STATISTICAL MODELLING OF RETURN ON CAPITAL EMPLOYED OF INDIVIDUAL UNITS

by

EMMANUEL CHAMUNORWA BUROMBO

Submitted in accordance with the requirements for

the degree of

MASTER OF SCIENCE

in the subject

STATISTICS

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: PROF PETER M. NJUHO

OCTOBER 2014

DECLARATION

Student number: **30628253**

I declare that **STATISTICAL MODELLING OF RETURN ON CAPITAL EMPLOYED OF INDIVIDUAL UNITS** is my own work and that all the sources that I have used or quoted have been indicated and 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:

Name: E. C. BUROMBO

Date: 12/10/2014

ACKNOWLEDGEMENTS

First, I thank God the Almighty for caring and protecting the various individuals who assisted me in carrying out this research project. In particular, I would like to thank the supervisor, Prof Peter M. Njuho for his guidance throughout the writing of this dissertation, of which it was going to be impossible to finish without his involvement. His high level professional maturity and commitment to duty/work was just second from none. Each time after talking to him, I would remain rejuvenated and eager to move forward. Not forgetting the departmental chair, Dr. Eeva Rapoo. Her swift intervention to finally nominate Prof Peter M. Njuho as supervisor of this dissertation has been a key to its successful completion. I am also thankful to Mr. Busanga Jerome Kanyama, another member of the department of statistics for providing advice for this dissertation. His co-operation throughout the conduct of this research was very influential to the writing of this dissertation.

Further, I wish to thank Mrs. Ashley Mutezo for her assistance with relevant data for this research project. Her co-operation was very marvellous. Special thanks go to Mr. Patrick Senderayi for editing this dissertation. My greatest thanks go to my friend Ian Ndlovu for his continuous prayers throughout the writing of this dissertation. He has been a constant source of inspiration and has been supportive every time I consulted him. I am extremely grateful to Joana and Richard Orton, and Anke and Vincent Wicherink for funding this research. Without their involvement I would be nothing. For the rest of my life, I will ever remain grateful to these two families for the support I got during my tenure as a masters student at UNISA.

ABSTRACT

Return on Capital Employed (ROCE) is a popular financial instrument and communication tool for the appraisal of companies. Often, companies management and other practitioners use untested rules and behavioural approach when investigating the key determinants of ROCE, instead of the scientific statistical paradigm. The aim of this dissertation was to identify and quantify key determinants of ROCE of individual companies listed on the Johannesburg Stock Exchange (JSE), by comparing classical multiple linear regression, principal components regression, generalized least squares regression, and robust maximum likelihood regression approaches in order to improve companies decision making. Performance indicators used to arrive at the best approach were coefficient of determination (R^2), adjusted R^2 (R^2_{adj}), and Mean Square Residual (MSE). Since the ROCE variable had positive and negative values two separate analyses were done.

The classical multiple linear regression models were constructed using stepwise directed search for dependent variable log ROCE for the two data sets. Assumptions were satisfied and problem of multicollinearity was addressed. For the positive ROCE data set, the classical multiple linear regression model had a R^2 of 0.928, an R^2_{adj} of 0.927, a MSE of 0.013, and the lead key determinant was Return on Equity (ROE),with positive elasticity, followed by Debt to Equity (D/E) and Capital Employed (CE), both with negative elasticities. The model showed good validation performance. For the negative ROCE data set, the classical multiple linear regression model had a R^2 of 0.666, an R^2_{adj} of 0.652, a MSE of 0.149, and the lead key determinant was Assets per Capital Employed (APCE) with positive effect,

followed by Return on Assets (ROA) and Market Capitalization (MC), both with negative effects. The model showed poor validation performance. The results indicated more and less precision than those found by previous studies. This suggested that the key determinants are also important sources of variability in ROCE of individual companies that management need to work with.

To handle the problem of multicollinearity in the data, principal components were selected using Kaiser-Guttman criterion. The principal components regression model was constructed using dependent variable log ROCE for the two data sets. Assumptions were satisfied. For the positive ROCE data set, the principal components regression model had a R^2 of 0.929, an R^2_{adj} of 0.929, a MSE of 0.069, and the lead key determinant was PC4 (log ROA, log ROE, log Operating Profit Margin (OPM)) and followed by PC2 (log Earnings Yield (EY), log Price to Earnings (P/E)), both with positive effects. The model resulted in a satisfactory validation performance. For the negative ROCE data set, the principal components regression model had a R^2 of 0.544, an R^2_{adj} of 0.532, a MSE of 0.167, and the lead key determinant was PC3 (ROA, EY, APCE) and followed by PC1 (MC, CE), both with negative effects. The model indicated an accurate validation performance. The results showed that the use of principal components as independent variables did not improve classical multiple linear regression model prediction in our data. This implied that the key determinants are less important sources of variability in ROCE of individual companies that management need to work with.

Generalized least square regression was used to assess heteroscedasticity and dependences in the data. It was constructed using stepwise directed search for dependent variable ROCE for the two data sets. For the positive ROCE data set, the weighted generalized least squares regression model had a R^2 of 0.920, an $\mathrm{R}^2_{\mathrm{adj}}$ of 0.919, a MSE of 0.044, and the lead key determinant was ROE with positive effect, followed by D/E with negative effect, Dividend Yield (DY) with positive effect and lastly CE with negative effect. The model indicated an accurate validation performance. For the negative ROCE data set, the weighted generalized least squares regression model had a R^2 of 0.559, an R^2_{adj} of 0.548, a MSE of 57.125, and the lead key determinant was APCE and followed by ROA, both with positive effects. The model showed a weak validation performance. The results suggested that the key determinants are less important sources of variability in ROCE of individual companies that management need to work with.

Robust maximum likelihood regression was employed to handle the problem of contamination in the data. It was constructed using stepwise directed search for dependent variable ROCE for the two data sets. For the positive ROCE data set, the robust maximum likelihood regression model had a R^2 of 0.998, an $\mathrm{R}^2_{\mathrm{adj}}$ of 0.997, a MSE of 6.739, and the lead key determinant was ROE with positive effect, followed by DY and lastly D/E, both with negative effects. The model showed a strong validation performance. For the negative ROCE data set, the robust maximum likelihood regression model had a R^2 of 0.990, an R^2_{adj} of 0.984, a MSE of 98.883, and the lead key determinant was APCE with positive effect and followed by ROA with negative effect. The model also showed a strong validation performance. The results reflected that the key determinants are major sources of variability in ROCE of individual companies that management need to work with.

Overall, the findings showed that the use of robust maximum likelihood regression provided more precise results compared to those obtained using the three competing approaches, because it is more consistent, sufficient and efficient; has a higher breakdown point and no conditions. Companies management can establish and control proper marketing strategies using the key determinants, and results of these strategies can see an improvement in ROCE.

KEY WORDS

Classical multiple linear regression, principal components regression, generalized least squares regression, robust maximum likelihood regression, Return on Capital Employed, stepwise directed search, Kaiser-Guttman criterion, key determinants.

TABLE OF CONTENTS

CHAPTER 3

CHAPTER 4

CHAPTER 1

INTRODUCTION

1.1 Background

The Johannesburg Stock Exchange (JSE) as the sole equities exchange in South Africa approves all companies seeking a listing (MacLeod, 2008). Companies across the range of mining, manufacturing, commerce, and service industry are listed on the JSE. The JSE initially established as a primary financing source to sustain the mining of gold, discovered in the South African mountain range, Witwatersrand, in 1886, make a considerable component of the economic activities of South Africa (MacLeod, 2008). Companies listed on the JSE contribute about 8.6 % of South Africa's Gross Domestic Product (GDP) indirectly and create an average monthly traded value of R63.99 million (Mkhize and Mbanga, 2006). In 2009, the listed companies sustained more than 137 million jobs throughout South Africa and added an estimated R10836.2 billion to state coffers (Econex and Quantec Research, 2010). Currently, the JSE is the best in Africa and is one of the top 20 exchanges in the world in terms of total market capitalization (World Federation of Exchanges, 2012). The positive performance of the JSE is an indicator of healthy economy of the South African.

The four markets offered by the JSE are equities, interest rate, active financial derivatives, and agricultural products. At the equities market, various companies meet to raise the public capital needed to expand their businesses. Likewise institutions (companies and government) and individuals in turn, approach the equities market to invest their money into these companies. Each listed company agrees on a certain amount of shares of their stock to be sold to the investing community in return for cash to finance its business. The investing community bet on the fact that the purchased stock will financially perform well and eventually produce a return for the investor. The improvement in the company's profits shows that the company is performing well overall. The measures of profitability include:

1) Gross Profit Margin (GPM).
2) Net Profit Margin (NPM).
3) Return on Investment (ROI).
4) Total Shareholder Return (TSR).
5) Return on Capital Employed (ROCE) (Damodaran, 2007; Eljelly, 2009; Nimalathasan and Brabete, 2010).

In this dissertation, we concentrate on ROCE.

Return on Capital Employed (ROCE) is a popular financial instrument and communication tool for the appraisal of companies (Singh and Yadav, 2013). It is a key parameter for measuring the performance of the management and value of a business (Damodaran, 2007). Return on Capital Employed (ROCE) indicates how well a company is utilising its capital to generate revenue and is calculated as $\frac{Profit\ before\ interest\ and\ tax}{Capital\ employed} \times 100\%$ (Mandal et al., 2010). Because ROCE measures profitability with respect to invested capital, it is important for firms that require large amounts of initial capital investment before they start producing goods/products. The advantage of this ratio is that it considers all the factors of a company and is widely used in 1) showing how much a company is gaining for its assets or losing for its liabilities, 2) proving the value the business gains from its assets and liabilities, 3) assessing whether a business generates enough returns to pay for its cost of capital, and 4) making intra- and inter-business comparisons (Damodaran, 2007; Singh and Yadav, 2013). While ROCE takes into consideration all the capital invested in the business, it does not deal with the sensitivity to capital structure and the fact that no account is taken of the cost of the capital employed. Return on Capital Employed (ROCE) is also distorted by the inclusion of non-operating assets such as surplus cash and investments on which the shareholders should not expect to earn more than a market rate of return (Singh and Yadav, 2013).

