y_T|Y_{T-1})f(y_{T-1}|Y_{T-2})...f(y_2|Y_1)f(y_1|\beta), \tag{6.17}$$

where $\beta = (\alpha_0, \alpha_1)'$ represents the parameters to be estimated. The conditional expression $f(y_1|\beta)$ can not be obtained easily. However, Tsay (2005) and Talke (2003) consider y_1 to take on it's observed value. Therefore, it is dropped from the original likelihood function and considered for conditioning, especially if the data set is big. This changes equation 6.17 to;

$$L(y_2, ..., y_T|\beta; y_1) = f(y_T|Y_{T-1})f(y_{T-1}|Y_{T-2})...f(y_2|Y_1). \tag{6.18}$$

Using the normal distribution function in 6.16, the final conditional likelihood function becomes;

$$L(y_2, ..., y_T|\beta; y_1) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\sum_{t=2}^{T} \frac{y_t^2}{2\sigma_t^2}\right), \tag{6.19}$$

$$= \prod_{t=2}^{T} (2\pi\sigma_t^2)^{-\frac{1}{2}} \exp\left(-\sum_{t=2}^{T} \frac{y_t^2}{2\sigma_t^2}\right). \tag{6.20}$$

We take natural logarithms for equation 6.20 in order to get the conditional log likelihood function.

$$l(y_2, ..., y_T|\beta; y_1) = -\frac{1}{2}\sum_{t=2}^{T} \ln(2\pi\sigma_t^2) - \frac{1}{2}\sum_{t=2}^{T} \frac{y_t^2}{\sigma_t^2}, \tag{6.21}$$

$$= -\frac{1}{2}\sum_{t=2}^{T} \left(\ln(2\pi) + \ln(\sigma_t^2) + \frac{y_t^2}{\sigma_t^2}\right). \tag{6.22}$$

The values of α_0 and α_1 that maximize the sample likelihood are called MLEs and can be obtained by maximizing the log likelihood function with respect to these parameters. Taking partial derivatives in the first term of equation 6.22 results

into a 0 since it does not have any of the parameters of interest. Therefore, we can remove it from the function and replace σ_t^2 by $\alpha_0 + \alpha_1 y_{t-1}^2$ from equation 6.2 .

$$l(y_2, ..., y_T|\beta; y_1) = -\frac{1}{2}\sum_{t=2}^{T} \ln(\alpha_0 + \alpha_1 y_{t-1}^2) - \frac{1}{2}\sum_{t=2}^{T} \frac{y_t^2}{\alpha_0 + \alpha_1 y_{t-1}^2}. \tag{6.23}$$

It is often very complicated to analytically find the parameter estimates from equation 6.23. However, different studies have used different approaches to solve this problem, for example, Talke (2003) suggested the numerical approach to find the estimates. They argued that if data was made available with a known maximum time T, then the parameter estimates could be computed. Some studies propose the use of algorithms. For example, Allal and Benmoumen (2014) proposed an estimate algorithm for the parameters of a ARCH(1) model. They did not consider any assumptions about initial values even if these assumptions are important in the Quasi-maximum likelihood estimation method (QMLE). They combined the maximum likelihood method, Kalman filter algorithm and the Simultaneous Perturbation Stochastic Approximation method (SPSA) to create the algorithm. Using simulated data, their results showed that the algorithm was very reliable, viable and promising in estimating the parameter values of a given model.

6.3.1.3 Forecasting with ARCH(1)

Unlike linear ARIMA models, when forecasting with volatility models we take into consideration the variance of the data. Assume a series $Y_T = \{y_1, ..., y_T\}$, 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;

$$\begin{aligned} y_T(l) &= E(y_{T+l}|Y_T) && (6.24)\\ &= E(\sigma_{T+l}\varepsilon_{T+l}|Y_T) \quad \text{For an ARCH(1)process,} && (6.25)\\ &= \sigma_{T+l}E(\varepsilon_{T+l}|Y_T) = 0. && (6.26) \end{aligned}$$

For this reason, 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;

$$y_T^2(l) = E(y_{T+l}^2|Y_T) \quad Y_T = \{y_1^2, ..., y_T^2\} \tag{6.27}$$
$$= E(\sigma_{T+1}^2\varepsilon_{T+1}^2|Y_T) \quad \text{For } l = 1 \tag{6.28}$$
$$= E(\sigma_{T+1}^2) \tag{6.29}$$
$$= \hat{\alpha_0} + \hat{\alpha_1}E(y_T^2) \tag{6.30}$$

But 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 calculated in part 6.3.1.2.

$$y_T^2(1) = \hat{\alpha_0} + \hat{\alpha_1}y_T^2 = \sigma_T^2(1) = E(\sigma_{T+1}^2|Y_T). \tag{6.31}$$

If $l = 2$, then the forecast is given as;

$$y_T^2(1) = E(y_{T+2}^2|Y_T) = E(\sigma_{T+2}|Y_T) \tag{6.32}$$
$$= E(\hat{\alpha_0} + \hat{\alpha_1}y_{T+1}^2|Y_T) = \hat{\alpha_0} + \hat{\alpha_1}E(y_{T+1}^2|Y_T) \tag{6.33}$$
$$= \hat{\alpha_0} + \hat{\alpha_1}(\hat{\alpha_0} + \hat{\alpha_1}y_T^2) \tag{6.34}$$
$$= \hat{\alpha_0} + \hat{\alpha_1}\sigma_T^2(1). \tag{6.35}$$

6.3.2 ARCH(q)

The ACRH(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,

$$\sigma_t^2 = \text{Var}(y_t|y_{t-1}, \cdots, y_{t-p}), \tag{6.36}$$
$$= \alpha_0 + \alpha_1 y_{t-1}^2 + ... + \alpha_q y_{t-q}^2, \tag{6.37}$$

where the restrictions $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1, 2, ..., q$ 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. Therefore, the

variance becomes;

$$\text{Var}(y_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i}. \tag{6.38}$$

The condition for second order stationarity is still assumed, therefore $\sum_{i=1}^{q} \alpha_i < 1$. All the other procedures, like estimation of parameters and forecasting in this process are similar to those discussed in the ARCH(1) process, with just a increase in the number of parameters included. An ARCH(q) process is uncorrelated with constant conditional and unconditional means, a constant unconditional variance, and a non constant conditional variance, just like an ARCH(1) process (Ruppert, 2010).

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 gradually increasing mean level can be handled better using transformation methods as discussed in section 3.6 (Eberly College Website). 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 example, from a financial point of view, reality shows that the price of a financial asset responds differently to positive and negative shocks.
- ARCH models do not provide any new ideas for understanding the source of variations of any given time series. They only help us understand the behaviour of the conditional variance.
- ARCH models are likely to over predict the volatility because they respond slowly to large isolated errors to the new developed series.

6.4 GARCH model

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. GARCH models allow the conditional variance,

σ_t^2, to depend on both previous conditional variances (σ_{t-i}^2) and previous squared values of the series (y_{t-i}^2). 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 (Efimova, 2013). GARCH models typically 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 similar to that behind an ARMA model. A high order AR or MA model may often be approximated by a mixed ARMA model, with fewer parameters (Chatfield, 2000).

Just like an ARCH process, a GARCH process is still defined using equation 7, where; ε_t is still assumed to be a sequence of iid random variables with mean 0 and variance 1. σ_t^2 is generally a function of previous conditional variances and previous observed values of the series. However, it specifically depends on the order of the model.

6.4.1 GARCH(1,1)

A GARCH(1,1) process is simply an extension 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;

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{6.39}$$

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. The introduced lagged variance reduces the initially many ARCH parameters to an easily dealt with quantity. This can be explained by expansion of the model 6.39;

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1(\alpha_0 + \alpha_1 y_{t-2}^2 + \beta_1 \sigma_{t-2}^2), \tag{6.40}$$

$$= \alpha_0 + \beta_1 \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \alpha_1 y_{t-2}^2 + \beta_1^2 \sigma_{t-2}^2. \tag{6.41}$$

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.

