Forecasting models for operational and tactical requirements in electricity consumption:
The case of the Ferrochrome Sector in South Africa

LIVHUWANI NEDZINGAHE

Student number: 210100602

Submitted in partial fulfilment of the requirements
for the degree

Master of Science

Department of Statistics and Operations Research
School of Computational and Mathematical Sciences
Faculty of Science and Agriculture

UNIVERSITY OF LIMPOPO
October 2010

Supervisor: Prof. 'Maseka Lesaoana
Co Supervisor: Mr Ozias Ncube

DECLARATION

I, **Livhuwani Nedzingahe,** declare that the submitted work on *Forecasting model for operational and tactical requirements in electricity consumption: The case of the Ferrochrome Sector in South Africa* is my own work, and that I have not used any other than permitted reference sources or materials nor engaged in any plagiarism. All references and other sources used or quoted by me have been appropriately and duly acknowledged by means of complete references. I further declare that the work has not been submitted for the purpose of academic examination, either in its original or similar form, anywhere else.

Signed: _____

  Ms Livhuwani Nedzingahe

Date: _____

ACKNOWLEDGMENTS

I wish to extend my sincere gratitude to all those who assisted me in completing this thesis materially, logistically and through word of advice or otherwise.

Firstly, I would like to thank God Almighty for protecting and guiding me throughout the research and writing of this thesis. If it was not for Him, this would not have been possible.

To my supervisors Prof .'Maseka Lesaoana and Mr Ozias Ncube, thank you for encouraging me and making me see that there is light at the end of the tunnel. Through your support, this research has been presented in two international conferences, (i.e. the International Statistical Institute Conference held in Durban in August 2009; and Forecasting in an ever changing environment Conference organised by Eskom Academy of Learning held in Midrand in November 2009).

The Department of Statistics and Operations Research at the University of Limpopo, thank you for your support and inputs through lunch hour talks that were professionally organised and made this thesis where it is today. Ms Managa Mulalo, thanks a lot for your unveiling support and the assistance you have rendered until the completion of this research.

With all due respect and pride I take this opportunity to thank my parents Baba Vho- M.B na Mma Vho-T.G Mphaphuli for their overwhelming support throughout all my life. Mum and Dad thank you, for today your growing weary has been answered. You have been a true inspiration.

I am sincerely grateful to my husband Rudzani Nedzingahe and our sons Mudanamvula and Mirunzini for the great support you have given me.

To my manager Ms Lona Manzana and Eskom, particularly Key Sales and Customer Services department (KSACS), thank you for encouraging me to pursue my studies and for allowing me to make use of resources available in carrying out the study.

Finally, I would like to thank everyone who supported me directly or indirectly until the finish line of my thesis. May God richly bless you.

ABSTRACT

Forecasting electricity consumption is a challenge for most power utilities. In South Africa the anxiety posed by electricity supply disruption is a cause for concern in sustainable energy planning. Accurate forecasting of future electricity consumption has been identified as an essential input to this planning process. Forecasting electricity consumption has been widely researched and several methodologies suggested. However, various methods that have been proposed by a number of researchers are dependent on environment and market factors related to the scope of work under study making portability a challenge. The aim of this study is to investigate models to forecast short term electricity consumption for operational use and medium term electricity consumption for tactical use in the Ferrochrome sector in South Africa. An Autoregressive Moving Average method is suggested as an appropriate tool for operational planning. The Holt-Winter Linear seasonal smoothing method is suggested for tactical planning.

Keywords: *Forecasting, electricity consumption, operational planning, tactical planning, ARIMA, Holt-Winter Linear seasonal smoothing, Ferrochrome sector*

Table of contents   **Page**

List of Tables

List of Figures

ACRONYMNS

ACF	Auto correlation function
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial Neural Networks
AR	Autoregressive
ARCH	Autoregressive conditional heterescasticity
ARIMA	Autoregressive integrated moving average
BIC	Bayes Information Criterion
BP	Back Propagation
BP ANN	Back Propagation Artificial Neural Networks
Cr	Chromium
C_r	Criterion
df	degree of freedom
DSM	Demand Side Management
DV	Dependent (predictor) variable
D-W	Durbin Watson
EC	Electricity Consumption
ENRC	Eurasian Natural Resources Corporation
FeCr	Ferrochrome
GARCH	Generalised autoregressive conditional heteroscedasticity
GDP	Gross Domestic Product
IFM	International Ferro Metals
IV	Independent variable
KSACS	Key Sales and Customer Services
MAD	Mean Absolute Deviation
MAPE	Mean absolute percent error
MSE	Mean squared of error
PACF	Partial Autocorrelation Function
r^2	Correlation coefficient
R^2	Coefficient of determination
RMSE	Root mean square of error
RMSFE	Root mean square of factored error
RSS	Residual sum of squares of error

SARIMA	Seasonal Autoregressive Integrated Moving Average
SASSC	South African Stainless Steel Consumption
SEE	Standard error of estimate
SSE	Sum of squares of error
SST	Total Sum of squares
USd	Us dollar /Rand change
VAL	Various Algorithms (a saturation of Algorithms by definition)
VIF	Variance-inflation factor
WSSC	World Stainless Steel Consumption
WFeCrS	World FeCr Suppliers

CHAPTER 1: INTRODUCTION

1.1 Forecasting: An Overview

Since growth in the energy demand shows an increasing trend over the years, fears of energy supply disruptions have triggered much debate geared towards the necessity for sustainable energy planning. Accurate forecasting of future electricity consumption has been identified as an essential input to this process. However forecasting electricity consumption is still a challenge in areas where those forecasts are used (Balnac and Bokhoree, 2008).

Electricity suppliers rely strongly on consumption forecast for planning purposes. Thus, without the forecast it is difficult to make proper energy plans for the future. In his paper on *Overview of Forecasting Methodology*, Walonick (1993) makes reference to the notion that inability to forecast technological futures is a failure of nerve. He argues that it is difficult to accept the implications of an unsuccessful forecast. It actually takes courage to accept the implications more especially when the truth points to an inaccurate forecast.

Most forecasters use historical behavior to predict the future. This may be considered as true due to widely accepted view that the past might not be correct. When historians write about the past, they often interpolate their own viewpoint and biases. Information becomes indistinct and altered over time. At the end the past is a reflection of the current conceptual reference of which time itself comes into question.

Walonick (1993) further argues that since the future is filled will uncertainty, facts give way to opinions. Thus, "the facts of the past provide the raw materials from which the mind makes estimates of the future". All forecasts represent future outlook. These (forecasts) are usually surrounded by events that might or might not occur and these uncertainties pose severe challenges when predicting the future. Nowadays, the rate of change in events is moving more rapidly than in the past and this has made forecasting more challenging. Trends are now more uncertain and cannot be sustained for a longer period like it was in the past.

Hyndman (2008) seconded that advance spreadsheets based systems is a complete solution for budgeting, forecasting, *what-if* scenarios planning, reporting and analysis. These capabilities include a built-in financial intelligence and a business rule builds the formulas for users ensuring 100 percent accuracy. Another capability is the profitability analysis which allows one to create a *what-if* scenario by modifying sales growth rate and all the other relevant accounts measured as a total percentage of total sales, and the results change immediately.

It becomes a challenge to operate with this method because of *what-if* scenarios:
- they are based on assumed and fixed future conditions,
- they are highly subjective,
- they are not replicable or testable,
- there is no possible way of quantifying probabilistic uncertainty, and
- this lack of uncertainty statements leads to false sense of accuracy which is largely guess work.

On the other hand, data mining prediction methods which include classification and regression trees, neural networks and nearest neighbour and naïve Bayes, have both benefits and shortcomings. With these methods it is rare to provide uncertainty statements about individual predictions leading to "false" sense of accuracy. In addition they have limited interpretability of many models and software can be up to 50 times the cost of comparable statistical software. Due to complex computations and automation, basic statistical principles and methods are ignored.

The silver jubilee of the International Institute of Forecasters provided an opportunity to review 25 years progress on time series forecasting. De Gooijer and Hyndman (2006) reviewed a number of techniques that gave a historical tour of a quarter of a century developments in this area. The reviewers state that exponential smoothing methods originated from the work of Brown in 1959 and 1963, Holt in 1957 and 2004, Winters in 1960 and Pegels in 1969. Muth, in 1960, was the first to introduce a statistical foundation for simple exponential smoothing methods. This development was enhanced by Box and Jenkins in 1970 and later by Robert in 1982. In 1983 and 1986 Abraham and Ledolter showed that additive exponential smoothing forecasts are

special cases of the ARIMA models. Further independent work on simple exponential smoothing by Gardner and later by Snyder, in 1985 has provided a basis for state space models, (De Gooijer and Hyndman, 2006). Since 1980 work on forecasting has gained momentum. Applications of exponential smoothing methods were expanded to computer components, air passengers and production planning.

Variations of the original forecasting methods include modifications that deal with discontinuity, constrained smoothing forecasts, and normalisation of seasonal components (De Gooijer and Hyndman, 2006). One notable variation is the Theta method for which further research is still needed. Multivariate versions of the exponential smoothing methods were applied to the Israel tourism data by Pfeffermann and Allon in 1989. In 1986 Johnston and Harrison studied prediction intervals through the equivalence between exponential smoothing methods and statistical models.

That "every time series can be regarded as the realization of a stochastic process" is a contribution by Yule in 1927 (De Gooijer and Hyndman, 2006). Other scholars joined Yule to formulate the concept of autoregressive (AR) and moving average (MA) models that have now become known as ARMA models. The existing knowledge that emerged from this early work was published by Box and Jenkins in 1970 in their book entitled *Time Series Analysis: Forecasting and Control*, hence the well-known Box-Jenkins approach to time series analysis and forecasting. It is this approach that has integrated and popularised the use of ARIMA (autoregressive integrated moving average) models. There are several techniques and methods suggested as criteria to minimize one-step-ahead forecast errors in the search process of an ARIMA model. These include: Akaike's information criterion, Akaike's final prediction error and Bayes information criterion.

The vector ARIMA (VARIMA) model, which is a multivariate generalization of the univariate ARMA model, was first derived by Quenouille in 1957. Artificial neural networks (ANNs) have been successfully applied in forecasting with notable successes especially in high frequency financial data (De Gooijer and Hyndman, 2006).

Applications of ANNs have been compared to traditional techniques in forecasting such as the random walk. In forecasting daily electricity load profiles, dynamic ANNs have been used. The ARCH (autoregressive conditional heteroscedastic) models were introduced by Engle in 1982; the generalized ARCH (GARCH) model was studied by Taylor in 1987, and later in 1994 by Bollerslev, Engle and Nelson (De Gooijer and Hyndman, 2006).

1.2 Philosophical perspective

"The post-modern world is characterised by rapid and high impact changes. Both the frequency and the magnitude of the changes are high. This is further complicated by the irregularities in both the high frequency and high amplitude changes" (Maseema, 2009). Consumption pattern of electricity has been affected by some social, economical and environmental factors by which the pattern will form various seasonal, monthly, daily and hourly complex variations and these had been leading to the extensions of various models. Many attempts have been made to find the best estimation for electricity consumption.

According to Keyno, Ghader, Azade and Razmi (2009), studies have tried to forecast the consumption in two levels: (1) macro economic decision making and (2) engineering and middle management. Over the years, finding the best estimation for electricity consumption through forecasting by time series analysis had been researched by many authors.

Keller (2005) defines time series as any variable that is measured over time in sequential order, and time series forecasting as forecasting that uses historical time series data to predict future values.

Univariate time series models make efficient use of available historical records of electricity consumption. Conventional forecasting methods generate forecasts with a margin of uncertainty, but of late it is imperative that the factors that have impact on the input to the forecasted profiles must be incorporated to improve the accuracy of the forecast. According to Maseema (2009), "forecasting in this complex environment requires past hard data, present hard data and possible future scenarios. The possible

future scenarios include both the worst possible and the best possible scenarios. The worst and best possible scenarios create limits/boundaries within which the forecasting is performed. The boundaries may include financial constraints, legal and (political) regulatory requirements, macro socio-economic constraints, etc.

The need for possible inputs from the future in order to make decisions about the future makes forecasting more than just a science. It has become a scientific art that requires knowledge and experience to translate weak signals starting to form from the macro-environment and interpret them accurately and correctly". On the same note, the result of a forecasting exercise is the commencement of a process to respond to the high frequency changes. The outcome of the model is a foundation from which the process started and should carry on as it is not the final goal since these models should take into account the many factors that can influence the forecast significantly. Forecasting only sets a direction in which way to go but does not determine the final destination. The strategic forecast of the organisation should set the ultimate goal in the longest term possible while interim tactical changes based on short-term forecast help the organisation deal with immediate changes.

Guerrero and Berumen (1998) suggest forecasting electricity consumption with extra-model information provided by consumers. When this was developed, the information provided by electricity consumers in an energy-saving survey, even though qualitative, was considered to be particularly important, because the consumers' perception of the future may take into account the changing economic conditions.

Souza, Barrows and Miranda (2007) forecast consumption using double seasonal exponential smoothing and interventions to account for holidays and temperature effects. Exponential smoothing methods, in particular the Holt-Winters method and its variations, have been recommended since they are highly adaptable and robust tools to forecast in different horizons.

Mohamed, Bodger and Hume (2004) investigate the influence of selected economic and demographic variables on the annual electricity consumption in New Zealand using multiple linear regression. The study uses gross domestic product, average price of electricity and population of New Zealand during the period 1965–1999. Mohamed

et al propose six forecasting models developed for electricity consumption in New Zealand. Three of these models (Logistic, Harvey Logistic and Harvey), are based on growth curves. A further model uses economic and demographic variables in multiple linear regression to forecast electricity consumption, while another model uses these factors to estimate future saturation values of the New Zealand electricity consumptions and combine the results with a growth curve model. The sixth model makes use of the Box-Jenkins ARIMA modeling technique. The developed models are compared using goodness of fit, forecasting accuracy and future consumption values. The future consumptions are also compared with the available national forecasts. The comparisons reveal that the best overall forecasts are given by the Harvey model for both domestic and total electricity consumption of New Zealand, while a specific form of the Harvey model, that is, the Harvey Logistic model, is the best in forecasting non-domestic electricity consumption.

Due to forecasting becoming more challenging because of the changing environment in the electricity market as well as other economic contributors globally, Gettler (2007) suggests steps to embrace in mind while forecasting. These steps are as follows:

"Defining a Cone of Uncertainty
This incorporates mapping out all the factors, including the relationships between them. Distinguish between the highly uncertain and outrageously impossible, the cone of uncertainty shouldn't be drawn too narrowly as it will lead to blindside and omit some factors behind which might lead to wrong output.

Looking for the S curve
Most changes follow an S curve pattern. Change starts slowly, and incrementally moves along quietly, then suddenly explodes, and eventually narrows off and drops back down. The trick is to define the S curve before the variation starts. One shouldn't expect the opportunities to be the same as those predicted by everyone.

Embracing the things that don't fit
Keep in mind indicators that don't fit into the data set.

Holding strong opinions weakly

The forecaster should not rely too much on seemingly strong information just because it supports the conclusion, good forecasters forecast often, and are always trying to prove they are wrong.

Looking back twice as far as you look forward

The recent past is rarely a good indicator of the future. But what need to be done, is to look for the patterns. It has been said that history doesn't repeat itself but sometimes it rhymes. The forecaster's job is to find the rhymes, not identical events. And it is important not to dismiss past events that don't fit in with one's view of the present and future.

Knowing when not to make a forecast "

There are moments when forecasting is impossible. The cone of uncertainty can contract and expand and sometimes it becomes so wide, that anything can happen. That's when the wise forecaster will refrain from making a forecast at all.

The above steps are useful in providing guidance to the application of forecasting in this study.

1.3 Aim and objectives

The aim of this study is to investigate models to forecast electricity consumption for short and medium term for operational and tactical use respectively using well-known methods and applying these in a practical situation, in particular, the Ferrochrome sector, Keller (2005).

The objectives of this study are:
- to investigate the market of the Ferrochrome sector and risk exposure of industry supplying electricity to this market;
- to investigate methods of forecasting electricity consumption with the aim of understanding the explicit factors that are more likely to affect the behaviour of the forecast;

- to identify models to forecast electricity consumption in the Ferrochrome sector for both short and medium term; and
- to recommend the best methods for adoption by electricity suppliers to the Ferrochrome sector.

1.4 Motivation for the study

The power utility has a challenge of supplying electricity to its key consumers on a continuous basis. In order to meet this requirement, forecasting short, medium and long term consumption is a necessity. Due to uncertain events that fluctuate rapidly over time, it has become a challenge to determine a proper method of predicting future forecasts. Traditional methods no longer handle the demand caused by the key consumers. One of the most intensive key consumers of electricity to the power utility is the Ferrochrome (FeCr) sector, which is the subject of this study.

This study would be valuable to the power utility because it will identify risks pertaining to the key consumers of electricity with respect to the FeCr sector. The study will focus on market analysis for the FeCr sector, statistical analyses of the data set and provide methods of forecasting electricity consumption investigated to forecast short and medium term for the FeCr sector.

The study seeks to make the power utility aware of factors that stand to affect the FeCr market as well as other efficient methods of forecasting consumption in addition to the one that is currently used by power utility. The study is also intended to serve as future source of reference to the power utility. It will assist in identifying different risks associated with the key consumers of electricity as well as other methods that can be used to analyse and forecast future electricity consumption for both operational and tactical use.

1.5 Data and assumptions

In this study, data analysis and findings will be based on the data set collected at Eskom Key Sales and Customer Service Department through Enerweb and Topline system, for electricity consumption in GWh for the period of at least five years. The data set is mainly for all the large FeCr customers at Eskom, only those that are in the

Key Sales and Customer Services database. The FeCr sector in this department consists of twelve large customers, and the data is collected daily as actual consumption is received.

The data set with electricity energy consumption for all the twelve large customers was reviewed. It was also noted that some of the customers are new in the system while others have been in the system for a longer period.

The scope of the study entails short and medium term forecasts for operational and tactical use respectively. *Short term forecasts consist of twelve months forecasts generated from monthly data which are derived from daily data. Medium term forecasts consist of ten year forecasts presented on a yearly basis derived from a monthly data.*

Assumptions used for each methodology are provided together with the description of the method.

1.6 Model comparison

Different models will be compared using a variety of statistics. There are varieties of tools provided for identifying potential forecasting models and for choosing the best fitting model. These varieties of tools allow one to decide how much control one can have over the process, from a hands-on approach to one that is completely automated (SAS user guide, 1989).

Managa (2006) indicates that "the accuracy of any model is measured by the lack of fit of the model at hand relative to operating model. The model, which is estimated to minimise the expected discrepancy, is the final ("best") model chosen. The overall discrepancy consists of two components: discrepancy due to approximation and discrepancy due to estimation".

Better known model selection methods include, Bayes Information Criterion (BIC), residual mean square error (MSE), coefficient of multiple determination (R^2), adjusted coefficient of multiple determination ($AdjR^2$), stepwise regression and mean absolute percent error (MAPE) (Managa, 2006).

1.7 Research report layout

The research report is divided into seven chapters. Chapter 1 gives a brief introduction to the problems on forecasting consumption in the energy sector as well as philosophical perspective. Aim and objectives, as well as motivation for the study are also stated, and how the models will be compared in order to select the best model to address the research question.

Chapter 2 gives a comprehensive literature review that seeks to highlight what other authors have established in relation to the topic under investigation, methods recommended and selection criterion used to select the best model fit. Chapter 3 studies the market analysis of electricity consumption in the Ferrochrome sector industry. In this chapter we give a detailed analysis of factors affecting this market and how these contribute to the outcome of the actual and forecasted data.

Methods that are used to forecast electricity consumption for short and medium term are found in Chapter 4 with all its derivations and characteristics. Chapter 5 gives the detailed analysis and findings of the data based on models explained and derived in Chapter 4. A comprehensive study of sensitivity analysis between the models that are widely used and those recommended by this study as well as the framework of what is accepted as the forecasting procedure in this context is found in Chapter 6.

Finally, in Chapter 7, the study concludes with recommendations, limitations and further work.

CHAPTER 2: LITERATURE REVIEW

2.1 Forecasting in General

Forecasting has long been in existence and continues to receive extensive attention in the literature. A number of authors provide the definition of forecasting in the literature based on the environment in which it is applied. Forecasting has been evolving over the years and saw many methods being established and some being developed.

According to Keller (2005), "forecasting is a tool to predict the future values of variables and trends by examining and analysing available information". Russell and Kratowicz (2004-2009), define *forecasting* as the "process of analyzing historical trends and current factors as a basis for anticipating market trends". The Business dictionary, (2007-2009), define "*forecasting* as the planning tool which assists management in its attempt to cope with uncertainty of the future". The Oxford dictionary (2006), define "*forecasting* as a method of predicting or estimating a future event or trend".

The Business dictionary (2007-2009), also explains that *forecasting* starts with allocating certain assumptions based on the management's experience, knowledge and judgement. The estimates are then projected into the coming months or years using one or more techniques such as Box-Jenkins models, Delphi method, exponential smoothing, moving averages, regression analysis, and/or trend projection. Thus, it could happen that the assumptions might result in a similar or overstated error in forecasting, for which the technique of sensitivity analysis is then used to assign a range of values to the uncertain factors.

Some authors define *forecasting* as a tool, while others describe it as a process. Based on Business dictionary (2007-2009) and Oxford dictionary (2006), the definition of a forecast can then be adopted as follows: "*a forecast is a planning tool which assists management in attempting to manage the uncertainty of the future by examining and*

analysing available information to predict future values through a set of assumptions". Forecasting shows what might happen, which implies that this could be coupled by uncertainties.

Any business needs forecasts for planning purposes and these can be categorized into three basic types. According to the article by Robert and Kugel (2008), a business needs to undertake the following:

Short term forecast for operational planning – where operational planning is defined as a subset of strategic work plan. It describes short-term ways of achieving milestones and explains how, or what portion of a strategic plan will be put into operation during a given operational period.

Medium forecast for tactical planning – where tactical planning is defined as the systematic determination and scheduling of immediate or short term activities required in achieving the objectives of strategic planning.

Long term forecast for strategic planning – where strategic planning is defined as the systematic process of envisioning a desired future, and translating this vision into broadly defined goals or objectives and a sequence of steps to achieve them (Business dictionary, 2007-2009).

The terms of reference for each type of planning defined above depend on the area of study.

Due to uncertainties posed by rapid changes in events surrounding forecasting electricity consumption, several authors have suggested methods and analyses concerning predicting the future of electricity consumption.

In the 1960s according to Bunn (1996), model building for forecasting remained very much in the conventional scientific method of formulating a particular theory, estimating its parameters from the available data and then validating its applicability. This was coupled with a strong judgemental input and the pragmatic emphasis upon estimation. The main purpose of pre-specified model and major thrust of research was upon the development of efficient methods of estimation. Examples of these models

are linear or exponential growth trends, a product life cycle, or a specific macroeconomic relationship.

In the 1970s general classes of models were created to deal with the model selection problem. This was such that most models could be seen as a special case of the overall class. Examples of such models are ARIMA (autoregressive integrated moving average) class of Box and Jenkins, the Bayesian multiprocessing model developed and established by Harrison and Stevens and the unified view of state space representation like Harvey (Bunn, 1996).

The 1980s saw the general data intensive and theory sparse techniques being developed. For the past recent years, this has accelerated more use of data with increasingly more emphasis on model specification and less judgemental input. Rather than propose one model, or select out of several models, the robustness of pursuing a combination possibly with the data determining the relative weights, has become well established in practice. Examples of these methods are multiple switching, combinations and the neural network techniques (Bunn, 1996).

The recognition of forecasting competitions cited an emphasis on finding methods which worked best in a generalisable sense as established by out-of-sample testing. Thus, data has to perform three roles: identification, estimation and validation (out-of-sample testing) (Bunn, 1996).

A number of different methods applied in forecasting such as multivariate forecasting methods and univariate forecasting methods can be found in the literature. According to Tabachnik and Fidell (1989), multivariate methods have been applied widely in economics and much less in other applications in forecasting. It has been proven in the literature on forecasting that multivariate models are not necessarily better than univariate ones.

2.2 Forecasting in the energy sector

Studies in forecasting procedures have been explored recently. Electricity consumption forecasting in particular continues to receive attention in the published statistical literature. All these are geared toward the elimination of *noise*, which is the main aim of time series analysis that comes as a result of outside disturbances.

The challenge of determining an electricity consumption forecasting model has long been of concern to forecasters and continues to receive attention in recent statistical literature (Hamzaçebi, 2007).

In real time, maintaining a particular voltage throughout an electricity grid, the amount of electricity drawn from the grid and the amount generated should balance. Forecasting electricity consumed from the grid is imperative. Short term forecasts (one to twelve months) are required to ensure system stability, medium term forecasts (one to ten years) are required for maintenance and scheduling, while long term forecasts (10 to 35 years) are required for capital planning (Smith, 2003).

Worldwide countries today are faced with challenges in electricity business planning due to lack of proper electricity consumption forecast that is lagging behind reality. Many models to forecast electricity consumption in different countries have been suggested although it can be argued that most of these methods still face challenges in minimising the margin of error compared to the actual situation.

The use of these forecasts becomes much more critical during the upswing and down swing movements of electricity demand, and this answers why there are numerous papers that can support and show the effectiveness of electricity consumption forecast in the industrial sector and how much these can influence a country's economic growth, and why models are still being developed even today above and beyond the many that have already been established.

According to Hamzaçebi (2007), electricity consumption forecasts are much more fundamental today because of the important function they depict in driving decisions

on capital intensive investments. They are important also for effective implementation of energy policies, price policies, capacity planning and geographical location of the plants.

2.3 Causal methods

Causal methods seek to define the relationship between the variable of interest and the value of related explanatory variables (Makridakis, Spyros, Steven and Hyndman, 1998). Makridakis *et al* (1998) use the cause and effect relationship between the variable whose future values are being forecasted and other related variables or factors. The widely known causal method is called *regression analysis* which is a statistical technique that is used to develop a mathematical model showing how a set of variables are related. In addition, regression analysis can be used to generate forecasts where a variable that is being forecasted is called a dependent variable, and variables that help in forecasting the values of the dependent variable are called the independent variables. A lot of literature exists where there is only one independent variable.

Regression can be divided into simple linear regression and multiple regression analysis, defined in the next section.

2.3.1 Simple Linear Regression and Multiple Regression

Simple linear regression is a regression analysis that makes use of one dependent variable and one independent variable, and approximates the relationship between these two by a straight line, while *multiple regression analysis* uses two or more independent variables to forecast values of the dependent variable (Sahu, 2007). According to Baker (2006), *multiple regression* is a method used to model the linear relationship between a dependent variable and two or more independent variables.

Multiple regression is used mainly for prediction and explanation or analysis. While simple regression allows one causal factor, multiple regression allows the use of more than one factor to make a prediction. Multiple regression also allows separate causal factors, analyzing each one's influence on what is being tried to explain. The model is

15

fit such that the sum-of-squares of differences of observed and predicted values is minimized.

Based on Hill and Lewiscki (2006), multiple regression is more to learn about the relationship between dependent variable and several independent variables for prediction. In recently published literature, multiple regression is one of the techniques that have been widely used to forecast electricity consumption.

Al-Ghandoor and Samhouri (2009), study two techniques, namely, multivariate linear regression and neuro-fuzzy models for modelling electricity consumption of the Jordanian industrial sector. Al-Ghandoor and Samhouri (2009) use the analysis of variance that is based on least square method to identify the most important variables that affect electricity consumption of the industrial sector, and those key variables were used to develop different models that were based on multivariate linear regression and neuro-fuzzy analyses.

Besides variables such as energy costs, production levels, number of employees, and number of establishments that other research papers have taken into consideration, the study by Al-Ghandoor and Samhouri (2009) utilizes advantages of such experience and introduces other new important variables such as structural effect (ES), the capacity utilization factor (G_l/G_N), electricity prices (E\$), fuel process (F\$), establishments (EM), capacity utilization (CU) and gross output (G).

After the key variables were used to develop different models that were based on multivariate linear regression, the proposed regression model is given by:

$$(E)_t = \mu_0 + \mu_{1,N}(G_!/G_N)_t + \mu_1(E\$)_t + \mu_2(F\$) + \mu_3(E\$) + \mu_4(CU) + \mu_5(EM) + \mu_6(G) + \varepsilon_t \ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots equation \ (2.1)$$

where,

      E is the electricity consumption,

      μ_0 is the regression model intercept,

$\mu_{1,N}$ the regression model coefficient of the gross output ratio between outputs of intensive (*I*) and non-intensive (*N*) industrial clusters,

μ_j represents the regression coefficients (j = 1,2,3,4,5,6),

t represents the year, and

ε_t is the difference between the actual and the predicted electricity consumption.

The second model is the neuro-fuzzy which is an associative memory system that consists of fuzzy nodes instead of simple input and output nodes, and it uses neural network learning functions to refine each part of the fuzzy knowledge separately. This model depict that learning in a separated network is faster than learning in a whole network. An adaptive neuro-fuzzy inference system (ANFIS) is a fuzzy inference system implemented in the framework of an adaptive neural network. By using a hybrid learning procedure, ANFIS can construct an input-output mapping based on both human-knowledge as fuzzy *If-Then* rules and stipulated input-output data pairs for neural networks training. Rahib, A, Vasif, HA, and Cemal A, (2005) present more background on neuro-fuzzy inference system and is cited under list of references of this study.

The multivariate model was tested by means of ANOVA for validity and significance of the model and was found not to be violating any assumptions. The same was done for the other model and it was found that the system is well trained to model the actual electricity energy consumption. Al-Ghandoor and Samhouri (2009) recommend that these two models be used to predict electricity energy consumption of the Jordanian industrial sector.

Bhargava, Singh and Gupta (2009), undertake a holistic view of growth of demand for electricity in Punjab, India. They apply multiple regression technique and use secondary data for the purpose of analysis. Their study concludes that demand for electricity in Punjab was price inelastic but income elastic for the majority of consuming sectors. Bhargava *et al* (2009) also establish that an important policy implication thereof is that price hike will be ineffective in regulating and managing demand, unless price would be varied in an hourly basis. Thus, Bhargava *et al* (2009),

recommend that the state resort to other demand-side management (DSM) measures such as improving efficiency of electricity use and its conservation.

Bhargava *et al* (2009), contemplate that the availability of electricity also has a bearing on its consumption and recommend that considering the high income elasticity of electricity demand sufficient electricity-generating capacity needs to be created, since demand is expected to grow at an accelerated rate in future. The study further concludes that in the long run, price-demand as well as income-demand relationship in case of electricity is likely to remain uncertain especially in the post-reform era.