A positive ROCE indicates that a company 1) offers products that command a return and 2) is growing, well managed and profitable (Pattabiraman, 2013; Singh and Yadav, 2013). This means that a greater portion of its profits is invested back into the business for the benefit of shareholders and generation of further growth. Companies with negative or zero ROCE are at risk of making loss if trading conditions deteriorate (Pattabiraman, 2013; Singh and Yadav, 2013). Often, such companies are subject to changes in management control and the immediate task of new management is to lift ROCE (Singh and Yadav, 2013).

1.2 Justification

For management to enhance ROCE, it must identify and control the key determinants of ROCE. Companies use a variety of methods to establish, organize and prioritize the key determinants in order to improve ROCE, and become more competitive. Often, management resorts to subjective methods, such as instinct, beliefs, trial and error, tradition, personal experience, and role modelling for establishing the key determinants (Hobarth, 2006). When data is available, management prefer to use financial statements to understand the relationship between the attributes and ROCE (Singh and Yadav, 2013). Appearance of a pattern in the analysis of financial statements suggests existence of variables which contribute to ROCE (Hillestad, 2007; Pattabiraman, 2013). However, this does not provide enough information for identifying variables as financial statements are not exhaustive, and is more meaningful as a ratio analysis instead of a modelling procedure. Some studies (Cameron, 2011; Steyn, 2012) have utilized the behavioural approach anchored on social science methodologies when investigating the key determinants of ROCE, in contrast to the scientific statistical paradigm. Companies management are interested in developing a procedure that will be applied in a broader way using a statistical technique. In searching for a statistical method that finds the key determinants of ROCE, simple linear regression was proposed and used to model ROCE (Kan and Robotti, 2009). Models were developed based on the available data in sectors such as finance, equity, retail, and construction. The simple linear regression approach models the relationship between a continuous dependent variable and one continuous or categorical independent variable (Hussien, 2010). While the simple linear regression model gave the effect of each variable on ROCE, it failed to detect the collective effects of the variables as the workings of systems are not represented by a simple formulation.

1.3 Problem Statement

This dissertation seeks to construct the best/optimal models which explain variability in positive ROCE and negative ROCE of individual companies by comparing classical multiple linear regression analysis, principal components regression, generalized least squares regression, and robust maximum likelihood regression approaches in order to improve companies decision making.

1.4 Motivation

The motivation behind the desire to apply the regression approaches to real life situation data lies in the general interest of the theory on parametric regression and passion to assess the strengths and weaknesses of the different regression approaches, and to infer appropriate conclusions.

1.5 Objectives

The objectives of this study are:

1) To assess the joint effect of the key determinants on positive ROCE and negative ROCE.

2) To identify the specific lead key determinants on positive ROCE and negative ROCE.

3) To evaluate the use value of the key determinants on positive ROCE and negative ROCE.

1.6 Significance

The regression models are used to guide company management, potential investors, shareholders, creditors, debtors, and policy planners to develop better decisions, and minimize risk. The models are useful because they give information about the variables which have the greatest impact on ROCE, and serve as strong guidelines to risk measurement in decision making. The variables are used in crafting marketing strategies that see a positive response in ROCE. When a company improves in

ROCE, it attracts more investors, shareholders, creditors, and debtors. This in turn, induces a positive influence on the performance of the JSE market and the broader economy of South Africa.

1.7 Dissertation Outline

This dissertation has five chapters. Following this introductory chapter is Chapter 2 which is the literature review. The chapter has seven sections. Section one discusses the measures of company performance and key determinants. The previous studies on predictive models relating to ROCE are presented in section two. This leads to the introduction of the potential key determinants of ROCE to be used in the research in section three. The last four sections review literature on the proposed regression techniques used to carry out this study. Section four delves into the relevant literature of classical multiple linear regression analysis. This is followed by the review of the motivational literature of principal components regression, generalized least squares regression, and robust maximum likelihood regression in sections five, six, and seven, respectively.

Chapter 3 presents the real data set used in this study. The data set came from the McGregor Bureau of Financial Analysis (BFA) database. The chronological procedures which are followed in order to perform classical multiple linear regression analysis, principal components regression, generalized least squares regression, and robust maximum likelihood regression are outlined and discussed in this chapter.

The results of the actual implementation of classical multiple linear regression analysis using real data are discussed in Chapter 4. The key conditions for modelling data using classical multiple linear regression analysis are:

1) Absence of multicollinearity among the independent variables.
2) Linearity of the coefficients.
3) Normal distribution for the residual terms.
4) Common distribution and independence of the residual terms.

Classical multiple linear regression analysis has a breakdown value of 0 % (Mutan, 2004). This implies that, any small distortion and contamination in the data cause violations of the four conditions, leading estimated regression coefficient values to vary in $(-\infty, +\infty)$. Hence, when the four conditions are not satisfied the classical multiple linear regression analysis technique is not optimal and give incorrect conclusions. In the respective situations, we demonstrate that principal components regression, generalized least squares regression, and robust maximum likelihood regression techniques are alternatives for modelling ROCE. The data analysis is done using SPSS and R statistical software packages. Lastly, some concluding remarks which include the summary of the dissertation and proposals for future research are presented in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1 Measures of Company Performance and Key Determinants

The various measures of company performance belong to one of the financial groups (Eljelly, 2009; Hobarth, 2006; Nimalathasan and Brabete, 2010).

For example:

1) Most returns belong to the profitability group.

2) Market value includes Market Capitalisation Growth (MCG) and Stock Price (SP) change.

3) Operating Profit Margin (OPM) and stock turnover are some of the indicators of operating efficiency.

4) Production is expressed in terms of sales efficiency.

5) Cash flow is indicated by dividends distributed to shareholders.

6) Growth is indexed by price to book value (Eljelly, 2009; Hobarth, 2006; Nimalathasan and Brabete, 2010).

Generally, most companies prefer customary means such as profitability to indicate performance (Eljelly, 2009).

The key determinants of a firm's performance belong to one of the following groups

1) Liquidity/solvency.

2) Asset utilisation.

3) Capital structure.

4) Profitability.

5) Growth ability.

6) Shareholder.

7) Size.

8) Market (Eljelly, 2009; Hobarth, 2006; Nimalathasan and Brabete, 2010).

Examples of each group are shown in Table 2.1.

Table 2.1. Examples of measures of key determinants

Measure	Examples
Liquidity/ Solvency	Average Collection Period, Average Payment Period, Expenses Income, Capital Adequacy Ratio, Cash Conversion Cycles, Cash Flow to Total Debt, Cash to Sales, Cash to Current Liabilities, Creditors/Payables Turnover Ratio, Current Assets, Current Assets Turnover, Current Liabilities, Current Ratio, Current Asset to Total Asset Ratio, Current Liabilities to Total Assets Ratio, Debtors Turnover Ratio, Funds Flow to Total Liabilities, Inventory Turnover Ratio, Liquidity Assets to Current Liabilities, Market Value Equity to Total Debt, Net Current Asset Turnover, Payable Ratio, Quick Ratio/Acid Test Ratio, Total Debt, Total Liabilities, Total Debt to Total Assets, Total Equity to Total Assets.
Asset utilisation	Assets per Capital Employed, Assets Turnover, Fixed Assets, Fixed Assets Turnover Ratio, Return on Assets, Total Assets per Employee, Working Capital per Employee.
Capital structure	Capital Gearing Ratio, Debt to Assets Ratio, Debt to Capital Ratio, Debt to Equity Ratio, Equity Capital, Indebtedness Ratio, Interest Coverage Ratio, Long Term Debt Ratio, Debt to Total Fund Ratio, Proprietary Ratio, Fixed Assets to Proprietor's Fund Ratio.
Profitability	Cash Flow Return on Investment, Cash Return Ratio, Expenses Ratio, Gross Profit Margin Ratio, Net Profit Margin Ratio, Net Income to Total Sales, Operating Profit/Earnings before Interest and Tax, Operating Revenue per Employee, Profit per Employee, Retained Earnings to Total Assets, Return on Capital Employed/ Return on Invested Capital, Return on Equity, Return on Investment, Return on Net Worth, Return on Risk-Adjusted Capital, Return on Sales, Sales Turnover, Tax to Total Assets, Total Shareholder Return, Equity Shareholder Return.
Growth ability	Retention Rate, Sustainable Growth Rate, Growth Rate on Net Profit, Growth Rate on Total Assets, Price to Book Value, Turnover Growth.
Shareholder	Book Value per Share, Cash Flow per Share, Dividend Yield, Dividend Payout Ratio, Dividend Cover, Dividend per Share, Dividend Rate, Earnings per Share, Earnings Yield, Price to Earnings Ratio, Price to Earnings Growth, Price to Cash Earnings per Share, Sales per Share, Shareholders' Equity.
Size	Capital Employed/Total Assets, Human Resource, Market Share, Sales Volume.
Market	Balanced Scorecard, Beta, Book Value per Market Value Ratio, Cash Value Added, Discounted Cash Flows, Economic Value Added, Free Cash Flows, Internal Rate of Return, Jensen's Alpha, Market Capitalization, Market Value Added, Market Value to Book Value Ratio, Net Present Value, Share Price, Shareholder Value Analysis, Tobin's Q Ratio, Total Business Return, Tracking Stocks.

Source: Modified from the literature of Eljelly (2009), Hobarth (2006), Nimalathasan and Brabete (2010)

The key determinants and performance are relative measures that apply differently (Hobarth, 2006). Some determinants of performance are performance indicators and vice-versa, and depend on the performance factor being measured (Hobarth, 2006). Identifying and quantifying key determinants of performance assist practitioners to take action with regard to the optimal variables that generate maximum output. This leads us to examine some previous studies on predictive models in our next section.

2.2 Previous Studies on Predictive Models

Models developed to investigate sources of variation in ROCE used the classical multiple linear regression analysis method (Cameron, 2011).

Nimalathasan and Brabete (2010) found that there is a significant moderate linear relationship between Interest Coverage and ROCE among the capital structure variables that were used in the model. The R^2 value of the relationship is 0.750.

According to Azhagaih and Gavoury (2011), variables Expenses Income, Debt to Assets, and Current Ratio had significant non-linear relationships with ROCE among the capital structure and liquidity variables that were fitted, with the model giving a R^2 value of 0.159.