As was the case in the ARCH model, the coefficients of a GARCH model must also be restricted to ensure that the conditional variances are uniformly positive. In a GARCH(1,1), these restrictions are; for a positive conditional variance, the parameters $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\beta_1 \geq 0$, and for the assumption of stationarity, $\alpha_1 + \beta_1 < 1$. The properties of a GARCH(1,1) model are not so different from those if an ARCH(1) process. The conditional mean is still 0, therefore the series is still a martingale difference. Due to the structure of the model in this case, the information set is, $Y_{t-1} = \{y_1, \sigma_1^2, ..., y_{t-1}, \sigma_{t-1}^2\}$. Therefore, to get the variance in this case, we still assume second order stationarity like for the ARCH process, and;

$$E(y_t^2) = E(\sigma_t^2 \varepsilon_t^2 | Y_{t-1}) \tag{6.42}$$

$$= E(\sigma_t^2 (E(\varepsilon_t^2 | Y_{t-1}))) \tag{6.43}$$

$$= E(\sigma_t^2) = E(\alpha_0 + \alpha_1(y_{t-1}^2) + \beta_1(\sigma_{t-1}^2)). \tag{6.44}$$

From the information set σ_{t-1}^2 is observed. Therefore, its expectation takes on its real value, which is the variance of y_{t-1}. Then, the assumption of second order stationarity shows that $E(y_t^2)$ is equal to $E(y_{t-1}^2)$ which is equal to the variance. Thus,

$$E(y_t^2) = \alpha_0 + \alpha_1(E(y_t^2)) + \beta_1(E(y_t^2)), \tag{6.45}$$

$$= \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}. \tag{6.46}$$

Forecasting the conditional variance one step ahead follows directly from the model. Forecasting more than one step ahead is carried out by replacing future values of σ_t^2 and of y_t^2 by their estimates.

6.4.2 GARCH(p,q)

A GARCH model of order (p, q) assumes the conditional variance depends 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 in practice to require the use of a GARCH model of order higher than (1,1).

In practice, if such a model is fitted to data and the stationarity condition is not satisfied, squared observations can be made stationary after taking first differences. This results into the Integrated GARCH (IGARCH) model. Other extensions of the basic GARCH model include Quadratic GARCH (QGARCH), which allows for negative shocks to have more effect on the conditional variance than positive shocks, and exponential GARCH (EGARCH) which also allows an asymmetric response by modelling $\log \sigma_t^2$, rather than σ_t^2 (Chatfield, 2002).

6.5 Non-normal distributions

All the ARCH models discussed earlier assume a normally distributed data set. Although this assumption is commonly used by researchers modelling volatility, it is not always obvious or appropriate. In fact, ARCH models with normal errors capture some but not necessarily all the nature of the distribution of data. Therefore, ARCH models involving errors that follow distributions with fatter tails than the normal distribution are also investigated. Commonly studied are ARCH models with errors following the Student-t distribution suggested by Bollerslev in 1987 (Bollerslev et al., 1992; Heracleous, 2003) and ARCH models with errors following the generalised error distribution suggested by Nelson in 1991 (Bollerslev et al.).

If the error term ε_t is assumed to follow a Student-t distribution with $n > 2$ degrees of freedom, then its probability distribution function (pdf) will be given by;

$$f(\varepsilon_t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{(n-2)\pi\sigma_t^2}}\left[1+\frac{y_t^2}{(n-2)\sigma_t^2}\right]^{-\frac{(n+1)}{2}}, \tag{6.47}$$

where $\Gamma()$ represents a gamma function. The Student-t distribution is symmetric about 0 and converges to the normal as n approaches infinity (Brase and Brase, 2011). Unlike the kurtosis of a normal distribution which is 3, the Student-t distribution has a bigger kurtosis of $3+\frac{6}{n-4}$. This explains the presence of heavier tails in a Student-t distribution (Rachev et al., 2008).

Another type of models developed by Nelson (1991) are the EGRACH models. These models assumed errors following the generalised error distribution with a

pdf given by;

$$f(\varepsilon_t) = \frac{n \exp\left[-(\frac{1}{2})|\frac{\varepsilon_t}{\lambda}|^n\right]}{\lambda 2^{(1+\frac{1}{n})}\Gamma(\frac{1}{n})}, \tag{6.48}$$

where Vasudeva and Kumari (2013) define λ as

$$\lambda = \left[2^{(\frac{-2}{n})}\frac{\Gamma(\frac{1}{n})}{\Gamma(\frac{3}{n})}\right]^{(\frac{1}{2})} \tag{6.49}$$

and n as the tail thickness parameter. When $n = 2$, the errors follow a standard normal distribution. When $n < 2$, the errors follow a distribution with thicker tails, and $n > 2$ gives the errors a distribution with thinner tails as compared to those of a normal distribution (Talke, 2003).

6.6 Model Specification

The best identification tool may be a time series plot of the series. It is usually 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 instance, 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 suggested. In practice, it is advisable to experiment with various ARCH and GARCH structures after realising the need to in the time series plot of the series (Eberly College Website).

Identifying an appropriate ARCH or GARCH model is not as easy as dealing with linear models, which partially explains 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 second-order properties alone are examined, and so non-linearity has to be assessed by examining 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 auto correlation 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 :

- A class of models is formulated assuming certain hypotheses. In this step, a general GARCH formulation is selected to model the available data. This

selection is carried out by careful 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 identified for the observed data. A trial model must be identified for the available data, as seen in the first step. In a first trial, the observation of the ACF and PACF plots of the data can help to make this selection. In successive trials, the same observation of the residuals obtained can refine the structure of the functions in the model.

- The model parameters are estimated. After the functions of the model have been specified, the parameters of these functions must be estimated. Good estimators of the parameters can 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.

- If the hypotheses of the model are validated, we can proceed to the next step. Otherwise, it is advisable to go back to the previous steps and refine the model. In this step, a diagnosis check is used to validate the assumptions of the GARCH model. Among the tests to validate the assumptions of the GARCH model chosen is a careful inspection of the ACF and PACF plots of the residuals.

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

Chapter 7

Applying ARCH models to ARIMA residual

In this chapter, we use the residuals from both ARIMA(0,0,0)(1,1,1)[12] and ARIMA(1,0,1)(0,1,0)[365] models for Uganda and South Africa respectively as new data. We test for the ARCH effect in both sets of data. We need to test whether these residuals display a change in variance before applying volatility models.

7.1 ARCH effect in ARIMA(0,0,0)(1,1,1)[12] 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 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:

$$Q = T(T+2)\sum_{k=1}^{L}(T-k)^{-1}r_k^2, \tag{7.1}$$

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; Patterson and Ashley, 2000; Wei, 2007)

While testing for autocorrelation in part 5.1.6.1, the residuals were statistically uncorrelated, which was emphasised by the ACF. However, we are not sure if these residuals are identically independently distributed through time. We use visual inspection of the time series plot of residuals in figure 5.10 and find no 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, we plot the ACF of squared residuals as shown in Figure 7.1. There are no significant spikes all through the 20 lags that are considered. Therefore, 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.

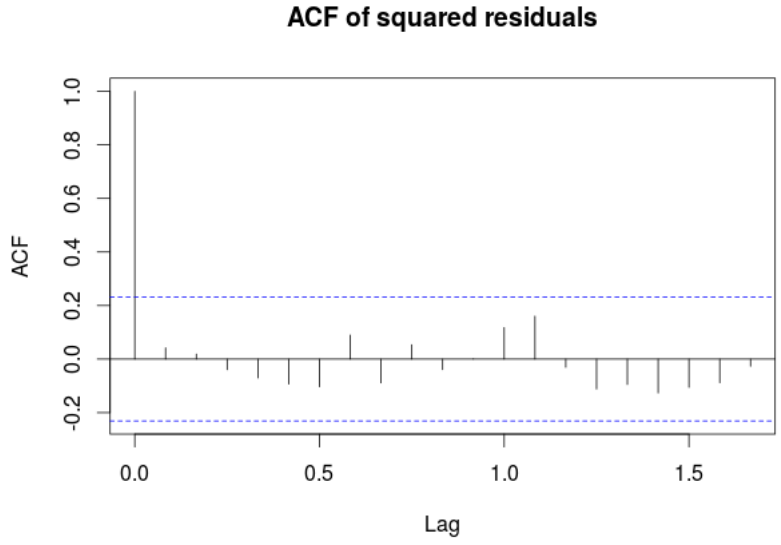

FIGURE 7.1: ACF of in-sample squared residuals.

We also consider a more "formal" method to test for ARCH effects. We use the *ArchTest()* function in R at lag 4. The null hypothesis of no ARCH is tested against the alternative hypothesis of ARCH effect at a 5% level of significance. A

p-value of 0.3299 is found which is greater than 0.05. We fail to reject the null hypothesis and conclude that there is no ARCH effect in the squared residuals of the ARIMA model.