Hirschhausen and Adres (2000), adopt a pragmatic approach of Cobb-Douglas function, which is based on policy oriented perspective to forecast electricity production and consumption for a ten year period at the national, sectorial and regional levels. The study introduces different scenarios in order to identify possible economic developments such as high growth, medium growth and low growth scenarios, to forecast the electricity consumption.

2.3.2 A combination of causal methods

In New Zealand, a comparison of alternative approaches for modelling and forecasting the demand for electricity have been suggested and developed by Fatai, Oxley and Scrimgeour (2003). The models are based upon either partial general equilibrium approach or constructed from spreadsheet models.

Tunç, Çamdali and Parmaksizoğlu (2006), predict Turkey's electricity energy consumption rates with regression analysis for the years 2010 and 2020. In the study, Tunç *et al* (2006), develop a linear mathematical optimization model to predict the distribution of future electrical power supply investments in Turkey.

Pao (2008), proposes two new hybrid nonlinear models that combine a linear model with artificial neural network to develop adjusted forecast, taking into account heteroscedasticity in the models input. He argues that both of the hybrid models can

decrease the round off and prediction errors for multi step ahead forecasts, and the inclusion of the heteroscedasticity variations in the input layer of the hybrid model contribute towards improving the modelling accuracy.

According to Hamzaçebi (2007), literature is enormous and is growing on forecasting. There are other techniques which have been developed that have originated from causal methods, and these include: Artificial Neural Networks, Artificial Intelligence, Neuro-fuzzy, Back Propagation, and the Cobb Douglas function which has been largely supported by recently published literature.

Metaxiotis, Kagiannas, Askounis and Psarras (2003), have given a literature review in the forecasting for short to medium term electricity consumption forecasts using the Artificial Intelligence (AI) technique. The study also review AI techniques used in short to medium term load forecasting under expert systems, artificial neural networks and genetic algorithms titles. On the same note, Feinberg and Genethliou (2005) suggested forecasting load for short, medium and long term using the AI taking into account important factors like weather data, the number of customers in different categories, the appliances in the area and their characteristics, the economic and demographic data and their forecasts, the appliance sales data, and other factors. According to Feinberg and Genethliou (2005), the AI technique is defined as "the capability of a device to perform functions that are normally associated with human intelligence, such as reasoning and optimization through experience. In other words it is the branch of computer science that attempts to approximate the results of human reasoning by organizing and manipulating factual and heuristic knowledge".

Feinberg and Genethliou (2005) suggested an additive model that takes the form of predicting load as the function of four components:

$$L = L_n + L_w + L_s + L_r \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots equation\ 2.2$$

Where,

L is the total load

L_n represents the normal part of the load, which is the standardized load shapes for each type of the day that has been identified as occurring throughout the year

L_w represents the weather sensitive part of the load

19

L_s is the special event component that create substantial deviation from the usual load pattern

L_r is a completely random term, the noise.

A multiplicative model is represented by:

$$L = L_n \times F_w \times F_s \times F_r, \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots equation\ 2.3$$

Where,

L_n is the normal (base) load and the correction factors F_w, F_s, and F_r are positive numbers that increase or decrease the overall load. These corrections are based on weather (F_w), special events (F_s), and random fluctuations (F_r)

After comparing several load models from statistical model based learning, Feinberg and Genethliou (2005), came to a conclusion that the following multiplicative model is the most accurate:

$$L\ (t) = F\ (d\ (t),\ h\ (t) \times f\ (w\ (t)) + R\ (t) \dots\dots\dots\dots\dots\dots\dots\dots equation\ 2.4$$

where,

$L(t)$ is the actual load at time t,

d(t) is the day of the week,

h(t) is the hour of the day,

$F(d,h)$ is the daily and hourly component,

$w(t)$ is the weather data that include the temperature and humidity,

$f(w)$ is the weather factor and

$R(t)$ is the random error.

To estimate the weather factor, Feinberg and Genethliou (2005) used the regression model,

$$f(w) = \beta_0 + \sum \beta_i\, \beta_j \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots equation\ 2.5$$

where,

X_j are the explanatory variables which are nonlinear functions of the current and past weather parameters and β_0, β_j are the regression coefficients.

Karaelmas (2006) is of the view that in recent studies AI techniques are commonly used as a forecasting tool. In order to obtain sectoral electricity energy consumption forecasts, Karaelmas developed three different Artificial Neural Network (ANN) models which change according to input neuron number. The performance of these models is tested using mean absolute errors, root mean square errors, and absolute percentage errors criterion. The model recommended for Turkey electricity consumption for industrial sector by ANN is:

$$\begin{bmatrix} Y_t \\ X_t \\ W_t \\ Z_t \end{bmatrix} = f \begin{bmatrix} Y_{t-1} \\ X_{t-1} \\ W_{t-1} \\ Z_{t-1} \end{bmatrix}$$

where,

Y_t, X_t, W_t and Z_t are the input neurons,

and

Y_{t-1}, X_{t-1}, W_{t-1} and Z_{t-1}, are the output neurons.

The above model was based on one year past observations. In this selected model, there were 4 input neurons, 2 hidden neurons and 4 output neurons for sectoral consumption forecasting.

Azadeh, Ghaderi and Sohrabkhani (2008), recommend ANN in forecasting electricity in high energy consuming industrial sectors for medium to long term purposes. The study reveals a 5-3-2-1 construction and recommend that the ANN approach had the better estimated values for electricity consumption in high energy consumption industrial sectors compared to regression models after testing the data through statistical techniques.

Zhang and Gu (2007) suggest an improvement on the Back Propagation Artificial Neural Networks (BP ANN) model for forecasting electricity consumption for the sector for China by modifying the standard Back Propagation learning algorithm through network connective weights and threshold to make the error function descending along negative grad direction. Adepoju G.A, Ogunjuyigbe S.O.A, and

Alawode K.O (2007), suggested the application of neural network to load forecasting in Nigerian Electrical power system where ANN is referred to a class of models inspired by the biological nervous system and the back propagation algorithm as a supervised learning algorithm used to change or adjust the weights of the neural network. The network inputs were the load of the previous hour, the load of the previous day, the load of the previous week, the day of the week and the hour of the day which resulted in an total of five ANN input values. When a back-propagation network with momentum and with an adaptive learning rate was trained and the neural network can forecast future load one hour ahead given the various inputs to the network, with a sigmoid transfer function used in the hidden layer while a linear transfer function was used in the output layer, the error values translate to an absolute mean error of 2.54 percent for the network.

Zhang and Gu (2007) further introduce additional additive momentum and adaptive learning which consider the moving trend in modification of ANN weight and not just the grad direction. Zhang and Gu (2007) argue that the industry electricity consumption is complex with multiple influencing factors which usually result in regular prediction model not fitting well for its prediction. The study recommends that the BP ANN model fairs better than traditional methods, a recommendation which was validated by simulation study. In this study the authors indicate that electricity is tightly correlated to the development of an economy.

Causal method has fairly been used in forecasting electricity consumption for both short and medium term as shown in the literature. Another commonly used method that has also received quite an enormous attention in forecasting electricity consumption is *time series method*. This method is reviewed in the next section.

2.4 Time series method

One of the most widely used forecasting techniques is the time series method of forecasting. According to McClave, Benson and Sincich (2001), the time series method relies on historical data and attempts to project historical patterns into the future, assuming that the same pattern will prevail. Patterns that appear in time series data are random (horizontal or stationary pattern), trend, cyclical and seasonal. These

patterns can be identified by fitting the data and then determining the kind of line the data resembles. Time series analysis can be defined as an ordered succession of values of a variable at equally spaced time intervals and can be used to obtain an understanding of the underlying forces and structure that produced the observed data.

Time series models, according to McClave *et al* (2001), can be represented in two forms, i.e. the additive and multiplicative models.

In using time series techniques, Garett and Leatherman (2000), identify several essential concepts that need to be considered prior to the selection of the technique. These include:

- **Trend**: – There are no proven *automatic* techniques to identify trend components in the time series data. However, as long as the trend is monotonous (consistently increasing or decreasing) that part of data analysis is typically not very difficult. If the time series data contain considerable error, then the first step in the process of trend identification is smoothing.
- **Cyclicality**: – In time series, *cyclicality* refers to the extent to which the consumption is influenced by general business cycles.
- **Seasonality**: – This is typically the case when the observations are monthly or quarterly. The mathematical formulas employed can be adjusted to determine both the degree of seasonality that may exist as well as whether seasonality is increasing or decreasing over time. Seasonality is formally defined as correlation dependency of order k between each i'th element of the series and the $(i-k)$'th element Hill and Lewicki (2006); and it is measured by autocorrelation (i.e., a correlation between the two terms); k is usually called the *lag*. If the measurement error is not too large, seasonality can be visually identified in the series as a pattern that repeats every k elements.
- **Randomness**: – This is another factor that affects time series data. Randomness refers to unexpected events that may distort trends that otherwise exist over the long-term. Events such as natural disasters, political crisis, and the outbreak of war can result in temporary distortions in trends. Randomness can also result from natural variations around average or typical behaviour.

The underlying assumption of time series techniques is that patterns associated with past values in a data series can be used to project future values. The data series should also be stationary.

Stationarity is when the data series have a constant mean and variance over time. This exists if the data series were divided into several parts and the independent averages of the means and variances of each part were about equal. If the averages of each mean or variance were substantially different, non-stationarity would be suggested. When randomness tends to characterize a data series, time series techniques do not perform very well. Performing time series analyses on non-stationary data can often result in biased estimates (Garett and Leatherman, 2000).

With the above in mind, many authors have looked at how these concepts can impact on the outcome of the forecast based on different suggested models. For instance, Yan and Choon (2009), use the back-propagation algorithm on a multi layered perceptron network to model and forecast the electricity consumption in Malaysia. They investigate the effectiveness of modelling the time series with both seasonal and trend patterns. They also perform data pre-processing which include deseasonilisation and detrending, neural network modelling and forecasting. The results are compared with the predictions obtained from Box-Jenkins Seasonal Auto Regressive Integrated Moving Average (SARIMA) model. The Mean Absolute Deviation (MAD), Mean Absolute Percent Error (MAPE) and the Root Mean Square Error (RMSE) were used as measurements for forecasting performance.

The simulated series was generated according to the multiplicative model as follows:

$$X_t = T_t S_t + \varepsilon_t \quad \text{.. } equation \text{ (2.6)}$$

where,

X_t is the Electricity Consumption

T_t is the trend factor at time t

S_t is the seasonality factor at time t

ε_t is the error term at time t.

After all the technical analysis had been performed for SARIMA models, the researchers concluded that errors for ANN models are smaller than for SARIMA models, and hence the detrending and deseasonalization of ANN gives the best results when compared to SARIMA models.

Souza *et al* (2007), recommend using the exponential smoothing methods, in particular the Multiplicative Holt-Winters method with double cycles, as the study found that the method variations were appropriate in forecasting short term consumption forecasts and were highly adaptable and robust tools that forecast in different horizons. Factors such as holidays and temperature effects were taken into account. The model performed well and did not seem to produce significant deterioration of the forecasts as the forecast horizon increases.

Andrews (1994), shows that Holt-Winters exponential smoothing method is similar to structural forecasting method. The only difference is that the structural model is based on a formal statistical model. The measurement equation is given by:

$$\gamma_t = \mu_t + \gamma_t + \varepsilon_t \quad \text{.. } equation \ (2.7)$$

where μ_t is a local linear trend component and γ_t is the seasonal component at time *t*.

If a white noise disturbance term is introduced, the seasonal component can be made stochastic. Thus,

$$\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t \quad \text{...} equation \ (2.8)$$

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t \quad \text{..} equation \ (2.9)$$

Al-Saba and EI-Amin (1999), also employ ANN and Box-Jenkins methods to forecast Saudi Arabia's peak load between 1997 and 2006. The study compares the performance of these two techniques and recommends that an ANN model is the best to use.

In some cases, a method that analyses a nonlinear structure is the generalized autoregressive conditionally heteroscedastic (GARCH) models. The multivariate GARCH models have been evaluated in the Nordic electricity markets by Malo and Kanto (2005). Malo and Kanto (2005) consider a variety of specification tests for multivariate GARCH models that were used in dynamic hedging in the electricity markets. Malo and Kanto (2005) argue that the test statistic included the robust conditional moments test for sign size bias along with the introduced copula tests for an appropriate dependence structure. In addition, hedging performance comparisons in terms of conditional and unconditional ex post variance portfolio reduction were conducted, and they consider such effort as worthwhile.

One of the methods that are receiving enormous attention is the use of algorithm based procedure where a combination of forecast methods is used. Mohamed et al (2004) carried out a comparison of models for New Zealand for forecasting electricity consumption. Six forecasting models are developed for electricity consumption in New Zealand, and three of these models (Logistic, Harvey Logistic and Harvey) are based on growth curves. A further model uses economic and demographic variables in multiple linear regression to forecast electricity consumption, while another uses these factors to estimate future saturation values of the New Zealand electricity consumption and combine the results with a growth curve model. The sixth model makes use of the Box-Jenkins ARIMA modelling technique. The results reveal that the best overall forecasts are provided by the Harvey model for both domestic and the total electricity consumption of New Zealand, while a specific form of Harvey Logistic model is the best in forecasting the non-domestic electricity consumption.

The Harvey Logistic model is based on Logistic model. The model is given by:

$$\ln Y_t = 2 \ln Y_{t-1} + \delta + \gamma_t + \varepsilon_t \ , \ T = 2 \dots T \quad \dots \dots \dots \dots \dots \dots \dots \dots equation(2.10)$$

where,

$\quad$ $Y_t = Y_t - Y_{t-1}$, is the electricity consumption at time t,

$\quad$ ε_t is the disturbance term with zero mean and constant variance,

$\quad$ δ and γ are constants to be found by regression.

Final results show that for total consumption forecasts, the best short term forecast is given by ARIMA model, the best medium term forecast is given by various logistic models (VAL) which is a saturation of logistic model and the best long term forecast was given by Harvey model.

2.5 Selection criterion

It is not possible for a forecast to be completely accurate. A forecast will always deviate from the actual requirement. The difference between the forecast and the actual is called *forecast error*. The idea in forecasting is that this error should be as small as possible. A relatively large degree of error may indicate that either the model that is being used is wrong, or the technique needs to be adjusted by changing its parameters (Russell and Taylor, 1995).

There is no common way of measuring forecast error and selecting the best model, but there are varieties of tools provided for identifying potential forecasting models and for choosing the best fitting model suggested in the literature. These varieties of tools allow one to decide how much control one can have over the process, from a hands-on approach to one that is completely automated (SAS user guide, 1989).

Many researchers have used different model selection tools to select the best method for the model recommended. For instance, Al-Ghandoor and Samhouri (2009) tested the multivariate model for validity and significance using ANOVA approach. Neuro-fuzzy model was also tested using the above approach. From the analysis, it was found that Neuro-fuzzy model was better in modelling the actual electricity energy consumption. However, they recommend that these two models can be used to predict electricity energy consumption of the Jordanian industrial sector.

Mohamed et al (2004), test and compare the performance of six models namely Logistic, Harvey Logistic, Harvey, multiple linear regression, Box-Jenkins ARIMA modelling technique and VAL using (MAPE).

Yan and Choon (2009) use MAD, MAPE and RMSE to measure the forecasting performance of back-propagation algorithm on a multi layered perceptron network compared to Box-Jenkins SARIMA model.

Costantini and Pappalardo (2008) apply the RMSE that shows that the forecast encompassing of a given model versus the other non-nested models is a sufficient condition for minimising the root mean square of factored error (RMSFE). This is used to improve the forecast accuracy to the seven time series models tested by ranking the overall forecasting models using RMSFE measure and eliminating those that are encompassed by others and combining the remaining forecasts.

There are different varieties of measures of forecast error. Better known model selection methods include, Bayes Information Criterion (BIC), root mean square error (RMSE), coefficient of multiple determination (R^2), adjusted coefficient of multiple determination (Adj R^2), stepwise regression which consists of backward and forward selection and mean absolute percent error (Managa, 2006). These statistical measures are commonly used by authors in order to select the best model that minimises bias and maximises accuracy. The formulas to these methods are highlighted in Chapter 4.

In both regression and time series analysis there are suggested ways of analysing the goodness of fit of the model. Garson (2009) suggests how regression and time series data analysis can be carried out for model check and to assess goodness of fit of the model. These checks are also highlighted in detail in chapter 4.

2.6 Overview

A lot of research around forecasting has been done and a number of conventional forecasting methods have been established. These methods (conventional) need to be tested against particular environments. The industrial market is surrounded by numerous factors that influence it and much is still to be researched and developed. In view of all these, from the literature, there is an indication from most of the studies suggested by various authors indicating that market dynamics does have an influence in the behaviour of electricity consumption in an industrial sector.

This chapter has given a comprehensive literature review on what other authors studied about forecasting in general, forecasting in an energy sector, causal method and time series techniques in forecasting electricity consumption and what they have recommended. Selection criterion and standard model checks of both casual and time series have also been reviewed.

The next chapter gives a comprehensive market analysis of electricity consumption in the ferrochrome sector which includes defining it and analysing factors that affect and influence the behaviour of how this sector consumes electricity.

CHAPTER 3: MARKET ANALYSIS

3.1 Electricity consumption

Forecasting electricity consumption is of national interest in any country. Future electricity forecasts are not only required for short and long term power planning activities but also in the structure of the national economy.

Eskom is a state owned supplier of electricity in South Africa, and since 2007 it experienced lack of capacity in the generation and reticulation of electricity. This resulted in severe electricity blackout all over the country during the first quarter of 2008 – with damaging effect of the economy. Both domestic and industrial consumers were severely affected (Inglesi, 2010).

South Africa is the world leader in ferrochrome production (by both tonnage and market value). The ferrochrome industry contributes a significant proportion to the country's gross domestic product and employs a considerable number of people.

Ferrochrome contributes 23.71%, followed by Aluminium with 20.00% and other industrial sectors ranging between 0.10% and 13.25%. Since the ferrochrome sector is one of the most intensive energy consumer, it is also faced with significant consumption challenges like load shedding and outages within the electricity market.

3.2 Ferrochrome sector

According to Silk (1988), ferrochrome (FeCr) is an alloy of chromium and iron containing between 50% and 70% chromium. FeCr is a carbothermic reduction operation taking place at high temperature. This is covered by the following:

- The ore (an oxide of chromium) and iron
    - Which is reduced by coal and coke to form an iron-chromium alloy called ferrochrome.

- The heat for this reaction comes from the electric arc
    - This is formed between the tips of the electrodes in the bottom of the furnace and the furnace hearth. This arc creates temperatures of about 2 800°C.
    - In the process of smelting, approximately 3 500 kilowatt-hours of electricity are consumed for each ton of ferrochrome produced.

FeCr is produced by electric arc melting of chromite, an iron magnesium chromium oxide and most important, chromium ore. Most of the world's FeCr is produced in South Africa (45%), China (14%), Kazakhstan (13%) and India (10%), all of which have large domestic chromite resources (Figure 3.1). These four countries account for 82% of the world's FeCr and the remaining 18% is produced by Russia (5%), Finland (4%), Zimbabwe (4%), Brazil (3%) and Sweden (2%).

The production of steel is the largest consumer of FeCr, especially the production of stainless steel with chromium content of 10% to 20%.

Figure 3.1:   Ferrochrome producers by country 2008
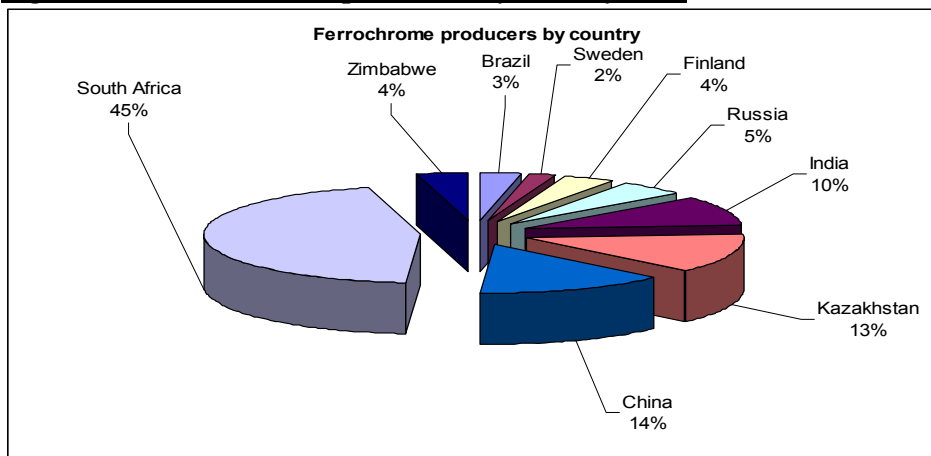
The FeCr sector consists of twelve (12) large industries which receive electricity directly from the grid and small customers that consume less than 100 GWh which receive electricity from the redistributors. In total, the FeCr sector contributes approximately 14% of total key customer sales and approximately 22% of industrial category. However, this sector has the most unpredictable market and high levels of uncertainties due to many drivers in the market compared to other sectors contributing to the total key customer sales.

Approximately 90% of the world's FeCr is used in the production of stainless and special steel. Over 80% of the world's FeCr is utilised in the production of stainless steel. According to sector analyses 2008, 28 million tons of stainless steel was produced in 2006. Stainless steel depends on chromium (an alloy within FeCr) for its appearance and its resistance to corrosion. The average chrome content in stainless steel is approximately 18%. It is also used when it is desired to add chromium to carbon steel.

FeCr from Southern Africa, known as 'charge chrome' and produced from a chromium (Cr) containing ore with a low Cr content, is most commonly used in stainless steel production.

Alternatively, high carbon FeCr produced from high grade ore found in Kazakhstan (among other places) is more commonly used in specialist applications such as engineering steels, and minimum levels of other elements such as Sulphur, Phosphorus and Titanium are important and production of finished metals takes place in small electric arc furnaces compared to large scale blast furnaces.

There are factors that influence the electricity consumption behaviour in the FeCr sector:

- The decline in demand of stainless steel eventually signifies a decline in the production of FeCr and will automatically have a negative impact in the consumption of electricity.
- Commodity prices can also have a negative impact on production in that if cost of production is higher than the spot price, producers usually intend to lower production until the market condition is favourable, this implies a reduction in electricity consumption.
- Growth due to expansion projects increases electricity consumption.
- FeCr producers that are looking forward to be self reliant and not depend on parastatals or anyone else for electricity entertain ideas to generate their own power. However, at the present moment only a few of these producers (for FeCr) will be able to meet the high and prohibitive costs associated with setting up generation units.

3.3 FeCr Market Dynamics

In any market, there are various dynamics that influence the behaviour of sales produced. These are mostly the factors that expose the market to risks that can impact negatively on sales. Within the FeCr market, the dynamics are explained in the sections that follow.

3.3.1 Demand

FeCr market is highly driven by demand. That means for a smelter to start producing FeCr, there must be sufficient orders. These orders are measured by Tons/US dollar. Since FeCr is used mostly in the production of stainless steel and precious metals, most demand comes from the stainless steel producers. South Africa accounts for close to 60% of the world's FeCr production and exports more than 90% of its FeCr. Local companies consume approximately 10% of the total production. Since 30 September 2008, there has been a decline in the amount of stainless steel produced and FeCr demand due to world economic meltdown. The world recession that was experienced had a severe negative impact on the demand for FeCr in the export market.

3.3.2 Commodity Price

Another underlying factor that drives FeCr market is the commodity price. Prices of FeCr are often quoted in terms of United States cents (US cents) per pound (lb) of chrome contained. Although producing companies will generally report production and sales in terms of metric tonnes of FeCr sold. In order to calculate the value of a metric tonne of FeCr from a price quoted in US cents, the percentage of chrome within the FeCr must be known. FeCr prices went down from $0.79/lb in the March 2009 quarter to $0.69/lb in the June 2009 quarter.

Price can contribute negatively or positively to FeCr market. On a positive note, when the contract price is low and the spot price is high, then the opportunity to make bigger profits is wide. This is also underpinned by electricity prices, the more favorable they become the more the industry cash flow becomes better. But if the

contract price of FeCr is high, and the spot price is low, it affects FeCr production to a point that they can switch off furnaces to prevent monetary loss, since it also becomes expensive to run furnaces when the market is down. For example, the two previous years saw International Ferro Metals (IFM), that is, FeCr producer making no material events or transactions in the period from 1 October 2008 to 12 November 2009 as reported by IFM analysts. The companies continued to experience extremely low FeCr demand as steelmakers suffered the global economic fallout.

3.3.3 Cost of Production

FeCr production is essentially a carbothermic reduction operation taking place at high temperatures. Cr Ore (an oxide of chromium and iron) is reduced by coal and coke to form the iron-chromium alloy. The heat for this reaction can come from several forms, but typically from the electric arc formed between the tips of the electrodes in the bottom of the furnace and the furnace hearth. This arc creates temperatures of about 2 800°C. In the process of smelting, huge amounts of electricity are consumed making production in countries with high power charges very costly.

Tapping of the material from the furnace takes place intermittently. When enough smelted FeCr has accumulated in the hearth of the furnace, the tap hole is drilled open and a stream of molten metal and slag rushes down into a chill or ladle. FeCr solidifies in large castings, which is crushed for sale or further processed.

FeCr is often classified by the amount of carbon and chrome it contains. The vast majority of FeCr produced is charge chrome from Southern Africa. With high carbon being the second largest segment followed by the smaller sectors of low carbon and intermediate carbon material. All these implicate to the cost of production.

The technology underlying these furnaces is different. Furnaces with new technology perform efficiently better than those with old technology. Thus the cost of production for producers with older technology becomes expensive.

If the price of FeCr is low and the cost of production is high, this presents a market condition that is unfavorable and that means producers might decide to switch off furnaces due to poor market conditions. Contributing to switching off of furnaces can

be due to seasonal changes. During winter period, the commodity prices of electricity are higher than in summer hence most of the producers will switch off furnaces and perform maintenances to the plant so as to contribute in mitigating risk in costs. Forecasters need to take these factors into account when forecasting.

3.4 Global market share of FeCr

As indicated in Figure 3.1, South Africa is by far, the leading manufacturer of chromite compared to other countries and a major supplier of FeCr with 45% of production. Among South Africa's major producers of chromite and/or ferrochrome are Assmang Ltd, Samancor Chrome, Xstrata South Africa and International Ferrometals Ltd (IFM). The disaggregation of FeCr market share is shown in Figure 3.2.

Figure 3.2: Estimated global market share 2008
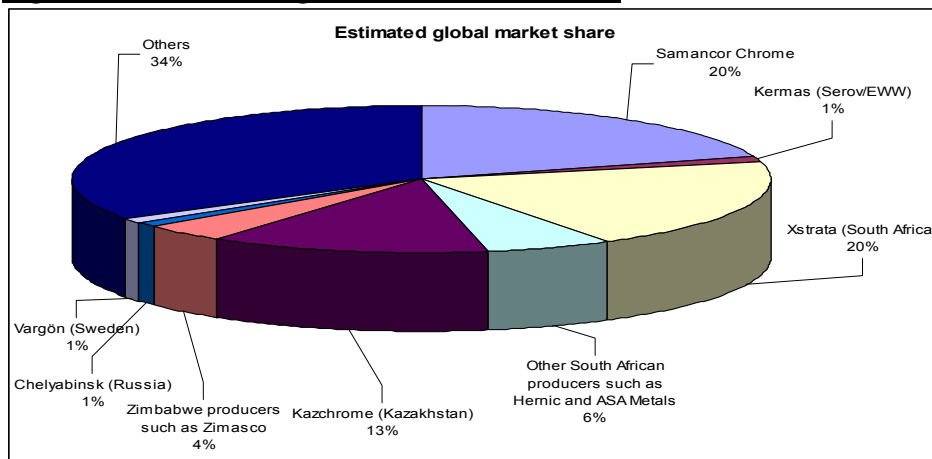
Other countries producing FeCr are:

China - The second leading manufacturer of chromite globally is China with 14% of global market share as shown in Figure 3.1. China is a country encircled by high population. This is also followed by enormous planned infrastructure for the coming years where demand for FeCr is expected to increase. Although 14% of the FeCr covers China market share, a significant quantity is its import that mostly comes from Turkey. However China also receives imports from other countries such as South Africa and others.

Kazakhstan – The third leading manufacturer of chromite globally is Kazakhstan. The country's top chromite producer is Eurasian Natural Resources Corp (ENRC), which operates the Donskoy Ore Mining & Processing unit. The company uses about 30% of its total chromite output to make chromium chemicals. Kazakhstan has a 13% of global market share which makes it one of the largest producers of FeCr as illustrated in Figures 3.1 and 3.2.

India – India is the fourth leading chromite ore producer globally with an output of about 3.5 to 4M tonnes making 10% of global market share as indicated in Figure 3.1. India recently decided to implement a significant export tax to ensure supply for domestic ferrochrome manufacture. In India, chromite ore is mainly produced in the state of Orissa, with a large portion of chromite production consumed by local FeCr makers.

Turkey – Turkey is emerging as one of the major suppliers of chromite to China's FeCr markets. Figure 3.1 indicates 14% of production of FeCr by China where the chromite is supplied by Turkey. Adverse winter conditions allow mining of chromite to be conducted only from May to the end of November. Bilfer Madencilik AS is one of Turkey's major chromite producers and primarily caters to the needs of the refractory and foundry industries. RHIAG sources about 2000 tonnes/y of refractory grade chrome for making sliding doors from Bilfer Madencilik.

Figure 3.3: Ferrochrome market share by company 2008
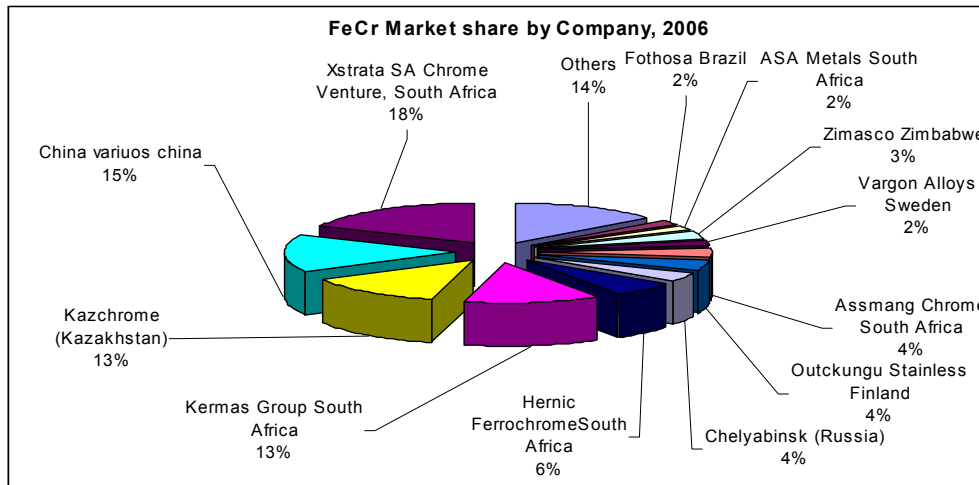
From Figures 3.1 through to Figure 3.3, it can be seen that there is a lot of production of ferrochrome that is taking place worldwide. However, South Africa dominates all the countries globally in reserves of chrome. South Africa exports chrome to countries like China, India, Kazakhstan and Turkey. FeCr as a product also get exported to countries where its demand is high.