On the other hand, Tudor (2009) observed significance of $\ln(\text{Earnings per Share})$ and $\ln(\text{Book Value per Market Value})$, and strong linear relationships of these variables with ROCE. The R^2 value of the model is 0.214.

Ray (2011) showed that there are significant linear relationships of Debt to Assets, Dividend Yield, Accumulated Depreciation to Gross Fixed Assets, Gross Fixed Asset to Net Fixed Asset, Current Asset Turnover, Net Current Asset Turnover, and Liquidity Assets to Current Liabilities with ROCE. The R^2 value of the model is 0.970.

In another study, Mohamad and Saad (2010) revealed that there are significant non-linear relationships of Cash Conversion Cycles, Current Assets to Current Liabilities, and Current Assets to Total Assets with ROCE among the liquidity variables that were fitted. The resulting model has a low R^2 value of 0.194.

Sulait (2010) showed that there exists a significant non-linear relationship between Social Relational Capital and ROCE among the variables that were fitted. The model showed a low R^2 value of 0.283.

A study by Aanu et al. (2014) provided significant non-linear relationships of audit committee independence and board size with ROCE. The model yielded a low R^2 value of 0.198.

The previous studies found that the classical multiple linear regression models are non-exhaustive and generally have poor fits. Most independent variables in the models depicted non-linear relationships with ROCE. It is possible to improve the predictability of ROCE by fitting principal components regression, generalized least squares regression, and robust maximum likelihood regression (Fox and Weisberg, 2010; Motyka, 2003; Myung, 2003; Sabbagh, 2003). The next section highlights the potential key determinants which are used in this study.

2.3 Potential Key Determinants

Thirteen variables are assessed on ROCE based on 1) findings gleaned from previous studies and 2) ease of computations. The variables on companies listed on the JSE are outlined as follows:

1) Assets per Capital Employed (APCE)

Assets per Capital Employed (APCE) shows how efficient a company is in generating assets from the money invested and is computed as $\frac{Total\ Assets}{Capital\ Employed}$ (Arafat and Shahimi, 2013).

2) Capital Employed (CE)

Capital Employed (CE) is the value of all the physical and material assets employed in running a company (Shourvarz and Sadeddin, 2011). It is determined by using the formula fixed assets + current assets–current liabilities (Shourvarz and Sadeddin, 2011).

3) Debt to Assets (D/A)

Debt to Assets (D/A) gives total liabilities per total assets of a company and is computed as $\frac{Liabilities}{Total\ Assets}$ (Nimalathasan and Brabete, 2010).

4) Debt to Equity (D/E)

Debt to Equity (D/E) is the relative proportion of shareholders' equity and debt used to finance a company's assets (Ray, 2011). Companies in heavy industries need investments in property and machinery resulting in a higher debt to equity ratio (Ray, 2011). It is used to assess the ability of the firm to meet its liabilities and is calculated by using the formula $\frac{Debt\ (liabilities)}{Equity}$ (Ray, 2011).

5) Dividend Yield (DY)

Dividend Yield (DY) also called dividend-price ratio is the return the investor gets from a stock (Shobhana and Karpagavalli, 2011). It is calculated by dividing the 12 month dividend per share by the last market share price (Erasmus, 2013). Investors, who want to get regular income, invest their money in stocks with high stable DYs.

6) Earnings per Share (EPS)

Earnings per Share (EPS) is an indicator of the economic performance of a firm (Mandal et al., 2010). It measures the financial position that looks at the cash flow generated by the company on a share basis (Sharma, 2011). It is computed by dividing the Net Income available to common shareholders by the number of Equity Shares outstanding (Erasmus, 2013).

7) Earnings Yield (EY)

Earnings Yield (EY) is earning per share per current market price and is determined by the ratio of reported earnings to share price (Wilcox, 2007). The advantage of this ratio is that it is easily compared with other forms of investments like fixed deposits and it grows over time.

8) Market Capitalization (MC)

Market Capitalization (MC) is the total market value of all outstanding shares publicly available and it measures company size (Comincioli et al., 2012). It is calculated as share price × number of shares outstanding (Comincioli et al., 2012).

9) Operating Profit Margin (OPM)

Operating Profit Margin (OPM) indicates the extent of sales that is absorbed by the cost of goods sold and operating expenses of a company, and it varies among individual companies of the same industry (Nimalathasan and Brabete, 2010). It is calculated by using the formula $\frac{Profit\ before\ interest\ and\ tax}{Turnover} \times 100\ \%$ (Nimalathasan and Brabete, 2010).

10) Price to Earnings (P/E)

Price to Earnings (P/E) measures the expensiveness of a stock, and it expresses the relationship between share price and earnings per share of a firm (Sharma, 2011). It is used to decide whether or not to buy shares in a company, with low P/E indicating better value of the company. Price to Earnings (P/E) is computed by using the formula $\frac{Market\ Share\ Price}{Earnings\ per\ Share\ (E/S)}$ (Sharma, 2011).

11) Return on Assets (ROA)

Return on Assets (ROA) indicates how profitable a company is relative to its total assets (Mandal et al., 2010). The parameter measures what the company can do with what it has got (Comincioli et al., 2012). It is generally computed as $\frac{Net\ Income}{Total\ Assets}$ (Comincioli et al., 2012).

12) Return on Equity (ROE)

Return on Equity (ROE) measures the profit that a company generates with the investor's money (Damodaran, 2007). This ratio provides information which is useful in estimating efficiency of a firm in generating earnings, meaning a high ROE ratio implies that equity shareholders are given a higher dividend (Comincioli et al., 2012). It is calculated as $\frac{Profit\ after\ tax}{Shareholder\ funds(capital+reserves)} \times 100\ \%$ (Damodaran, 2007).

13) Share Price (SP)

Share Price (SP) is the market price at which a share is bought or sold, and is generally calculated as $\frac{Total\ Net\ Assets}{Number\ of\ Shares}$ (Sharma, 2011).

A brief review of classical multiple linear regression analysis follows in the next section.

2.4 Classical Multiple Linear Regression Analysis

Classical multiple linear regression analysis (Fielding and Gilbert, 2000) is a flexible statistical technique. It is informative when applied to any process or system. Classical multiple linear regression analysis uses ordinary least squares (OLS) to identify and model the relationship between measurable attributes (Mutan, 2004). The OLS is an estimation procedure which seeks the estimated regression coefficients that provide the most accurate description of the data (Myung, 2003). The method was developed by Adrien Marie Legendre (1752-1833) and Carl Friedrich Gauss (1777-1881). Legendre and Gauss started applying the OLS method in determining the orbits of bodies about the sun from astronomical data and this led to the first publication of the method as an appendix to a book by Legendre in 1805 (Mutan, 2004). More recent applications of the OLS method have been in agriculture, engineering, management, manufacturing, and psychology. For example, see Hussien (2010), Myung (2003), Person (2011), Pierce et al. (2010), and Shaffer (2007) for some of these applications.

2.4.1 Selection Procedures

Modelling is done to find a powerful model that explains the dependent variable. In developing a classical multiple linear regression model, too many independent variables cause a higher prediction variance, whereas too few independent variables give a biased prediction (Ryan, 2008). A problem arises on which variables to exclude or include for prediction. Ryan (2008), pointed to three major techniques for solving the problem namely:

1) Stepwise regression.
2) All possible subsets regression.
3) Alternative sub-regressions.

The stepwise regression procedures are:

1) Backward elimination (BE).
2) Forward selection (FS).
3) Stepwise directed search.

The alternative sub-regressions include:

1) Variation of the principal components method.
2) Latent root regression analysis (LRRA).

The BE procedure starts with the whole regression model and removal of all unnecessary independent variables without increasing the prediction variance (Faraway, 2002). The FS procedure works from the opposite direction of BE and involves less computations than BE (Bowerman and O'Connel, 1990). The variables are inserted one at a time, starting with the independent variable that is most highly correlated with the dependent variable. However, FS gives predictors which are poor and inadequate. Stepwise directed search is the best and common variable screening method. It is a combination of FS and BE. A combination of FS and BE finds better subsets than the separate methods (Shaffer, 2007). However, stepwise directed search stops prematurely before all the important variables are captured hence the formulation of the best subset is not guaranteed. In addition, it fails to give room for the existence of other possible sub regressions. The all possible regression method fits every possible regression equation out of a total number of possible regressions (Ryan, 2008). However, it is not feasible in case of many independent variables.

2.4.2 Assumptions

Classical multiple linear regression model requires that statistical assumptions be fulfilled. The assumptions are:

1) Linearity of the coefficients.
2) Normal distribution for the residual terms.
3) Common distribution (homoscedasticity) of the residual terms.
4) Independence among the residual terms.

When modelling ROCE of companies listed on a stock exchange market using financial data, it is possible that these assumptions are not satisfied (Nimalathasan and Brabete, 2010).

2.4.3 Diagnostics

Nimalathasan and Brabete (2010) used the following multiple decision criteria:

1) Highest coefficient of determination R^2 and adjusted R^2.

2) Lowest Mean Square Residual (MSE).

3) P-value for the F-statistic<0.05 for significance of the overall estimated regression.

4) P-value for the T-statistic<0.05 for significance of individual estimated regression coefficients.

The R^2 statistic is the coefficient of multiple determinations. It measures the proportion of variation in the dependent variable, which is obtained by using the independent variables in the model. The problem of R^2 is that it will always increase even if the added variable is not significant, hence adjusted R^2 is used. The adjusted R^2 statistic has the advantage that if insignificant variables are added to the model the value of adjusted R^2 decreases (Draper and Smith, 1998). A model with a good fit is indicated by R^2 and adjusted R^2 values close to 1 and a value of MSE close to 0 (Bowerman and O'Connel, 1990; Draper and Smith, 1998).