Due to these realisations, we do not find it necessary to proceed with volatility testing. Results show that the monthly electricity demand data in Uganda does not exhibit changing variance and autocorrelation. Therefore, the originally chosen ARIMA(0,0,0)(1,1,1)[12] is the best model for making predictions for upto 12 months for the monthly electricity demand in Uganda. Policy makers in Uganda can use this model to make predictions of monthly electricity demand in Uganda.

7.2 ARCH effect in ARIMA(1,0,1)(0,1,0)[365] residuals

By visual inspection of the time series plot of residuals in Figure 5.18 there exists a tendency of large (small) absolute values of the residual process being followed by other large (small) absolute values. This is evidence of an ARCH processes. To certainly confirm presence of ARCH effects, we plot the ACF of squared residuals as shown in Figure 7.2. There are many significant spikes, indicating the existence of dependency in the residuals. This means the variance of residuals is conditional on its history.

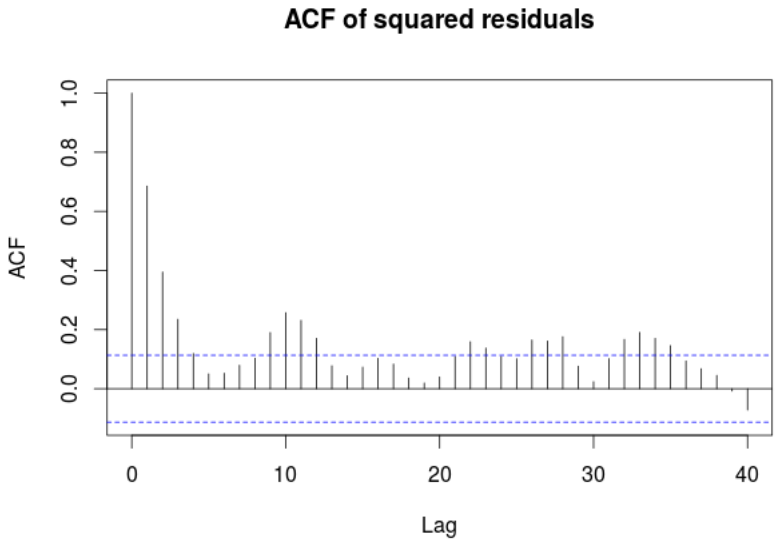

FIGURE 7.2: ACF of out-of-sample squared residuals.

Using the formal ARCH test in R, a p-value $< 2.2e-16$ is returned which is less than 0.05. Therefore, we reject H_0 and conclude 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 further analysis using a volatility model such as GARCH where the variances themselves are modelled as an AR(p) model. Since both positive and negative changes are observed in the daily electricity load this is another factor for considering a non linear model to analyse the errors.

7.2.1 Model order selection

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

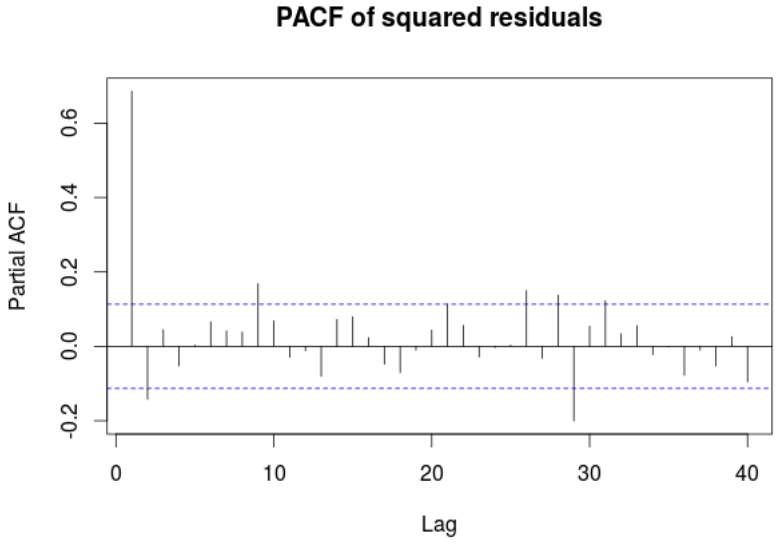

FIGURE 7.3: PACF of squared residuals.

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;

Model	A	B	S	H-Q
ARCH(1)	-0.60208	-0.56505	-0.60228	-0.58726
ARCH(2)	-0.59576	-0.54637	-0.59610	-0.57599
GARCH(1,1)	-0.58731	-0.53792	-0.58765	-0.56754
EGARCH(1,1)	-0.59044	-0.52871	-0.59098	-0.56574
Standard GARCH(1,1)	-1.4002	-1.3261	-1.4010	-1.3706

TABLE 7.1: Information criteria for different models

The best model from the analysis is the standard GARCH(1,1) relative to all the other suggested models. It has the lowest values for all information criteria. The coefficients of the model are shown in Table 7.2:

	μ	ω	α_1	β_1
Parameter	-0.020456	0.001184	0.342303	0.656697
s.e.	0.010219	0.001115	0.073618	0.072525

TABLE 7.2: Coefficients for the standard GARCH(1,1) model

Therefore, the best chosen model is:

$$\sigma_t^2 = 0.001184 + 0.342303\varepsilon_{t-1}^2 + 0.656697\sigma_{t-1}^2. \tag{7.2}$$

7.2.2 Forecasting volatility with standard GARCH(1,1)

We then use the identified model to make volatility forecasts. The interest of the study is to make forecasts that cover a time horizon of 30 days ahead. The sigma and series forecasts developed are shown in Table 7.3

The chosen standard GARCH(1,1) model forecasts the volatility in residuals for the demand of electricity as displayed in Figure 7.4.

7.2.3 Autocorrelation of standard GARCH(1,1) residuals

When we fit the standard GARCH(1,1) model to the data, among the results returned is the test for autocorrelation on the model's standardised residuals. Like in the case of ARIMA models, autocorrelation is tested with a null hypothesis of

Time	Series	Sigma	Time	Series	Sigma
T+1	0.003799	0.1062	T+16	-0.018753	0.1152
T+2	-0.005095	0.1068	T+17	-0.018758	0.1157
T+3	-0.010483	0.1075	T+18	-0.018761	0.1163
T+4	-0.013747	0.1081	T+19	-0.018762	0.1168
T+5	-0.015725	0.1088	T+20	-0.018764	0.1173
T+6	-0.016923	0.1094	T+21	-0.018764	0.1178
T+7	-0.017649	0.1100	T+22	-0.018765	0.1183
T+8	-0.018089	0.1106	T+23	-0.018765	0.1188
T+9	-0.018356	0.1112	T+24	-0.018765	0.1193
T+10	-0.018517	0.1118	T+25	-0.018765	0.1198
T+11	-0.018615	0.1124	T+26	-0.018765	0.1203
T+12	-0.018674	0.1129	T+27	-0.018765	0.1208
T+13	-0.018710	0.1135	T+28	-0.018765	0.1213
T+14	-0.018732	0.1141	T+29	-0.018765	0.1218
T+15	-0.018745	0.1146	T+30	-0.018765	0.1222

TABLE 7.3: Standard GARCH(1,1) volatility forecasts for 30 days ahead.

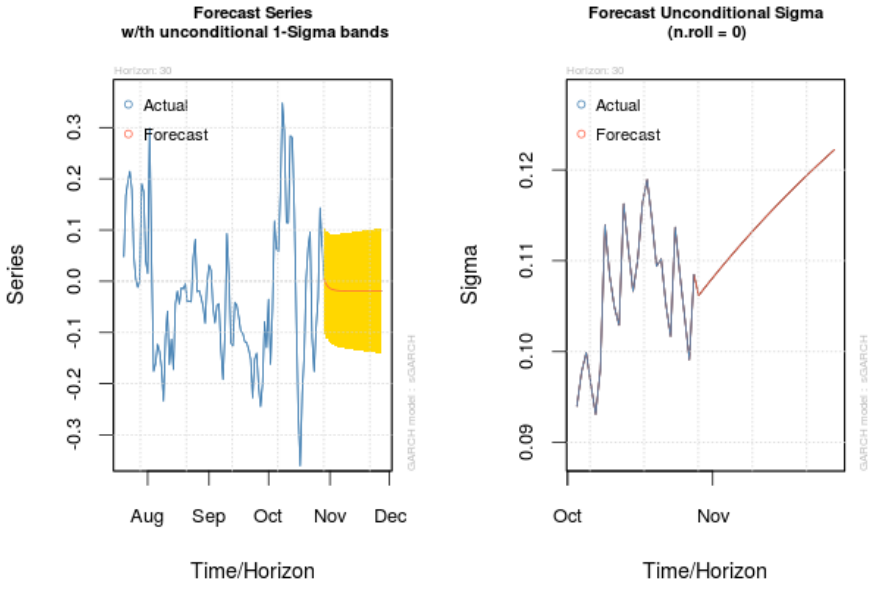

FIGURE 7.4: Volatility forecasts from the standard GARCH(1,1) model.

no autocorrelation, against the alternative of autocorrelation. Table 7.4 shows a sample of the tested lags and their p-values all greater than 0.05.