CHAPTER 4: MATHEMATICAL FORMULATION

This chapter reviews the mathematical formulation of the methods used in forecasting electricity consumption for both operational and tactical use. The methodology applied in this study is Multiple Linear Regression and Time Series analysis which are discussed below.

4.1 Multiple and Linear regression (Causal method)

Regression analysis measures the extent of weight of the independent variables on a dependent variable. In the case of a single independent variable, the dependent variable could be predicted from the independent variable by the following equation:

$$y = a + bx \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots equation\ (4.1)$$

where,

    a is constant,

    b is the parameter estimate,

    y is the dependent variable, and

    x is the independent variable.

The multiple regression model is given by the following equation:

$$y_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \cdots + b_n x_{i,n} + e_i \dots \dots \dots \dots equation\ (4.2)$$

where,

    $x_{i,n}$ is the value of n'^{th} predictor in year i,

    b_0 is the regression constant,

    b_n is the coefficient on the n'^{th} predictor,

    N is the total number of predictors,

    y_i is the predictand in the year i, and

    e_i is the error term.

The predictor equation takes the form:

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{i,2} + \ldots\ldots\ldots\ldots \hat{b}_N x_{i,N} \ldots\ldots\ldots\ldots\ldots\ldots\ldots equation\ (4.3)$$

where, the variables are defined as in equation (4.2) above, except that the hat denotes estimated values.

4.2 Time Series method

Time series method consists of exponential smoothing and Box Jenkins models. These can either take a multiplicative or additive form.

Additive form is given by:

$$Y_t = T_t + C_t + S_t + R_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots equation\ (4.4)$$

and multiplicative form is given by

$$Y_t = T_t * C_t * S_t * R_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots equation\ (4.5)$$

where,

T_t represents the secular trend also known as the long trend, which describes the long term movements of Y_t

C_t represents the cyclical trend, which describes the fluctuations of the time series about the secular trend that is attributable to business and economic conditions.

S_t represents the seasonal effect that describes the fluctuations in the time series that recur during specific time periods.

R_t represents the residual effect which is what remains of Y_t after the secular, cyclical and seasonal components have been removed.

4.2.1 Exponential Smoothing model

Exponential smoothing method consists of the following:

4.2.1.1 Single exponential smoothing model

The model is a moving average of forecasts that have been corrected for the error observed in preceding forecasts. In the first smoothing model, there is no trend or seasonal pattern assumed. The method is given by:

$$F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1}) \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots equation\ (4.6)$$

where,

F_t is the forecast at time t,

A_{t-i} is the actual value at time t-i, and

N is the number of time periods averaged.

The parameter α is the smoothing coefficient and has an estimated value between zero and one. It is referred to as an exponential smoothing model because the value of α tends to affect past values exponentially.

As α approaches one, the forecast resembles a short-term moving average, while α closer to zero tends to resemble long-term moving averages.

Again, α is typically estimated using trial and error to secure the best fitting model and to find the model that minimizes forecast error.

4.2.1.2 The Holt model

The single parameter smoothing model presented in equation (4.6) can be adapted to take into account trends that may be presented in the data. When the trend parameter is included in the single parameter smoothing model, this form is called the Holt model. The method is given by:

$$F_{t+k} = S_t + KT \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots equation\ (4.7)$$

$$S_t = \alpha A_t + (1-\alpha)(S_{t-1} - T_{t-1}) \dots \dots \dots \dots \dots \dots \dots \dots \dots equation\ (4.8)$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots equation\ (4.9)$$

where,

F_{t+k} is the forecast at time k periods in the future,

A_t is the actual value at time t,

S_t is the level of the series at time t,

T_t is the trend at time t, and α and β are smoothing parameters.

The forecast at time t for k periods into the future equals the level of the series at t plus the product of k and the trend at time t. The level of the series is estimated as a function of the actual value of the series at time t, the level of the series at a previous time, and the estimated trend at a previous time. The parameter is a smoothing coefficient. The trend at time t is estimated to be a function of the smoothed value of the change in level between the two time periods and the estimated trend for the previous time period. The values for the smoothing parameters, α and β are between 0 and 1.

4.2.1.3 Damped Trend Exponential Smoothing

While the Holt model takes into consideration the trend that may be inherent in the data series, it somewhat unrealistically assumes the trend continues in perpetuity. This means it can overshoot estimates several time periods in the future. A variation known as damped trend exponential smoothing has the effect of dampening the trend into subsequent periods as time continues. Damped trend exponential smoothing includes a third parameter, ϕ, with a value between zero and one that specifies a rate of decay in the trend. The method is given by:

$$F_{t+k} = S_t + \sum \emptyset^i T_t \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots equation\ (4.10)$$

$$S_t = \alpha A_t(1 - \alpha)(S_{t-1} + \emptyset T_{t-1}) \ldots \ldots \ldots \ldots \ldots \ldots equation\ (4.11)$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)\emptyset T_{t-1} \ldots \ldots \ldots \ldots \ldots \ldots equation\ (4.12)$$

where,

F_{t+k} is the forecast at time k periods in the future,

A_t is the actual value at time t,

S_t is the level of the series at time t,

T_t is the trend at time t, and

α, β, ϕ are the smoothing parameters.

4.2.1.4 Holt-Winter's Linear Seasonal Smoothing

Holt-Winter's Linear Seasonal Smoothing model adapts Holt's method to include a seasonal component in addition to a smoothing coefficient and a trend parameter. The first variant of the model is additive. This assumes that the seasonality is constant over the series being forecast. The method is given by:

$$F_{t+k} = S_t + kT_t + I_{t-p-k} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots equation\ (4.13)$$

$$S_t = S_{t-1} + T_{t-1} + \alpha(A_t - S_{t-1} - T_{t-1} - I_{t-s}) \dots\dots equation\ (4.14)$$

$$T_t = T_{t-1} + \alpha\beta(A_t - S_t - T_{t-1} - I_{t-s}). \dots\dots\dots\dots\dots equation\ (4.15)$$

$$I_t = I_{t-s} + \delta(1 - \alpha)(A_t - S_{t-1} - T_{t-1} - I_{t-s}) \dots\dots equation\ (4.16)$$

where,

F_{t+k} is the forecast at time k periods in the future,

A_t is the actual value at time t,

S_t is the level of the series at time t,

T_t is the trend at time t,

I_t is the seasonal index at time t,

s is the seasonal index counter, and

α, β, and δ are the smoothing parameters.

The multiplicative variant of this model assumes that the seasonality is changing over the length of the series. The method is then given by:

$$F_{t+k} = (S_t + kT_t)I_{t-p-k} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots equation\ (4.17)$$

$$S_t = S_{t-1} + T_{t-1} + \frac{\alpha}{I_{t-s}}\left(A_t - I_{t-s}(S_{t-1} + T_{t-1})\right) \ldots \ldots equation\ (4.18)$$

$$T_t = T_{t-1} + \frac{\alpha\beta}{I_{t-s}}\left(A_t - I_{t-s}(S_{t-1} + T_{t-1})\right) \ldots \ldots \ldots \ldots \ldots equation\ (4.19)$$

$$I_t = I_{t-s} + \frac{S(1-\alpha)}{S_t}\left(A_t - I_{t-s}(S_{t-1} + T_{t-1})\right) \ldots \ldots \ldots equation\ (4.20)$$

where,

F_{t+k} is the forecast at time k periods in the future,

A_t is the actual value at time t, S_t is the level of the series at time t,

T_t is the trend at time t,

I_t is the seasonal index at time t, s is the seasonal index counter, and

α and β are the smoothing parameters.

4.2.2 Box Jenkins models

Box Jenkins models consist of Autoregressive model AR(p) and Moving average model MA(q) . Below is the mathematical representation of the ARMA (p,q) Box Jenkins model.

4.2.2.1 Autoregressive moving average model (p,q)

The Box Jenkins ARMA(p,q) model is denoted by the following equation

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \cdots$$
$$- \theta_p a_{t-p} \ldots \ldots \ldots \ldots \ldots \ldots \ldots equation\ (4.21)$$

Where,

$\phi_1 y_{t-1} + \cdots + \phi_p y_{t-p}$ represents the autoregressive part of the model (AR)

$-\theta_1 a_{t-1} - \cdots - \theta_p a_{t-p}$ represents the moving average part of the model (MA)

ϕ_0 is the ,model intercept and

a_t is the "white noise" or error

$\phi_{0,1},...,\phi_p, \theta_1,...\theta_p$ are the parameters (coefficients) of the model

The parameters are determined by the method of moments, least squatters, the method maximum likelihood, or some other method that is consistent. The white noise errors terms a_t are assumed to have the following properties:

1. $E(a_t) = 0, \forall_t$ (zero mean assumption)
2. $E(a_t{}^2) = \sigma_a^2, \forall t$ (constant varience assumption)
3. $E(a_s a_t) = 0, \forall_t \neq t$ (independence of errors assumption
4. a_t are normally distributed

Equation 4.21 is referred as the intercept – form of the Box Jenkins ARMA (p,q) model. An algebraically equivalent form is as follows:

$$y_t - \mu = \phi_1(y_{t-1}) - \mu + \cdots + \phi_p(y_{t-p} - \mu) + a_t - \theta_1 a_{t-1} - \cdots -$$
$$\theta_p a_{t-p} \ldots \ldots \ldots \ldots \ldots \ldots equation\ (4.22)$$

In this form the mean of y denoted by μ, is related to the intercept ϕ_0 of equation 4.21 by the formula

$$\mu = [\frac{\phi_0}{(1-\phi_1 - \cdots \phi_p)}]$$

Thus, the most compact way to write the Box –Jenkins ARMA(p,q) model is by using "backshift operator polynomials

$$(B) = 1 - \phi_1 B - \phi_2 B - \phi_2 B^2 - \ldots - \phi_p B^p \text{ and}$$

$$(B) = 1 - \theta_1 B - \theta_2 B - \theta_2 B^2 - \ldots - \theta_p B^p$$

Where (B) is called the autoregressive backshift operator and (B) is called the moving average backshift operator. Using these polynomials, equation 4.22 can be written as

$$(B)(y_t - \mu) = \theta(B)a_t \ldots \ldots \ldots \ldots \ldots \ldots equation\ (4.23)$$

In order to be able to estimate Box Jenkins models, stationarity and invertibility conditions must hold.

Replacing the backshift operators $B, B^2, \ldots, B^p$ in the autoregressive polynomial and the moving average polynomial with corresponding powers of $z, z^2, \ldots, z^p$, and set these

polynomials to zero, the result is the autoregressive polynomial of the ARMA(p,q) Box- Jenkins model

$$\phi(z) = 1 - \phi_1 z - \phi_1 z^2 - \cdots \phi_p z^p = 0 \ldots \ldots \ldots \ldots \ldots \ldots \text{ equation } (4.24)$$

$$\theta(z) = 1 - \theta_1 z - \theta_1 z^2 - \cdots \theta_p z^p = 0 \ldots \ldots \ldots \ldots \ldots \ldots \text{ equation } (4.25)$$

Treating the parameters, $\phi_1, \phi_2,.., \phi_p$ and $\theta_1, \theta_2,...,\theta_p$ as known, let the $z_1^{AR}, z_2^{AR},...,z_p^{AR}$ denote the p roots (zeroes) of the autoregressive polynomial (4.24) and $z_1^{MA}, z_2^{MA},...,z_q^{MA}$ denote the q roots of the moving average polynomial (4.25), for the Box-Jenkins model (4.21) to be stationary ,it must be the case that all of the roots of the autoregressive polynomial (4.24) must be greater than one in magnitude. For the Box-Jenkins model (4.21) to be invertible it must be the case that all of the roots of the moving average polynomial (4.25) must be greater than one in magnitude.

When the Box-Jenkins model is stationary, its observations y_t satisfy the following three properties:

1. $E(y_t) = \mu \ \forall t$ (i.e the mean of y_t is constant for all time periods)
2. $Var(y_t) = \sigma^2_y \forall t$ (i.e. the variance of y_t is constant for all time periods)
3. $\text{Cov}(y_t, y_{t-j}) = \gamma_j$ (i.e. the covariance of y_t and y_{t-1} is constant for all time periods and fixed j, j=1,2...)

4.2.2.2 Seasonal Autoregressive Moving Average (SARIMA) model

To account for seasonal effects in the data, for example, when there are long-term cycles in the data, one can extend the ARIMA model by adding another set of Orders and specifying the seasonal period that results to a SARIMA model. SARIMA of equation (p, d, q) (P, D, Q) $_{12}$ is represented as:

$$\phi(B)\Phi(B^{12})(1 - B)^d(1 - B^{12})^D(y_t - \mu) = \theta(B)\Theta(B^{12})a_t \ldots.. \text{ equation } (4.26)$$

where,

ϕ and θ are the parameter estimates,

45

Φ and Θ are the seasonal parameter estimates,

ϕ(B) is a polynomial of degree p in the backshift operator B,

Φ (B^{12}) is a polynomial of degree p in (B^{12}),

θ(B) is a polynomial of degree q,

Θ (B^{12}) is a polynomial of degree seasonal (Q) in B^{12}, and

D is the degree of difference.

4.2.3 Diagnostics checking of the models

The following is an overview of standard diagnostics checking to confirm the goodness of fit of the model in regression and time series methods.

4.2.3.1 Regression diagnostics (model checks)

In regression, it is important to confirm the goodness of fit of the model and the statistical significance of the estimated parameters. Commonly evaluation of goodness of fit includes the R^2, analyses of the pattern of residuals and hypothesis testing. Statistical significance can be checked by an F-test of the overall fit, followed by t-tests of individual parameters.

Interpretations of these diagnostic tests heavily depend on the model assumptions. Although examination of the residuals can be used to invalidate a model, the results of a t-test or F-test are sometimes more difficult to interpret if assumptions of the model are violated.

The following are the regression diagnostics to be checked for the goodness of fit of the model:

4.2.3.1.1 Assumptions

According to Osborne & Waters (2002) and Russell & Mackinnon (1993), when assumptions are not met, the results may not be trustworthy, resulting in a Type I or Type II error, or over- or underestimation of significance or effect size. Assumptions for the regression coefficient for the j'th independent variable, b_j, j= 1,...,m include normality, linearity, reliability of measurement and homoscedasticity, and are independent and identically distributed meaning the observations are taken from a random sample.

4.2.3.1.1.1 Normality Assumption:

Regression assumes that variables have normal distributions conditional on the regressors. Thus,

$$\varepsilon \mid X \sim N(0,\sigma^2 I_n)$$

Non-normally distributed variables (highly skewed or kurtotic variables, or variables with substantial outliers) can distort relationships and significance tests. Visual inspection of data plots, skew, kurtosis, and P-P plots give researchers information about normality, and Kolmogorov-Smirnov tests provide inferential statistics on normality.

4.2.3.1.1.2 Linearity Assumption:

Standard multiple regression can only accurately estimate the relationship between dependent and independent variables if the relationships are linear in nature. Thus if X is a random variable, then the regressors in X must all be linearly independent.

Thus, Pr [rank(X) = p] =1 where p is represents finite moment.

When this assumption is violated, the regressors become multicollinear. If the relationship between independent variables and the dependent variable is not linear, the results of the regression analysis will underestimate the true relationship. In multiple regression, underestimation carries a risk of Type I errors (overestimation)

for other independent variables that share variance with that independent variable. A preferable method of detection is to examine residual plots or run regression analyses that incorporate curvilinear components. If the relationship is nonlinear, the data should be transformed or an alternative statistical model should be considered.

4.2.3.1.1.3 Nonstochastic:

The errors are uncorrelated with the individual predictors. That is Var $[\varepsilon | X]=\sigma^2 I_n$ where I_n is an $n \times n$ identity matrix, and σ^2 is a parameter which determines the variance of each observation. This assumption is examined by performing residual analysis with scatter plots of the residual against individual predictors. Violation of this assumption might mean transformation of the predictors.

4.2.3.1.1.4 Zero mean:

The expected value of the residual is zero. $E[\varepsilon | X]=0$. The immediate assumption is that the erros have zero mean: $E[\varepsilon] = 0$, and that the regressors are uncorrelated with errors $E[X'\varepsilon] =0$. This assumption is critical for OLS theory. The regressors are called exogenous if this assumption hold; otherwise it is called endogenous.

4.2.3.1.1.5 Nonautocorrelation:

The errors are uncorrelated between observations. $E[\varepsilon_i \varepsilon_j | X]=0$ for $i \neq j$.

4.2.3.1.1.6 Homoscedasticity:

$$E [\varepsilon^2{}_j | X] = \sigma^2$$

This means that the error terms has the same variance σ^2 in each observation. It is indicated when the residuals are not evenly scattered around the line. This assumption can be verified by visual examination of a plot of the standardized residuals (the errors) by the regression standardized predicted value.

4.2.3.1.2 Regression coefficients

The regression coefficient, b, is the average amount the dependent variable (DV) increases when one independent variable (IV) increases by one unit and the other IVs are held constant. In other words, the b coefficient is the slope of the regression line: the larger the b, the steeper the slope, the more the DV changes for each unit change in the IVs. The b coefficient is the unstandardized simple regression coefficient for the case of one IV. When there are two or more IVs, the b coefficient is a partial regression coefficient.

4.2.3.1.3 Multicollinearity

Multicollinearity refers to excessive correlation of the predictor variables. When correlation is excessive (some use the rule of thumb of r >0.90), standard errors of the beta coefficients become large, making it difficult or impossible to assess the relative importance of the IVs. Multicollinearity is less important where the research purpose is sheer prediction since the predicted values of the DV remain stable, but it is a severe problem when the research purpose includes causal modelling.

Indicators that multicollinearity may be present in a model are large changes in the estimated regression coefficient when a predictor variable is added or deleted. The other indicator is insignificant regression for the affected variable in a multiple regression but a rejection of the hypothesis that those coefficients are insignificant as a group (using F test).

Tolerance, T is given as follows:

$$T = 1 - R^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots equation \ (4.27)$$

where R^2 is the coefficient of determination.

The tolerance equation above is for the regression of one IV on all the other IVs, ignoring the dependent variable. There will be as many tolerance coefficients as there

are IVs. The higher the intercorrelation of the IVs, the more the tolerance will approach zero. As a rule of thumb, if tolerance is less than 0.20, a problem with multicollinearity is indicated. When tolerance is close to 0 there is high multicollinearity of that variable with other IVs and the beta coefficients will be unstable. The more the multicollinearity, the lower the tolerance, and the more the standard error of the regression coefficients.

It should be noted that tolerance is part of the denominator in the formula for calculating the confidence limits on the b (partial regression) coefficient.

The VIF is simply the reciprocal of tolerance. Therefore, when VIF is high there is high multicollinearity and instability of the b and beta coefficients.

4.2.3.1.4 Significance testing

One of the most common diagnostic checking in regression analysis is the dynamic inference which signifies drawing the interpretation that the DV changes i units because the IV changes one unit.

The t-tests are used to assess the significance of individual i coefficients, i.e. specifically testing the null hypothesis that the regression coefficient is zero. A common rule of thumb is to drop from the equation all variables not significant at the 0.05 level or better.

4.2.3.1.5 Standard error of estimate, confidence intervals, and prediction intervals

For large samples, the standard error of estimate (SEE), approximates the standard error of a predicted value. SEE is the standard deviation of the residuals. In a good model, SEE will be markedly less than the standard deviation of the dependent variable. In a good model, the mean of the dependent variable will be greater than 1.96 times SEE.

4.2.3.1.5.1 The confidence interval of the regression coefficient

Based on t-tests, the confidence interval is the plus/minus range around the observed sample regression coefficient, within which we can be, say, 95% confident the real regression coefficient for the population regression lies. Confidence limits are relevant only to random sample datasets. If the confidence interval includes 0, then there is no significant linear relationship between x and y. We do not reject the null hypothesis that x is independent of y.

4.2.3.1.5.2 The confidence interval of the dependent variable, y

The confidence interval of y (DV) is also called the standard error of mean prediction. Some 95 times out of a hundred, the true mean of y will be within the confidence limits around the observed mean of n sampled cases. That is, the confidence interval is the upper and lower bounds for the mean predicted response.

4.2.3.1.5.3 The prediction interval of the dependent variable, y

For the 95% confidence limits, the prediction interval on a fitted value is plus/minus the estimated value plus or minus 1.96 times $SQRT(SEE + S^2_y)$, where S^2_y is the standard error of the mean prediction as shown from the equations above. Prediction intervals are upper and lower bounds for the prediction of the dependent variable for a single case. Thus repeated independent random samples taken from the same population in an identical manner will yield confidence intervals that contain the true value of the parameter being estimated in 95% of the samples. The prediction interval will be wider (less certain) than the confidence interval, since it deals with an interval estimate of cases, not means.

4.2.3.1.6 F test

The F test is used to test the significance of R, which is the same as testing the significance of R^2, which is the same as testing the significance of the regression model as a whole. If prob(F) < α, then the model is considered significantly better than would be expected by chance and we reject the null hypothesis of no linear

relationship of y to the IVs. F is a function of R^2, the number of IVs, and the number of cases.

Thus, F is computed with k and (n - k - 1) degrees of freedom, where k is the number of terms in the equation, excluding the constant.

F is given as follows;

$$F = \frac{\left[\frac{R^2}{k}\right]}{\left[\frac{1-R^2}{n-k-1}\right]} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots equation(4.28)$$

Alternatively, F is the ratio of the mean square for the regression model divided by the mean square for error (residual).

4.2.3.1.7 Partial F test

The Partial F test can be used to assess the significance of the difference of two R^2 for nested models. Nested means that one is a subset of the other. Also, unique effects of individual IVs can be assessed by running a model with and without a given IV, then taking partial F to test for the difference. In this way, partial F plays a critical role in the trial-and-error process of model-building.

Let there be q, a larger model and let p be a nested smaller model.

Let RSSp be the residual sum of squares (deviance) for the smaller model.

Let RSSq be the residual sum of squares for the larger model.

Partial F has df($_1$) and df($_2$) degrees of freedom, where

$$df_1 = d_f \ for \ RSS_p - RSS_q \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots equation \ (4.29)$$

$$df_2 = d_f \ for \ RSS \ in \ the \ larger \ model \dots \dots \dots \dots \dots \dots equation \ (4.30)$$

$$Partial \ F = \left(\frac{(RSS_p - RSS_q)}{(df_1 * RSS_q/df_2)}\right) \dots \dots equation \ (4.31)$$

4.2.3.1.8 Effect size measures

The beta weights are the regression b coefficients for standardized data. Beta is the average amount the DV increases when the IV increases by one standard deviation when all other IVs are held constant. It is perfectly possible for some or even all beta weights to be greater than 1.0.

The model comparison method, sometimes called the dropping method, of assessing the relative importance of IVs is an alternative to the beta weight method. It is often preferred when the purpose is to build a model with fewer independent variables that add value or have an effect.

4.2.3.1.9 Correlation

Pearson's r^2 is the percent of variance in the DV explained by the given IV when (unlike the beta weights) all other IVs are allowed to vary. The result is that the magnitude of r^2 reflects not only the unique covariance it shares with the DV, but uncontrolled effects on the DV attributable to covariance the given IV shares with other IVs in the model. A rule of thumb is that multicollinearity may be a problem if a correlation is greater than 0.90 or several correlations are greater than 0.70 in the correlation matrix formed by all the IVs.

The intercept, variously expressed as e, c, or x-sub-0, is the estimated Y value when all the IVs have a value of 0. Sometimes this has real meaning and sometimes it does not, that is, sometimes the regression line cannot be extended beyond the range of observations, either back toward the Y axis or forward toward infinity.

4.2.3.1.10 R^2

R^2, multiple correlation or the coefficient of multiple determination means that the repeated independent random samples taken from the same population in an identical manner will yield confidence intervals that contain the true value of the parameter being estimated in 95% of the samples. R^2 can also be interpreted as the proportion of the variation of error in estimating the DV when knowing the IVs. That is, R^2 reflects

the number of errors made when using the regression model to estimate the value of the DV, in ratio to the total errors made when using only the mean of the DV as the basis for estimating all cases.

Mathematically,

$$R^2 = \left(1 - \left(\frac{SSE}{SST}\right)\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots equation\ (4.32)$$

$$SSE = \sum\left((Y_i - \hat{Y}_i)^2\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots equation\ (4.33)$$

where

SSE is error sum of squares

Y_i is the actual value of Y for the ith case

$\hat{Y}_i$ is the regression prediction for the ith case, and

SST is the total sum of squares.

4.2.3.1.11 Adjusted R^2

When comparing models with different numbers of IVs, Gujarati (2006) recommends that, it is a good practice to find the adjusted R^2 value because it explicitly takes into account the number of variables included in the model. The latter can be applied even when we are not comparing two regression models.

$$Adj\ R^2 = 1 - [(1 - R^2)(N - 1)/(N - k - 1)] \dots\dots\dots\dots\dots\dots equation\ (4.34)$$

where

n is sample size

k is the number of terms in the model excluding the constant.

4.2.3.1.12 Residuals

Residuals are the differences between the observed values and those predicted by the regression equation. Residuals measure the closeness of fit between the predicted and actual values. The method to calculate residual values are indicated below.

Residuals represent error. Residual analysis is used for three main purposes:

(1) to spot heteroscedasticity (i.e., increasing error as the observed Y value increases),

(2) to spot outliers (influential cases), and

(3) to identify other patterns of error (i.e., the error associated with certain ranges of X variables).

Five main types of residuals that are analysed in the data are:

Unstandardized residuals refer in a regression context to the linear difference between the locations of an observation (point) and the regression line (or plane or surface) in multidimensional space.

Standardized residuals are residuals after they have been constrained to a mean of zero and a standard deviation of 1. A rule of thumb is that an outlier is a point whose standardized residual is greater than 3.3 (corresponding to the .001 alpha level).

Deleted residuals, also called "jacknife residuals," compute the standard deviation omitting the given observation prior to standardizing or studentizing the residual. Deletion does not apply to unstandardized residuals, so "deleted residuals" are actually standardized deleted residuals. Analysis of outliers usually focuses on deleted residuals.

Studentized residuals are constrained only to have a standard deviation of 1, but are not constrained to a mean of 0.

Studentized deleted residuals are residuals which have been constrained to have a standard deviation of 1, after the standard deviation is calculated leaving the given

case out. Studentized deleted residuals are often used to assess the influence of a case and identify outliers.

The error term from the regression equation is unknown because the true model is unknown. Once the model has been estimated, the regression residuals are defined as:

$$\hat{e}_i = y_i - \hat{y}_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots equation\ (4.35)$$

where

y_i is the observed value of predictand in the year i

$\hat{y}_i$ is the predicted value in the year i.

Thus the regression equation consists of the following variables:

- The unknown parameters denoted as β or b, may be a scalar or a vector of length k.
- The independent variables X, are the predictor variables in the regression equation. Predictors are assumed to be continuous, interval variables, but also, it is common to notice the use of ordinal data in linear regression.
- The dependent variable *Y*, also known as *predictor* variable, is the predicted variable in the regression equation and is assumed to be continuous, interval variable. The regression equation is a function of variables X and β, and it is given by

$$Y = f(X, \beta) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots equation\ (4.36)$$

The simple regression equation is given by:

$$Y = \beta_0 + \beta_1 X_{1i} + \varepsilon_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots equation\ (4.37)$$

The estimates of beta weights β*₁* and β₀ are given by:

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots equation\ (4.38)$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \, equation \ (4.39)$$

where

$\bar{x}$ is the mean (average) of the x values and $\bar{y}$ is the mean of the y values.

Under the assumption that the error term has a constant variance, the estimate of that variance is given by:

$$\sigma_\tau^2 = \frac{SSE}{N-2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \, equation \ (4.40)$$

This is called the root mean square error (RMSE) of the regression.

The standard errors of the parameter estimates are given by:

$$\hat{\sigma}_{\beta_0} = \hat{\sigma}_\tau * \sqrt{\left(\frac{1}{N} + \frac{\bar{x}^2}{\Sigma(x_i - \bar{x})^2}\right)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \, equation \ (4.41)$$

and

$$\sigma_{\beta_1} = \sigma_\tau * \sqrt{\frac{1}{\Sigma \, (x_i - \bar{x})^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \, equation \ (4.42)$$

In matrix notation, the normal equations are written as

$$(X^T X)\hat{\beta} = X^T y \, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \, equation \ (4.43)$$

4.2.3.1.13 Stepwise regression

Stepwise regression is also called *statistical regression*. It is a way of computing regression in stages. In stage one, the IV best correlated with the DV is included in the equation. In the second stage, the remaining IV with the highest partial correlation with the DV, controlling for the first IV, is entered.

This process is repeated, at each stage partial-ling for previously-entered IVs, until the addition of a remaining IV does not increase R^2 by a significant amount (or until all variables are entered). Alternatively, the process can work backward, starting with all variables and eliminating IVs one at a time until the elimination of one makes a significant difference in R^2.

Thus, each variable is entered in sequence and its value is assessed. If adding the variable contributes to the model then it is retained, but all other variables in the model are then retested to see if they are still contributing to the success of the model. If they no longer contribute they are removed. This method ensures we ultimately end up with the smallest possible set of predictor variables to be included in the model.

Although several authors have widely used stepwise regression in selecting the best regression model while eliminating the insignificant IVs, some authors, however, criticised stepwise regression for various reasons. What follows are the criticisms of stepwise regression as highlighted by some authors.

Criticism of stepwise regression

Some statisticians criticize stepwise regression although it has been widely used in the literature. According to Hocking (1976), Draper and Smith (1981); stepwise regression includes regression models in which the choice of independent variables is carried out by an automatic procedure. Generally, this takes the form of a sequence of F-tests, but other techniques possible are t-tests, adjusted R^2, Akaike information criterion, Bayesian information criterion, Mallows' Cp, or false discovery rate. In statistics, the coefficient of determination R^2 is the proportion of variability in a data set that is accounted for by a statistical model. Stepwise regression consists of forward selection, backward and a combination of the two.

Several points about this method have been made:

1. A sequence of F-tests is often used to control the inclusion or exclusion of variables, but these are carried out on the same data and so there will be problems of multiple comparisons for which many correction criteria have been developed.

2. It is difficult to interpret the p-values associated with these tests, since each is conditional on the previous tests of inclusion and exclusion and so the p-values are compromised.

3. According to Rencher and Pun (1980), and Copas (1983), the tests themselves are biased, since they are based on the same data.