The major weakness of the classical multiple linear regression model is the instability of OLS estimates when the independent variables are involved in multicollinearity (Shaffer, 2007). Multicollinearity is a statistical phenomenon in which two or more independent variables in a classical multiple linear regression model are heavily related (Bowerman and O'Connel, 1990; Hair et al., 2010). Two variables exhibit complete collinearity if their correlation coefficient is 1 and complete lack of collinearity if their correlation coefficient is 0 (Hair et al., 2010). A diagnostic statistic called Variance Inflation Factor (VIF) is often used to identify multicollinearity (Shaffer, 2007). In addition to the aforementioned decision criteria, we propose 1) fulfilment of all statistical assumptions, 2) VIF <10 for lack of multicollinearity among the independent variables, and 3) Cook's Distance< $\frac{4}{\text{sample size}}$ for absence of influential data points on the prediction model. A data point is influential if its removal causes large changes in the estimated coefficients and fitted values (Hussien, 2010). In this study, we relate situations when outliers and leverages are classified as influential points. An outlier is a data point which is not typical of the rest of the data and a leverage point shows how much an individual point influences its own predicted value (Faraway, 2002). Variance Inflation Factors (VIFs) and Cook's Distances hedge the prediction model from multicollinearity and influential data points, respectively (Bowerman and O'Connel, 1990).

Validation concludes the construction of a classical multiple linear regression model. The purpose of validating the prediction model is to assess its performance. Validation ensures that the prediction model is reliable if it is used outside the vicinity of the original data (Shaffer, 2007). Bowerman and O'Connel (1990), and Hair et al. (2010) divided validation into 1) the split-sample to assess the predictability of the model, 2) use of the results of the prediction model in comparison with other previously validated results, and 3) collection of new data to validate the prediction model. The current study focuses on the last one of the three methods. The existence of multicollinearity leads to unstable estimates due to inflated variance. To address this problem, principal components regression is considered in the next section.

2.5 Principal Components Regression

Principal components regression is a means of regressing principal components on the dependent variable using OLS. Principal components analysis attempts to overcome the ill-conditioned situation of multicollinearity. It is a permanent solution for multicollinearity as it replaces original variables with their linear combinations which are uncorrelated and of maximum variance (Soosova, 2005). By their nature, there is no multicollinearity which exists between the principal components. The first principal component accounts for much of the variability in the data, and each principal component that follows accounts for most of the remaining variability than the subsequent ones (Andreica, 2009). It is the first few subsets of all the principal components which are used in regression, and thus reducing the dimensionality of the data without loss of information. Including all the principal components result in a model with large variance of regression coefficient estimates leading to insignificant of some of the principal components.

Principal components regression is criticised for using regressors that have no relevance with the dependent variable (Grosswindhager, 2009). Thus, there is a possibility for a less important principal component to be very important in predicting the dependent variable. More criticism is ascribed to a lack of meaning attached to the principal components as compared to their original variables (Faraway, 2002). Overall, it is concluded that principal components are meaningless unless interpreted.

Multicollinearity exists when the number of independent variables exceeds the sample size because the data tend to crowd at one place. Multicollinearity is unavoidable because most variables in a system are correlated. A typical data set is found in a financial system of companies listed on a stock exchange market. When predicting ROCE of companies listed on a stock exchange market the independent variables are often near collinear (Azhagaih and Gavoury, 2011).

Introduced by Adcock (1878) and developed by Pearson (1901), Spearman (1904) and Hotelling (1933); the first motivations for principal components analysis started in classical analytic geometry and then found applications in other fields of study (Cook, 2007). The first regression models implicating principal components were applied in calibration, chemometrics, climatology, image recognition, measurement error, and microarray data (Cook, 2007).

Recent studies of multicollinearity data have applied principal components regression (Cook, 2007). Principal components regression for reliability prediction is investigated in Shirgaokar (2009) who developed principal components regression models based on thermo-mechanical data by examining a wide range of variables. Shirgaokar (2009) then demonstrated the validity of the models predictors by using validation datasets which were not used for model development.

For linear regression, Li (2010) developed a principal components regression model using Fourier Transform Infrared (FTIR) data. Fourier Transform Infrared (FTIR) data is characterised by several hundred multicollinear variables which are more than the sample size. It is not possible to give reliable results when using the classical multiple linear regression model. Regression models developed using principal components of the variables are useful (Li, 2010). Li (2010) further illustrated the principal components regression technique using a simulated study.

Maitra and Yan (2008) discussed the computational procedures, theoretical properties and interpretations of the principal components regression algorithm in a case that the independent variables are correlated; and illustrated the procedures using a simulated data set.

Similarly, Motyka (2003) described the principal components regression algorithm to analyse financial mortgage market data, the results which showed that the technique significantly reduced the residual variance after each component was captured in the model.

The application of the principal components regression model in analyzing dendroecological data is illustrated in Fekedulegn et al. (2002). In estimating a response function, Fekedulegn et al. (2002) compared five principal components selection methods, some of which are discussed in this study. Fekedulegn et al. (2002) noted that the differences in sign, magnitude and statistically significance of the estimated coefficients were attributable to the selection method.

Ma (2007) applied the principal components analysis method to lymphoma data and used the resulting principal components in the linear regression survival model. Lymphoma data is right censored survival data with high dimensional microarray measurements (Ma, 2007). Ma (2007) compared the principal components regression approach with other model reduction techniques and found that it was relatively insensitive to the number of covariates.

Functional linear modelling has undergone the most development since the late1990s theoretically and practically (Tran, 2008). In Tran (2008), the principal components based method, as used in functional data analysis is discussed, assuming a linear regression model. An observation represents a curve instead of a point in a functional data set (Tran, 2008). Tran (2008), noted that functional data are infinite dimensional and measurement on the same curve displays high correlation invalidating the classical multiple linear regression model. In this dissertation, we address multicollinearity amongst independent variables by using a principal components regression model. The problem of non-constant variances and dependences among the residuals in classical multiple linear regression analysis is solved by using generalized least square regression. Thus, we consider generalized least square regression in our next section.

2.6 Generalized Least Squares Regression

Technical advances in regression analysis during the 1970s acknowledged and embraced the problems caused by heterogeneous of the residual term (Howarth, 2001). This led, to the introduction of generalized least square regression to overcome such problems and other techniques for fitting residuals in variables and mixture models (Howarth, 2001).

Generalized least squares regression is an extension of classical multiple linear regression. It provides a possibility for unequal residual variances and dependences between different residuals of the classical multiple linear regression model (Fox and Weisberg, 2010). In certain instances, these violations in assumptions are not corrected by basic transformations. Consequently, classical multiple linear regression often gives a false result. Thus, generalized least square estimation is intended to provide a linear model when the assumptions of homoscedasticity and independence of the residual terms fail efficaciously (Faraway, 2002). Its main thrust is to get better independent estimates of the residual variance than those for the presumed poor OLS estimates. However, the many elements in the residual variance-covariance matrix; sometimes make its estimation difficult without specifying further structure for the dependent residuals (Fox and Weisberg, 2010).

A common application of generalized least square estimation is in time series regression where it is very unlikely that residuals are independent, because observations denote different equally spaced intervals of time (Fox and Weisberg, 2010). Fox and Weisberg (2010) discussed the application of generalized least square compared to OLS. They considered the assumption of independence between different residuals of the classical multiple linear regression model, illustrating with an example on crime rate distributions and admitting inappropriate distribution modelling. Kariya and Kurata (2004) compared generalized least square and OLS methods for the simple linear regression model with an auto-regressive residual distribution, to model emission data. The study revealed that generalized least square estimates performed better than OLS estimates. In a related study, Smadi and Abu-Afouna (2012) derived generalized least square estimates and OLS estimates for the simple linear regression model with periodic auto-regressive residuals. They investigated the relative efficiency of the estimates using real data and Monte-Carlo simulation. The results showed that generalized least square estimates out-performed OLS estimates in terms of estimating the actual variances of the estimates.

Orsini et al. (2006) used generalized least square technique for trend estimation of summarized dose-response epidemiology data. In epidemiology studies the concern of the practitioner lies in constructing a variance-covariance estimate for the log relative risks; and whether the relationship between increasing levels of exposure and the risk of diseases follows a linear dose-response pattern. Orsini et al. (2006) examined many real examples citing the consequences of using different models.

Previous applications include the use of linear regression methods in the repeated measures settings. In most repeated measures designs, measurements are correlated and the dependent variable has non-constant variances because observations appear in duplicate or replicate. Generalized least square analysis allows for more flexibility in the configuration of the experiment and is helpful to repeated measures designs, because it accounts for unstructured variance-covariance between measurements than OLS (Holsclaw, 2007, p. 23). Holsclaw (2007) provided three types of the generalized least square regression method in terms of variance-covariance matrices of the residual terms, focusing on procedures for testing the models for homoscedasticity and independence of the residual distribution in repeated measures designs. The paper presented several real examples and results of simulations, evaluating the performance of the three generalized least square regression models. Given that in our data set each observational unit has a single reading of ROCE, the focus of this study is in the first and simplest type of the generalized least square method. When the assumptions are not satisfied in classical multiple linear regression analysis, we use robust maximum likelihood regression. This is considered in the next section.

2.7 Robust Maximum Likelihood Regression

Classical multiple linear regression analysis perform badly when the assumptions are violated especially the normality assumption. The presence of a heavy/fat tailed residual distribution due to contamination in the data is common. Under such conditions, robust maximum likelihood regression becomes a viable option. It is not vulnerable to unusual data, hence restrains the influence of outlying data points. Robust maximum likelihood seeks the estimated regression coefficient values that are mostly likely to have produced the data (Myung, 2003). It gives reduced weights at the tails of the distribution whilst OLS gives weight one to all observations (Howell, 2013). In this way, larger residuals caused by influential data have a smaller effect on maximum likelihood estimates than OLS estimates. This results in a more precise representation of the relationship between independent variables and dependent variable, giving a deeper insight into the system being studied. By extracting as much information as possible, practitioners make better informed decisions.

Although maximum likelihood regression is robust in the face of data contamination and demands no distributional assumptions, making it to have a wider generalizability than the aforementioned techniques, the method has no basis for testing hypotheses and thus making it possible for a rise in complexities when testing the significance of the regression coefficients (Mutan, 2004; Myung, 2003).

Robust maximum likelihood regression was introduced by R.A. Fisher (1922-1925) and was first applied by P.J. Huber, F. R. Hampel and J.W. Tukey in the context of estimating the centre of a distribution (Fox, 2002; Howell, 2013). Huber, Hampel and Tukey showed that maximum likelihood estimates are consistent and asymptotically normal under certain conditions (Fox, 2002; Howell, 2013).