Graphically, figure 7.5 shows the ACF of the residuals. All significant spikes are within the estimated boundaries. Therefore, there is no evidence of autocorrelation in the residuals of the chosen standard GARCH(1,1) model.

Lag	statistic	p-value
1	0.1554	0.6934
5	0.5903	1.0000
9	2.4577	0.9555

TABLE 7.4: Autocorrelation test for different lags

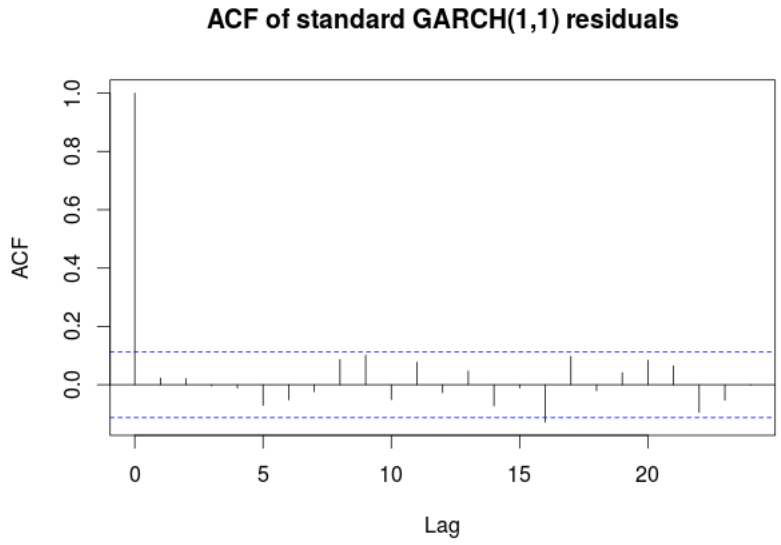

FIGURE 7.5: ACF showing no evidence of autcorrelation.

7.2.4 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, using two degrees of freedom. Therefore we fail to reject the null and conclude that there is no dependency in the squared residuals. Table 7.5 shows a sample of the tested lags.

Lag	statistic	p-value
1	0.0003894	0.9843
5	2.2068301	0.5704
9	4.1850309	0.5578

TABLE 7.5: Dependency test for different lags

Graphically, Figure 7.6 shows the ACF of squared residuals. None of the lags shows significant spikes. Therefore, there is no evidence of dependency in the squared residuals of the chosen standard GARCH(1,1) model.

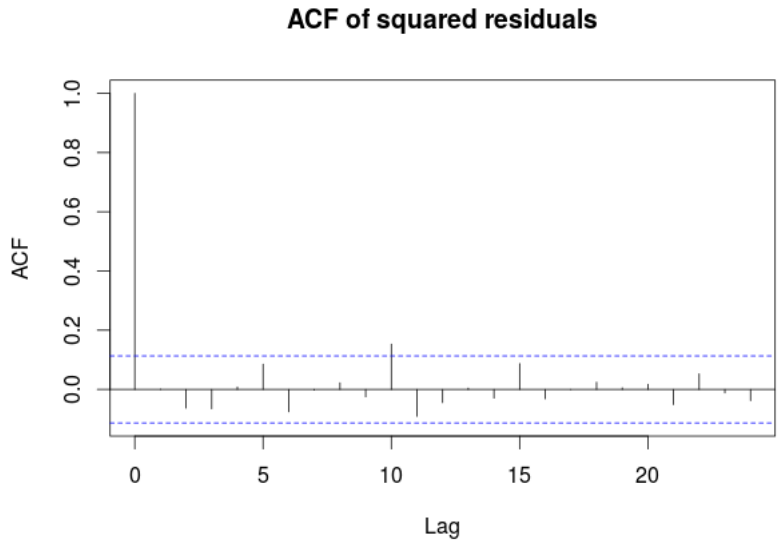

FIGURE 7.6: ACF showing no evidence of dependency.

Likewise, a time series plot of the residuals as shown in Figure 7.7 shows no evidence of clustering; low variations followed by low variations and high variations followed by high variations. The variance in the plot is fairly constant. Therefore, there is no evidence of ARCH effects in the residuals.

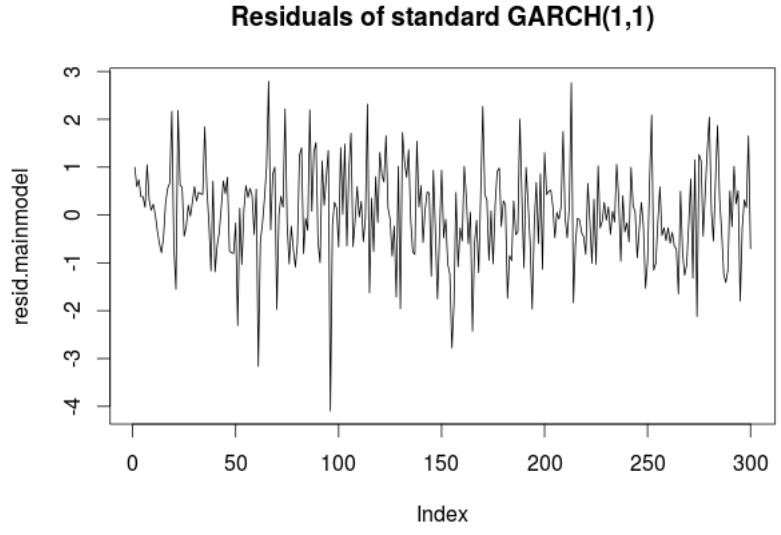

FIGURE 7.7: Time series plot of residuals.

7.2.5 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 main objective of the study was to find a model that fits the data best and can be used by policy and decision makers in the electricity sector to make the most accurate forecasts possible. If the aim 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 but a few), work best when dealing with financial data. This is because deeper features such as, leverage, convergence, persistent variance, are considered when using these models. In our study, after fitting a standard GARCH model, the residuals returned passed all diagnostic tests. That is why we did not find it fit to consider more complicated versions.
- From the tried models, an EGARCH was considered as seen in Table 7.1 , which returned information criteria values higher than the standard GARCH model.

Chapter 8

Holt-Winters Exponential Smoothing

Normally when forecasting, developed models are compared with a simple benchmark model. In this study, we compare the developed models with Holt-Winters model because of the seasonality components exhibited by both data sets. Most of the literature about the Holt-Winters Exponential Smoothing method used in this study is from Hyndman and Athanasopoulos (2014). This model was developed by both Holt (1957) and Winters (1960) who extended Holt's original Exponential Smoothing method to capture seasonality. The Holt-Winters seasonal method is made up of the forecast equation and three smoothing equations; for the level, trend, and the seasonal component, with their respective smoothing parameters. In this study, m is used to denote the period of the seasonality, for example, 4 for quarterly data, 12 for monthly data, and 52 for weekly data.

Seasonality can be viewed in two different ways;

- Additive seasonality: When using the Holt-Winters model, the additive method is used if the seasonal variation is roughly constant throughout the data. Here the seasonal component is expressed in absolute terms in the scale of the observed series, and in the level equation the series is seasonally adjusted by subtracting the seasonal component. The additive model is written as;

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t-m+h_m^+}, \tag{8.1}$$

where; the level equation (l_t) with smoothing parameter α shows a weighted average between the seasonally adjusted observation ($y_t - s_{t-m}$) and the non-seasonal forecast ($l_{t-1} + b_{t-1}$) for time t. It is defined as;

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}), \tag{8.2}$$

the trend equation (b_t) with smoothing parameter β^* is identical to Holt's linear method. The seasonal equation shows a weighted average between the current seasonal index, ($y_t - l_{t-1} - b_{t-1}$), and the seasonal index of the same season the year before (m time periods ago). It is defined as;

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}, \tag{8.3}$$

and the seasonal equation (s_t) with smoothing parameter γ is defined as;

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}. \tag{8.4}$$

Estimates of the seasonal indices used for forecasting should be from the final year of the sample, that is the purpose of having $h_m^+ = \lfloor (h-1) \text{mod m} \rfloor + 1$ as part of the equation.