4.2.3.2 Model checks in time series method

Time series method consists of Smoothing and Box Jenkins methods which are discussed below.

4.2.3.2.1 Smoothing methods

Smoothing methods assess the model fit through various values for the parameter and select a value that looks reasonable either from the plot of data and forecasts, or from statistics such as the MSE, MPE, and MAPE as explained in the literature review chapter.

4.2.3.2.2 Box Jenkins models

4.2.3.2.2.1 Assumptions

Stationarity of the mean assumption: the data varies about a mean level. Another assumption in these models is that the variance is constant over the period of available data. Further to assume is that there is no strong seasonality present in the data.

Thus if the model is adequate, then it should not show any patterns and the residuals should not be seriously auto correlated. This can be checked by plotting residuals versus time.

4.2.3.2.2.2 Plot of autocorrelation and partial autocorrelation function

The function of ACF and PACF is to check the adequacy of the model. The plots of ACF and PACF include lines indicating which of the autocorrelations are significant. The ACF of a time series is a bar chart of the coefficients of correlation between a time series and lags of itself and the PACF is the plot or the partial correlation between a time series coefficients between the series and lags of itself. That is, ACF of a time series Y at lag 1 is the coefficient of the correlation between Y (t) and Y (t-1), which is presumably also the correlation between Y (t-1) and Y (t-2). Correlation can also be expected in Y (t) and Y (t-2) if Y (t) is correlated with Y (t-1) and Y (t-1) is correlated with Y (t-2). The PACF at all lags can be computed by fitting a succession of AR (k) where k=1, 2 ...n models with increasing numbers of lags. The PACF at lag k is equal to the estimated AR (k) coefficient in an autoregressive model with k terms. By inspection of the PACF, one can determine how many AR terms one need to use to explain the autocorrelation pattern in a time series. This is done to test if the above assumptions are met. If the assumptions are violated, then the data should be transformed either by differencing or by taking logs.

4.2.3.2.2.3 Box-Lyung statistic

Some problems may arise in testing time series plots partly due to many lags being tested simultaneously or partly due to the fact that the estimates of the correlations are themselves correlated, then the overall test should be done. Box-Lyung statistic is the test to be performed for cases like these. Box-Lyung statistic tests whether the autocorrelations at the first k lags are in accordance with the null hypothesis that they are all zero, that is, are consistent with the residuals forming a white noise process.

4.2.3.2.2.4 t-Statistic

If the model is not just adequate but also parsimonious, then all the parameters included in the model should be significant. This may be checked using the t statistic provided for each of the parameter estimates.

4.3 Summary

The mathematical formulation of the method applied in the data series has been reviewed in this chapter. The multiple regression and the time series mathematical representation of these methods have all been reviewed.

CHAPTER 5: DATA ANALYSIS AND RESULTS

5.1 Introduction

In this section, several methods are studied for operational electricity consumption forecast for the FeCr sector. Of the methods examined, the model that best fits the data and meets the assumption will be applied in forecasting electricity consumption in the FeCr sector.

For short term forecasting, the recommended model will be responsible for predicting daily electricity consumption forecasts for the FeCr sector to inform the monthly projection over a twelve month period (one year). For medium term the recommended model will be responsible for predicting electricity consumption forecasts for the ferrochrome sector over a ten year period. Short term forecasts are for operational planning while medium term forecasts are for tactical planning as explained in Chapter 1 of this study.

5.2 Data Analysis

The data set available was fitted for both short and medium term purposes. In this section, the data is analysed for short and medium term using the best standards of analysis explored in Chapter 2. The analysis is based on causal and time series analysis.

5.2.1 Short Term Forecasting

An analysis of short term data set is as follows:

5.2.1.1 The data set

The electricity consumption data set of the FeCr sector used in this study consists of 1827 data points collected on a daily basis from April 2003 to March 2008 as indicated in Figure 5.1. The period for which the estimation of parameters and other specifications are done covers the 1827 observations. In order to achieve comparability with the previous actual figures, the year to be forecasted started from April 2008 to March 2009. The forecasts produced are for the period of twelve months generated from daily outcomes. These forecasts are for short term (one month – twelve months forecast).

Figure 5.1: Ferrochrome daily electricity consumption data from 2003 to 2008
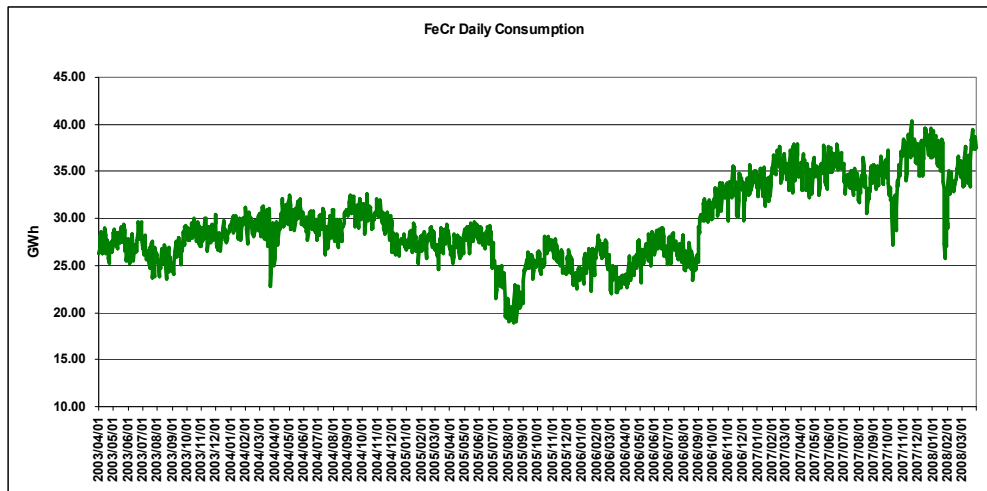
5.2.1.2 Descriptive statistics

The descriptive statistics in Table 5.1 and Table 5.2 shows an increasing growth in demand of FeCr sector. The FeCr sector started with a consumption of as low as 18 GWh per day and within a period of five years it had reached a maximum consumption of 40 GWh per day (Table 5.1 and Table 5.2).

Table 5.1: Summary Descriptive statistics of daily electricity consumption of the FeCr sector

Variable	Mean	Std.Dev	Minimum	Maximum	N	Missing
Consumption FeCr (GWh)	29.48	4.08	18.94	40.35	1827	0

The data does not show any significant skewness and kurtosis (Table 5.2). Skewness ranges between -1 and +1. Thus skewness of 0.37 is within range and therefore not significant. If kurtosis is less than -1.2 then it is significant .Thus in Table 5.2, kurtosis is equal -0.44 and thus is insignificant.

Table 5.2: Detailed Descriptive statistics of daily electricity consumption of the FeCr sector

Descriptive Statistics (FeCr Main Data)	
Valid N	1827.00
Mean	29.48
Confidence -95%	29.30
Confidence +95%	29.66
Geometric Mean	29.20
Median	28.57
Minimum	18.94
Maximum	40.34
Lower Quartile	26.73
Upper Quartile	32.52
Percentile 10'th	24.87
Percentile 90'th	35.60
Range	21.40
Quartile Range	5.79
Variance	16.61
Std.Dev.	4.08
Standard Error	0.09
Skewness	0.37
Std.Err. Skewness	0.06
Kurtosis	-0.44
Std.Err. Kurtosis	0.11

5.2.1.3 Box and Whisker plot

The average electricity consumption of the FeCr sector per day is 29 GWh as illustrated in Figure 5.2. The simple plot in Figure 5.1 also illustrates that consumption of electricity by FeCr sector has been fluctuating over the five year period in this study. This behaviour is attributable to market dynamics shared in

Chapter 3 of this study, which include the demand, outages, commodity prices, cost of production, and unforeseen world events. An increasing growth in consumption is also due to expansions and the completion of new projects.

Figure 5.2: <u>**Box and Whisker plot of FeCr daily electricity consumption**</u>
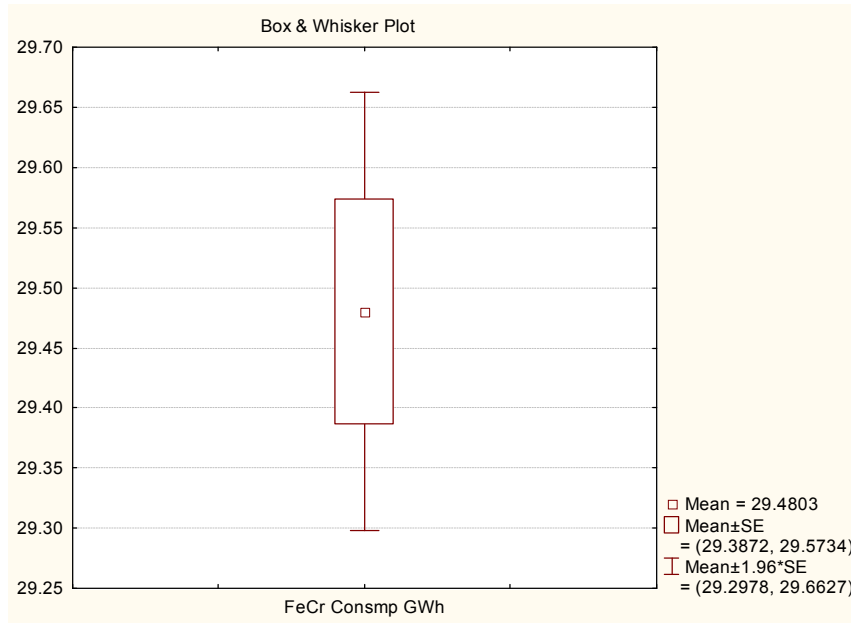
5.2.1.4 Normality

The data follows a normal distribution as demonstrated in Figure 5.3 and Figure 5.4. In this regard, no transformation is required for this data set.

Figure 5.3: <u>Distribution of daily electricity consumption for the FeCr sector</u>
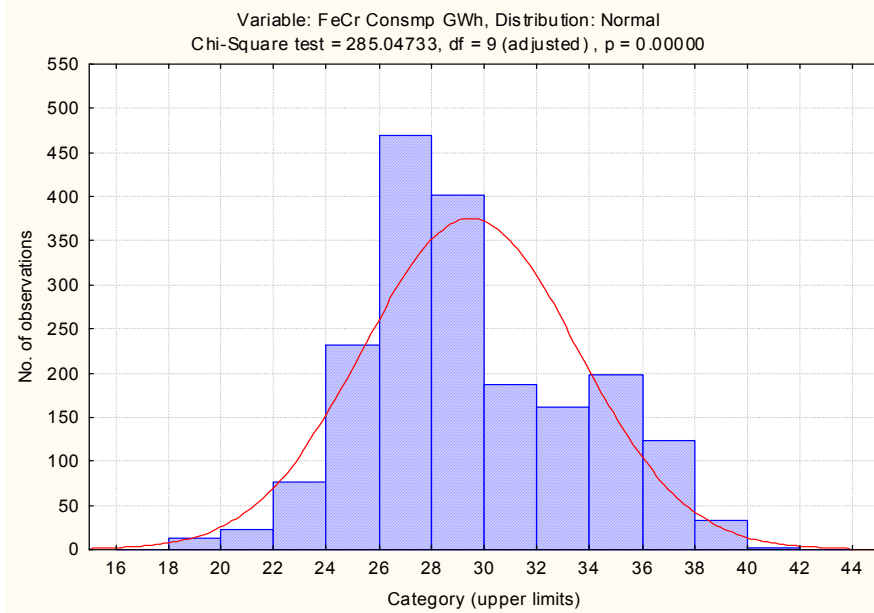
Figure 5.4: <u>Normal probability plot for daily electricity consumption of the FeCr sector</u>
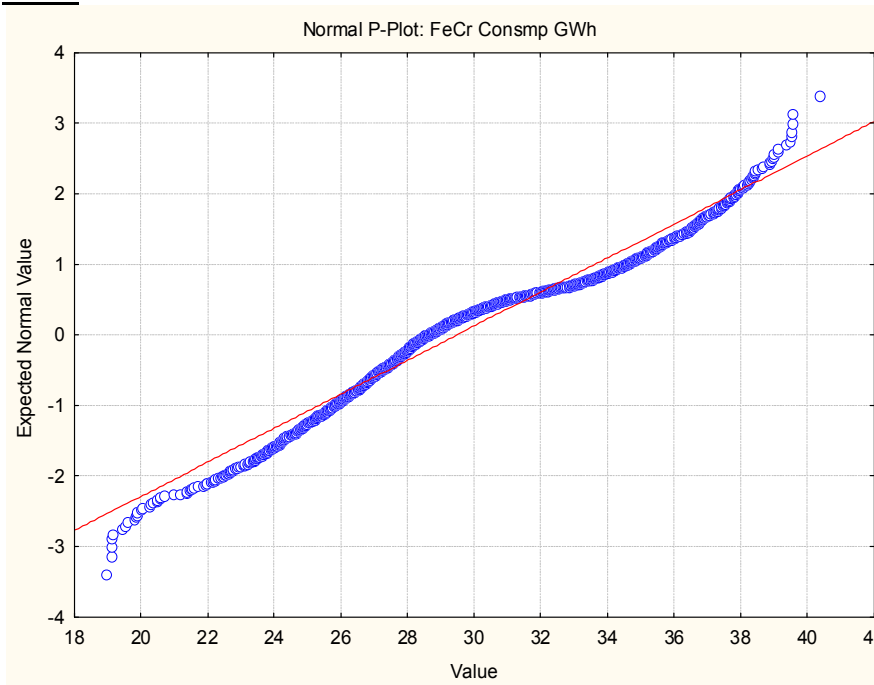
5.2.1.5 Model checks: Autocorrelation and Partial Autocorrelation functions

The purpose of the ACF (Figure 5.5) and PACF (Figure 5.6) is to identify potential models based on the patterns of these functions. ARIMA(2,0,2)(1,0,0), (2,1,1)(1,0,0)

and (2,0,2)(1,1,0) models have been fitted to the data. D(1), is the first order of differencing that has been applied to the models.

From the data analysis section, we noted that the data set is normally distributed although there are some small seasonality signs. Thus the analysis shows that the data can be used as it is, in coming up with a forecasting model. To select the model that best fits the data, several models were tried (Figures 5.5, 5.6, 5.7, 5.8, 5.9 and Figure 5.10). Most of the residual values are within the upper control limit and lower control limit band which implies that they are random and are not correlated. The other two models fitted are good but the better estimates emerge from the model ARIMA(2,0,2)(1,0,0).

Figure 5.5: **ACF of ARIMA(2,0,2)(1,0,0)**

Autocorrelation Function
FeCr Consmp KWh: ARIMA (2,0,2)(1,0,0) residuals;
(Standard errors are white-noise estimates)

Lag	Corr.	S.E.	Q	p
1	-.003	.0227	.02	.8841
2	+.006	.0227	.09	.9542
3	-.024	.0227	1.18	.7575
4	-.053	.0227	6.68	.1540
5	-.012	.0227	6.97	.2227
6	+.059	.0227	13.66	.0337
7	+.137	.0227	50.00	.0000
8	+.041	.0227	53.23	.0000
9	-.033	.0227	55.34	.0000
10	-.122	.0227	84.25	.0000
11	-.091	.0227	100.3	.0000
12	+.008	.0227	100.5	.0000

--- Conf. Limit
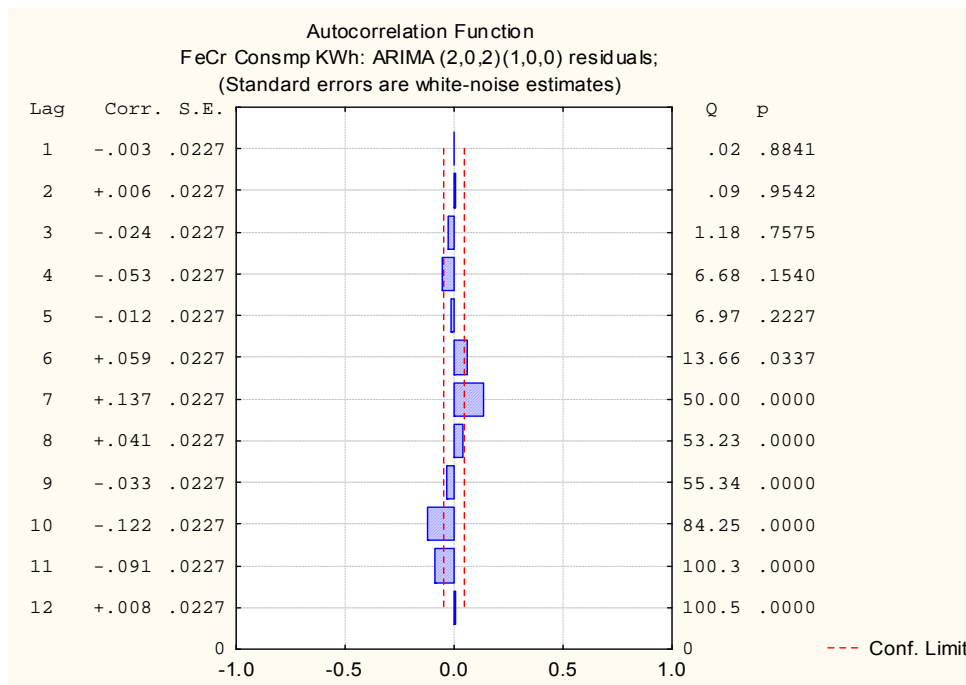
Figure 5.6: ACF of ARIMA(2,1,2)(1,0,0)

Autocorrelation Function
FeCr Consmp KWh: ARIMA (2,1,2)(1,0,0) residuals;
(Standard errors are white-noise estimates)

Lag	Corr.	S.E.		Q	p
1	-.179	.0228		61.38	.0000
2	-.080	.0228		73.65	.0000
3	+.039	.0228		76.58	.0000
4	-.031	.0228		78.37	.0000
5	-.091	.0228		94.15	.0000
6	-.083	.0228		107.4	0.000
7	+.036	.0228		109.8	0.000
8	-.011	.0228		110.0	0.000
9	+.019	.0228		110.7	0.000
10	-.033	.0228		112.8	0.000
11	-.018	.0228		113.5	0.000
12	-.000	.0228		113.5	.0000

...... Conf. Limit

Figure 5.7:ACF of ARIMA(2,0,2)(1,1,0)

Autocorrelation Function
FeCr Consmp KWh: ARIMA (2,0,2)(1,1,0) residuals;
(Standard errors are white-noise estimates)

Lag	Corr.	S.E.		Q	p
1	-.179	.0228		61.38	.0000
2	-.080	.0228		73.65	.0000
3	+.039	.0228		76.58	.0000
4	-.031	.0228		78.37	.0000
5	-.091	.0228		94.15	.0000
6	-.083	.0228		107.4	0.000
7	+.036	.0228		109.8	0.000
8	-.011	.0228		110.0	0.000
9	+.019	.0228		110.7	0.000
10	-.033	.0228		112.8	0.000
11	-.018	.0228		113.5	0.000
12	-.000	.0228		113.5	.0000

- - - Conf. Limit

Figure 5.8: PACF of ARIMA(2,0,2) (1,0,0)

Partial Autocorrelation Function
FeCr Consmp KWh: ARIMA (2,0,2)(1,0,0) residuals;
(Standard errors assume AR order of k-1)

Lag	Corr.	S.E.
1	−.003	.0228
2	+.006	.0228
3	−.024	.0228
4	−.053	.0228
5	−.013	.0228
6	+.059	.0228
7	+.136	.0228
8	+.040	.0228
9	−.034	.0228
10	−.116	.0228
11	−.081	.0228
12	+.012	.0228

--- Conf. Limit

-1.0   -0.5   0.0   0.5   1.0
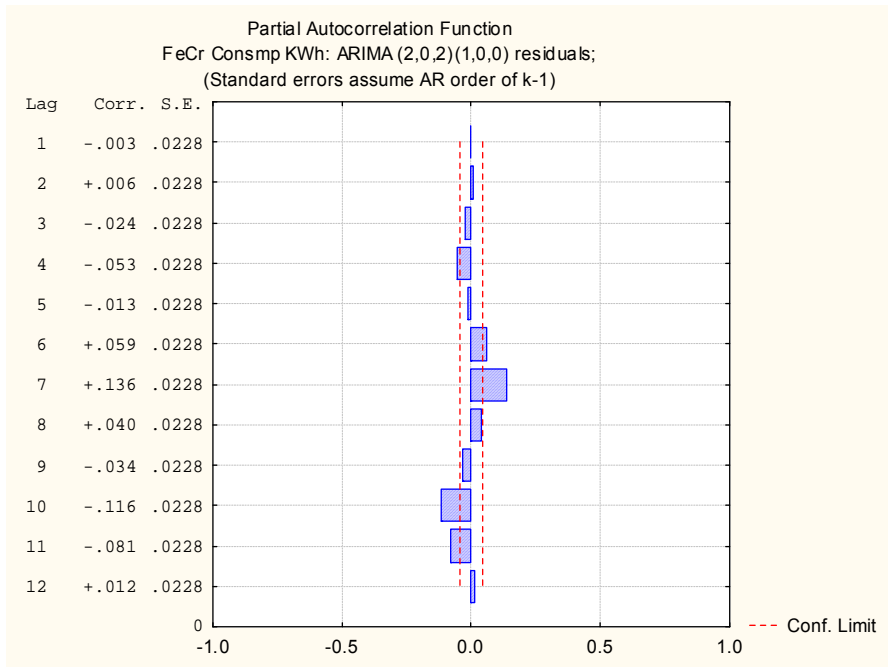
Figure 5.9: <u>PACF of ARIMA(2,1,2)(1,0,0)</u>

Partial Autocorrelation Function
FeCr Consmp KWh: ARIMA (2,1,2)(1,0,0) residuals;
(Standard errors assume AR order of k-1)

Lag	Corr.	S.E.
1	-.179	.0228
2	-.116	.0228
3	+.003	.0228
4	-.033	.0228
5	-.103	.0228
6	-.134	.0228
7	-.028	.0228
8	-.031	.0228
9	+.006	.0228
10	-.056	.0228
11	-.060	.0228
12	-.041	.0228
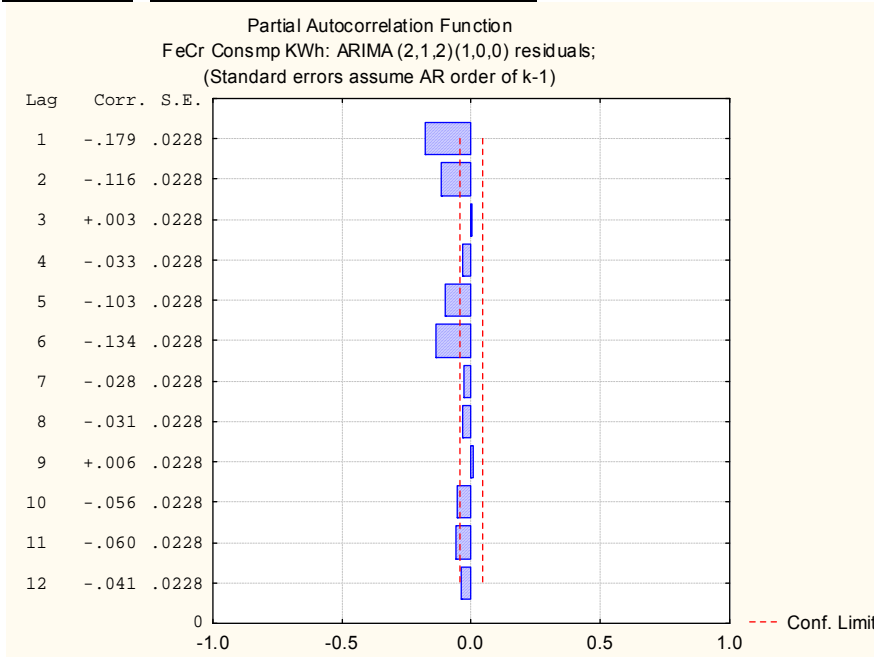
Figure 5.10: <u>PACF of ARIMA(2,0,2)(1,1,0)</u>

Partial Autocorrelation Function
FeCr Consmp KWh: ARIMA (2,0,2)(1,1,0) residuals;
(Standard errors assume AR order of k-1)

Lag	Corr.	S.E.
1	-.179	.0228
2	-.116	.0228
3	+.003	.0228
4	-.033	.0228
5	-.103	.0228
6	-.134	.0228
7	-.028	.0228
8	-.031	.0228
9	+.006	.0228
10	-.056	.0228
11	-.060	.0228
12	-.041	.0228
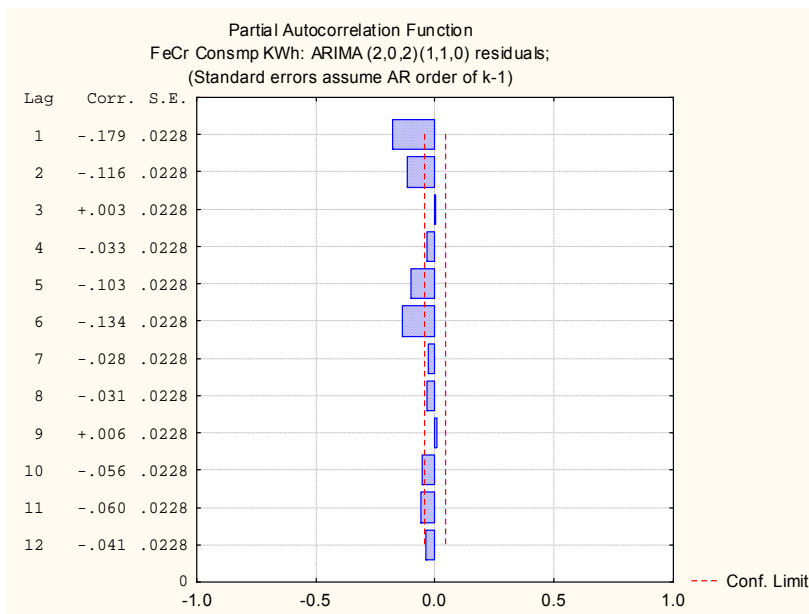
5.2.1.6 Residual Analysis

Figures 5.11, 5.12, and Figure 5.13 show the residual plot of ARIMA(2,02)(1,0,0), ARIMA(2,1,2)(1,0,0) and ARIMA(2,0,2)(1,1,0). It is observed that all the ARIMA

methods show residual values that are constant around the mean confirming that the data is normal.

Figure 5.11: Residual plots of ARIMA(2,0,2)(1,0,0)
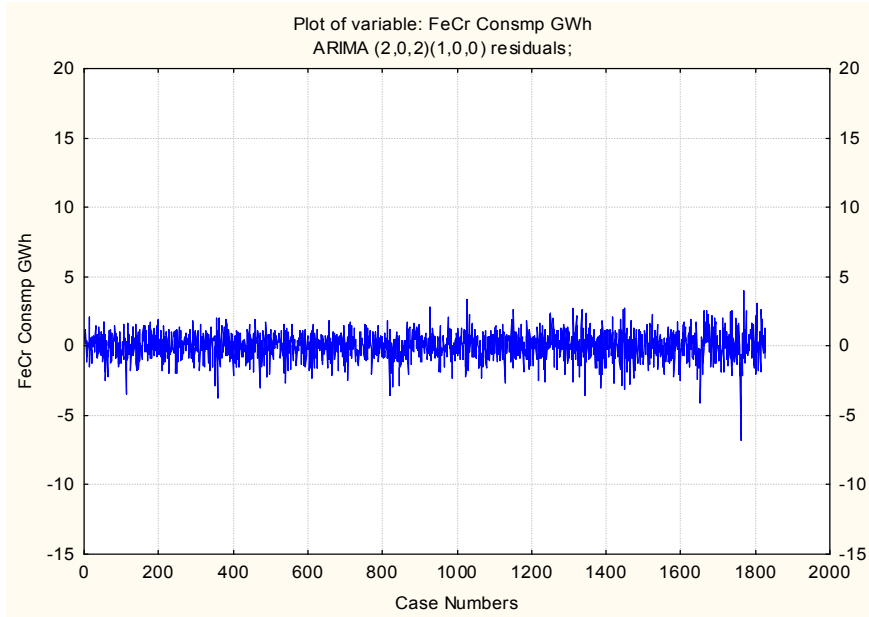
Figure 5.12: Residual plots of ARIMA(2,1,2)(1,0,0),
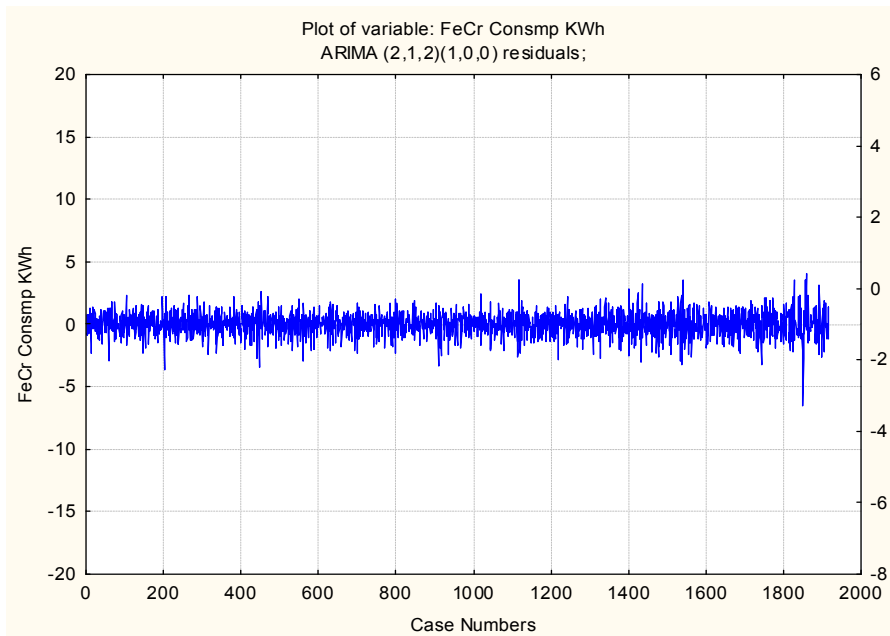
Figure 5.13: Residual plots of ARIMA(2,0,2)(1,1,0)

Plot of variable: FeCr Consmp KWh
ARIMA (2,0,2)(1,1,0) residuals;
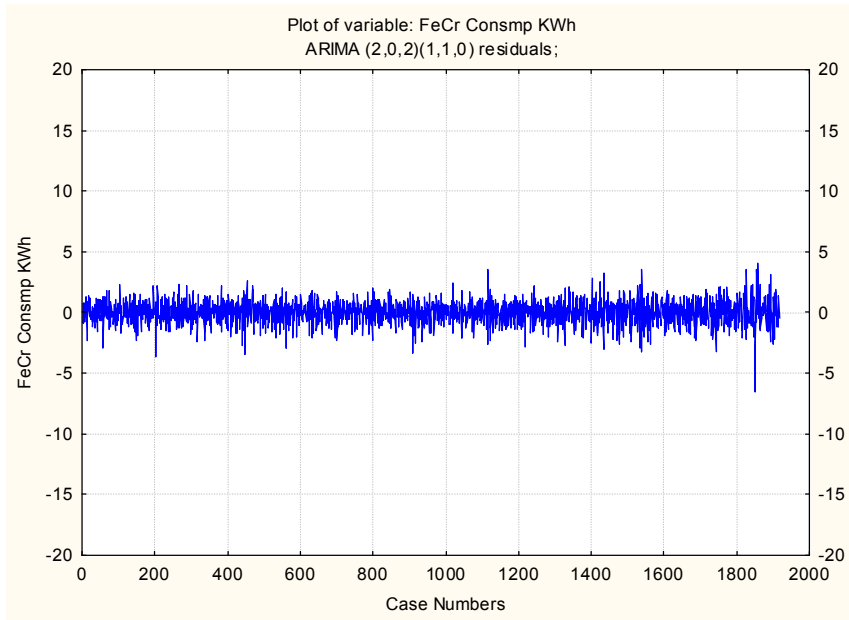
Figure 5.14: Plot of FeCr consumption kWh

Plot of variable: FeCr Consmp KWh
D(1);

5.2.1.7 Model fit

The goodness of fit for ARIMA and exponential smoothing models is demonstrated in Table 5.3 and Table 5.4 respectively. The goodness of fit test indicate that ARIMA(2,0,2)(1,0,0) is the best model when compared to the other two ARIMA

72

models. All the parameter estimates for ARIMA(2,0,2)(1,0,0) are significant. Although ARIMA(2,1,2)(1,0,0), has a good Bayes information criterion (BIC), two parameter estimates are not significant, and this can lead to an unreliable forecast. ARIMA(2,0,2)(1,1,0) is the worst model in this case with the highest BIC. ARIMA(2,0,2)(1,0,0) also shows that there is no autocorrelation between residual values since the Durbin Watson (D-W) statistic of 2 is significant.