More recently, Sabbagh (2003) investigated the method of maximum likelihood estimation as applied to discrete-time single-input single-output (SISO) Hammerstein models. Hammerstein models are applied to model dynamic systems having nonlinearity as its input (Sabbagh, 2003). Two maximum likelihood based identification methods to a Hammerstein model are presented. Sabbagh (2003) then demonstrated the findings using several real and Monte Carlo simulated examples.

In 2003, Myung applied the method of robust maximum likelihood to the psychological science field of study to estimate the parameters under exponential and power distributions for a multiple linear regression model. Maximum likelihood estimates of regression coefficients were derived and compared with OLS estimates. Myung (2003) showed that maximum likelihood estimates are more consistent, sufficient, and efficient compared to OLS estimates in the two distributions.

In a study performed by Chen (2002), the robust maximum likelihood regression method was used to model growth data. Growth data are time series data, and thus maximum likelihood regression was applied to model the data as they vary over time. Chen (2002) found that maximum likelihood estimates have high sample efficiency to OLS estimates when the residual terms are heteroscedastic and dependent. Thus, improving classical multiple linear regression model using maximum likelihood regression model is promising for our data set.

Lokshin and Sajaia (2004) discussed maximum likelihood method as regards to fitting the endogenous switching regression model. In this model, the switching equation sorts individuals over two different states with one regime observed (Lokshin and Sajaia, 2004). An example is presented using data obtained from the finance sector.

Karlsson (2002) presented a maximum likelihood method for Gaussian/normal distribution, Student-t distribution and Generalized Residual distribution in the context of fitting the Generalized AutoRegressive Conditional Heteroscedadticity (GARCH) model to the financial return series. The intention was to examine the robustness property of maximum likelihood in fitting the GARCH model. According to Karlsson (2002), implementing the maximum likelihood method to the GARCH model leads to two problems namely 1) the unrealistic modelling of noise by a Gaussian distribution and 2) the calculation of unobservable values from the observed sample. In the last section we give a summary of this chapter.

2.8 Summary

The literature review encapsulates:

- Measures of company performance which belong to profitability, market, operating efficiency, sales efficiency, shareholders, and growth.

- Measures of key determinants of a firm’s performance which belong to liquidity/solvency, asset utilisation, capital structure, profitability, growth ability, shareholder, size, and market.
- Potential key determinants of ROCE which are APCE, CE, D/A, D/E, DY, EPS, EY, MC, OPM, P/E, ROA, ROE and SP.
- Strengths and weaknesses/limitations of classical multiple linear regression analysis in modelling ROCE.
- Principal components regression as a technique considered to address the problem of multicollinearity among the independent variables in classical multiple linear regression analysis.
- Generalized least square regression as a technique proposed to solve the problem of non-constant variances and dependences among the residuals in classical multiple linear regression analysis.
- Robust maximum likelihood regression as a technique proposed when the assumptions are violated in classical multiple linear regression analysis.
- Strengths and weaknesses/limitations of principal components regression, generalized least squares regression, and robust maximum likelihood regression in modelling ROCE.
- Applications of principal components regression, generalized least squares regression, and robust maximum likelihood regression by various scholars.
- Possibilities for improving classical multiple linear regression model using principal components regression, generalized least squares regression, and robust maximum likelihood regression models for our data set.

The next chapter provides statistical methods for modelling ROCE of companies listed on the JSE.

CHAPTER 3

MATERIALS AND METHODS

3.0 Introduction

In addressing the research problem indicated in this study, secondary data on the audited performance from JSE for the period 2010 was used to demonstrate various statistical techniques. Statistical software packages used were SPSS and R. Various statistical methods were assessed namely classical multiple linear regression analysis, principal components regression, generalized least squares regression, and robust maximum likelihood regression. The following section gives a description of the data set.

3.1 The Data Set

The secondary data used in this study is on the audited performance results of companies listed on the JSE at annual company financial statements for the period 2010 (Johannesburg Stock Exchange, 2010). In general, the financial statements provide information to the potential and existing investors. The information assists investors in decision making with respect to buying and selling of shares. The data set was sourced from the McGregor BFA database (http://www.mcgregorbfa.com). A total of 387 listed companies constituted the data set. However, some of the data on the companies contained missing values or incomplete data. Because of missing values and incomplete data, 30 companies were deleted from the data set. Thus, the cleaned data set consisted of 357 listed companies on 14 variables. The variables considered are:

Dependent variable: Return on Capital Employed (ROCE).

Independent variables: Assets per Capital Employed (APCE), Capital Employed (CE), Debt to Assets (D/A), Debt to Equity (D/E), Dividend Yield (DY), Earnings per Share (EPS), Earnings Yield (EY), Market Capitalization (MC), Operating Profit Margin (OPM), Price to Earnings (P/E), Return on Assets (ROA), Return on Equity (ROE), and Share Price (SP).

Table 3.1 presents these variables and codes associated with them in the subsequent discussions.

Table 3.1. Variables and codes associated with the variables

ROCE	APCE	CE	D/A	D/E	DY	EPS	EY	MC	OPM	P/E	ROA	ROE	SP
Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}

The ROCE data has positive and negative values. The values are percentages. To address the problem of losing data/information, two separate analyses were done. To avoid statistical problems when using negative values, modulus transformation (Hair et al., 2010) was used for the negative ROCE data analysis. There are 278 companies in the data set for positive ROCE and 79 companies in the data set for negative ROCE. The classical multiple linear regression analysis procedures are discussed in the following section.

3.2 Classical Multiple Linear Regression Model

3.2.1 Model

The classical multiple linear regression model takes the form (Shaffer, 2007)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + \varepsilon_i, \text{ with } i = 1,2,\ldots,n \qquad (3.1)$$

where Y_i is the value of the dependent variable in the ith observation, β_0 is the intercept parameter representing the value when all independent variables are set to zero, β_j is the slope parameter representing the effect in the dependent variable Y associated with a unit change in X_j holding other independent variables constant, X_{ij} is the value of the jth independent variable in the ith observation, ε_i is a random error term of the ith observation, k is the number of independent variables, and n is the number of observations. According to Person (2011), each independent variable in the model requires 10 to 20 observations. To handle the problem of the small data set (n=79), correlation matrix was used to remove some of the independent variables which had significance correlations with the dependent variable. Correlation matrix provides the evidence regarding the association between the dependent variable and independent variables, and also among the independent variables. A conservative significance level 5 % was used in determining whether the correlation was significant or not.

We write (3.1) in matrix form as

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (3.2)$$

where, $\boldsymbol{Y} = [Y_1, Y_2, \ldots, Y_n]'$ is an *nx1* data vector and an *n* x (*k+1)* incidence matrix

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}$$

consisting of the independent variables whose (i, j) th element is the value of the ith observation for the jth independent variable. The columns of $\boldsymbol{X}$ are linearly independent with $rank(\boldsymbol{X}) = k + 1$, $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \dots, \beta_k]'$ is a *(k+1) x 1* parameter vector and $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]'$ is an *nx1* vector of random error.

3.2.2 Estimation

Various estimation methods exist, but discussed here is the ordinary least square (OLS) estimation procedure. The estimates of regression parameters in (3.1) are obtained using OLS which operates by minimizing the residual sum of squares. Thus

$$\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{3.3}$$

is minimized where $\hat{Y}_i$ is the ith predicted value (Faraway, 2002). We write (3.3) in matrix notation as

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \left(\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right)'\left(\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right) = \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right) = \boldsymbol{Y}'\boldsymbol{Y} - 2\hat{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{Y} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}. \tag{3.4}$$

Consider partial derivatives with respect to components of $\hat{\boldsymbol{\beta}}$ in (3.4) and equate results to zero to obtain the OLS estimates of parameter vector. Solving the following normal equations

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{0} \tag{3.5}$$

for $\hat{\boldsymbol{\beta}}$ gives a vector of OLS estimated regression parameter values

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{3.6}$$

We obtain $\hat{\boldsymbol{\beta}}$ explicitly from the data. The fitted OLS regression function is

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{3.7}$$

with the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ given by

$$\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{3.8}$$

where, σ^2 is estimated by

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1} = \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n-k-1}. \quad (3.9)$$

3.2.3 Variable Selection Procedure

Three stepwise regression methods namely 1) Backward Elimination (BE), 2) Forward Selection (FS), and 3) Stepwise directed search are widely discussed in many textbooks [see, for example, Bowerman and O'Connel (1990), Draper and Smith (1998), Faraway (2002), Ryan (2008)]. Stepwise regression is vulnerable because slight variation in the data causes selection of some variables in place of others and fails to consider the underlying knowledge of the process. Other relevant variables find chance to be added in the model.

The BE method enters all the independent variables and finds the contribution of each variable by checking p-to-remove or partial regression sum of squares (Bowerman and O'Connel, 1990). The FS method enters the independent variables one after the other until all variables significantly explain the dependent variable (Bowerman and O'Connel, 1990). At each stage of stepwise directed search, the variable with highest significant partial correlation coefficient is added to the model, and p-to- remove for all variables now in the model is computed to check if any of the variables previously added contribute insignificantly to the regression and be removed (Bowerman and O'Connel, 1990). In the development of the model, stepwise directed search was preferred due to its property that allows for reassessment of a dropped variable.

3.2.4 Model Assumptions

Model (3.1) is based on normality, independence, constant variances and linearity assumptions. The validity of these assumptions is assessed by conducting analysis of residuals, $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$. For direct comparison purposes, standardized residuals $r_i = \frac{\hat{\varepsilon}_i - E(\hat{\varepsilon})}{s.e\ (\hat{\varepsilon})}$ apply, instead of the raw values due to different measurement units (Hair et al., 2010). The residual plots are commonly used and have the advantage of showing a variety of features quickly and clearly (Ryan, 2008). We assessed the data on normality, constant variance and independence assumptions using residuals from a classical multiple linear regression analysis.