- Multiplicative seasonality: With seasonal variations changing proportional to the level of the data, the multiplicative method is preferred. Here the seasonal component is expressed in percentages and the series is seasonally adjusted by dividing through by the seasonal component. A multiplicative model is written as;

$$\hat{y}_{t+h|t} = (l_t + hb_t)s_{t-m+h_m^+}, \tag{8.5}$$

where;

$$l_t = \alpha\frac{y_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1}), \tag{8.6}$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}, \tag{8.7}$$

$$s_t = \gamma\frac{y_t}{(l_{t-1} + b_{t-1})} + (1 - \gamma)s_{t-m}. \tag{8.8}$$

8.1 Application of Holt-Winters model to data

A time series described using an additive model with increasing or decreasing trend and seasonality can be forecast using Holt-Winters exponential smoothing for short-term forecasts. Holt-Winters exponential smoothing estimates the level, slope and seasonal component at the current time point, using their respective smoothing parameters. The parameters α, β and γ all have values between 0 and 1. Values close to 0 mean that relatively little weight is placed on the most recent observations when making forecasts of future values.

8.1.1 Uganda Monthly demand data.

Uganda monthly electricity demand data tests positive for both trend and monthly seasonality, therefore can be described using an additive model. To make forecasts of this data, we fit a predictive model to the data using the *HoltWinters()* function in R. We face a problem when applying Holt-Winters model to the available data. Although the data tested positive for monthly seasonality when using ARIMA models, the Holt-Winters models does not recognise the seasonality. When tested, the data returns an error *time series has no or less than 2 periods*, which is not actually true.

We try fitting an ordinary Holt's exponential smoothing model to the data, by setting the seasonality smoothing parameter to "false". The estimated value of $\alpha = 0.3968748$, and $\beta = 0.308989$. These are both low values, implying that both estimates of the current value of the level and slope of the trend component are based mostly upon observations far in the past in the time series. This is realistically true because from the time series plot in Figure 3.2, the level and the slope of the time series both do not change quite a lot over time.

We can see from Figure 8.1 that the in-sample forecasts do not match the observed values well. They tend to take a different nature all together. This means the fitted model does not define the nature of the data well. This concurs with the initial problem faced where the model fails to capture the seasonal nature of the data. With this limitation, we do not go ahead to evaluate the accuracy of the model.

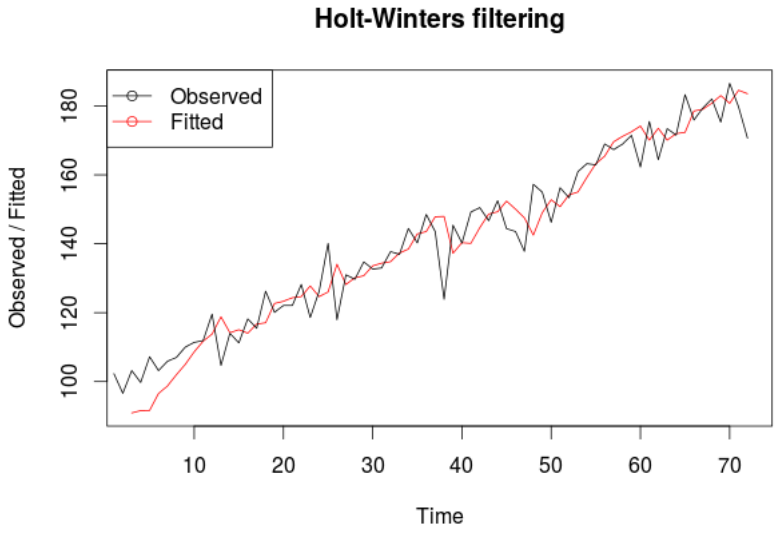

FIGURE 8.1: Forecasting monthly electricity demand in Uganda using Holt's exponential smoothing.

8.1.2 South Africa daily demand data.

Although the data does not show a trend component, it shows clear seasonality and tests positive for annual seasonality. Therefore, an additive model can be employed to make forecasts. Testing the data using the *HoltWinters()* function in R returns the same error as it does for Uganda monthly data. We also fit an ordinary Holt's exponential smoothing model to the test set of South Africa daily electricity demand data.

The estimated value of $\alpha = 1$, and $\beta = 0.01366412$. This means that the estimate of the current value of the level is based on recent observations but that of the trend component is based mostly upon observations very far in the past. This is true because the data shows no evidence of a trend in Figure 3.7, but the level changes quite a lot over time.

Figure 8.2 shows that the in-sample forecasts match the observed values perfectly. This shows evidence that the fitted model defines the nature of the data well. We go ahead to forecast demand for the next 300 days and compare it with the test set. All forecasts for 300 days can not be displayed in a table but they follow a decreasing pattern as we go further in time. The maximum demand is 5334310 which appears at step $h = 1$ and the minimum is 4242225 at $h = 300$. This is

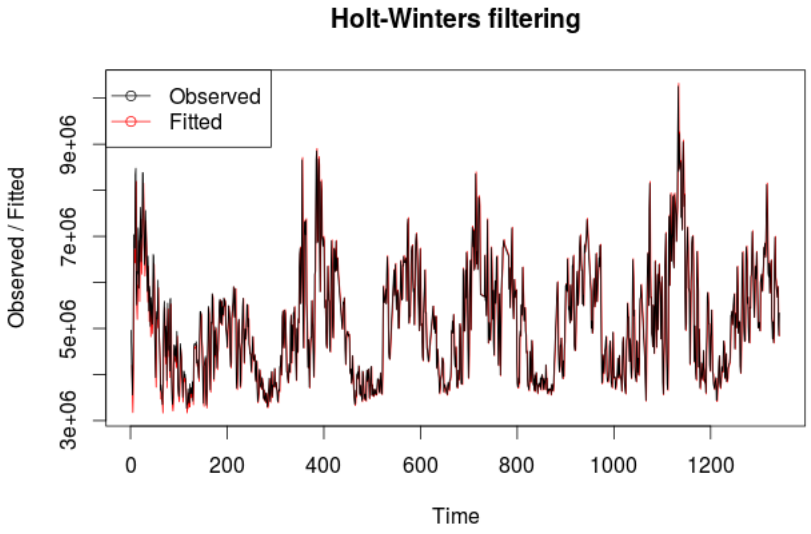

FIGURE 8.2: Forecasting daily electricity demand in South Africa using Holt's exponential smoothing.

a clear characteristic of exponential smoothing but it is not a realistic pattern, especially with the given seasonal data. Forecasts are displayed in Figure 8.3.

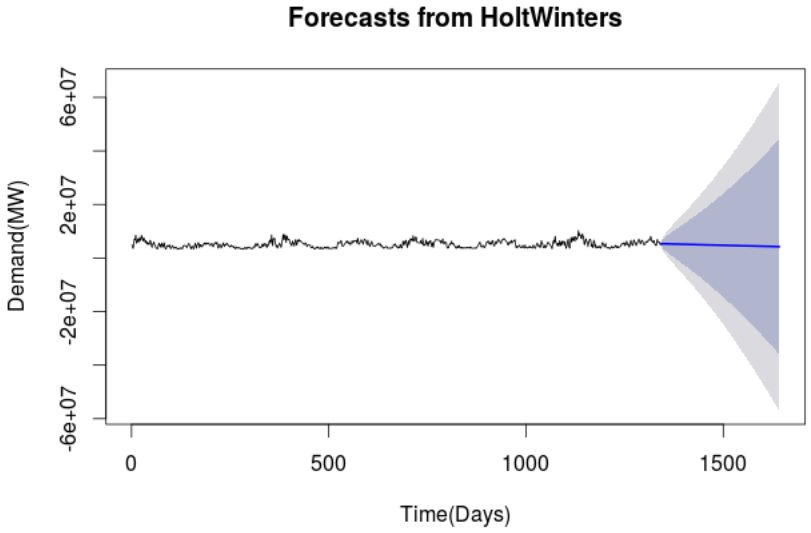

FIGURE 8.3: Out-of-sample forecasts for 300 days ahead using Holt's exponential smoothing.