Thus with MAPE of 2.6% which is significant at α = 5%, BIC of 0.99, R^2 of 94.4% and Mean Absolute deviation of 0.76 it can be concluded that ARIMA(2,0,2)(1,0,0) fits the data well when compared to the other two ARIMA models.

Table 5.3: Fitted Box Jenkins ARIMA Models

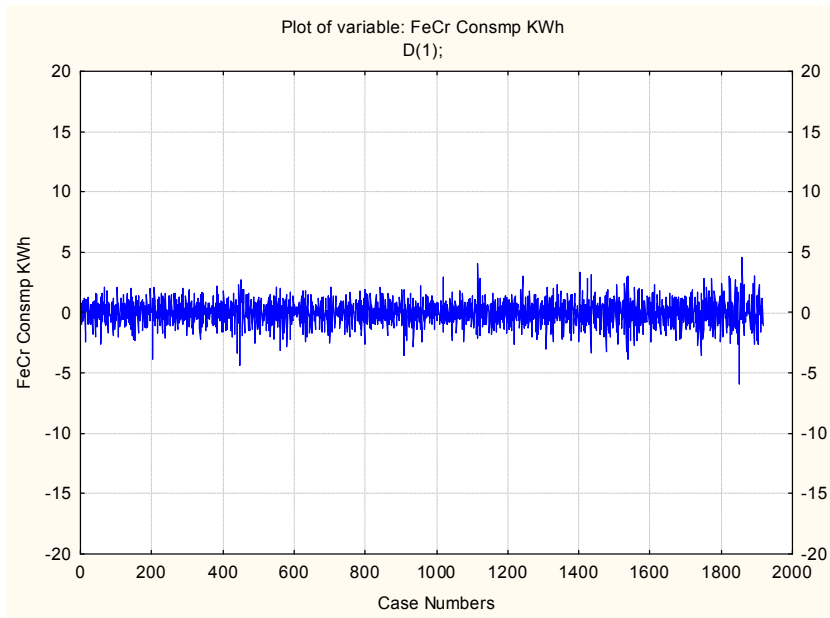| Forecast Model for FeCrConsmpKWh ARIMA(2,0,2)*(1,0,0)					Forecast Model for FeCrConsmpKWh ARIMA(2,1,2)*(1,0,0)					Forecast Model for FeCrConsmpKWh ARIMA(2,0,2)*(1,1,0)				
Term	Coefficient	Std. Error	t-Statistic	Significance	Term	Coefficient	Std. Error	t-Statistic	Significance	Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	1.4811	0.1038	14.2703	1.0000	a[1]	0.8820	0.1932	4.5658	1.0000	a[1]	1.4388	0.1339	10.7449	1.0000
a[2]	-0.4813	0.0698	-6.8980	1.0000	a[2]	-0.2964	0.1119	-2.6494	0.9919	a[2]	-0.4396	0.0927	-4.7427	1.0000
b[1]	0.7063	0.0517	13.6625	1.0000	b[1]	1.1016	0.1970	5.5917	1.0000	b[1]	0.6634	0.0694	9.5551	1.0000
b[2]	0.1087	0.0327	3.3225	0.9991	b[2]	-0.2688	0.1669	-1.6103	0.8927 <-	b[2]	0.1414	0.0413	3.4258	0.9994
A[365]	0.2812	0.0263	10.7023	1.0000	A[365]	0.2614	0.0268	9.7439	1.0000	A[365]	-0.6649	0.0226	-29.4422	1.0000
_CONST	0.0034				Insignificant MA terms are harmless.									

Within-Sample Statistics			Within-Sample Statistics			Within-Sample Statistics		
Sample size 1827	Number of parameters 5		Sample size 1827	Number of parameters 5		Sample size 1827	Number of parameters 5	
Mean 29.57	Standard deviation 4.146		Mean 29.57	Standard deviation 4.146		Mean 29.57	Standard deviation 4.146	
R-square 0.9437	Adjusted R-square 0.9436		R-square 0.9438	Adjusted R-square 0.9436		R-square 0.9418	Adjusted R-square 0.9416	
Durbin-Watson 2.002	** Ljung-Box(18)=229.4 P=1		Durbin-Watson 2.01	** Ljung-Box(18)=198.5 P=1		Durbin-Watson 1.982	** Ljung-Box(18)=363.1 P=1	
Forecast error 0.9844	BIC 0.9932		Forecast error 0.9842	BIC 0.993		Forecast error 1.002	BIC 1.011	
MAPE 0.02596	RMSE 0.983		MAPE 0.02593	RMSE 0.9828		MAPE 0.02307	RMSE 1	
MAD 0.7575			MAD 0.7566			MAD 0.6801		

Table 5.4: Fitted Exponential Smoothing models

Model: Exponential Smoothing Forecast Model for FeCrConsmpKWh Exponential smoothing: No trend, Additive seasonality			Forecast Model for FeCrConsmpKWh Simple exponential smoothing: No trend, No seasonality Confidence limits proportional to level		
Component	Smoothing Weight	Final Value	Component	Smoothing Weight	Final Value
Level	0.81965	37.994	Level	0.84375	37.699
Seasonal	0.57639				

Within-Sample Statistics		Within-Sample Statistics	
Sample size 1827	Number of parameters 2	Sample size 1827	Number of parameters 1
Mean 29.57	Standard deviation 4.146	Mean 29.57	Standard deviation 4.146
R-square 0.9419	Adjusted R-square 0.9419	R-square 0.9356	Adjusted R-square 0.9356
Durbin-Watson 1.901	** Ljung-Box(18)=646.5 P=1	Durbin-Watson 1.934	** Ljung-Box(18)=681 P=1
Forecast error 0.9993	BIC 1.003	Forecast error 1.052	BIC 1.054
MAPE 0.0264	RMSE 0.9988	MAPE 0.02803	RMSE 1.052
MAD 0.7725		MAD 0.8175	

Validity of the two exponential smoothing models is tested as shown in Table 5.4. It can be observed that exponential smoothing with no trend but additive seasonality fits the data better when compared to the other model (exponential smoothing model with no trend and no seasonality). Exponential smoothing with no trend but additive seasonality exhibits a lower MAPE (2.6%). Furthermore, exponential smoothing with no trend and additive seasonality has a lower BIC but comes out with a higher R^2 of 94.2%.

Based on this analysis we can conclude that exponential smoothing with no trend but additive seasonality has a better fit compared to exponential smoothing model with no trend and no seasonality. This is also re-affirmed by Figure 5.1 which shows no trend.

5.2.1.8 Results and discussion

From the Box and Jenkins method, ARIMA(2,0,2)(1,0,0) fitted the data well while for exponential smoothing models, Winter's exponential smoothing fitted the data well too. We next make a comparative analysis of the two models so that we find the best model that best fit the data set. The purpose is to find a model that can produce electricity consumption forecast effectively for short term (one to twelve months forecast) for operational use in the FeCr sector.

The adequacy of the two models, ARIMA(2,0,2)(1,0,0) and Winter's exponential smoothing is compared as shown in Table 5.5. It is observed that ARIMA(2,0,2)(1,0,0) is a better model when compared to Winter exponential smoothing as it has a lower BIC criterion (0.993) and also exhibits a lower MAPE (2.596%). Above all this, it has a marginally higher R^2 (94.4%) compared to the Winter's exponential smoothing model.

The D-W statistic for ARIMA(2,0,2)(1,0,0) is also significant at 2.001. Based on this analysis, ARIMA(2,0,2)(1,0,0) is the best model to forecast electricity in the short term for operational use in the FeCr sector.

Table 5.5: Findings ARIMA(2,0,2)(1,0,0) Winters Exponential Smoothing

Forecast Model for FeCrConsmpKWh ARIMA(2,0,2)*(1,0,0)					Model: Exponential Smoothing Forecast Model for FeCrConsmpKWh Exponential smoothing: No trend, Additive seasonality		
Term	Coefficient	Std. Error	t-Statistic	Significance	Component	Smoothing Weight	Final Value
a[1]	1.4811	0.1038	14.2703	1.0000	Level	0.81965	37.994
a[2]	-0.4813	0.0698	-6.8980	1.0000	Seasonal	0.57639	
b[1]	0.7063	0.0517	13.6625	1.0000			
b[2]	0.1087	0.0327	3.3225	0.9991			
A[365]	0.2812	0.0263	10.7023	1.0000			
CONST	0.0034						
Within-Sample Statistics					**Within-Sample Statistics**		
Sample size 1827	Number of parameters 5				Sample size 1827	Number of parameters 2	
Mean 29.57	Standard deviation 4.146				Mean 29.57	Standard deviation 4.146	
R-square 0.9437	Adjusted R-square 0.9436				R-square 0.9419	Adjusted R-square 0.9419	
Durbin-Watson 2.002	** Ljung-Box(18)=229.4 P=1				Durbin-Watson 1.901	** Ljung-Box(18)=646.5 P=1	
Forecast error 0.9844	BIC 0.9932				Forecast error 0.9993	BIC 1.003	
MAPE 0.02596	RMSE 0.983				MAPE 0.0264	RMSE 0.9988	
MAD 0.7575					MAD 0.7725		

Figure 5.15: Daily electricity consumption forecast versus actual
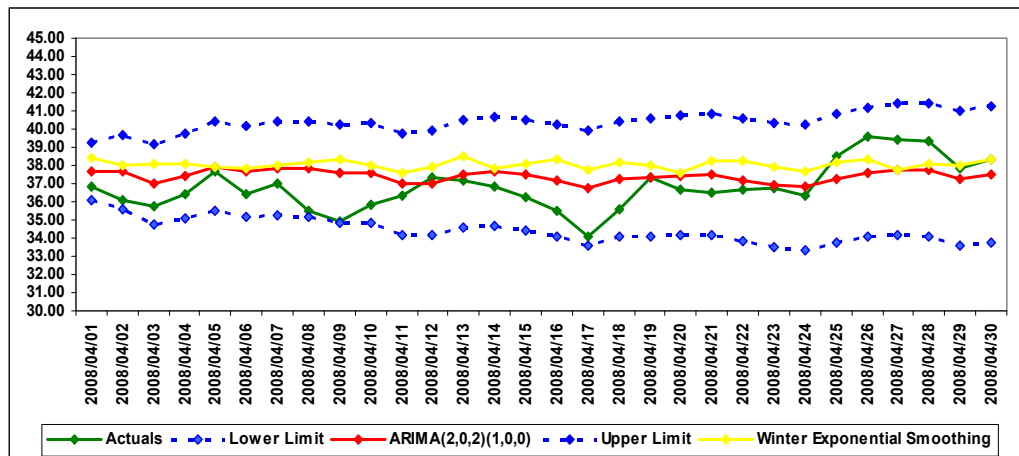
Actual values and estimated figures using the two models (ARIMA(2,0,2)(1,0,0) and Winter's exponential smoothing) are shown in Figure 5.15. This also confirms that ARIMA(2,0,2)(1,0,0) is more superior than the Winter's exponential smoothing since it exhibits lower deviations from the actual. Based on chapter 4, section 4.2.3.1.5, the upper and lower limit of 95% confidence limit calculated from ARIMA(2,0,2)(1,0,0) method as the most efficient model, illustrates that the forecasts being estimated, will yield the confidence interval that contain the true value in 95% of the samples.

75

Figure 5.16: <u>Monthly electricity consumption forecast versus actual</u>
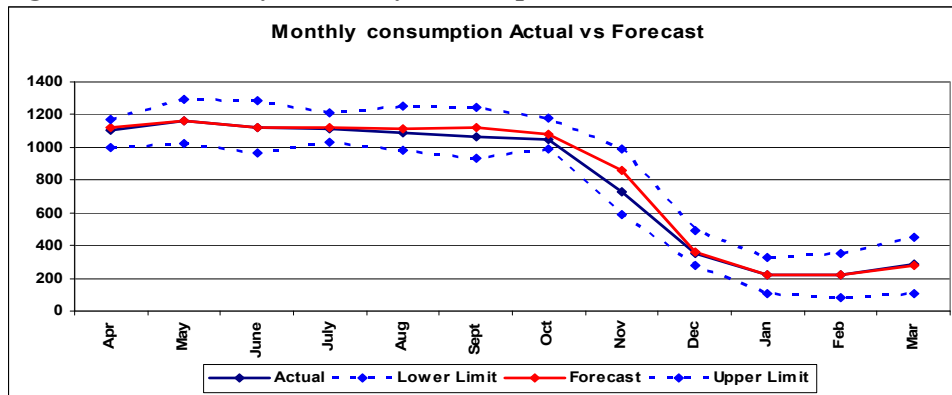
The results shown in Figure 5.16 are of one year forecast from April to March (twelve months period) derived from applying ARIMA(2,0,2)(1,0,0) model. The model shows good performance over the period of twelve months since the forecasts show minimal deviations from the actual values and the forecast values are within the lower and upper bounds. The noticeable gap between the months of October to December is due to immeasurable changes brought by the global melt down that has vigorously affected the behaviour of the electricity consumption sector. However after the melt down, the model is showing a strong upward trend as the economy picks up.

In Figure 5.17 the monthly MAPE for the forecasts are compared to 5% level of significant. Figure 5.17 indicates that during the month of November, the forecast was above 5% level of significance. This corresponds to the start of the global melt down. Thus, the ARIMA(2,0,2(1,0,0) method is recommended to produce twelve months forecasts for operational use in the FeCr sector.

76

Figure 5.17: **Monthly MAPE versus significant α = 0.05**
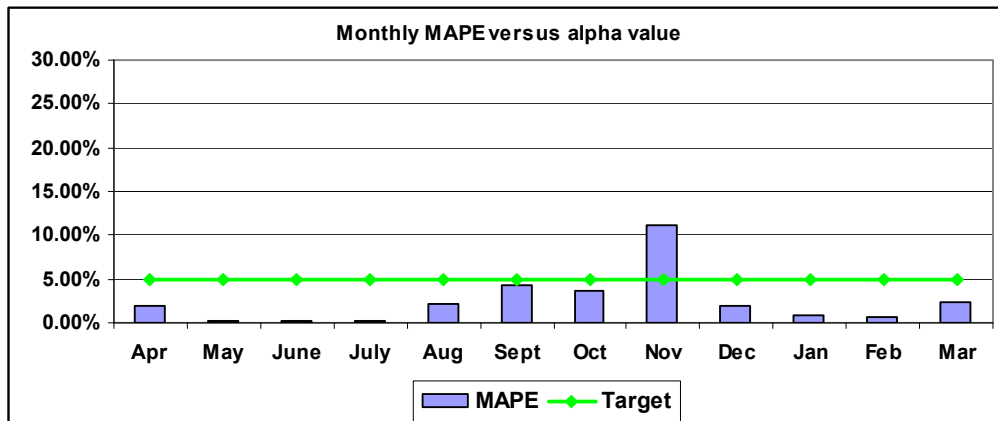
5.2.2 Medium Term Forecasting

In analysing the medium term forecast, we consider the data set, analyse descriptive statistics, check for normality, derive model that fits the data, check model adequacy and finally study the results for all the models that are fitted.

In this section, there are two sets of data. The data set available is the yearly data set that consists of 29 actual values from 1980 – 2008. The other data set consists of 80 actual values from 2003 – 2008 captured monthly. Several methods were examined in order to find the best model that can forecast electricity consumption for medium term use. The methods applied are the Multiple Linear regression using the yearly data set, and the Box Jenkins and exponential smoothing using monthly data set. Findings from all these models are presented in sections that follow.

5.2.2.1 Multiple Linear Regression

5.2.2.1.1. Data set

Figure 5.18 shows electricity consumed on a yearly basis from 1980 to 2008 by the FeCr sector. The data reveals an increasing trend over the period. An attempt is made to fit multiple regression model. Regrettably, only 29 data points are considered. Other models such as exponential smoothing will be fitted to the same data but

77

monthly data points will be considered. This data set is too small to apply the Box and Jenkins method.

Figure 5.18: <u>Yearly electricity consumption for the FeCr sector</u>
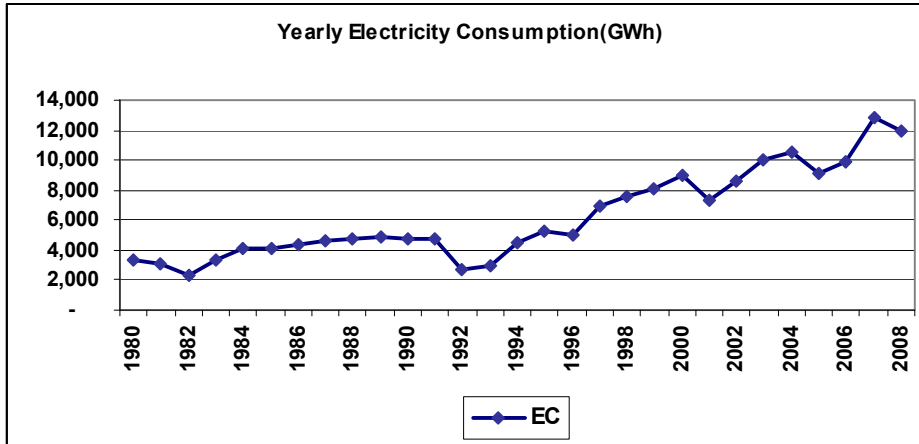
In applying multiple regression model, electricity consumption was the dependent variable while independent variables that affect electricity consumption (EC) were identified as follows;

      Independent variable – Gross Domestic Product (GDP)

      Independent variable – World Stainless Steel Consumption (WSSC)

      Independent variable – South Africa Stainless Steel Consumption (SASSC)

      Independent variable – World Ferrochrome Supply (WFeCrS)

      Independent variable – US dollar (USd).

The independent variables selected to be used in the model are shown in Figure 5.19.

All the independent variables indicate a strong relationship to electricity consumption (dependent variable) with respective R^2 values well above 50% (Figure 5.19).

Correlation between variables is shown in Table 5.6. It can be observed that there is a strong correlation between independent variables and dependent variable. In addition there is also a strong correlation between independent variables indicating multicollinearity. This means that all the independent variables can be used for analysis and prediction of electricity consumption forecast.

Figure 5.19: **Scatter plot of dependent versus independent variables**

Scatter plot EC versus GDP   y = 0.0162x - 7706.7
R² = 0.9028

Scatter plot EC versus WSSC   y = 0.4533x - 1058.5
R² = 0.8992

Scatter plot EC versus SASSC   y = 8.1368x + 3514.3
R² = 0.7498

Scatter plot EC versus WFeCrS   y = 2.0663x - 2491.3
R² = 0.9018

Scatter plot EC versus USd   y = 955.17x + 2138.3
R² = 0.7348
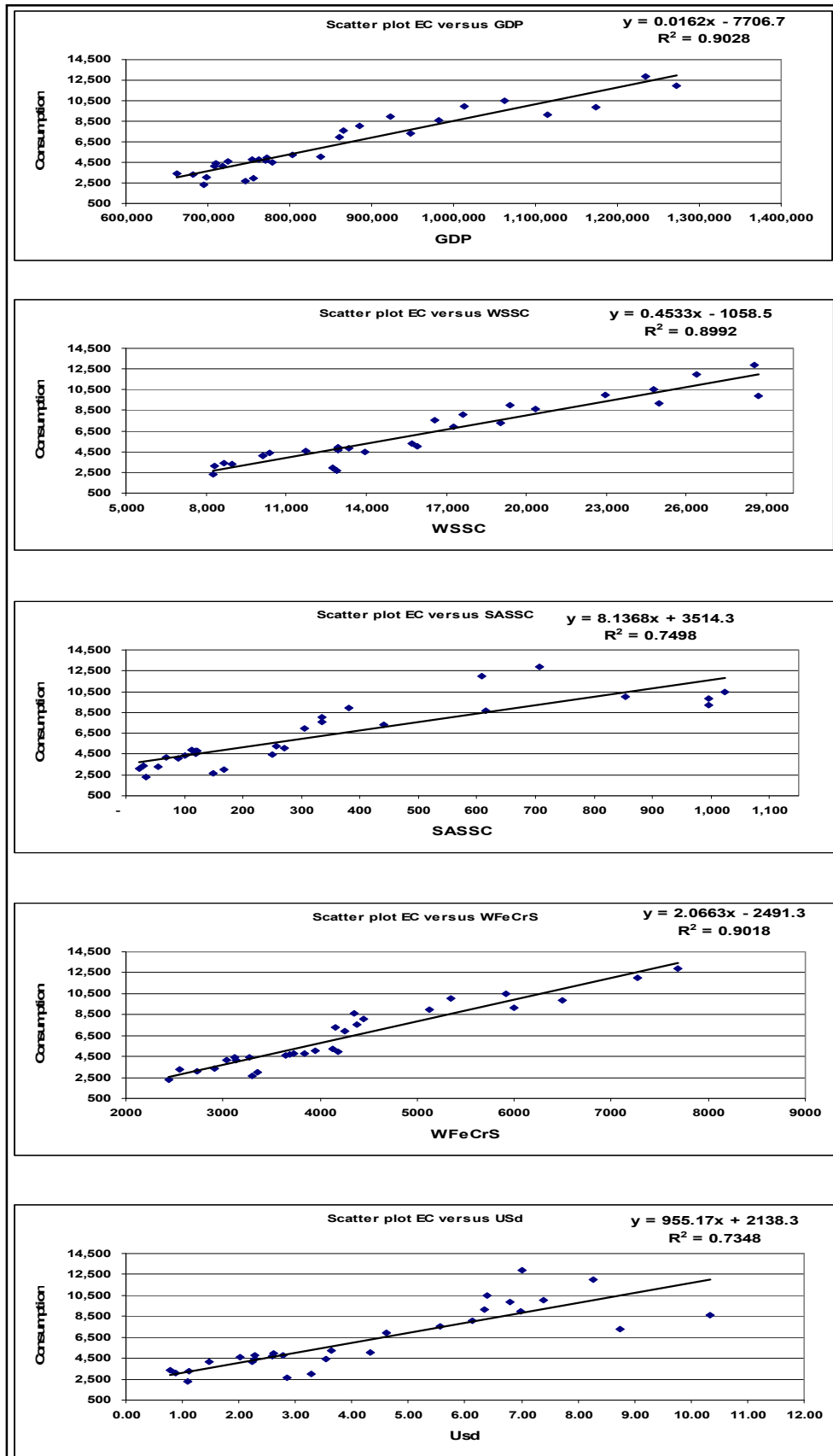
Some of the independent variables were not normal and had to be transformed so as to satisfy the important assumption that the DV and the IVs are linearly related. This is done in detail under diagnostic checking (Section 5.2.2.1.2).

Table 5.6: Correlation matrix between variables

	GDP	WSSC	SASSC	WFeCrS	Usd	EC
GDP	1.00	0.98	0.90	0.97	0.84	0.95
WSSC	0.98	1.00	0.94	0.96	0.86	0.95
SASSC	0.90	0.94	1.00	0.85	0.80	0.87
WFeCrS	0.97	0.96	0.85	1.00	0.76	0.95
Usd	0.84	0.86	0.80	0.76	1.00	0.86
EC	0.95	0.95	0.87	0.95	0.86	1.00
Means	859025.1	16088.6	334.3	4222.9	4.3	6234.3
Std.Dev.	173483.9	6198.8	315.3	1361.8	2.7	2963.2
No.Cases	29					
Matrix	1					

5.2.2.1.2 Model checks

To check for model adequacy, diagnostic checking using multiple regression analysis tools is explored.

The first part of regression analysis is to test if the data does not violate any of the assumptions illustrated in the literature review in Chapter 2 of this study. This includes normality assumption, constant residuals and homoscedasticity.

5.2.2.1.3 Normality assumption

To test for this assumption for both dependent and independent variables, Kolmogorov–Smirnov test was applied.

Kolmogorov–Smirnov test is used to test the null hypothesis that the data comes from a normally distributed set against the alternative hypothesis that the data do not come from a normally distributed set. The null hypothesis is rejected if the calculated α is less than $\alpha=0.05$ significant level.

Based on the analysis in Figure 5.13, we reject the null hypothesis (P<0.05) and conclude that the data (for the medium term) is normal.

The data for both the independent and dependent variables were not normal except for US dollar. Transformation was applied as follows:

Dependent variable Electricity Consumption (EC), independent variables World Stainless Steel Consumption (WSSC) and World Ferrochrome Supply (WfeCrS), were transformed by taking the logarithm function.

The independent variable Gross Domestic Product (GDP) was transformed using the reciprocal transformation (multiplicative inverse).

No transformation was necessary for Usdollar (Usd) as it was normal.

5.2.2.1.4 Test for residuals

The main aim of testing for residuals is so that heteroscedasticity (variance of residuals not constant over time) and outliers are spotted, as well as identifying other patterns of error.

The assumption of homoscedasticity is that the residuals are approximately equal for all predicted dependent variable scores. Alternatively, the variability in scores for dependent variable is the same for all values of the independent variable. Heteroscedasticity is usually shown by a cluster of points that is wider as the values for the predicted energy consumption get larger.

The residual plots that have been examined (Table 5.7) show the data is fairly homoscedastic. In fact, residual plots show that the data meets the assumptions of homoscedasticity, linearity, and normality because the residual plot is rectangular, with a concentration of points along the centre.

The data also depict few outliers that are within range. Thus we can conclude that the data do not have worrying outliers that can risk the outcome of the analysis.

Table 5.7: Residual test

	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance
1	8.128881	8.083519	0.045362	-1.19096	0.29937	0.112097	14.35866	0.100200	0.039888
2	8.032360	8.020243	0.012117	-1.32913	0.07997	0.107795	13.20511	0.024534	0.002211
3	7.747165	7.895129	-0.147964	-1.60234	-0.97650	0.093674	9.73552	-0.239493	0.159123
4	8.097426	7.926081	0.171346	-1.53475	1.13081	0.063869	4.00927	0.208366	0.055995
5	8.331586	8.182181	0.149405	-0.97551	0.98601	0.061995	3.72151	0.179442	0.039126
6	8.322394	8.300935	0.021460	-0.71619	0.14162	0.068306	4.72446	0.026933	0.001070
7	8.388678	8.280548	0.108130	-0.76071	0.71361	0.055594	2.80372	0.124950	0.015256
8	8.427487	8.427363	0.000124	-0.44012	0.00082	0.060436	3.48883	0.000147	0.000000
9	8.475746	8.395995	0.079751	-0.50861	0.52632	0.057053	3.00404	0.092925	0.008886
10	8.504513	8.645535	-0.141022	0.03630	-0.93068	0.070188	5.04225	-0.179546	0.050209
11	8.454253	8.436184	0.018069	-0.42085	0.11925	0.039792	0.96546	0.019408	0.000189
12	8.478453	8.516011	-0.037559	-0.24654	-0.24787	0.046056	1.62122	-0.041382	0.001148
13	7.881560	8.274945	-0.393385	-0.77295	-2.59618	0.059368	3.33269	-0.464724	0.240658
14	7.996654	8.348469	-0.351815	-0.61239	-2.32183	0.044533	1.45297	-0.385076	0.092973
15	8.402905	8.274094	0.128811	-0.77481	0.85010	0.073797	5.67599	0.168865	0.049099
16	8.566745	8.579424	-0.012679	-0.10807	-0.08368	0.059653	3.37409	-0.015005	0.000253
17	8.524566	8.571085	-0.046519	-0.12628	-0.30701	0.044466	1.44577	-0.050903	0.001620
18	8.842894	8.659674	0.183220	0.06717	1.20917	0.045064	1.51101	0.200998	0.025939
19	8.931419	8.829753	0.101667	0.43857	0.67096	0.043911	1.38589	0.110987	0.007509
20	8.995413	8.858474	0.136939	0.50129	0.90374	0.040314	1.01651	0.147371	0.011160
21	9.098962	9.101893	-0.002932	1.03283	-0.01935	0.066670	4.45505	-0.003635	0.000019
22	8.897956	8.932784	-0.034828	0.66356	-0.22985	0.081282	7.09166	-0.048899	0.004995
23	9.062884	9.099126	-0.036242	1.02679	-0.23918	0.106851	12.95784	-0.072088	0.018758
24	9.210140	9.092921	0.117219	1.01324	0.77360	0.050377	2.12938	0.131786	0.013935
25	9.259035	9.120444	0.138591	1.07334	0.91464	0.054867	2.70572	0.159504	0.024215
26	9.124891	9.143014	-0.018123	1.12263	-0.11960	0.067723	4.62772	-0.022647	0.000744
27	9.196850	9.214508	-0.017658	1.27875	-0.11654	0.076377	6.14859	-0.023673	0.001034
28	9.463974	9.524893	-0.060919	1.95652	-0.40204	0.088974	8.68871	-0.092977	0.021636
29	9.392662	9.503228	-0.110566	1.90922	-0.72969	0.071794	5.32038	-0.142573	0.033126
Minimum	7.747165	7.895129	-0.393385	-1.60234	-2.59618	0.039792	0.96546	-0.464724	0.000000
Maximum	9.463974	9.524893	0.183220	1.95652	1.20917	0.112097	14.35866	0.208366	0.240658
Mean	8.628912	8.628912	0.000000	0.00000	0.00000	0.065961	4.82759	-0.002973	0.031751
Median	8.504513	8.571085	0.000124	-0.12628	0.00082	0.061995	3.72151	0.000147	0.013935

5.2.2.1.5 Correlation and multicollinearity

Correlation coefficient is a statistic that gives a measure of how closely two variables are related, and it can vary from -1 indicating perfect negative correlation to +1 indicating perfect positive correlation. Multicollinearity refers to excessive correlation of the predictor or independent variables.