Normality

The residuals are assumed to be normally distributed, $\hat{\boldsymbol{\varepsilon}} \sim N(\mathbf{0},\ \sigma^2 \boldsymbol{I}_n)$. Testing for non-normality of the residuals was done using normal quantile-quantile (Q-Q) plot, normal probability plot (NP-P), frequency distribution plot (histogram), coefficients of skewness and kurtosis values, Kolmogorov-Smirnov test, and Shapiro-Wilk test (Hair et al., 2010). A straight 45 degree line is formed in normal Q-Q plot and normal P-P plot if the normality assumption is satisfied. Skewness and kurtosis values equal or close to 0 and 3 respectively, show that the normality assumption is fulfilled. Violation of the normality assumption (p-value< 5 %) was tested using Kolmogorov-Smirnov and Shapiro-Wilk.

Constant Variance

The homogeneity of variance (Ho: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2 = \sigma^2$) against Ha: At least two variances are not equal was conducted. A systematic pattern of r_i versus standardized $\hat{Y}_i$ plot instead of a random scatter of points shows violation of the assumption.

Independence

Testing of no autocorrelation among the residuals was conducted using Durbin-Watson (D-W) (Fox and Weisberg, 2010). The terms $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$ are mutually independent if $E(\hat{\varepsilon}_t \hat{\varepsilon}_{t-m}) = 0$, for m$\neq 0$. The D-W test statistic is given by

$$d = \frac{\sum_{t=2}^{n}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{n}\hat{\varepsilon}_t^2}. \quad (3.10)$$

The decision criteria are shown in Table 3.2, where d_L and d_U are the lower and upper bounds for the critical values at 5 % level of significance, respectively. Rejection of the null hypothesis means that there is autocorrelation. Doubt is thus cast on the fitted model and the data is reconsidered in light of this new information.

Table 3.2. Decision criteria for testing autocorrelation of residuals

Positive autocorrelation	Negative autocorrelation
If $d < d_L$ reject null hypothesis	If $d > 4 - d_L$ reject null hypothesis
If $d > d_U$ do not reject null hypothesis	If $d < 4 - d_U$ do not reject null hypothesis
If $d_L \leq d \leq d_U$ test is inconclusive	If $4 - d_U \leq d \leq 4 - d_L$ test is inconclusive

The distribution of F and T test statistics depends on these assumptions. When the assumptions are violated the test statistics values become inflated. The statistics cease to be valid (Holsclaw, 2007). This problem was resolved by applying suitable transformations. Common transformation methods are:

1) Logarithmic.

2) Square root.

3) Inverse.

4) Box-Cox.

The Box-Cox transformation is given by the formula

$$y_i^{\lambda} = \begin{cases} (y_i^{\lambda} - 1)/\lambda;\ \lambda \neq 0 \\ log y_i;\ \lambda = 0 \end{cases} \quad (3.11)$$

where λ is a transformation parameter estimate which is chosen by maximum likelihood (Faraway, 2002). We used the logarithm (base 10) transformation in this study. Often, however, the violations in assumptions are not corrected by basic transformations. Under these circumstances some alternative methods to classical multiple linear regression become viable options. In this dissertation, we investigated generalized least square and robust maximum likelihood regression procedures. These are explained in sections 3.4 and 3.5 respectively.

3.2.5 Tests of Statistical Significance

The procedure followed in classical multiple regression analysis involves 1) determining if the independent variables affect the dependent variable, 2) how each independent variable affects the dependent variable given that it does affect, and 3) developing simple prediction model. Part (1) corresponds to

Ho: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ against Ha: $\beta_j \neq 0$ for at least one j. The test was performed through an analysis of variance (ANOVA) table. In general, a ANOVA table format is shown in Table 3.3.

Table 3.3. Analysis of variance for fitting classical multiple linear regression

Source of Variation	**DF**	**SS**	**MS**	**F-statistic**	**P-value**
Regression	k	SSR	$MSR = \frac{SSR}{k}$	$F(R) = \frac{MSR}{MSE}$	Sig if <0.05
Residual	$n - k - 1$	SSE	$MSE = \frac{SSE}{n-k-1}$		
Total	$n - 1$	SST			

SS is sum of squares term, DF is degrees of freedom for SS term and MS is mean of square term. The F test statistic for the hypothesis is

$$\mathrm{F(R)} = \frac{MSR}{MSE} \sim F_{k,n-k-1}. \quad (3.12)$$

We reject Ho in favour of Ha if p-value of the F-statistic is less than ∝-level. We conclude that at least one of the independent variables is related to the dependent variable. A ∝-level 5 % was used in this dissertation.

Part (2) corresponds to testing for significance of the individual regression coefficients. The hypothesis for testing the significance of β_j is presented as

Ho : $\beta_j = 0$ against

Ha : $\beta_j \neq 0$.

The test was done using a T test. A T test statistic for the hypothesis is given by

$$T = \frac{\hat{\beta}_j}{se\ (\hat{\beta}_j)} \sim T_{n-k-1} \tag{3.13}$$

where se $(\hat{\beta}_j)$ is the standard error of $\hat{\beta}_j$. Rejection of Ho means that $\hat{\beta}_j$ is significantly different from zero and thus X_j is used as a predictor of Y. Non rejection of Ho means that X_j is not a significant explanatory variable. The variable is removed and another model is estimated without this variable.

3.2.6 Goodness of Fit Tests

The R^2 statistic was used to assess the goodness of fit of the model for the data. The formula for R^2 is

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{SSE}{SST} = \frac{SSR}{SST} \tag{3.14}$$

and its value ranges from 0 to 1 inclusively (Bowerman and O'Connel, 1990; Draper and Smith, 1998). A model with the R^2 value close to 1 for the data, was regarded the best model because the independent variables explain the variation in the dependent variable almost perfectly well (Bowerman and O'Connel, 1990). However, an additional variable to the model will always increase R^2, irrespective of whether the added variable is statistically significant or not (Draper and Smith, 1998). Thus, it is possible for a model that has a large value of R^2 to give poor estimates of the dependent variable. To solve this problem, an adjusted R^2 statistic was used. It is given by

$$R_{adj}^2 = 1 - \frac{MSE}{MST} = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} = 1 - \frac{n-1}{n-k-1}(1 - R^2) \quad (3.15)$$

and also its value ranges from 0 to 1 inclusively (Bowerman and O'Connel, 1990; Draper and Smith, 1998). If variables of no explanatory power are added to the model, the value of R_{adj}^2 does not increase (Draper and Smith, 1998). The MSE is a function of R^2. It is expressed as $MSE = (1 - R^2)\hat{\sigma}_Y^2$ where $\hat{\sigma}_Y^2$ is an estimate of the variance of Y and is obtained by using the formula

$$\hat{\sigma}_Y^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1} = \frac{SST}{n-1}. \quad (3.16)$$

A MSE of 0 means data points lie on the regression plane/hyperplane (Bowerman and O'Connel, 1990). A low MSE value was considered to indicate a reliable model.

3.2.7 Multicollinearity

The presence of multicollinearity increases the standard errors of the estimated regression coefficients, causing the p-values of the test Ho:$\beta_j = 0$ to be greater than 5 % (Shaffer, 2007). This makes it difficult to obtain accurate estimates of the individual effects of the independent variables, hence reducing the degree of confidence that one can place in the coefficient estimates values. It becomes difficult to assess the individual effects because the estimated regression coefficient values have dubious interpretations (Bowerman and O'Connel, 1990). Incorrect conclusions are drawn about the effect of the independent variables on the model. Thus, multicollinearity must be eliminated.

Variance Inflation Factor (VIF) statistic was used to identify multicollinearity. Each independent variable is regressed against other independent variables in the model. The VIF for X_j is

$$VIF_j = \frac{1}{(1-R_j^2)} = (1-R_j^2)^{-1} \tag{3.17}$$

where R_j^2 is the coefficient of multiple determinations from regressing X_j on the other independent variables (Faraway, 2002). A high VIF_j (>10) implies high levels of multicollinearity for the effect of X_j (Hair et al., 2010). Multicollinearity was eliminated by dropping variables with VIFs>10. However, discarding some of the highly correlated variables leads to loss of information. Bowerman and O'Connel (1990), and Soosova (2005) proposed several methods to avoid this problem. The methods include:

1) Use of a different data set which is often difficult to get.
2) Employ ridge regression.
3) Convert the variables into uncorrelated factors using principal components analysis.

In this dissertation, we focused on principal components regression. Section 3.3 gives a full discussion of the procedure.

3.2.8 Influential Observations

The influence of a point is assessed by examining the effects it produces on the fit when it is removed in the fitting process (Ryan, 2008). We used Cook's Distance to assess influential observations. The statistic is sensitive to changes in the fitted model if observations are removed (Hussien, 2010). Cook's Distance is given as

$$C_i = \frac{(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)})}{k\hat{\sigma}^2} \tag{3.18}$$

where, $\hat{\boldsymbol{\beta}}$ is the vector of estimated regression coefficient values obtained using all the observations and $\hat{\boldsymbol{\beta}}_{(i)}$ is the vector of estimated regression coefficient values which results from removing the ith observation (Faraway, 2002). Thus, Cook's Distance reveals which observations are influential in the sense that they affect the fitted equation's coefficients. If a point is influential, its removal causes large changes and the value of C_i will be large. An observation with C_i greater than $\frac{4}{n}$ is regarded as an influential observation (Bowerman and O'Connel, 1990). A scatter plot of C_i versus i was used to identify influential observations. Influential points bias the prediction and distort the significance of estimated regression coefficients by pushing or pulling the estimated regression plane/hyperplane in a certain direction (Bowerman and O'Connel, 1990). In the construction of the model, conclusions were not based on influential points.

Identifying Outliers and Leverage Points as Influential Observations

A $|(standardized\ residual)_i| > 2$ indicates that the i^{th} observation is an outlier (Hair et al., 2010). If an outlier affects the values of the estimated coefficients of the model, it becomes an influential point and has a major influence on the position of the fitted model on its own territory (Hussien, 2010). The fact that an observation is a large outlier is generally not good, and at the same time it does not necessarily mean that the observation is influential in fitting the chosen model. Outliers that occurred due to gross errors during the recording of the data were deleted. However, outliers that are genuine observations suggest considerable attention (Hussien, 2010). These data points were followed and examined.