Visual inspection of Figure 8.3 shows that forecasts do not follow the original pattern of the data. Therefore, we do not go on to conduct model diagnosis, but rather conclude that a Holt's exponential smoothing model is still not a good choice for South Africa daily electricity demand data.

Chapter 9

Conclusion

In this study, we have investigated models for forecasting electricity demand in both Uganda and South Africa. For Uganda, we used peak monthly data collected from UMEME, from January 2008 to December 2013. We faced a problem of limited data from Uganda because of the economic condition of the country. It is characterised with poor data management and storage policies. The furthest UMEME could go back in time was 2008, and the data was received in May 2014. Hence the duration the data covered. For South Africa, daily data from 1-January-2004 to 30-June-2008 was collected from ESKOM, for an anonymous region in the country. We faced a problem of constrained details about the data. We could not explain explicitly why the data had its features because of the confidentiality that was required of us when receiving the data.

The main purpose of the study was to investigate how better volatility models forecast electricity demand, compared to linear models. Linear models were applied to electricity demand data from the two countries under study, and residuals were modelled using volatility models like GARCH. The linear models were used to forecast 12 months ahead electricity demand for Uganda and 300 days ahead electricity demand for South Africa, after dividing the data into the training and test sets. Both data sets were tried with a Holt's exponential smoothing model but it was not a good choice because it did not capture the nature of the data clear enough for it too be considered for forecasting.

A seasonal ARIMA(0,0,0)(1,1,1)[12] model with a drift was found to be the most suitable model to make 12 months-ahead predictions of peak monthly electricity demand for Uganda. The residuals of the seasonal ARIMA(0,0,0)(1,1,1)[12] model

were tested for autocorrelation, normality, and ARCH effects. These residuals showed no evidence of autocorrelation, non-normality, and ARCH effects. Therefore, we did not proceeded with the application of volatility models. This indicated that the seasonal ARIMA(0,0,0)(1,1,1)[12] model (with a drift) with RMSE of 4.872027 and MAPE of 2.347028 and homoskedastic error terms is the most appropriate model for forecasting monthly electricity in Uganda. The monthly behaviour of forecast values of Ugandan data depicts that the electricity demand will increase in the following year with fluctuations between 175GW and 195GW through the tested months. The forecast model and the forecast graph reveal that electricity demand is increasing with time, with the highest demand (195GW) predicted during October and December. This means that the government of Uganda must take effective steps to increase the electricity production through construction of many power plants and implementing different energy sources. This will help improve the economical status of the country by meeting the increasing demand of electricity.

For South African data, a seasonal ARIMA(1,0,1)(0,1,0)[365] (with a drift) was found to be the most suitable model. However, it returned autocorrelated non-normal residuals which also tested positive for ARCH effects. Therefore, there was room to improve 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 predict the conditional variances. The non linear issues of variances were handled appropriately through the fitted standard GARCH models. This is because a model diagnosis run after fitting the standard GARCH model returned homoskedastic, non-autoocorrelated residuals. Therefore, we conclude that it is always a good practice to test the volatility of variances and standard deviations after fitting linear models to improve the accuracy of the forecasts. These models provide flexibility to coexist with other models. The combination of Seasonal ARIMA and GARCH gives more accurate forecasts than just a linear model. R-programming is well suited for modelling and forecasting electricity demand in this case.

This study is similar to one carried out by Yasmeen and Sharif (2014), where monthly electricity consumption (EC) for Pakistan was studied and a model developed to forecast four years ahead. 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 EC in Pakistan. However, this is similar to the results obtained using Uganda monthly demand data; linear models out performed non linear volatility models. Another similar study was carried out by Sigauke and Chikobvu (2011), who studied the prediction of daily peak electricity demand in South Africa using three volatility forecasting models; a seasonal auto regressive 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. Similar to our results, the non linear volatility model out performed linear models when dealing with daily demand data. The Reg-SARIMA-GARCH model produced better forecast accuracy with a MAPE of 1.42%, for out of sample prediction of daily peak demand.

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. On the other hand however, weekly, daily, hourly or any other shorter-period historical data requires application of volatility models. This is because during short periods 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.

Areas for further study would include;

- Fitting models whose distributions accommodate non-normality in case the available time series data is not 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 develope one model that suits the data perfectly.

Appendix A

Codes used in R

This was the code used to plot the histogram with an overlaid normal curve as developed by Coghlan (2014)

```
Avril Coghlan's function
plotForecastErrors <- function(forecasterrors) {
    # make a histogram of the forecast errors:
    mybinsize <- IQR(forecasterrors)/4
    mysd <- sd(forecasterrors)
    mymin <- min(forecasterrors) - mysd * 5
    mymax <- max(forecasterrors) + mysd * 3
    # generate normally distributed data with mean 0 and standard deviation
    # mysd
    mynorm <- rnorm(10000, mean = 0, sd = mysd)
    mymin2 <- min(mynorm)
    mymax2 <- max(mynorm)
    if (mymin2 < mymin) {
        mymin <- mymin2
    }
    if (mymax2 > mymax) {
        mymax <- mymax2
    }
    # make a red histogram of the forecast errors, with the normally
    # distributed data overlaid:
    mybins <- seq(mymin, mymax, mybinsize)
```

```
    hist(forecasterrors, col = "red", freq = FALSE, breaks = mybins)
    # freq=FALSE ensures the area under the histogram = 1 generate normally
    # distributed data with mean 0 and standard deviation mysd
    myhist <- hist(mynorm, plot = FALSE, breaks = mybins)
    # plot the normal curve as a blue line on top of the histogram of forecast
    # errors:
    points(myhist$mids, myhist$density, type = "l", col = "blue", lwd = 2)
}
```

Appendix B

Other relevant tables

Table B.1 below represents the ACF and PACF coefficients for lag 1 to 20 of the differenced Uganda monthly electricity demand data. To calculate these coefficients, we used the acf() and pacf() functions in R and set "plot=FALSE" in both functions.

ACF and PACF coefficients

Lag	ACF	PACF	Lag	ACF	PACF
1	-0.624	-0.624	11	0.063	-0.045
2	0.205	-0.301	12	0.022	-0.011
3	-0.047	-0.123	13	-0.133	-0.131
4	-0.041	-0.137	14	0.176	-0.039
5	0.084	-0.017	15	-0.046	0.180
6	-0.195	-0.246	16	-0.082	-0.071
7	0.251	-0.017	17	0.162	0.139
8	-0.219	-0.099	18	-0.191	-0.064
9	0.203	0.075	19	0.099	-0.098
10	-0.159	0.022	20	-0.040	-0.020

TABLE B.1: ACF and PACF coefficients for first order differenced data.

References

Ratnadip Adhikari and RK Agrawal. An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*, 2013.

South Africa. *Energy Security Master Plan: Electricity, 2007-2025, Department of Minerals and Energy (1996)*. Department of Minerals and Energy, 2007. URL https://books.google.com.gh/books?id=exvPoAEACAAJ.

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.

H Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, pages 267–281. Akademinai Kiado, 1973.

Jelloul Allal and Mohammed Benmoumen. Parameter estimation for arch (1) models based on kalman filter. *Applied Mathematical Sciences*, 8(56):2783–2791, 2014.

Miguel A Arranz. Portmanteau test statistics in time series. *Time Orientated Language*, pages 1–8, 2005.

James Baanabe. Energy supply in uganda. In *Presentation by James Baanabe, Ministry of Energy and Mineral Resources Uganda. National workshop on promoting sustainable transport solutions for East Africa*, volume 1, 2012.

Naval Bajpai. *Business statistics*. Pearson Education India, 2009.

Peter Bartlett. Introduction to time series analysis. lecture 6.

Dale E Berger. Introduction to multiple regression. *Claremont Graduate University. Retrieved on Dec*, 5:2011, 2003.

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

Alok Bhargava. On the theory of testing for unit roots in observed time series. *The Review of Economic Studies*, 53(3):369–384, 1986.

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

M Bögl, P Filzmoser, T Gschwandtner, S Miksch, W Aigner, A Rind, and T Lammarsch. Visually and statistically guided imputation of missing values in univariate seasonal time series. 2015.

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

Tim Bollerslev, Robert F Engle, and Daniel B Nelson. Arch models. *Handbook of econometrics*, 4.