From Tables 5.6 and 5.9, it can be observed that (after data transformation), there is a strong positive correlation between the dependent and independent variables ranging between 70% and 94%.

Based on Table 5.9, there is a high potential of multicollinearity amongst variables as the tolerance level (Table 5.8) is less than 0.2. When the tolerance level is close to 0 there is high multicollinearity of that variable with other independent variables and

beta coefficients will be unstable. It is common that the variable with the most variance inflation factor (VIF) which is calculated by taking the reciprocal of tolerance level will be dropped. This is worked out by applying the stepwise forward regression to find the most capable independent variables that provides a model that is adequate to predict proper forecast for this sector.

Table 5.8: Tolerance test

Variable	Toleran.	R-square	Partial Cor.	Semipart Cor.
GDP	0.013684	0.986316	-0.049202	-0.014150
WSSC	0.009446	0.990554	-0.222918	-0.065685
SASSC	0.182314	0.817686	-0.100339	-0.028968
WFeCrS	0.045520	0.954480	0.609668	0.220935
Usd	0.122058	0.877942	0.507907	0.169367

Table 5.9: Multicollinearity test

	GDP	WSSC	SASSC	WFeCrS	Usd	EC
GDP	1.000000	-0.983464	0.678153	-0.954585	-0.897040	-0.925672
WSSC	-0.983464	1.000000	-0.771496	0.964264	0.891041	0.925943
SASSC	0.678153	-0.771496	1.000000	-0.696578	-0.702653	-0.690423
WFeCrS	-0.954585	0.964264	-0.696578	1.000000	0.796504	0.932695
Usd	-0.897040	0.891041	-0.702653	0.796504	1.000000	0.863558
EC	-0.925672	0.925943	-0.690423	0.932695	0.863558	1.000000

5.2.2.1.6 Autocorrelation

The test for autocorrelation function is based on the D-W Statistic. In simple terms, we test the null hypothesis that there is no autocorrelation against the alternative H_i that there is autocorrelation. H_0 is rejected if the test statistic approaches a value of 2, implying that there is no autocorrelation. If the error terms are highly positively correlated, the statistic would be less than 1 and could get near zero, which would indicate autocorrelation, and hence we do not reject H_0. If the error terms are highly negatively correlated, the statistic would be greater than 3 and could get near the upper limit of 4.

Based on Table 5.10, the D-W Statistic is equal to 1.44. Thus H_0 is not rejected. We conclude that there is no autocorrelation between the residual values of energy consumption and predictor variables since the D-W Statistic is greater than 1.

However, D-W value should be closer to 2 as much as possible so that there is no autocorrelation between the variables.

Table 5.10: Autocorrelation test

	Durbin-Watson d	Serial Corr.
Estimate	1.444624	0.270425

5.2.2.1.7 Model Adequacy

The analysis of the dependent variable with each predictor variable is shown in Tables 5.11 to 5.16. It can be observed from these tables that all the predictor variables show a good adjusted R^2 of between 85% and 94%, except for SASSC (with an adjusted R^2 4 6%). This is indicative that the model fits the data well.

Table 5.11: Regression Results between Dependent

EC and Independent GDP

Multiple Regression Results						
Dependent: EC Multiple R = .92567179 F = 161.6374						
R²= .85686825 df = 1,27						
No. of cases: 29 adjusted R²= .85156708 p = .000000						
Standard error of estimate: .184195067						
Intercept: 11.153360126 Std.Error: .2014862 t(27) = 55.355 p = 0.0000						
GDP beta=-.93						
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(27)	p-level
Intercept			11	0.2	55.3555	0.000000
GDP	-0.925672	0.072809	-2095513	164823.6	-12.7137	0.000000
Effect	Sums of Squares	df	Mean Squares	F	p-level	
Regress.	5.484005	1	5.484005	161.6374	0.000000	
Residual	0.916051	27	0.033928			
Total	6.400056					

Table 5.12: Regression results between Dependent EC and Independent WSSC

Multiple Regression Results					
Dependent: EC　　　Multiple R = .92594280　F = 162.3011					
R²= .85737006　df = 1,27					
No. of cases: 29　　　adjusted R²= .85208747　p = .000000					
Standard error of estimate: .183871899					
Intercept: -2.387371054 Std.Error: .8653917 t(27) = -2.759 p = .0103					
WSSC beta=.926					

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(27)	p-level
Intercept			-2.38737	0.865392	-2.75872	0.010289
WSSC	0.925943	0.072681	1.14506	0.089881	12.73974	0.000000

Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	5.487217	1	5.487217	162.3011	0.000000
Residual	0.912840	27	0.033809		
Total	6.400056				

Table 5.13: Regression Results between Dependent EC and Independent SASSC

Multiple Regression Results					
Dependent: EC　　　Multiple R = .69042338　F = 24.59411					
R²= .47668445　df = 1,27					
No. of cases: 29　　　adjusted R²= .45730239　p = .000034					
Standard error of estimate: .352202011					
Intercept: 8.904289973 Std.Error: .0857953 t(27) = 103.79 p = 0.0000					
SASSC beta=-.69					

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(27)	p-level
Intercept			8.9043	0.085795	103.7853	0.000000
SASSC	-0.690423	0.139219	-32.4889	6.551182	-4.9592	0.000034

Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	3.050807	1	3.050807	24.59411	0.000034
Residual	3.349249	27	0.124046		
Total	6.400056				

Table 5.14: Regression results between dependent EC and Independent WFeCrS

Multiple Regression Results					
Dependent: EC　　　Multiple R = .93269470　F = 180.5636					
R²= .86991940　df = 1,27					
No. of cases: 29　　　adjusted R²= .86510160　p = .000000					
Standard error of estimate: .175596672					
Intercept: -3.687748658 Std.Error: .9171756 t(27) = -4.021 p = .0004					
WFeCrS beta=.933					

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(27)	p-level
Intercept			-3.68775	0.917176	-4.02077	0.000419
WFeCrS	0.932695	0.069410	1.48342	0.110395	13.43740	0.000000

Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	5.567533	1	5.567533	180.5636	0.000000
Residual	0.832523	27	0.030834		
Total	6.400056				

Table 5.15: Regression results between Dependent EC and Independent USd

Multiple Regression Results					

Dependent: EC   Multiple R = .86355805   F = 79.18738
               R^2= .74573250   df = 1,27
No. of cases: 29   adjusted R^2= .73631518   p = .000000
       Standard error of estimate: .245502228
Intercept: 7.963151645 Std.Error: .0876107 t(27) = 90.892 p = 0.0000

      Usd beta=.864

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(27)	p-level
Intercept			7.963152	0.087611	90.89242	0.000000
Usd	0.863558	0.097043	0.155251	0.017446	8.89873	0.000000

Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	4.772730	1	4.772730	79.18738	0.000000
Residual	1.627326	27	0.060271		
Total	6.400056				

Table 5.16: Regression results between Dependent variable EC and overall independent variables

Overall						
Multiple Regression Results						

Dependent: EC   Multiple R = .95785641   F = 51.15007
           R^2= .91748890   df = 5,23
No. of cases: 29   adjusted R^2= .89955170   p = .000000
         Standard error of estimate: .151524956
Intercept: 2.978279345 Std.Error: 7.917545 t(23) = .37616 p = .7102

     GDP beta=-.12   WSSC beta=-.68   SASSC beta=-.07
    WFeCrS beta=1.04   Usd beta=.485

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(23)	p-level
Intercept			3	8	0.37616	0.710244
GDP	-0.120964	0.512015	-273836	1159087	-0.23625	0.815329
WSSC	-0.675851	0.616274	-1	1	-1.09667	0.284135
SASSC	-0.067844	0.140276	-3	7	-0.48365	0.633207
WFeCrS	1.035533	0.280732	2	0	3.68869	0.001215
Usd	0.484782	0.171439	0	0	2.82773	0.009541

Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	5.871981	5	1.174396	51.15007	0.000000
Residual	0.528076	23	0.022960		
Total	6.400056				

5.2.2.1.8 Goodness of fit

To check for model adequacy, stepwise regression is applied as explained in the literature review chapter.

Forward selection method is used for selecting variables to include in the model. At stage one of this method, the independent variable that correlates best with the dependent variable is included in the equation. In the second stage, the remaining independent variable with the highest partial correlation to the dependent variable, is entered. This process is repeated, at each stage partial-ling for previously entered independents.

Tables 5.11 to Table 5.16 give the analysis of independent variable with the each predictor variable electricity consumption (EC). It can be observed from these tables that all the predictor variables show a good adjusted R^2 of between 85% and 94% with the exception of SASSC which shows a poor R^2 value of 46%.

Table 5.16 shows the model with all the independent variables included. The model in this case is significant (F=51.2) with a remarkable R^2 of 92%.

The aim of stepwise regression is to eliminate all the variables that when added, they do not impact on R^2 significantly.

With stepwise regression, analysis is done by steps until the procedure stops with no dependent variable with a better R^2 than the previous results. In this section there are three steps, and we next provide the analysis of each step.

Step 0

This step indicated in Table 5.17 has no variables added to the dependent variable (energy consumption). Variability between the data is at 0.47 and not influenced by any independent variable.

Table 5.17: Regression results at Step 0

```
   Multiple Regression Results (Step   0)

 Dependent: EC                 Multiple R = 0.00000000    F = 0.000000
                                    R²= 0.00000000   df =   0,28
 No. of cases: 29              adjusted R²= 0.00000000   p = -0.00000
              Standard error of estimate:  .478093542


      Step 0:  No variables in the regression equation
```

Step 1

First predictor variable added is the WFeCrS (Table 5.18). The R^2 and its adjusted value are at 86% which shows that there is a strong relationship between the dependent variable and this predictor variable. The beta coefficient of 0.933 indicates that this is a good estimate. The model is also significant (F =180.56), implying that the model fits the data well.

Table 5.18: Regression results at Step 1

```
Multiple Regression Results (Step 1)

 Dependent: EC                 Multiple R =  .93269470    F = 180.5636
                                    R²=  .86991940   df =   1,27
 No. of cases: 29              adjusted R²=  .86510160   p =  .000000
              Standard error of estimate:  .175596672
 Intercept: -3.687748658 Std.Error: .9171756  t (27) = -4.021  p = .0004

WFeCrS beta=.933
```

Step 2

The second predictor variable added is the USd (Rand dollar change). Addition of this variable improves the R^2 and its adjusted value to 90%. The model is still significant at F=131 which implies that the model with these two predictor variables fits the data better compared to the previous model with one variable. The beta coefficients are 0.67 and 0.33 for WFeCrS and Usd respectively as shown in Table 5.19.

Table 5.19: Regression results at Step 2

```
   Multiple Regression Results (Step 2)

 Dependent: EC                 Multiple R =  .95380564    F = 131.0367
                                    R²=  .90974520   df =   2,26
 No. of cases: 29              adjusted R²=  .90280253   p =  .000000
              Standard error of estimate:  .149052878
 Intercept:  -.470592126  Std.Error: 1.228112  t(26) = -.3832  p = .7047

WFeCrS beta=.670          Usd beta=.330
```

Step 3 (Final Solution)

The last predictor variable entered is the WSSC. This also improves the R^2 and its adjusted value to 91%. The overall model is still significant at F=91.55 while the beta coefficient has improved to 1.01 for WFeCrS, the USd beta is 0.49 and WSSC beta is now at -0.49 (Table 5.20). Between the residual values of the dependent variables and predictor variables there is no autocorrelation (Table 5.22).

Thus step three presents the simplest possible set of predictor variables that is included in the model.

From this final step, it can be observed that two variables were dropped namely, the Gross Domestic Product (GDP) and the South Africa Stainless Steel Consumption (SASSC) due to high multicollinearity impact and normality assumption not met.

Table 5.20: Regression results at Step 3

```
  Multiple Regression Results (step   3, final solution)
  no other F to enter exceeds specified limit
 Dependent: EC              Multiple R =  .95737474    F = 91.54648
                                     R²=  .91656639    df =   3,25
 No. of cases: 29          adjusted R²=  .90655435     p =  .000000
            Standard error of estimate:  .146147844
 Intercept:   .718273807  Std.Error: 1.463408  t(25) = .49082  p =  .6278

      WFeCrS beta=1.01            Usd beta=.494            WSSC beta=-.49
```

Table 5.21: Regression results summary of the final results

	Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variables included
WFeCrS	1	0.932695	0.869919	0.869919	180.5636	0.000000	1
Usd	2	0.953806	0.909745	0.039826	11.4728	0.002257	2
WSSC	3	0.957375	0.916566	0.006821	2.0439	0.165196	3

Table 5.22: Autocorrelation between variables residuals

	Durbin-Watson d	Serial Corr.
Estimate	1.384488	0.302621

From Table 5.23, the model is significant with F=91.5. Therefore the three independent variables namely, WFeCrS, USd and WSSC are the best variables to predict electricity consumption for FeCr sector for medium term.

Table 5.23: Overall goodness of fit

Effect	Sums of Squares	df	Mean Squares	F	p-level	
Regress.	5.866076	3	1.955359	91.54648	0.000000	
Residual	0.533980	25	0.021359			
Total	6.400056					

Table 5.24 indicates that WSSC is not significant when applying multiple regression model to the data. Therefore WFECRS and USd are the best predictor variables for predicting electricity consumption for medium term using multiple linear regression for the FeCr. The R^2 of 92% also shows a strong relationship between the predictor variables with the independent variables.

The regression equation provided by these results will be used to predict future electricity consumption for the FeCr sector and the results will be evaluated in the results section with other model estimations.

Table 5.24: Parameter estimates for Multiple Regression Model

Forecast Model for EnergyConsumptionActual
Regression(3 regressors, 0 lagged errors)

Term	Coefficient	Std. Error	t-Statistic	Significance
WFECRS	1.035102	0.439490	2.355236	0.973664
USD	456.793858	143.606689	3.180868	0.996222
WSSC	0.003105	0.141282	0.021974	0.017364 <-

Marked regressors are insignificant.

Within-Sample Statistics	
Sample size 29	Number of parameters 3
Mean 6234	Standard deviation 2963
R-square 0.9163	Adjusted R-square 0.9099
Durbin-Watson 0.9613	Ljung-Box(18)=28.17 P=0.9405
Forecast error 889.5	BIC 1002
MAPE 0.1295	RMSE 842.2
MAD 625.7	

5.2.2.2 Exponential Smoothing

Several models for exponential smoothing were tried in order to select the one(s) that best fit the data. The analysis is provided in the sections below.

5.2.2.2.1 Data set and descriptive statistics

Time series analysis is applied in order to arrive at the model that best fit the data. The data set consists of 80 (Table 5.25) data points. Figure 5.20 illustrates monthly consumption data used to forecast consumption for medium term. The sector consumed an average electricity of 868.09 GWh per month. For a period of seven years, the FeCr sector consumed maximum electricity of 1160.82 GWh and a minimum of 217 GWh, the latter being attributable to the global economic meltdown that affected the FeCr sector between October 2008 and August 2009. The 95% confidence interval is assumed for upper and lower limits. This data set is not normal and is transformed using logarithms.

Figure 5.20 : Monthly electricity consumption for the FeCr sector
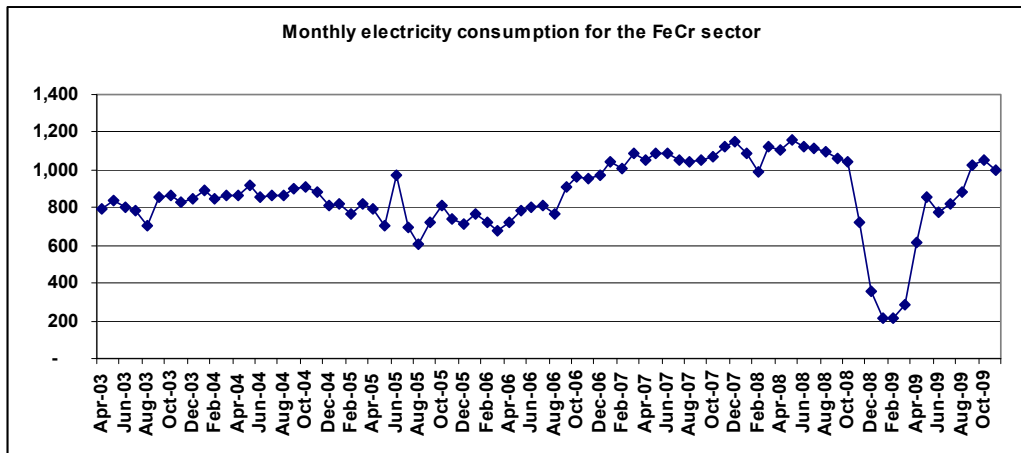
Table 5.25: Descriptive statistics of the monthly FeCr electricity consumption data

Discriptive Statistics	
Valid N	80
Mean	868.09
-95% Confidence	824.19
95%Confidence	911.99
Minimum	217.69
Maximum	1160.82
Variance	38920.44
Std.Dev.	197.28
Skewness	-1.20
Kurtosis	2.45

5.2.2.2.2 Model checks: Assumptions and data analysis

Distribution of the monthly electricity consumption data after transformation follows a normal distribution as shown in Figure 5.21.

Figure 5.21: Distribution of the monthly electricity consumption
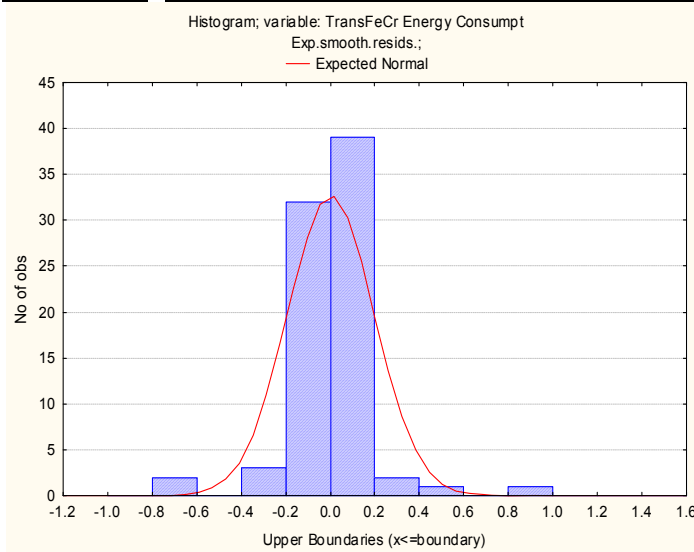
The residual plot in Figure 5.22 shows a constant mean and variance over time. This indicates that the data is stationary after transformation and hence confirms normality.

Figure 5.22: <u>Residual plot of the monthly electricity consumption</u>
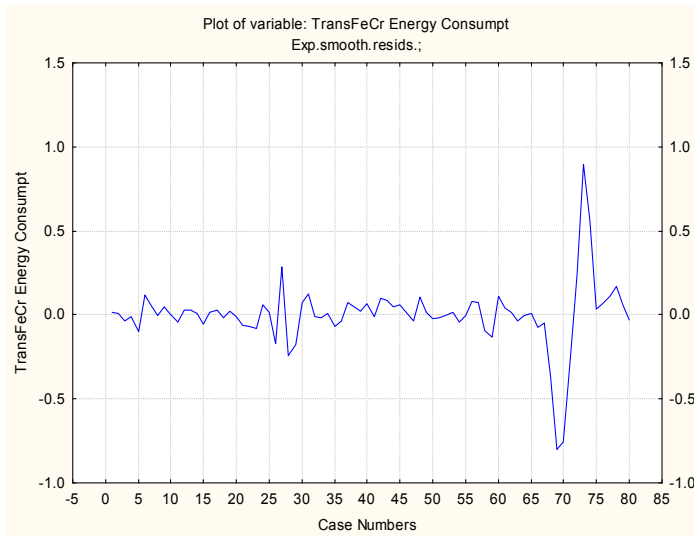
All exponential smoothing models that were considered here are shown in Table 5.26. Exponential smoothing with multiplicative seasonality and trend is the best model compared to other exponential smoothing models. It has the least MAPE (2.5%). Model adequacy and goodness of fit is examined in Table 5.23.

Table 5.26: Parameter estimates for all fitted exponential smoothing models

Exponential Smoothing Forecast Model for FeCrEnergyConsumpt Exponential smoothing: No trend, Additive seasonality	Simple exponential smoothing: No trend, No seasonality
Component **Smoothing Weight** **Final Value**	**Component** **Smoothing Weight** **Final Value**
Level 0.93562 1003.8 Seasonal 0.99948	Level 1.00000 1003.0
Within-Sample Statistics	**Within-Sample Statistics**
Sample size 80 Number of parameters 2 Mean 868.1 Standard deviation 197.3 R-square 0.7557 Adjusted R-square 0.7525 Durbin-Watson 1.495 Ljung-Box(18)=18.05 P=0.5473 Forecast error 98.14 BIC 102.4 MAPE 0.03041 RMSE 96.91 MAD 60.59	Sample size 80 Number of parameters 1 Mean 868.1 Standard deviation 197.3 R-square 0.7383 Adjusted R-square 0.7383 Durbin-Watson 1.612 Ljung-Box(18)=12.85 P=0.1995 Forecast error 100.9 BIC 103.1 MAPE 0.03416 RMSE 100.3 MAD 65.

Holt exponential smoothing: Linear trend, No seasonality			Additive Winters: Linear trend, Additive seasonality		
	Smoothing	Final		Smoothing	Final
Component	Weight	Value	Component	Weight	Value
Level	0.99960	1003.1	Level	0.99995	1003.0
Trend	0.00016	0.063617	Trend	0.00000	0.00017852
			Seasonal	0.00861	
Within-Sample Statistics			Within-Sample Statistics		
Sample size 80	Number of parameters 2		Sample size 80	Number of parameters 3	
Mean 868.1	Standard deviation 197.3		Mean 868.1	Standard deviation 197.3	
R-square 0.7383	Adjusted R-square 0.7349		R-square 0.7383	Adjusted R-square 0.7315	
Durbin-Watson 1.611	Ljung-Box(18)=13.34 P=0.2288		Durbin-Watson 1.612	Ljung-Box(18)=13.52 P=0.2405	
Forecast error 101.6	BIC 105.9		Forecast error 102.2	BIC 108.9	
MAPE 0.03418	RMSE 100.3		MAPE 0.03416	RMSE 100.3	
MAD 65.36			MAD 65.35		

Exp. smoothing: Multipl. season (4) S0=6.692 T0=1.000 (Yearly Data FeCr 2003_2008 monthlynew.sta) Expon.trend,mult.season; Alpha= .700 Delta=.200 Gamma=.008 TransFeCr Energy Consumpt	
Summary of Error	Error
Mean error	0.002
Mean absolute error	0.100
Sums of squares	3.020
Mean square	0.038
Mean abs. perc. error	0.025

5.2.2.2.3 Model Adequacy of the Fifth model

Most of the residuals are within the upper and lower control limits, based on Figure 5.23 and Figure 5.24. Therefore, it can be concluded that no autocorrelation exists within the residuals of the data series. This model fits the data well compared to all the others that were tested in this study.

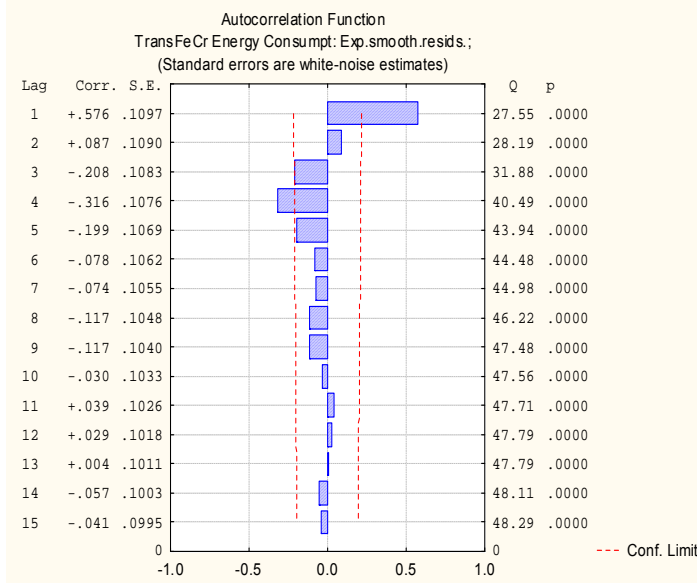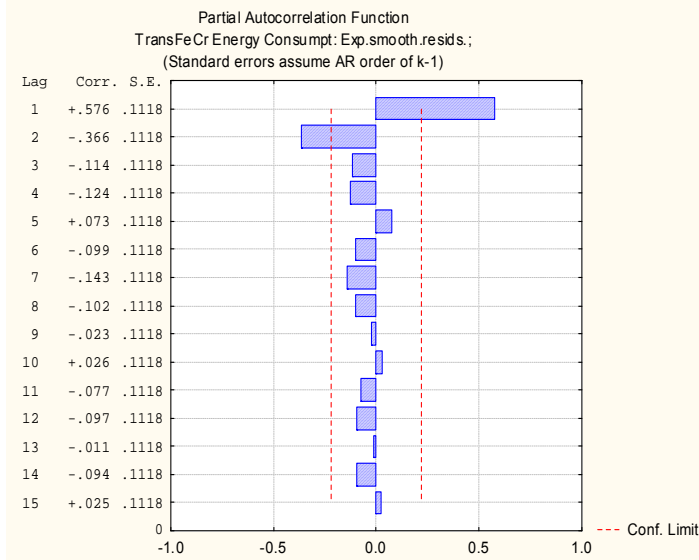**Figure 5.23:** **ACF of residuals of the monthly electricity consumption**

```
                      Autocorrelation Function
                TransFeCr Energy Consumpt: Exp.smooth.resids.;
                   (Standard errors are white-noise estimates)
  Lag   Corr.  S.E.                                          Q      p
   1   +.576  .1097                                        27.55  .0000
   2   +.087  .1090                                        28.19  .0000
   3   -.208  .1083                                        31.88  .0000
   4   -.316  .1076                                        40.49  .0000
   5   -.199  .1069                                        43.94  .0000
   6   -.078  .1062                                        44.48  .0000
   7   -.074  .1055                                        44.98  .0000
   8   -.117  .1048                                        46.22  .0000
   9   -.117  .1040                                        47.48  .0000
  10   -.030  .1033                                        47.56  .0000
  11   +.039  .1026                                        47.71  .0000
  12   +.029  .1018                                        47.79  .0000
  13   +.004  .1011                                        47.79  .0000
  14   -.057  .1003                                        48.11  .0000
  15   -.041  .0995                                        48.29  .0000
         0                                                  0
       -1.0    -0.5     0.0     0.5     1.0      --- Conf. Limit
```

Figure 5.24: **PACF of residuals for monthly electricity consumption**

```
                   Partial Autocorrelation Function
                TransFeCr Energy Consumpt: Exp.smooth.resids.;
                   (Standard errors assume AR order of k-1)
  Lag   Corr.  S.E.
   1   +.576  .1118
   2   -.366  .1118
   3   -.114  .1118
   4   -.124  .1118
   5   +.073  .1118
   6   -.099  .1118
   7   -.143  .1118
   8   -.102  .1118
   9   -.023  .1118
  10   +.026  .1118
  11   -.077  .1118
  12   -.097  .1118
  13   -.011  .1118
  14   -.094  .1118
  15   +.025  .1118
         0                                                  --- Conf. Limit
       -1.0    -0.5     0.0     0.5     1.0
```

5.2.2.2.4 Goodness of fit

Multiplicative exponential smoothing model with trend and seasonality is observed to fit this data well. Therefore, this method can be used to forecast electricity consumption in the medium term.

5.2.2.3 Box Jenkins models

5.2.2.3.1 Data set

The data set applicable here is the monthly data from 2003 to 2009 (Figure 5.20) and the analyses of the data follows that of exponential smoothing in Section 5.2.2.2.