We refer to leverage points with regard to X and not Y. $\boldsymbol{H}$ is $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and is the hat matrix. The ith diagonal element of $\boldsymbol{H}$ is H_{ii}. H_{ii} is the leverage of the ith data point, and lies between 0 and 1 inclusively (Hair et al., 2010). Points with $H_{ii} > \frac{2k+2}{n}$ are points with high leverage (Bowerman and O'Connel, 1990). However, the points may not be influential observations. An observation with $|standardized\ residual| > 2$ and leverage $> \frac{2k+2}{n}$ shows considerable influence on the model fit (Hair et al., 2010). Outliers and high leverage points were scrutinized carefully to see if they are influential points.

3.2.9 Model Validity

The performance of the prediction model was assessed by fitting it in a new data called validation data. The regression parameter estimates, standard errors of parameter estimates, R^2, R^2_{adj}, VIF, Mean Square Residual of the Prediction ($MSEP$), and p-values were calculated for the validation data. Closeness of their values to those of the training data means that the prediction model is satisfactory/accurate (Hair et al., 2010). The $MSEP$ is computed as

$$MSEP = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}. \tag{3.19}$$

The Validation Data Set

The validation data used in this study is on the audited performance results of companies listed on the JSE at annual company financial statements for the period 2011 (Johannesburg Stock Exchange, 2011). The data set was sourced from the McGregor BFA database (http://www.mcgregorbfa.com). The cleaned validation data set consisted of 322 listed companies. The validation data set was divided into positive ROCE and negative ROCE. There are 239 companies in the validation data set for positive ROCE and 83 companies in the validation data set for negative ROCE. The classical multiple linear regression procedures were implemented using

SPSS software. We utilized principal components regression to handle the problem of multicollinearity in the data.

3.3 Principal Components Regression

Principal components analysis was applied first on the data and then OLS regression model was constructed with the principal components. We consider a vector of k random variables $\boldsymbol{x}=(x_1, x_2, \dots, x_k)'$ and a coefficient vector $\boldsymbol{\vartheta}_1 = (\vartheta_{11}, \vartheta_{12}, \dots, \vartheta_{1k})'$. Our aim is to find $\boldsymbol{\vartheta}_1$ such that $Var(\boldsymbol{\vartheta}_1'\boldsymbol{x}) = \boldsymbol{\vartheta}_1'\boldsymbol{V}\boldsymbol{\vartheta}_1$ is maximum subject to the normalization constraint $\boldsymbol{\vartheta}_1'\boldsymbol{\vartheta}_1 = 1$, where $\boldsymbol{V}$ is a *k x k* positive semi definite symmetric variance-covariance matrix of $\boldsymbol{x}$ and

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0 \tag{3.20}$$

are the k eigenvalues of $\boldsymbol{V}$ (Faraway, 2002; Sopipan et al., 2012). Using the technique of Lagrange multipliers for the optimization problem we maximize

$$\boldsymbol{\vartheta}_1'\boldsymbol{V}\boldsymbol{\vartheta}_1 - \lambda(\boldsymbol{\vartheta}_1'\boldsymbol{\vartheta}_1 - 1) \tag{3.21}$$

where λ is a Lagrange multiplier (Sopipan et al., 2012). Differentiating (3.21) with respect to $\boldsymbol{\vartheta}_1$ and equating to the *k x 1* zero vector gives

$$\boldsymbol{V}\boldsymbol{\vartheta}_1 - \lambda\boldsymbol{\vartheta}_1 = \boldsymbol{0} \text{ or } (\boldsymbol{V} - \lambda\boldsymbol{I}_k)\boldsymbol{\vartheta}_1 = \boldsymbol{0} \tag{3.22}$$

where $\boldsymbol{I}_k$ is the *k x k* identity matrix. Given that $\boldsymbol{\vartheta}_1 \neq \boldsymbol{0}$, there exists a solution only if $|\boldsymbol{V} - \lambda\boldsymbol{I}_k| = 0$ such that if λ is an eigenvalue of $\boldsymbol{V}$ then $\boldsymbol{\vartheta}_1$ is its corresponding eigenvector. Consequently, λ is an eigenvalue of $\boldsymbol{V}$ and $\boldsymbol{\vartheta}_1$ is its corresponding eigenvector $\boldsymbol{\vartheta}_1$. We maximize

$$\boldsymbol{\vartheta}_1'\boldsymbol{V}\boldsymbol{\vartheta}_1 = \boldsymbol{\vartheta}_1'\lambda\boldsymbol{\vartheta}_1 = \lambda\boldsymbol{\vartheta}_1'\boldsymbol{\vartheta}_1 = \lambda \tag{3.23}$$

and by (3.22), λ has to be as large as possible. Thus $Var(\boldsymbol{\vartheta}_1'\,\boldsymbol{x}) = \boldsymbol{\vartheta}_1'\lambda\boldsymbol{\vartheta}_1 = \lambda = \lambda_1$, the largest eigenvalue by (3.20) corresponding to the eigenvector $\boldsymbol{\vartheta}_1$. The first principal component of $\boldsymbol{x}$ is $p_1 = \boldsymbol{\vartheta}_1'\boldsymbol{x} = \sum_{h=1}^{k} \vartheta_{1h}\, x_h$. The next step is to maximize

$Var\ (\boldsymbol{\vartheta}_2'\ \boldsymbol{x})= \boldsymbol{\vartheta}_2'\boldsymbol{V}\boldsymbol{\vartheta}_2$ subject to $\boldsymbol{\vartheta}_2'\boldsymbol{\vartheta}_2 = 1$ and $\boldsymbol{\vartheta}_2'\boldsymbol{\vartheta}_1 = 0$ where $\boldsymbol{\vartheta}_2 = (\vartheta_{21}, \vartheta_{22}, \dots, \vartheta_{2k})'$ (Faraway, 2002; Sopipan et al., 2012). Applying the method of Lagrange multipliers for the optimization problem we maximize

$$\boldsymbol{\vartheta}_2'\boldsymbol{V}\boldsymbol{\vartheta}_2 - \lambda(\boldsymbol{\vartheta}_2'\boldsymbol{\vartheta}_2 - 1) - \theta\boldsymbol{\vartheta}_2'\boldsymbol{\vartheta}_1 \quad (3.24)$$

where λ and θ are Lagrange multipliers (Sopipan et al., 2012). Differentiating (3.24) with respect to $\boldsymbol{\vartheta}_2$ and equating to the *k x 1* zero vector we get

$$\boldsymbol{V}\boldsymbol{\vartheta}_2 - \lambda\boldsymbol{\vartheta}_2 - \theta\boldsymbol{\vartheta}_1 = \boldsymbol{0} \quad (3.25)$$

Multiplying (3.25) by $\boldsymbol{\vartheta}_1'$ from the left gives

$$\boldsymbol{\vartheta}_1'\boldsymbol{V}\boldsymbol{\vartheta}_2 - \lambda\boldsymbol{\vartheta}_1'\boldsymbol{\vartheta}_2 - \theta\boldsymbol{\vartheta}_1'\boldsymbol{\vartheta}_1 = 0 \quad (3.26)$$

which results in $\theta = 0$, because $\boldsymbol{\vartheta}_1'\boldsymbol{V}\boldsymbol{\vartheta}_2 = 0$ and $\boldsymbol{\vartheta}_1'\boldsymbol{\vartheta}_2 = 0$ from Cov $(\boldsymbol{\vartheta}_1'\ \boldsymbol{x}, \boldsymbol{\vartheta}_2'\ \boldsymbol{x}) = \boldsymbol{\vartheta}_1'\boldsymbol{V}\boldsymbol{\vartheta}_2 = \mathrm{Cov}\ (\boldsymbol{\vartheta}_2'\ \boldsymbol{x}, \boldsymbol{\vartheta}_1'\ \boldsymbol{x}) = \boldsymbol{\vartheta}_2'\boldsymbol{V}\boldsymbol{\vartheta}_1 = \boldsymbol{\vartheta}_2'\lambda_1\boldsymbol{\vartheta}_1 = \lambda_1\boldsymbol{\vartheta}_2'\boldsymbol{\vartheta}_1 = \lambda_1\boldsymbol{\vartheta}_1'\boldsymbol{\vartheta}_2 = 0$. Thus,

$$\boldsymbol{V}\boldsymbol{\vartheta}_2 - \lambda\boldsymbol{\vartheta}_2 = \boldsymbol{0} \text{ or } (\boldsymbol{V} - \lambda\boldsymbol{I}_k)\boldsymbol{\vartheta}_2 = \boldsymbol{0} \quad (3.27)$$

making λ an eigenvalue of $\boldsymbol{V}$ and $\boldsymbol{\vartheta}_2$ its corresponding eigenvector. Once more, $\boldsymbol{\vartheta}_2'\lambda\boldsymbol{\vartheta}_2 = \lambda$ hence by (3.27), λ has to be as large as possible. If we assume that $\boldsymbol{V}$ does not have recurring eigenvalues, then $\lambda \neq \lambda_1$. However, if they are the same then $\boldsymbol{\vartheta}_2 = \boldsymbol{\vartheta}_1$, defying the constraint $\boldsymbol{\vartheta}_2'\boldsymbol{\vartheta}_1 = 0$. Thus, λ is the second largest eigenvalue of $\boldsymbol{V}$, and $\boldsymbol{\vartheta}_2$ is its corresponding eigenvector. The second principal component of $\boldsymbol{x}$ with $Var(\boldsymbol{\vartheta}_2'\boldsymbol{x}) = \lambda_2$ is $p_2 = \boldsymbol{\vartheta}_2'\boldsymbol{x} = \sum_{h=1}^{k} \vartheta_{2h}\, x_h$ and the process goes on. As the algorithm continues and since $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$, we find principal components of maximum variation orthogonal to the principal components we have obtained before, until the total variance has been decomposed. In general, the jth principal component of $\boldsymbol{x}$ with $Var(\boldsymbol{\vartheta}_j'\boldsymbol{x}) = \lambda_j$, where λ_j is the jth largest eigenvalue of $\boldsymbol{V}$ and $\boldsymbol{\vartheta}_j = (\vartheta_{j1}, \vartheta_{j2}, \dots, \vartheta_{jk})'$ is its corresponding eigenvector is $p_j = \boldsymbol{\vartheta}_j'\boldsymbol{x} = \sum_{h=1}^{k} \vartheta_{jh}\, x_h$. The original set of variables $x_1,\ x_2, \dots, x_k$ is changed to a new set of uncorrelated variables $p_1,\ p_2, \dots,\ p_k$. Kaiser-Meyer Olkin (KMO) index and Bartlett's test of sphericity were used to assess the sampling adequacy of how the correlations are for principal components analysis. For an optimal solution, KMO index ≥ 0.50 is considered adequate (Ul-Saufie et al., 2011). A p-value<5 % level of significance of Bartlett's test of sphericity shows that there are sufficient correlations for principal components analysis (Ul-Saufie et al., 2011). However, Bartlett's test of sphericity tests for correlation among the independent variables, but fail to identify the specific variables involved.