Tim Bollerslev, Ray Y Chou, and Kenneth F Kroner. Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52(1): 5–59, 1992.

George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252, 1964.

George EP Box and Gwilym M Jenkins. *Time series analysis: forecasting and control, revised ed.* Holden-Day, 1976.

George EP Box, Gwilym M Jenkins, and Gregory C Reinsel. *Time series analysis: forecasting and control*, volume 734. John Wiley & Sons, 2011.

Charles Henry Brase and Corrinne Pellillo Brase. *Understandable statistics: concepts and methods*. Cengage Learning, 2011.

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

C. Lee, 2006. Maximum likelihood estimation. http://econ.nsysu.edu.tw/ezfiles/124/1124/img/Chapter17_MaximumLikelihoodEstimation.pdf, Accessed July 2015.

C. Lee (2006). Maximum likelihood estimation. http://econ.nsysu.edu.tw/ezfiles/124/1124/img/Chapter17_MaximumLikelihoodEstimation.pdf, Accessed July 2015.

Nunzio Cappuccio and Diego Lubian. The fragility of the kpss stationarity test. *Statistical Methods and Applications*, 19(2):237–253, 2010.

Rui Castro. Lectures 2 and 3-goodness-of-fit (gof) tests. 2013.

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

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

Christopher Chatfield. The analysis of time series : an introduction. 1996. Includes indexes.

S Gonzales Chavez, J Xiberta Bernat, and H Llaneza Coalla. Forecasting of energy production and consumption in asturias (northern spain). *Energy*, 24(3):183–198, 1999.

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

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

Onyechi Pious Chukwukelue, Ezeliora Chukwuemeka Daniel, Chinwuko Emmanuel Chuka, and Aguh Patrick Sunday. Time series and statistical analysis of plastic yield: A case study of finoplastika industries ltd, nigeria. 2013.

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

Avril Coghlan. A little book of r for time series. *Wellcome Trust Sanger Institute: Cambridge, UK*, 2014.

Steve R Conradie and LJM Messerschmidt. *A Symphony of Power: The Eskom Story*. C. van Rensburg Publications, 2000.

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

Artur CB da Silva Lopes. Deterministic seasonality in dickey–fuller tests: should we care? *Empirical Economics*, 31(1):165–182, 2006.

Richard A Davis and William Dunsmuir. Maximum likelihood estimation for ma (1) processes with a root on or near the unit circle. *Econometric theory*, 12(01): 1–29, 1996.

José Ruy Porto de Carvalho, Eduardo Delgado Assad, Aryeverton Fortes de Oliveira, and Hilton Silveira Pinto. Annual maximum daily rainfall trends in the midwest, southeast and southern brazil in the last 71 years. *Weather and Climate Extremes*, 5:7–15, 2014.

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.

Larry Dwyer, Alison Gill, and Neelu Seetaram. *Handbook of research methods in tourism: Quantitative and qualitative approaches*. Edward Elgar Publishing, 2012.

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

Benjamin Eden. *A Course in Monetary Economics: Sequential Trade, Money, and Uncertainity*. John Wiley & Sons, 2008.

Olga Efimova. *Energy Commodity Volatility Modelling using GARCH*. PhD thesis, University of Calgary, 2013.

Alan C Elliott, Linda S Hynan, Joan S Reisch, and Janet P Smith. Preparing data for analysis using microsoftexcel. *Journal of investigative medicine*, 54(6): 334–341, 2006.

Emma Onyango. Uganda to increase power generation, east african business week. http://www.busiweek.com/index1.php?Ctp=2&pI=470&pLv=3&srI=53&spI=20&cI=11, Accessed August 2014.

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

EPRC. Economic policy research centre-the ugandan economy today. http://www.eprc.or.ug/pdf_files/ugandaneconomytodayvol15no4.pdf, Accessed February 2015.

ERA website. Electricity regulatory authority. http://www.era.or.ug/, Accessed July 2015.

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

Julian Faraway and Chris Chatfield. Time series forecasting with neural networks: a comparative study using the air line data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(2):231–250, 1998.

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.

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

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

Sajal Ghosh. Univariate time-series forecasting of monthly peak demand of electricity in northern india. *International Journal of Indian Culture and Business Management*, 1(4):466–474, 2008.

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

Michael Haag. Autocorrelation of random processes. *Connexions, Tech. Rep,[Online]. Available: http://cnx. rice. edu/content/m10676/latest*, 2005.

James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

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.

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

M.S. Heracleous. *Volatility Modeling Using the Studentst Distribution*. PhD thesis, Virginia Polytechnic Institute and State University, 2003.

Hansjörg Herr, Milka Kazandziska, and Silke Mahnkopf-Praprotnik. The theoretical debate about minimum wages. Technical report, Global Labour University Working Paper, 2009.

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

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

Charles C Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *Carnegie Institute of Technology, Pittsburgh*, 52, 1957.

James Honaker and Gary King. What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581, 2010.

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

Joop J Hox and Hennie R Boeije. Data collection, primary vs. secondary. *Encyclopedia of social measurement*, 1:593–599, 2005.

C. Hurlin. *Maximum Likelihood Estimation*. PhD thesis, University of Orleans, 2013.

Clifford M Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.

William R Huss. Comparative analysis of company forecasts and advanced time series techniques using annual electric utility energy sales data. *International Journal of Forecasting*, 1(3), 1985.

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

Rob J Hyndman, George Athanasopoulos, Slava Razbash, Drew Schmidt, Zhenyu Zhou, Yousaf Khan, Christoph Bergmeir, Earo Wang, and Maintainer Rob J Hyndman. Package forecast. 2015.

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

Investopidia-Volatility. Volatility definition. http://www.investopedia.com/terms/v/volatility.asp, Accessed April 2015.

Mortaza Jamshidian, Siavash Jala Jalal, and Camden Jansen. Missmech: an r package for testing homoscedasticity, multivariate normality, and missing completely at random (mcar). *Journal of Statistical Software*, 56(6), 2014.

G Janacek. Time series analysis forecasting and control. *Journal of Time Series Analysis*, 31(4):303–303.

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.

JRC Website. European commission, quantitative versus qualitative methods. http://forlearn.jrc.ec.europa.eu/guide/4_methodology/meth_quanti-quali.htm, Accessed August 2014.

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

Kabiza Website. The weather and climate in uganda - east africa. http://kabiza.com/kabiza-wilderness-safaris/the-weather-and-climate-in-uganda-east-africa/, Accessed February 2016.

Mohammad Karamouz, Sara Nazif, and Mahdis Falahi. *Hydrology and hydroclimatology: principles and applications*. CRC Press, 2012.

Michael F Keating. Global best practice (s) and electricity sector reform in ugandaf. 2006.

Oliver N Keene. The log transformation is special. *Statistics in medicine*, 14(8): 811–819, 1995.

Chandra Bhushan Kesavabhotla and Vinay Babu. Application of analytics using r programming for forecasting day-ahead electricity demand. *IJCSI*, 2012.

J Kiche, J Aduda, KHO Athiany, Apudo BO, and A Wanjoya. Comparison of survival models and estimation of their parameters with respect to mortality in a given population. *Mathematical Theory and Modeling*, 4(7):164–176, 2014.

Paul DW Kirk and Michael PH Stumpf. Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data. *Bioinformatics*, 25(10):1300–1306, 2009.

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.

Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1):159–178, 1992.

Pengfei Li. Box cox transformations: An overview. *presentation, http://www. stat. uconn. edu/˜ studentjournal/index_files/pengfi_s05. pdf*, 2005.

Roderick JA Little. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404): 1198–1202, 1988.

Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2014.

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

Helmut Lütkepohl and Fang Xu. The role of the log transformation in forecasting economic variables. *Empirical Economics*, 42(3):619–638, 2012.

David Maradiaga, Aude Pujula, and Hector Zapata. Exploring time series data properties in sas®. In *SAS Global Forum 2013, Paper 456-2013, Statistics and Data Analysis*. Citeseer, 2013.

Eric Martinot et al. *Renewables 2005: Global status report*. Worldwatch Institute Washington, DC, 2005.

Joseph Mawejje, Ezra Munyambonera, and Lawrence Bategeka. Uganda electricity sector reforms and institutional restructuring. *Economic Policy Research Centre Research Series*, (89), 2012.