Model 1

Table 5.25: Parameter estimates for ARIMA(2,0,0) model

```
        Variable: Consumpt
 Transformations:
          Model: (2,0,0)
 No. of obs.:80  Initial SS=6336E4  Final SS=1411E3 (2.227%)  MS=18094.
  Parameters (p/Ps-Autoregressive, q/Qs-Moving aver.); highlight: p<.05
              p(1)   p(2)
   Estimate: 1.1324 -.1387
   Std.Err.: .11295 .11396
```
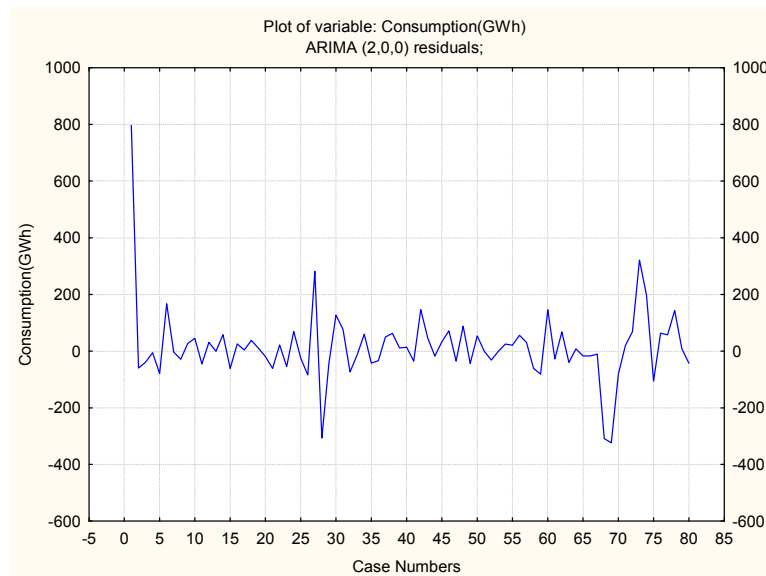
Figure 5.25: Residual plot of ARIMA(2,0,0) model

Plot of variable: Consumption(GWh)
ARIMA (2,0,0) residuals;

Figure 5.26: ACF for ARIMA(2,0,0) model

```
                    Autocorrelation Function
               Consumption(GWh): ARIMA (2,0,0) residuals;
                 (Standard errors are white-noise estimates)
 Lag   Corr.  S.E.                                   Q      p
  1   -.000  .1097                                  .00  .9984
  2   -.051  .1090                                  .22  .8970
  3   +.004  .1083                                  .22  .9745
  4   -.099  .1076                                 1.06  .9008
  5   +.014  .1069                                 1.08  .9561
  6   -.036  .1062                                 1.19  .9771
  7   -.015  .1055                                 1.22  .9906
  8   -.029  .1048                                 1.29  .9957
  9   -.063  .1040                                 1.65  .9958
 10   -.049  .1033                                 1.88  .9972
 11   +.049  .1026                                 2.10  .9981
 12   +.020  .1018                                 2.14  .9992
 13   +.040  .1011                                 2.30  .9995
 14   -.088  .1003                                 3.07  .9989
 15   -.016  .0995                                 3.10  .9995
         0                              0       ····· Conf. Limit
        -1.0    -0.5    0.0    0.5    1.0
```

Figure 5.27: PACF ARIMA(2,0,0) model

```
                Partial Autocorrelation Function
            Consumption(GWh): ARIMA (2,0,0) residuals;
              (Standard errors assume AR order of k-1)
 Lag   Corr.  S.E.
  1   -.000  .1118
  2   -.051  .1118
  3   +.004  .1118
  4   -.101  .1118
  5   +.015  .1118
  6   -.048  .1118
  7   -.013  .1118
  8   -.045  .1118
  9   -.062  .1118
 10   -.063  .1118
 11   +.040  .1118
 12   +.005  .1118
 13   +.033  .1118
 14   -.103  .1118
 15   -.008  .1118
         0                              0      --- Conf. Limit
        -1.0    -0.5    0.0    0.5    1.0
```
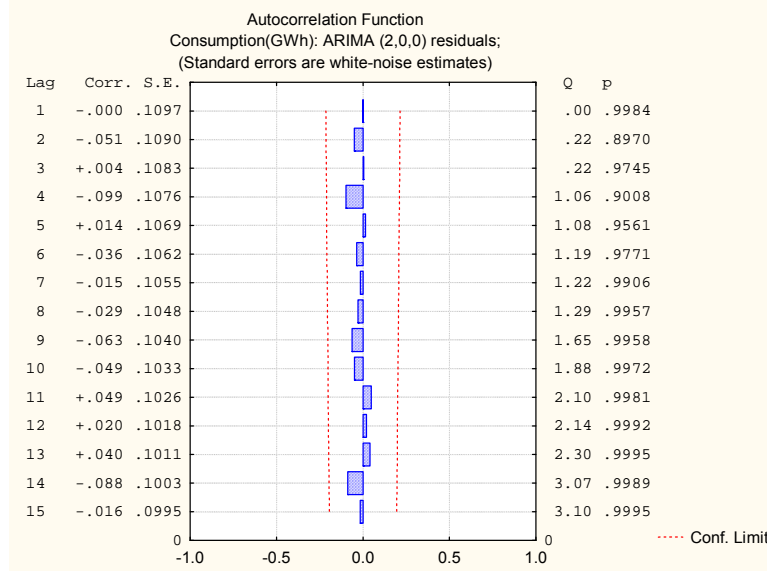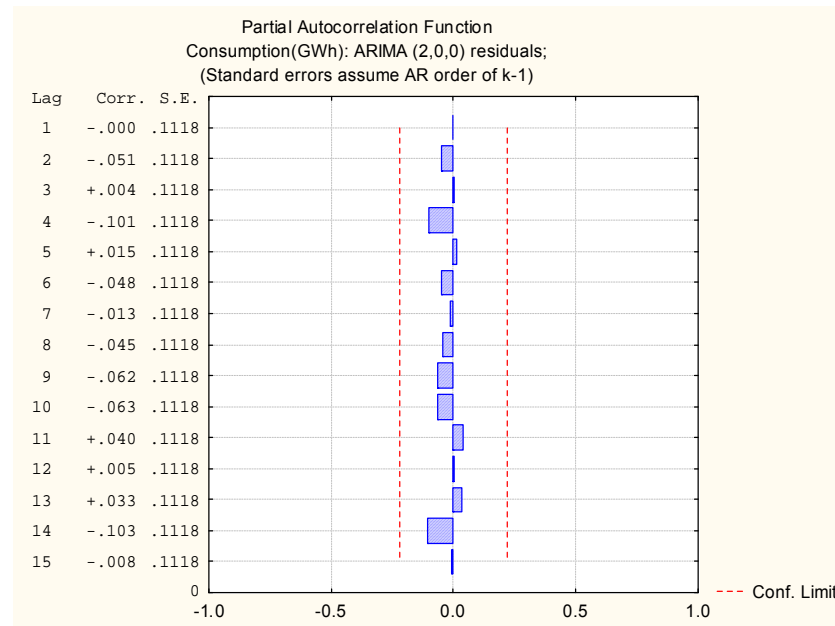
It can be noted that the ARIMA(2,0,0) model seems to be an adequate fit based on autocorrelation function (Figure 5.26) and partial autocorrelation function (Figure 5.27). Based on D-W statistic in Table 5.29, it can be concluded that there is no autocorrelation between the residual values. Parameter estimates are only given by the Autoregressive of order 2 represented as AR(2) with no differencing and zero moving average estimates.

Despite the fact that there is no autocorrelation and the model has a good fit, the setback is with the MAPE of 9% which is greater than α=5%. Thus, it can be concluded based on Table 5.29 that this model will not be an appropriate one for forecasting medium term electricity consumption for the FeCr sector.

Model 2

Table 5.26: Parameter estimates for ARIMA(2,1,0)(1,0,0)

```
        Variable: Consumpt
 Transformations: D(1)
          Model: (2,1,0)(1,0,0)  Seasonal lag:  12
 No. of obs.:79  Initial SS=8045E2  Final SS=7701E2 (95.72%)  MS=10133.
  Parameters (p/Ps-Autoregressive, q/Qs-Moving aver.); highlight: p<.05
              p(1)   p(2)  Ps(1)
   Estimate: .19974 -.0467 .06460
   Std.Err.: .11608 .11776 .14711
```

Figure 5. 28: Residual plot for ARIMA(2,1,0)(1,0,0)
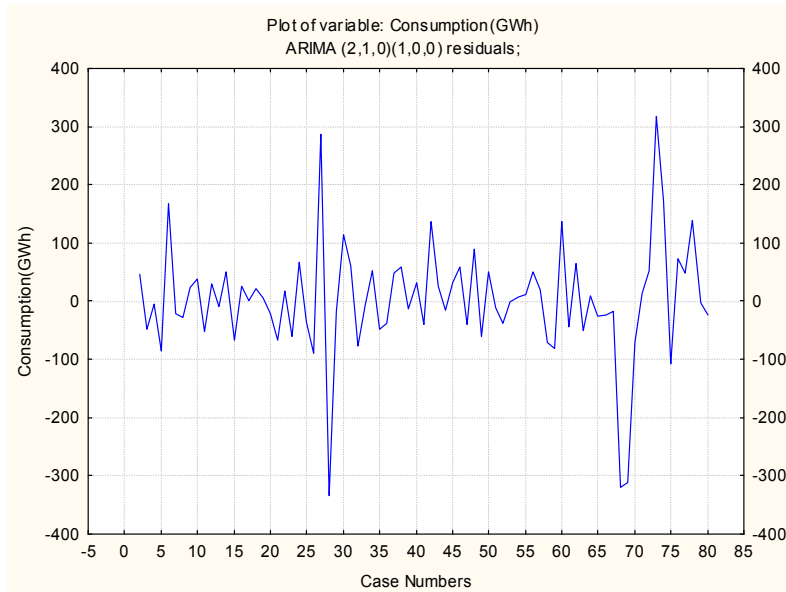
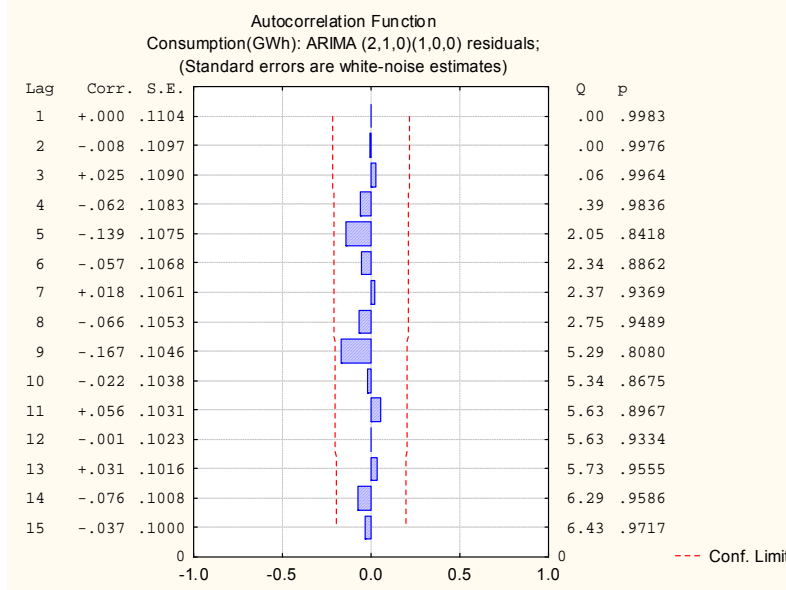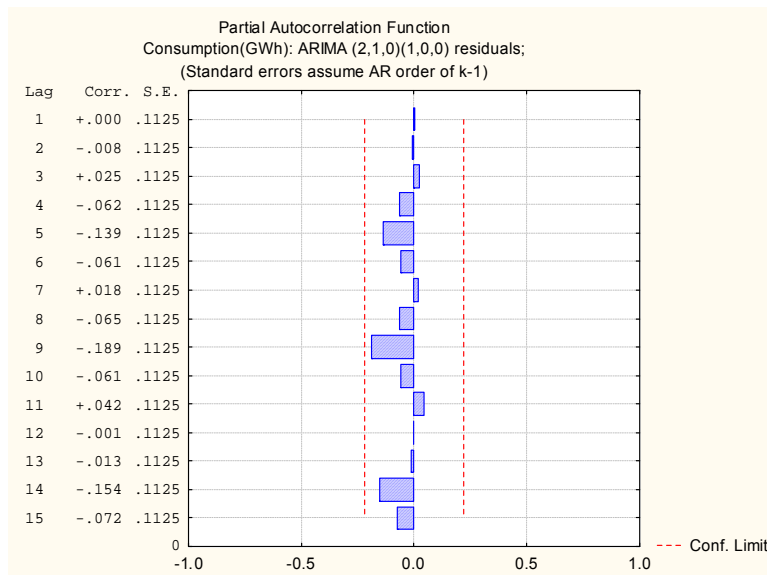**Figure 5.29: The ACF of ARIMA(2,1,0)(1,0,0) model**

Autocorrelation Function
Consumption(GWh): ARIMA (2,1,0)(1,0,0) residuals;
(Standard errors are white-noise estimates)

Lag	Corr.	S.E.	Q	p
1	+.000	.1104	.00	.9983
2	-.008	.1097	.00	.9976
3	+.025	.1090	.06	.9964
4	-.062	.1083	.39	.9836
5	-.139	.1075	2.05	.8418
6	-.057	.1068	2.34	.8862
7	+.018	.1061	2.37	.9369
8	-.066	.1053	2.75	.9489
9	-.167	.1046	5.29	.8080
10	-.022	.1038	5.34	.8675
11	+.056	.1031	5.63	.8967
12	-.001	.1023	5.63	.9334
13	+.031	.1016	5.73	.9555
14	-.076	.1008	6.29	.9586
15	-.037	.1000	6.43	.9717

--- Conf. Limit

Figure 5.30: The PACF for ARIMA(2,1,0)(1,0,0) model

Partial Autocorrelation Function
Consumption(GWh): ARIMA (2,1,0)(1,0,0) residuals;
(Standard errors assume AR order of k-1)

Lag	Corr.	S.E.
1	+.000	.1125
2	-.008	.1125
3	+.025	.1125
4	-.062	.1125
5	-.139	.1125
6	-.061	.1125
7	+.018	.1125
8	-.065	.1125
9	-.189	.1125
10	-.061	.1125
11	+.042	.1125
12	-.001	.1125
13	-.013	.1125
14	-.154	.1125
15	-.072	.1125

--- Conf. Limit

The ARIMA(2,1,0)(1,0,0) model seems to fit the data adequately. However, there is also no autocorrelation between the residual values. All the parameter estimates are not significant (Table 5.27). The MAPE of 8% which is greater $\alpha=5\%$ simply means that the error margin is too wide and forecasts will deviate significantly. Therefore, this model will not be appropriate for forecasting medium term electricity consumption for the FeCr sector.

Table 5.27: Overall goodness of fit for models fitted

Forecast Model for FeCrEnergyConsumpt ARIMA(2,0,0)				Forecast Model for FeCrEnergyConsumpt ARIMA(2,1,0)*(1,0,0)			
Term	Coefficient	Std. Error	t-Statistic	Significance			

Term	Coefficient	Std. Error	t-Statistic	Significance	Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	1.1106	0.1072	10.3568	1.0000	a[1]	0.2007	0.1125	1.7846	0.9217 <-
a[2]	-0.2768	0.1078	-2.5680	0.9879	a[2]	-0.0481	0.1126	-0.4270	0.3294 <-
_CONST	144.3305				A[12]	0.0672	0.1345	0.4999	0.3814 <-

Within-Sample Statistics		Within-Sample Statistics	
Sample size 80	Number of parameters 2	Sample size 80	Number of parameters 3
Mean 868.1	Standard deviation 197.3	Mean 868.1	Standard deviation 197.3
R-square 0.774	Adjusted R-square 0.7711	R-square 0.7497	Adjusted R-square 0.7432
Durbin-Watson 2.025	Ljung-Box(18)=6.808 P=0.008351	Durbin-Watson 2	Ljung-Box(18)=8.769 P=0.03514
Forecast error 94.39	BIC 98.45	Forecast error 99.97	BIC 106.5
MAPE 0.09154	RMSE 93.2	MAPE 0.08848	RMSE 98.08
MAD 63.91		MAD 64.06	

Model 1 (ARIMA(2,0,0)) and Model 2 (ARIMA(2,1,0)(1,0,0)), both have MAPE that is greater than 5% level of significance. Model 2 parameter estimates are all not significant. Thus we can conclude that Box Jenkins method is not a good method to use for medium term when forecasting electricity consumption for the FeCr sector.

5.2.2.2.3 Results and discussion

The results on how efficient the models that were found to have the best fit to the data discussion thereof are discussed next. These outcomes are only for those models that were found to have the best fit on each of the methods that were examined. After fitting the models to the data set, the outcomes will be compared so that the appropriate method to forecast the FeCr sector electricity consumption for medium term or tactical use is selected. The methods fitted include Multiple Linear Regression, Exponential Smoothing and Box Jenkins method. Only Multiple Linear Regression and Exponential Smoothing results are shown since these methods had the potential to forecast electricity consumption for medium term in the FeCr sector.

Table 5.28 and Table 5.29 demonstrate results of the multiple regression and exponential smoothing goodness of fit. It is observed that exponential smoothing model fits the data better with a MAPE of 2.5%, while multiple regression model is worse off with a MAPE of 12.95%. The plot of actual values against estimated figures for exponential smoothing and regression models (Figure 5.31, 5.33 and Figure 5.34) further confirms that exponential smoothing model fits the data better compared to multiple regression model.

Table 5.28: Overall results, Stepwise Regression in Multiple Regression Method

Forecast Model for EnergyConsumptionActual Regression(3 regressors, 0 lagged errors)				
Term	Coefficient	Std. Error	t-Statistic	Significance
WFECRS	1.035102	0.439490	2.355236	0.973664
USD	456.793858	143.606689	3.180868	0.996222
WSSC	0.003105	0.141282	0.021974	0.017364 <-

Marked regressors are insignificant.

Within-Sample Statistics

Sample size 29	Number of parameters 3
Mean 6234	Standard deviation 2963
R-square 0.9163	Adjusted R-square 0.9099
Durbin-Watson 0.9613	Ljung-Box(18)=28.17 P=0.9405
Forecast error 889.5	BIC 1002
MAPE 0.1295	RMSE 842.2
MAD 625.7	

Figure 5.31: Yearly Forecast versus Actuals using Multiple Regression
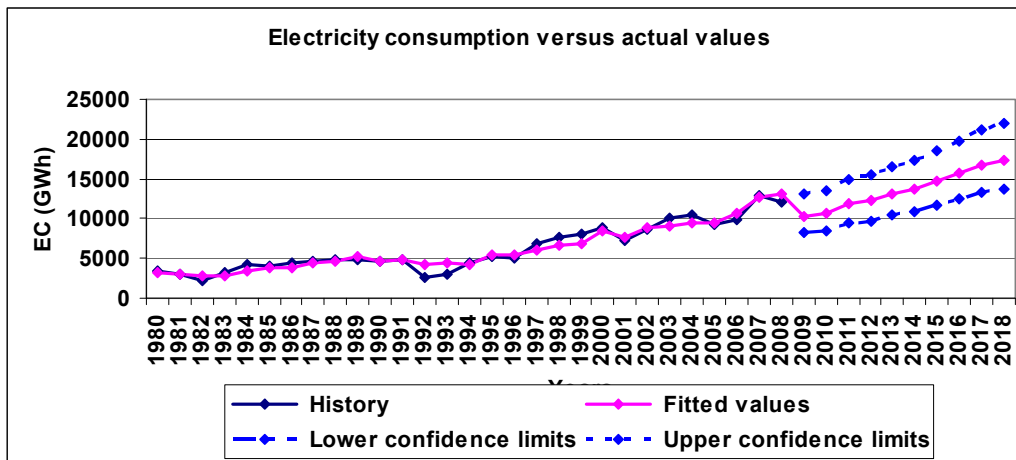
Exponential smoothing produces the best medium term forecast on a monthly basis that can be estimated yearly as shown in Figures 5.33 and 5.34 after being back transformed. Based on the seasonality factor, the results show a gradual increase in consumption over the months. These monthly forecasts can inform the yearly forecasts as shown in Figure 5.34. Parameters such as α, δ or β and γ can be used to adjust for randomness that can exist as a result of unforeseen world events that can

have an impact on the behaviour of the forecast. This is done through market analysis presented in Chapter 3 of this study.

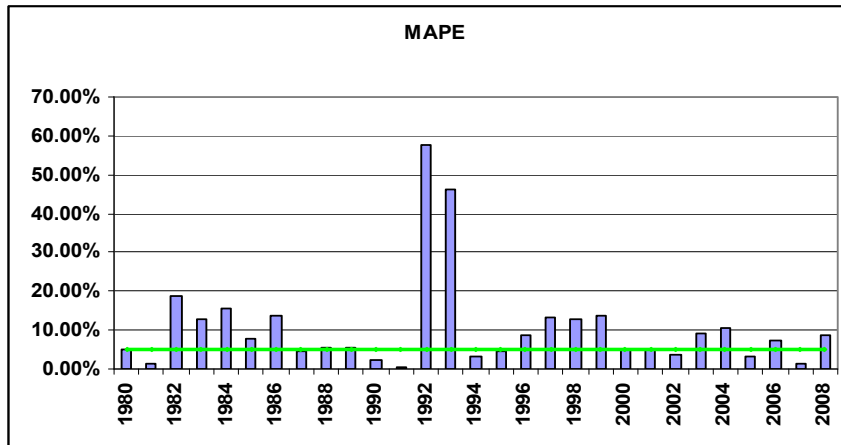**Figure 5.32:** <u>**Yearly MAPE versus α = 0.05 level of significance**</u>

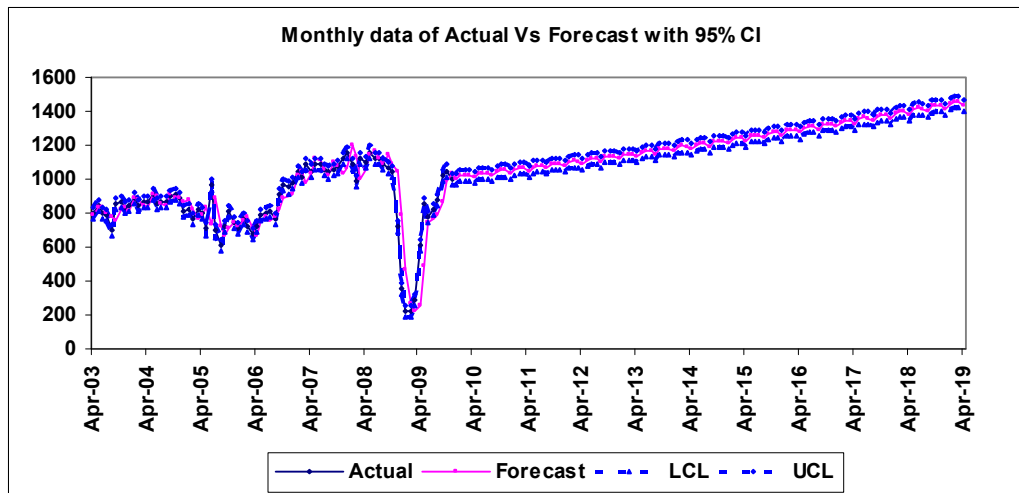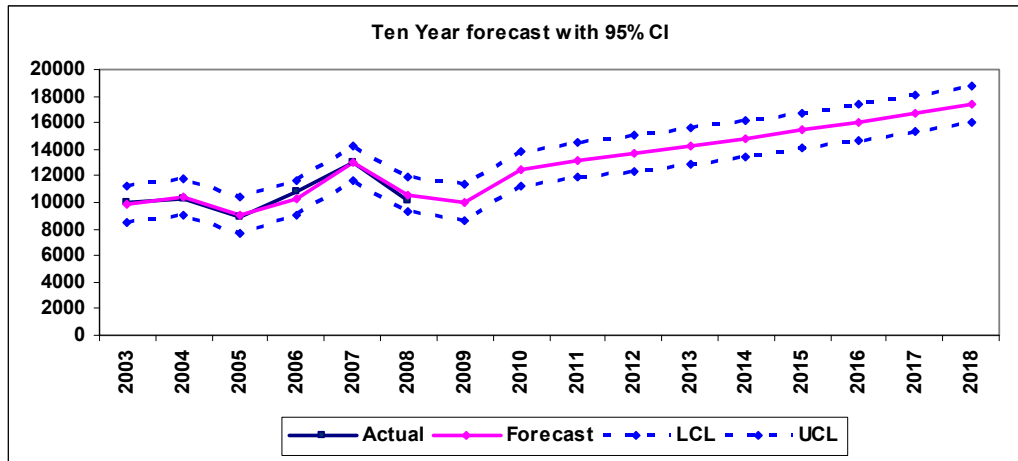**Figure 5.33:** <u>**Monthly forecast versus Actuals using Exponential Smoothing**</u>

Figure 5.34: **Yearly forecast versus Actuals using Exponential Smoothing method**

The results that compare the previous actual with the forecasts for each year are shown in Figure 5.35 where the MAPE over all the years are below 5% level of significance. These MAPEs on a yearly basis imply that the error in GWh is between 24GWh and 500GWh (i.e. about $\pm$(7GWh, 44GWh) per month on average). This shows that this method is appropriate for forecasting the medium term electricity consumption forecast for the FeCr sector. Average MAPE over the five year period is only 2%, which is consistently below the 5% level of significance.

Figure 5.35: **Yearly MAPE versus α = 0.05 level of significance**
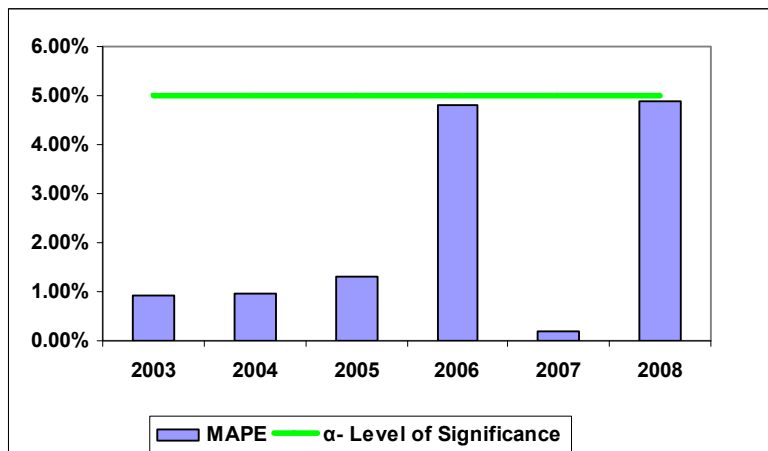
Table 5.29: Overall results for Exponential smoothing model

Exp. smoothing: Multipl. season (4) S0=6.692 T0=1.000 (Yearly Data FeCr 2003_2008 monthlynew.sta) Expon.trend,mult.season; Alpha= .700 Delta=.200 Gamma=.008 TransFeCr Energy Consumpt	
Summary of Error	Error
Mean error	0.002
Mean absolute error	0.100
Sums of squares	3.020
Mean square	0.038
Mean abs. perc. error	0.025

The multiple regression model is derived from stepwise forward regression where each variable is entered in sequence and its value is assessed. If adding the variable contributed to the model then it is retained, but all other variables in the model are then retested to see if they are still contributing to the success of the model. If they no longer contribute they are removed. This method ensures that the end result is the smallest possible set of predictor variables included in the model.

Based on Table 5.28, it can be observed that WSSC is not significant while the other two parameters are significant at 5% level. The MAPE (Figure 5.32) for the multiple regression model on average is 12.95% which is way above the 5% level. Therefore, multiple regression is not as reliable as exponential smoothing in forecasting electricity consumption for the FeCr sector on medium term basis.

Another shortcoming with the multiple regression method is that for the forecast to be produced, all the forecasted data for the dependent variable should be available which can be difficult to find. Looking at the predictor variable such as the US dollar, it can be difficult to find medium term forecast that is reliable since it has many factors contributing to its fluctuations. Thus, this method can be unreliable since it depends on forecasts of many other predictor variables to predict the future.

The results of the multiple regression fit in Figure 5.31 illustrate that the forecast and the actual values are not as close together as those in Holt-Winters Linear seasonal smoothing in Figure 5.34.

CHAPTER 6: SENSITIVITY ANALYSIS

6.1 Introduction

The traditional, statistical approach to forecasting has been based upon the identification, specification and estimation of a single model. Recently, computationally intensive methods which depart from tradition have become popular. The purpose of this chapter is to carry out sensitivity analysis by comparing models that best fitted the data set studied in this research for both short term and medium term with models that have gained popularity in recent years (as a result of being adopted by many researchers), and then finding the method that works best. The data set used in the previous chapter will be used to identify, estimate and validate the method that works best.

One of the models that have gained much popularity that will be used in comparison to those fitted is the GARCH model. These will be compared with Box Jenkins ARIMA and multiplicative seasonality Holt-Winter Linear smoothing.

6.2 GARCH Model

To compute GARCH (p, q) model, the following steps were taken:

The first point is to estimate the best fitting AR(q) model such that

$$y_t = a_0 + a_1 y_{t-1} + \cdots + a_q y_{t-q} + \varepsilon_t = a_0 + \sum_{i-1}^{q} a_i y_{t-i} + \varepsilon_t \ldots.. \; equation \; (6.1)$$

where,

a_0 is the intercept of the model

ε_t is the error terms and

y_t is the (AR) model

Secondly, the autocorrelation of residual errors are computed and plotted using

$$\rho = \frac{\sum_{t-i+1}^{T}\left(\hat{\epsilon}_t^2 - \hat{\delta}_t^2\right)\left(\hat{\varepsilon}_{t-1}^2 - \hat{\delta}_{t-1}^2\right)}{\sum_{t-1}^{2}\left(\hat{\varepsilon}_t^2 - \hat{\delta}_t^2\right)^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots equation \; (6.2)$$

where $\epsilon_t \sim \text{IN}(0,1)$

For large samples, the standard deviation of $\rho(i)$ is given by:

$$\rho(i) = \frac{1}{\sqrt{T}} \quad ...\,... \; equation \; (6.3)$$

where,

  T is the total number of samples.

Thus, individual values that are larger than $\rho(i)$ indicate the GARCH errors. LJung-Box test is used to estimate the number of lags and to test the hypothesis that there exists GARCH error in the conditional variance.

This model is tested for both short term and medium term. Thus the dataset applied for short term is the daily ferrochrome electricity consumption while for medium term we use the monthly ferrochrome electricity consumption data set.