Joseph Mawejje, Ezra Munyambonera, and Lawrence Bategeka. Powering ahead: The reform of the electricity sector in uganda. *Energy and Environment Research*, 3(2):p126, 2013.

Alex Maynard. Econometric theory, by James Davidson, Blackwell Publishers, 2000, 2003.

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

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.

Richard Meese and John Geweke. A comparison of autoregressive univariate forecasting procedures for macroeconomic time series. *Journal of Business & Economic Statistics*, 2(3):191–200, 1984.

Sitansu S Mittra. *Database Performance Tuning and Optimization: Using Oracle*, volume 1. Springer Science & Business Media, 2002.

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

Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

Steffen Moritz, Alexis Sardá, Thomas Bartz-Beielstein, Martin Zaefferer, and Jörg Stork. Comparison of different methods for univariate time series imputation in r. *arXiv preprint arXiv:1510.03924*, 2015.

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

Daniel B Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.

New Vision Website. Demand for electricity up-by ibrahim kasita, uganda. http://www.newvision.co.ug/news/634884-Demand-for-electricity-up.html, Accessed July 2014.

Heino Bohn Nielsen. Non-stationary time series and unit root testing. 2006.

Salah A Nusair et al. Testing the validity of purchasing power parity for asian countries during the current float. *Journal of economic development*, 28(2): 129–147, 2003.

Ludger Odenthal, Zbigniew Zimny, and Karl P Sauvant. *Foreign direct investment in Africa: Performance and potential*. UN, 1999.

Francis Kwabena Oduro-Gyimah. *Time Series Analysis of Microwave Transmission on the Yeji-Salaga Line-Of-Sight Telecommunication Link*. PhD thesis, Department of Mathematics, Kwame Nkrumah University of Science and Technology, 2011.

Robert B O'hara and D Johan Kotze. Do not log transform count data. *Methods in Ecology and Evolution*, 1(2):118–122, 2010.

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.

J Osborne. Notes on the use of data transformations. *Practical Assessment, Research and Evaluation*, 9:42–50, 2005.

Jason W Osborne. Improving your data transformations: Applying the box cox transformation. *Practical Assessment, Research & Evaluation*, 15(12):1–9, 2010.

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 & Business Media, 2000.

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

Pierre Perron and Gabriel Rodríguez. Searching for additive outliers in nonstationary time series*. *Journal of Time Series Analysis*, 24(2):193–220, 2003.

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

WH Press, SA Teukolsky, WT Vetterling, and BP Flannery. Statistical description of data. *Numerical Recipes in C: The Art of Scientific Computing*, page 636, 1992.

DHR Price and JA Sharp. A comparison of the performance of different univariate forecasting methods in a model of capacity acquisition in uk electricity supply. *International Journal of Forecasting*, 2(3):333–348, 1986.

Svetlozar T Rachev, John SJ Hsu, Biliana S Bagasheva, and Frank J Fabozzi. *Bayesian methods in finance*, volume 153. John Wiley & Sons, 2008.

Srinivasa Rao Rallapalli and Sajal Ghosh. Forecasting monthly peak demand of electricity in india-a critique. *Energy Policy*, 45:516–520, 2012.

Suhasini Subba Rao. A course in time series analysis. Technical report, Technical report, Texas A&M University, 2008.

Nancy Reid. Likelihood inference. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5):517–525, 2010.

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

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

Peter M Robinson. Conditional-sum-of-squares estimation of models for stationary time series with long memory. *Lecture Notes-Monograph Series*, pages 130–137, 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.

Patrick Royston. Remark as r94: A remark on algorithm as 181: The w-test for normality. *Applied Statistics*, pages 547–551, 1995.

Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Donald B Rubin. Multiple imputation after 18+ years. *Journal of the American statistical Association*, 91(434):473–489, 1996.

Eva Rubliková and Manulea Magalhăes Hill. Modelling time series with conditional heteroscedasticity. *Scientific papers of the University of Pardubice. Series D, Faculty of Economics and Administration. 10 (2006)*, 2006.

David Ruppert. *Statistics and data analysis for financial engineering*. Springer Science & Business Media, 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.

Safari Website. Uganda - weather & climate. https://www.safaribookings.com/uganda/climate, Accessed February 2016.

RM Sakia. The box cox transformation technique: a review. *The Statistician*, 1992.

SAPower. List of power stations in south africa — wikipedia, the free encyclopedia, 2015. URL https://en.wikipedia.org/w/index.php?title=List_of_power_stations_in_South_Africa&oldid=681621652. [Online; accessed 19-September-2015].

Peter Saundry. Energy profile of uganda. *The Encyclopedia of Earth*, 2009.

Joseph L Schafer. Multiple imputation: a primer. *Statistical methods in medical research*, 8(1):3–15, 1999.

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

Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*.

Neil Shephard. Distribution of the ml estimator of an ma (1) and a local level model. *Econometric Theory*, 9(03):377–401, 1993.

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

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

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

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

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.

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

Skyscrapercity. Uganda business, economy and infrastructure news. http://www.skyscrapercity.com/showthread.php?t=1137943&page=60, Accessed July 2015.

Jitkomut Songsiri, Joachim Dahl, and Lieven Vandenberghe. Maximum-likelihood estimation of autoregressive models with conditional independence constraints. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 1701–1704. IEEE, 2009.

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

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

Yong Tao, Xiangjun Wu, and Changshuai Li. Rawls' fairness, income distribution and alarming level of gini coefficient. *arXiv preprint arXiv:1409.3979*, 2014.

Thorsten Thadewald and Herbert Büning. Jarque–bera test and its competitors for testing normality–a power comparison. *Journal of Applied Statistics*, 34(1): 87–105, 2007.

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

John Turkson and Norbert Wohlgemuth. Power sector reform and distributed generation in sub-saharan africa. *Energy Policy*, 29(2):135–145, 2001.

UEDCL website. Uganda electricity distribution company limited. www.uedcl.co.ug/, Accessed July 2015.

UETCL website. Uganda electricity transmission company limited. www.uetcl.com/, Accessed July 2015.

Uganda. Annual report of the republic of uganda 2013. *Ministry of energy and mineral development*, page 13.

University of South Carolina. Forecasting and time series by joshua m. tebbs, department of statistics. http://www.stat.sc.edu/~tebbs/stat520/f13notes.pdf, Accessed February 2015.

Ilana Van Wyk. tata ma chance: On contingency and the lottery in post-apartheid south africa. *Africa*, 82(01):41–68, 2012.

Joachim Vandekerckhove, Dora Matzke, and Eric-Jan Wagenmakers. Model comparison and the principle. *The Oxford Handbook of Computational and Mathematical Psychology*, page 300, 2015.

R Vasudeva and J Vasantha Kumari. On general error distributions. 2013.

Rangaswamy Vedavalli. *Energy for Development: Twenty-first Century Challenges of Reform and Liberalization in Developing Countries*. Anthem Press, 2007.

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

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

VSC website. Vanguard software corporation, forecasting and planning solutions. http://www.vanguardsw.com/101/what-is-forecasting.htm, Accessed July 2014.

Njeri Wamukonya. Power sector reform in developing countries: mismatched agendas. *Energy Policy*, 31(12):1273–1289, 2003.

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

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

Yanjun Wang and Qun Liu. Comparison of akaike information criterion (aic) and bayesian information criterion (bic) in selection of stock–recruitment relationships. *Fisheries Research*, 77(2):220–225, 2006.

Simon P Washington, Matthew G Karlaftis, and Fred L Mannering. *Statistical and econometric methods for transportation data analysis*. CRC press, 2010.

Jinfeng Wei. *Time series modeling for terrain profiles*. ProQuest, 2007.

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

E. W Weisstein. Least squares fitting. 2002.

Peter R Winters. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342, 1960.

Arch G Woodside and Drew Martin. *Tourism management: analysis, behaviour and strategy*. Cabi, 2008.

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

Diethelm Wuertz and Helmut G Katzgraber. Precise finite-sample quantiles of the jarque-bera adjusted lagrange multiplier test. *arXiv preprint math/0509423*, 2005.

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

Muhammad Imran Yousuf. Using experts' opinions through delphi technique. *Practical assessment, research & evaluation*, 12(4):1–8, 2007.

Victor Zarnowitz. The accuracy of individual and group forecasts from business outlook surveys. *Journal of forecasting*, 3(1):11–26, 1984.

Eric Zivot and Jiahui Wang. *Modeling financial time series with S-Plus®*, volume 191. Springer Science & Business Media, 2007.