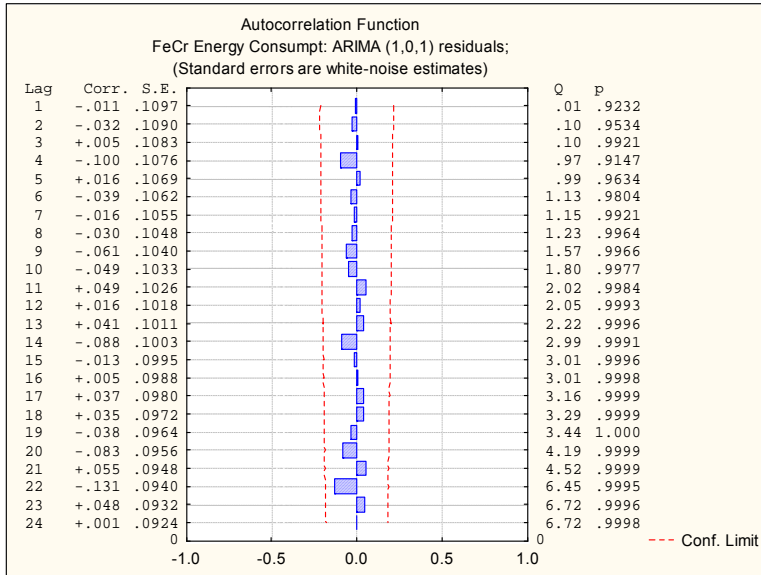**Figure 6.1: ACF of the residual values for medium term GARCH(1,1)**

Table 6.1: Summary of medium term analysis for GARCH(1,1) model

Forecast Model for FeCrEnergyConsumpt ARIMA(1,0,1)				
Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	0.8075	0.0746	10.8234	1.0000
b[1]	-0.2867	0.1204	-2.3802	0.9803
CONST	167.0950			
Within-Sample Statistics				
Sample size 80		Number of parameters 2		
Mean 868.1		Standard deviation 197.3		
R-square 0.7717		Adjusted R-square 0.7688		
Durbin-Watson 1.97		Ljung-Box(18)=6.988 P=0.009777		
Forecast error 94.86		BIC 98.94		
MAPE 0.09212		RMSE 93.66		
MAD 63.1				

Figure 6.2: ACF of the residual values for short term GARCH(1,1)

Autocorrelation Function
FeCr Consmp GWh: ARIMA (1,0,1) residuals;
(Standard errors are white-noise estimates)
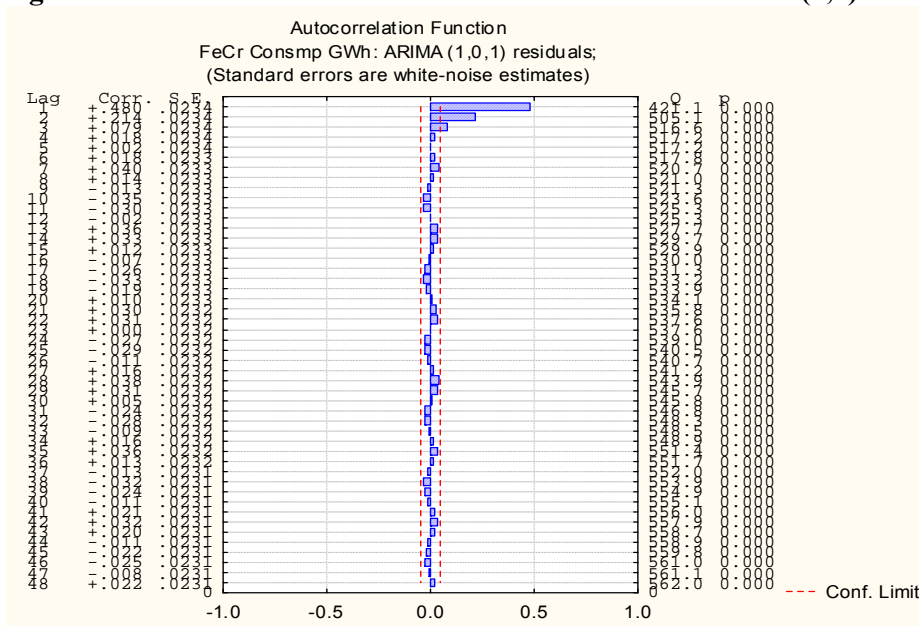
Figure 6.3: Residual values for short term GARCH(1,1)

Plot of variable: FeCr Consmp GWh
ARIMA (1,0,1) residuals;
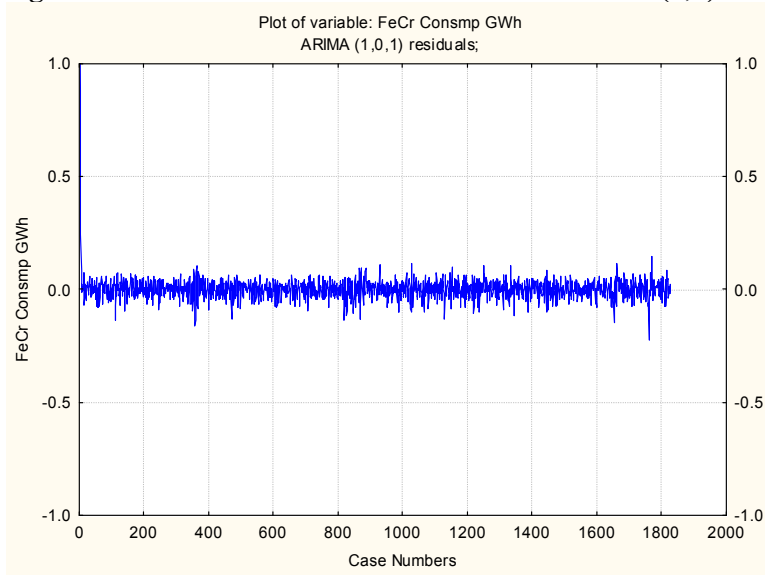
Table 6.2: Summary of short term analysis for GARCH(1,0) model

Forecast Model for FeCrConsmpGWh
ARIMA(1,0,0)

Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	0.9685	0.0059	163.7522	1.0000
CONST	0.9325			

Within-Sample Statistics

Sample size 1827	Number of parameters 1
Mean 29.57	Standard deviation 4.146
R-square 0.9357	Adjusted R-square 0.9357
Durbin-Watson 2.155	** Ljung-Box(18)=431.7 P=1
Forecast error 1.052	BIC 1.053
MAPE 0.02798	RMSE 1.051
MAD 0.8163	

Having computed the GARCH models, the next step is to test if there exist the GARCH errors such that the models can be extracted further from the residual errors of the AR(q) or ARIMA(p, q) models.

108

Hypothesis testing

We test the following hypotheses

H_o : There are existing GARCH errors in the conditional variance

H_i : There are no GARCH errors

The asymptotic standard deviation is given by:

$$\rho(i) = \frac{1}{\sqrt{T}}$$

where T is the total number of samples. Individual values that are larger than $\rho(i)$ indicate the GARCH errors. Ljung - Box test is used to estimate for the number of lags until the values are less that 10 percent significant.

Thus, for medium term data where the data points were equals to 80,

$$\rho(i) \quad = 1/\sqrt{80}$$
$$= 0.111803$$
$$\rho(i) \quad = 1/\sqrt{1827}$$
$$= 0.023395$$

When comparing $\rho(i)$ with the residuals of the autocorrelation function for both the short term and medium term in Figure 6.2 and Figure 6.3 respectively, it can be seen that there is no residual value which is greater than $\rho(i)$. This indicates that we reject the null hypothesis that there exist the GARCH errors in the conditional variance. This implies that there are no GARCH errors in the conditional variance of both data sets.

6.3 Sensitivity Analysis

The GARCH model is one of the widely applied models in forecasting electricity consumption. In this context the data maintain the homoscedastic behaviour within the residual values of the data sets. Thus this model is better applied in the data set

that has heteroscedastic condition in the variance such that the residual values from the autocorrelation function will have to be modelled in order to get the parameter estimates of the GARCH model.

The GARCH, ARIMA and the Holt-Winter Exponential Smoothing model has the following properties:

Properties of GARCH model

The primary interest of the GARCH model encompasses the following:
- Modelling changes in the variance
- Providing improved estimations of the volatility
- The model is not necessarily concerned with better forecasts but can improve forecasts already generated so as to reduce the margin of error
- The model can be integrated into ARMA
- The GARCH is more useful in modelling financial time series (De Gooijier & Handyman, 2006).

Properties of ARIMA models

- The ARIMA model consists of unit-root non-stationary time series which can be rendered stationary by the difference operation
- ARIMA models are useful in real applications
- The models are applied in stable data that has regular correlation
- The ARIMA needs a minimum of 40 data points.

Properties of the Holt-Winter Exponential Smoothing model

- The Holt-Winter Exponential Smoothing model is used when the data shows trend and seasonality
- The data can depict multiplicative or additive seasonality
- Parameters are initialized only once, hence once the parameters have been established, the forecasting can proceed without any delay in re-computation of the parameters
- Past data need not be remembered

- The model is suitable for series where the parameters remain more or less constant over a period of time (Kalekar, 2004).

Other models such as the multivariate data analysis have a short coming in that more dependent variables and independent variable should be examined in order to arrive at the proper model which is not robust to the interest of this study.

On the other hand, neural networks which with their ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. This method processes information in a similar way that the human brain does. The network is composed of a large number of highly interconnected processing elements working in parallel to solve a specific task. It learns by example. These elements cannot be programmed to perform a specific task. The examples must be selected carefully otherwise useful time is wasted, or even worse the network might be functioning incorrectly.

Thus for this method (neural networks), it is easy to make a mistake. The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable. In addition, this method is not in competition with others but complimentary to other methods. For this study, it is a task that is more suitable to an algorithmic approach operation while there are more tasks that are suited to neural networks. Hence a combination of two approaches, normally conventional forecasting methods and neural networks is more likely to perform at maximum efficiency (Stergiou and Siganos, 2008).
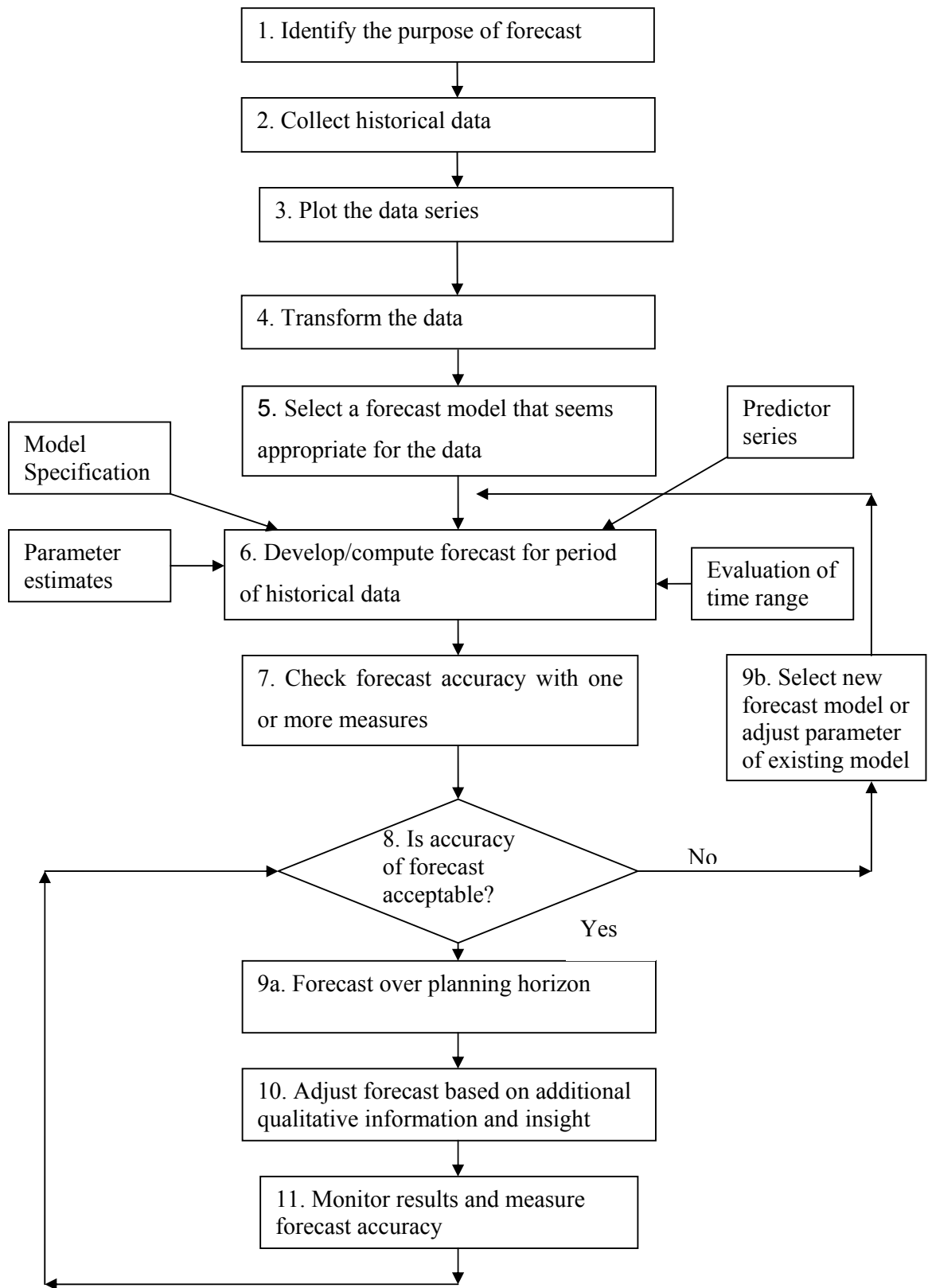
Thus comparing the results of the models computed in Chapter 5 to the most widely used methods such as GARCH, shows that the models that are recommended to forecast electricity consumption in the Ferrochrome sector for this study are as the findings points out.

6.4 Analysis and post optimality

The forecasting process starts with patterns of consumption behaviour which require one to plot the available historical consumption data , observe and analyse, and then attempt to determine the forecasting method that best fits the patterns exhibited by the data. Once a method has been identified, there are several measures available for comparing the historical data with the forecasts as discussed in the literature review of this study to determine how accurate the forecasts are.

If the forecast is not accurate, another method can be tried or judgement, experience, or knowledge of the market could be used as discussed in Chapter 3 of this study, or even intuition to adjust the forecast in order to enhance further accuracy. The actual consumption should then be monitored to assess the performance of the forecast method over the planning period. If the forecast is accurate, it is appropriate to continue using the forecast method, otherwise consideration must be given to selecting a new model or adjusting the existing one. The flow chart diagram that summarises all the work done in the previous chapters as stipulated by Russell and Taylor (1995) is indicated in Figure 6.4.

Figure 6.4: <u>Flow chart diagram for the forecasting process</u>

```
┌─────────────────────────────────────────┐
│  1. Identify the purpose of forecast     │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│  2. Collect historical data              │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│  3. Plot the data series                 │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│  4. Transform the data                   │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐      ┌──────────────┐
│  5. Select a forecast model that seems   │      │  Predictor   │
│  appropriate for the data                │      │  series      │
└─────────────────────────────────────────┘      └──────────────┘
```

┌──────────────┐
│ Model   │
│ Specification│
└──────────────┘

┌──────────────┐   ┌───┐   ┌──────────────┐
│ Parameter   │ ───► │ 6. Develop/compute forecast for period │ ◄─── │ Evaluation of│
│ estimates   │   │ of historical data   │   │ time range   │
└──────────────┘   └───┘   └──────────────┘

```
┌─────────────────────────────────────────┐      ┌──────────────────┐
│  7. Check forecast accuracy with one     │      │ 9b. Select new   │
│  or more measures                        │      │ forecast model or│
└─────────────────────────────────────────┘      │ adjust parameter │
                    │                             │ of existing model│
                    ▼                             └──────────────────┘
              ╱─────────────╲
             ╱  8. Is accuracy ╲
            ╱   of forecast     ╲  ──── No ────►
            ╲   acceptable?     ╱
             ╲                 ╱
              ╲───────────────╱
                    │ Yes
                    ▼
┌─────────────────────────────────────────┐
│  9a. Forecast over planning horizon      │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│  10. Adjust forecast based on additional │
│  qualitative information and insight      │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│  11. Monitor results and measure         │
│  forecast accuracy                        │
└─────────────────────────────────────────┘
```

CHAPTER 7: CONCLUSION AND RECOMMEMDATIONS

7.1 Discussion and recommendations

The challenge of determining electricity consumption forecasting model is of concern to forecasters and continues to receive attention in statistical literature. Forecasting is becoming more challenging because of the changing environment in the electricity market as well as other economic contributors globally. The strategic forecast of the power utility company should then set the ultimate goal in the longest term possible while interim tactical changes based on short-term forecast help deal with immediate changes.

The main objective of this research is to identify models to forecast the short and medium term forecasts for operational and tactical use using quantitative time series analysis models. The FeCr sector is examined for this study. Short term refers to a period of up to twelve months (one year) and medium term refers to a period of two to ten years.

Fitted for short term is the Autoregressive Moving Average model and Simple Exponential Smoothing with no trend and seasonality. Comparing these two models based on the best model selection tools like the R^2, the Bayes Information Criterion and the MAPE; the ARIMA(2,0,2)(1,0,0) model is recommended. From the analyses this model has the best fit to the data series and can thus be used to forecast short term electricity consumption for the ferrochrome sector.

Fitted for medium term is the Multiple Linear regression using stepwise regression and Holt-Winter's Linear Seasonal Smoothing. The Holt-Winter's Linear Seasonal Smoothing was the best fit for forecasting medium term. The Multiple Linear regression using stepwise regression cannot be recommended as the best model because its error is too wide. Another disadvantage is that it is difficult to make any conclusion with the Multiple Linear regression using stepwise regression because of

high variability due to multicollinearity and uncertainties surrounding the independent variables forecast for medium term. To some predictor variables, the data set that is used to produce a forecast will have to be forecasted. This means more risk surrounding the final forecast of the dependent variable since the data set involved is also a forecasted one.

The Holt-Winter's Linear Seasonal Smoothing exhibits both trend and seasonality. The two main Holt–Winter models are the additive model for time series exhibiting additive seasonality and the multiplicative model for time series exhibiting multiplicative seasonality. The Holt-Winter's Linear Seasonal Smoothing with multiplicative seasonality is the one recommended for forecasting electricity consumption for the ferrochrome sector for medium term. This model is effective in forecasting medium term forecasts as well as providing the confidence limits which measure the level of uncertainty into the future.

Another advantage of Holt-Winter's Linear Seasonal Smoothing is that it is a highly adaptable and robust tool to forecast in different horizons. Unforeseen world events that convey uncertainties into the future predictions can be built-in to this model since it has many parameter estimates that take such factors into account. In order to select the best parameter estimates the MAPE is used.

7.2 Limitations

The following are identified as limitations of the study:
- The study is based on the power utility's experience and therefore it might not necessarily hold true for other organisations in similar business.
- The study is based on data collected from power utility customers, which is highly confidential.
- The study does not provide or design new forecasting tools but recommends proper usage of some of the available tools.
- Tools available might limit the application of more complex methodology and findings of the methods applied are based on the available tools provided.

- Forecasts for other variables used other than from power utility, remain at the source and will not be used for any other interest except the one for which it was requested.

7.3 Further work

The models recommended in this study for both short and medium term were based on 2003 to 2008 data This means that the performance of the models should be monitored and compared against the actuals over time to assess if indeed the models are appropriate. All the measures that estimate appropriate models have been presented, but the truth still has to prevail.

This is a call to forecasters, researchers or statisticians for more research in sensitivity analysis to be carried out in order to best select the method that forecast well with fewer risks attached to it.

Like most studies there is still more work to be done. The following are the delimitations of the study.

- This research focuses on fewer causal variables and there is still a need to study many other variables to determine if they depict the same patterns or provide a better model than the one recommended.
- The study focuses on time and not much on world qualitative factors although it did touch on that, but an intensive study that looks at qualitative factors that can enhance the quality of the forecast is required.
- This study invites researchers to critique the models recommended and hence prove or disapprove their relevancy.
- This study only focuses on developing models that are applicable to FeCr sector. It will be interesting as a further study to investigate models that are appropriate to forecast electricity consumption in other sectors such as coal, platinum, traction etc.
- To ensure that models developed from this study remain relevant, proper monitoring and periodic reviews need to be enforced against changes in the environment.

- It is evident that multicollinearity is a common problem from this study, there is a need to perform analysis using statistical analysis that eliminates the presence of multicollinearity like principal components analysis which this study did not include. The use of principal components may not necessarily produce a better model but eliminates multicollinearity. It will be interesting as a further study to investigate use of principal components if it does eliminates multicollinerity

- In this study, there was no attempt to combine methods to improve the efficiency of the forecast. Hence, another focus may be to investigate the possibility of combining the models determined in the study with other conventional methods in an effort to improve efficiency

7.4 Conclusion

The main aim of this study is to find forecasting models to forecast electricity consumption for operational and tactical planning in the ferrochrome sector. In order to achieve this, a compressive literature review was carried out to review what other authors have found when carrying similar studies. In addition, it is of interest to find out other methods gaining popularity as far as this topic is concerned. It is interesting to establish that this topic continues to receive extensive attention in the literature.

The mathematical formulation of the methods and models investigated in this study has been provided to enhance the quality of the forecasts found. While the investigated methods of forecasting electricity consumption are technical, reliability of the forecasts is compromised if the use of judgment, experience and knowledge of the market is not facilitated.

One of the objectives of this study is to investigate the market of the ferrochrome sector and risk exposure of the power utility that is supplying electricity to this sector. In this study the market has been extensively investigated and findings discussed. The emphasis has been on identifying attribute/factors that have an impact on the behaviour of electricity consumption in the sector. In addition to stipulating the characteristics of the sector, the study also added more quality to the market dynamics

identified to influence the behaviour of electricity consumption. This is vital information for the power utility as it is the main supplier of electricity to this sector.

In this study a new method of forecasting electricity consumption in the ferrochrome sector for short term and medium term to increase efficiency is provided. Different methods that produced different models have been tried in order to select the best model that fitted the data adequately. Assumptions for the data sets provided were clearly indicated. Different measures to test the forecast accuracy after models identification were applied in order to establish which forecast model is more accurate than the others. For operation and tactical planning, models to forecast electricity consumption in the ferrochrome sector have been identified and recommended. To assess the robustness of these models, the data set for the ferrochrome sector electricity consumption was used. The results confirm goodness of the models proposed.

While there was no attempt to combine methods, a sensitivity analysis was conducted instead. The sensitivity analysis did not focus on the sensitivity of the model parameters as expected but on the sensitivity of the methods. In particular the methods investigated were GARCH, ARIMA and the Holt- Winter seasonal exponential smoothing. However, the literature review reveals other models that are gaining popularity in forecasting electricity consumption. These include neural network and GARCH models among others. In the study, a summary of the strength and weakness of the neural network and GARCH models is provided. However, only and in-depth analysis of the GARCH model was done.

In this study, an ARIMA method is suggested as an appropriate tool for operational forecasting of electricity consumption. The Holt-Winter Linear seasonal smoothing method is suggested for tactical planning.

REFERENCES

1. Al-Ghandoor, A & Samhouri, M. (2009). Electricity consumption in the industrial sector in Jordan: Application of multivariate linear regression and adaptive neuro-fuzzy techniques. *Journal of Mechanical and Industrial Engineering*, 3(1):69-76.

2. Al-Saba, T & El-Amin, I. (1999). Artificial neural networks as applied to long-term demand forecasting. *Artificial Intelligence in Engineering*, 13:189-197.

3. Adepoju GA, Ogunjuyigbe SOA, and Alawode KO. (2007). Application of Neural Network to Load Forecasting in Nigerian Electrical Power System. *The Pacific Journal of Science and Technology*.8(1):68-72

4. Andrews RL. (1994). Forecasting performance of structural time series models. *Journal of Business and Economic Statistics*, 12(1):129-133.

5. Azadeh, A, Ghaderi, SF and Sohrabkhani, S. (2008). Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management*, 49(8):2272-2278.

6. Baker, S. (2006). Multiple Regression Theory. http://hadm.sph.sc.edu/ courses/J716/pdf/716-3%20Multiple%20Regression.pdf

7. Balnac, K and Bokhoree C. (2008). Forecasting electricity demand of Mauritius using support vector machines for sustainable energy planning. *The International Journal of Environmental, Cultural, Economic and Social Sustainability*, 5:271-284.

8. Bhargava, N, Singh, B and Gupta, S. (2009). Consumption of electricity in Punjab: Structure and growth. *Energy Policy*, 37:2385-2394.

9. Bunn, DW. (1996). Non-traditional methods of forecasting. *European Journal of Operational Research*, 92:528-536.

10. Business dictionary. (2007-2009). http://www.businessdictionary.com

11. Ceylan, H and Ozturk, HK. (2004). Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach. *Energy Conversion and Management*, 45:2525-2537.

12. Copas, JB. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society,* 45(3): 311-354.

13. Costantini, M and Pappalardo, C. (2008). *Combination of forecast methods using encompassing tests*: *An algorithm-based procedure*. http://www.ihs.ac.at (accessed 04 February 2010).

14. De Gooijer JG and Hyndman, RJ. (2006). 25 years of time series forecasting. *International Journal of Forecasting,* 22:443-473

15. Draper, N & Smith, H. (1981). Applied Regression Analysis, (2nd Edition). New York: John Wiley & Sons, Inc.

16. Fatai, K, Oxley, L and Scrimgeour, FG. (2003). Modeling and forecasting the demand for electricity in New Zealand: a comparison of alternative approaches. *The Energy Journal*, 24:75-102.

17. Feinberg EA & Genehliou D, (2005). Applied mathematics for structural electric power systems: Load Forecasting. Power Electronics and Power systems, 2005: 269-285

18. Garett, TA and Leatherman, JC. (2000). An introduction to state and local public finance. The Web book of Regional Science. http://www.rri.wvu.edu/WebBook/Garrett/chapterfour.htm

19. Garson, DG. (2009). Quantitative Research in Public Administration: Multiple Regression. Caldwell: ncsu.edu

20. Gettler, L. (2007). Sox First: Six steps for forecasting. *Management & Compliance*.http://www.soxfirst.com/50226711/six_steps_for_forecasting.php (accessed 20 July 2009).

21. Ghader, SF, Azedeh, MA and Mohammadzadeh, S. (2006). Modeling and forecasting the electricity demand for major economic sectors of Iran. *Information Technology Journal*, 5(2):260-266.

22. Guerrero, VM and Berumen, E. (1998). Forecasting electricity consumption with extra-model information provided by consumers. *Journal of Applied Statistics,* 25(2):283-299.

23. Gujarati, DN. (2006). Essentials of Econometrics, (3rd Edition). U.S. Military Academy, West Point: McGraw-Hill Higher Education.

24. Hamzaçebi, C. (2007). Forecasting of Turkey's net electricity energy consumption on sectoral bases. *Energy Policy,* 35:2009-2016.

25. Hill, T and Lewicki, P. (2006). Statistics: Methods and Applications: a Comprehensive Reference for Science, Industry, and Data Mining. Tulsa Ok: StatSoft Inc.

26. Hirschhausen, C and Andres, M. (2000). Long-term electricity demand in China – From quantitative to qualitative growth. *Energy Policy,* 28:231-241.

27. Hocking, RR. (1976). The analysis and selection of variables in linear regression. *Biometrics, 32 (1):1-49, March 1976 (A Biometrics Invited Paper).*

28. Hyndman, RJ. (2008). Forecasting and the importance of being uncertain. *Indian Institute of Management Calcutta slides.* Australia: Centage publisher.

29. Inglesi, R. (2010). Aggregate electricity demand in South Africa: Conditional forecasts to 2030. Applied Energy: 87(2010):197-204

30. Kalekar, PS. (2004). Time series forecasting using Holt-Winters exponential smoothing.
http://www.it.iitb.ac.in/~praj/acads/seminar/04329008ExponentialSmoothing.pdf (accessed 04 January 2010).

31. Keller, G. (2005). Statistics for Management and Economics, (7th edition). United States of America, Duxbury: Thompson Brook/Cole publishing.

32. Keyno HRS, Ghader, F, Azade, A and Razmi, J. (2009). Forecasting electricity consumption by clustering data in order to decrease the periodic variable's effects and by simplifying the pattern. *Energy Conversion and Management,* 50(3):829-836.

33. Makridakis, S, Spyros, M, Steven C. W, and Hyndman R. J, (1998). 3rd edition. Forecasting: Methods and Applications. New York: John Wiley and sons.

34. Malo, P and Kanto, A. (2005). Evaluating multivariate GARCH models in the Nordic electricity markets. *Communications in Statistics: Simulation and Computation*, 35:117-148.

35. Managa, A. (2006). *Model selection criterion*. Unpublished thesis. Pretoria: University of Limpopo.

36. Maseema, MP. (2009). *Forecasting in an ever changing environment*. Paper presented at the forecasters forum meeting of Eskom on the forecasting in an ever changing environment conference: Midrand, South Africa.

37. McClave, JT, Benson, PG & Sincich, T. (2001). Statistics for Business and Economics, (8th edition). United States of America, New Jersey: Prentice-Hall, Inc.

38. Metaxiotis, K, Kagiannas, A, Askounis, D and Psarras, J. (2003). Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher. *Energy Conversion and Management,* 44:1525-1534.

39. Mohamed, Z, Bodger, PS, Hume, DJ. (2004). Forecasting electricity consumption: A comparison of models for New Zealand and the Maldives. Kathmandu, Nepal: *International Conference on Power Systems*, 3-5 November 2004: 6 pp.

40. Osborne, JW & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, 8(2).

41. Oxford dictionary. (2006). Cape Town: Oxford university Press Southern Africa

42. Ozturk, HK, Ceylan, H, Canyurt, OE & Hepbasli, A. (2005). Electricity estimation using genetic algorithm approach: a case study of Turkey. *Energy,* 30:1003-1012.

43. Pao, HT. (2009). Forecasting energy consumption in Taiwan using hybrid nonlinear models. *Energy*, 34(10):1438-1446.

44. Rahib, A,Vasif, HA, &Cemal A. (2005). Electricity Consumption Prediction Model using Neuro-Fuzzy System. World academy of science, Engineering and Technology 8(24):128-131

45. Rencher, AC & Pun, FC. (1980). Inflation of R^2 in best subset regression. *Technometrics,* 22:49-53.

46. Robert, D & Kugel, C. (2008). Intelligent Enterprise: Six steps to better sales forecasting and demand planning. Available from Intelligent Enterprise: www.intelligententerprise.com/showarticle.jtml?article ID=209000118

47. Russell, L & Kratowics, PE. (Editors). (2004-2009). Plant services: Forecasting. Available from plant services database: http://www.plantservices.com/articles/2003/28.html

48. Russell, D & Mackinnon JG. (1993). *Estimation and inference in econometrics*. Cape Town: Oxford University press.

49. Russell, RS & Taylor, BW. (1995). Production and Operations Management. New Jersey: Prentice-Hall, Inc.

50. Sahu, A. (2007). Forecasting. Available from reference for business database: www.referenceforbusiness.com

51. SAS user guide. (1989). *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2,* Cary, NC: SAS Institute Inc

52. Silk, MH. (1988). World chromite resources and ferrochromium production. Mintek, Randburg: Council for Mineral Technology publishing.

53. Smith, M. (2003). Electricity load and price forecasting using statistical methods and models. Available from http://www.secondmoment.org/articles/electricity.php (accessed 26 January 2010).

54. Souza, RC, Barros M, Miranda, CVC. (2007). Short term load forecasting using double seasonal exponential smoothing and interventions to account for holidays and temperature effects. *Journal of Operational Research Society*, 3-5 October 2007.

55. Stergiou, C. & Siganos, D. (2008).Neural Networks.Available from http://www.doc.ic.ac.uk/-nd/surprise_96/journal/vol4/ccs11/report.html (accessed 09 March 2010)

56. Tabachnik, BG & Fidell, LS. (1989). Using Multivariate Statistics, (3rd edition). New York: HaperCollins.

57. Tunç, M, Çamdali, Ü & Parmaksizoğlu, C. (2006). Comparison of Turkey's electrical energy consumption and production with some European countries and optimization of future electrical power supply investments in Turkey. *Energy Policy,* 34:50-59.

58. Walonick, DS. (1993). *An overview of forecasting methodology*. Available from http://www.statpac.com/research-papers/forecasting.htm (accessed 18 July 2009).

59. Yan, LM & Choon, OH. (2009). *Neural networks forecasting on electricity consumption in Malaysia*. Available from http://math.usm.my/research/OnlineProc/CS10.pdf (accessed 25 July 2009).

60. Zhang, X & Gu R. (2007). *Improved BP neural network for forecasting industrial electricity consumption in China*. Available from: http://ieeexplore.ieee.org/Xplore (accessed 25 July 2009).