To construct the principal components regression model, we assume that the independent variables in the classical multiple linear regression model (3.2) have been mean centred and standardized (Grosswindhager, 2009). This ensures the equivalence of $\frac{X'X}{n-1}$ to the correlation matrix of the independent variables. The values of the principal component for each observation are given by

$$\boldsymbol{P} = \boldsymbol{XU} \tag{3.28}$$

where, the (i,j)th element of $\boldsymbol{P}$ is the value of the ith observation for the jth principal component and $\boldsymbol{U}$ is *k x k* loading matrix where the unit normalised eigenvector of $\boldsymbol{X'X}$ are the columns of $\boldsymbol{U}$ (Fekedulegn et al., 2002). This means $\boldsymbol{U}$ is an orthonormal matrix. The classical multiple linear regression model (3.2) is written as (Grosswindhager, 2009)

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} = \boldsymbol{XUU'\beta} + \boldsymbol{\varepsilon} = \boldsymbol{PU'\beta} + \boldsymbol{\varepsilon} = \boldsymbol{P\delta} + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{U'\beta} = \boldsymbol{\delta} \tag{3.29}$$

In this way, the principal components take the place of the independent variables in the classical multiple linear regression model. In (3.29), OLS is used to find $\hat{\boldsymbol{\delta}}$ as follows

$$\hat{\boldsymbol{\delta}} = (\boldsymbol{P'P})^{-1}\boldsymbol{P'Y} = \left((\boldsymbol{XU})'(\boldsymbol{XU})\right)^{-1}(\boldsymbol{XU})'\boldsymbol{Y} = (\boldsymbol{U'X'XU})^{-1}\boldsymbol{U'X'Y} =$$

$$\boldsymbol{U}^{-1}(\boldsymbol{X'X})^{-1}(\boldsymbol{U}')^{-1}\boldsymbol{U'X'Y} = \boldsymbol{U}^{-1}(\boldsymbol{X'X})^{-1}\boldsymbol{X'Y} = \boldsymbol{U}^{-1}\hat{\boldsymbol{\beta}}. \quad (3.30)$$

Equation (3.30) implies that

$$\hat{\boldsymbol{\beta}} = \boldsymbol{UU}^{-1}\hat{\boldsymbol{\beta}} = \boldsymbol{U}\hat{\boldsymbol{\delta}}. \quad (3.31)$$

This is as good as finding $\hat{\boldsymbol{\beta}}$ in the classical multiple linear regression model (3.2). However, the relevance of principal components regression is seen when independent variables are involved in multicollinearity in the classical multiple linear regression model. It is possible to find better stable estimates of $\boldsymbol{\beta}$ if we remove a subset of the principal components with small variances (Grosswindhager, 2009). From (3.31), $\hat{\boldsymbol{\beta}} = \boldsymbol{U}\hat{\boldsymbol{\delta}} = \boldsymbol{U}(\boldsymbol{P'P})^{-1}\boldsymbol{P'Y} = \boldsymbol{U}\left((\boldsymbol{XU})'(\boldsymbol{XU})\right)^{-1}(\boldsymbol{XU})'\boldsymbol{Y} = \boldsymbol{UD}^{-2}(\boldsymbol{XU})'\boldsymbol{Y}$

$$= \boldsymbol{UD}^{-2}\boldsymbol{U'X'Y} = \boldsymbol{D}^{-2}\boldsymbol{UU'X'Y} = \sum_{j=1}^{k} d_j^{-2}u_j\, u_j'\boldsymbol{X'Y} = \sum_{j=1}^{k}\frac{u_j u_j'}{d_j^2}\boldsymbol{X'Y} \quad (3.32)$$

where d_j^2 and u_j are the jth diagonal of $\boldsymbol{D}^2$ and jth column of $\boldsymbol{U}$, respectively. The jth diagonal of $\boldsymbol{D}$ is the jth largest eigenvalue of $(\boldsymbol{X'X})^{-1}$.

$$\begin{aligned} \text{Var}\,(\hat{\boldsymbol{\beta}}) &= \text{Var}\,(\boldsymbol{U}\hat{\boldsymbol{\delta}}) = \text{Var}\,(\boldsymbol{U}(\boldsymbol{P'P})^{-1}\boldsymbol{P'Y}) \\ &= \boldsymbol{U}(\boldsymbol{P'P})^{-1}\boldsymbol{P'}Var(\boldsymbol{Y})(\boldsymbol{U}(\boldsymbol{P'P})^{-1}\boldsymbol{P'})' \\ &= \sigma^2\boldsymbol{U}(\boldsymbol{P'P})^{-1}\boldsymbol{P'P}(\boldsymbol{P'P})^{-1}\boldsymbol{U'} \\ &= \sigma^2\boldsymbol{U}(\boldsymbol{P'P})^{-1}\boldsymbol{U'} = \sigma^2\boldsymbol{UD}^{-2}\boldsymbol{U'} \\ &= \sigma^2\sum_{j=1}^{k} d_j^{-2}u_j\, u_j' = \sigma^2\sum_{j=1}^{k}\frac{u_j u_j'}{d_j^2} \end{aligned} \quad (3.33)$$

Equation (3.33) shows that independent variables with large coefficients in any of the principal components with small eigenvalues have large variance in the elements of $\hat{\boldsymbol{\beta}}$. This problem is solved by removing terms in (3.32) that relate to a small d_j^2. This gives $\hat{\boldsymbol{\beta}} = \sum_{j=1}^{t}\frac{u_j u_j'}{d_j^2}\boldsymbol{X'Y}$ where t<k.

A variety of selection criteria have been developed to determine the number of principal components t. We restricted to Kaiser-Guttman of eigenvalues and scree test. The Kaiser-Guttman criterion stipulates that principal components which have eigenvalues ≥1 bring more information than the original variables, and are considered important (Andreica, 2009). However, this method is less useful with smaller numbers of variables. A Scree test is performed by plotting eigenvalues with their corresponding principal components. The point where the plot levels off indicates that eigenvalues beyond this point are small compared to the ones before them, and the corresponding principal components are not retained (Andreica, 2009). $|Loadings| \geq 0.5$ were considered very significant in the principal component retained (Hair et al., 2010). Principal components regression was performed in SPSS. Generalized least square regression was used to assess heteroscedasticity and dependences in the data.

3.4 Generalized Least Squares Regression

Generalized least square regression has a model of the form (3.2) (Waterman, 2002). However, it has variation on the variance-covariance structure of the residual vector from classical multiple linear regression. The vector of residuals is assumed to have the following *n x n* symmetric, non-singular and positive definite variance-covariance matrix (Holsclaw, 2007)

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{bmatrix}. \quad (3.34)$$

In Σ, $\sigma_i^2 = Var(\hat{\varepsilon}_i) = Cov(\hat{\varepsilon}_i, \hat{\varepsilon}_i)$ and $\rho_{il}\sigma_i\sigma_l = \sigma_{il} = Cov(\hat{\varepsilon}_i, \hat{\varepsilon}_l)$ where, ρ_{il} is the correlation coefficient of $\hat{\varepsilon}_i$ and $\hat{\varepsilon}_l$ for $i \neq l$ and σ_i is the square root of σ_i^2 (Holsclaw, 2007). Equation (3.34) depicts that different diagonal and nonzero off-diagonal entries in Σ, correspond to non-constant residual variances and dependent residuals respectively (Fox and Weisberg, 2010). This implies that the residuals are independent and heteroscedastic when Σ is diagonal (Fox and Weisberg, 2010). In this situation weighted generalized least square regression is used. We write $\Sigma = diagonal\left(\frac{1}{z_1}, \ldots, \frac{1}{z_n}\right)$, where the z_i are the weights (Faraway, 2002). Thus $\boldsymbol{S} = diagonal\left(\sqrt{\frac{1}{z_1}}, \ldots, \sqrt{\frac{1}{z_n}}\right)$, with $\Sigma = \boldsymbol{SS}'$ where $\boldsymbol{S}$ is a triangular matrix using Choleski Decomposition (Faraway, 2002). So $\sqrt{z_i}X_i$ is regressed on $\sqrt{z_i}Y_i$, where the column of ones in $\boldsymbol{X}$ is replaced with $\sqrt{z_i}$ (Faraway, 2002). Observations with low and high variability get high and low weight, respectively.

Generalized least square estimates of $\boldsymbol{\beta}$ are obtained by minimizing the generalized sum of squares

$$\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)' \Sigma^{-1} \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \qquad (3.35)$$

over all $\widehat{\boldsymbol{\beta}}$ (Waterman, 2002). Taking the derivative of (3.35) with respect to the components of $\widehat{\boldsymbol{\beta}}$ and requiring this expression to be *(k+1) x 1* zero vector, gives the generalized normal equations

$$\boldsymbol{X}' \Sigma^{-1} \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}' \Sigma^{-1} \boldsymbol{Y} = \boldsymbol{0}. \qquad (3.36)$$

By solving (3.36), generalized least square (GLS) estimates are

$$\widehat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{X}' \Sigma^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}' \Sigma^{-1} \boldsymbol{Y}. \qquad (3.37)$$

The fitted values are

$$\widehat{\boldsymbol{Y}}_{GLS} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{GLS} = \boldsymbol{X}(\boldsymbol{X}' \Sigma^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}' \Sigma^{-1} \boldsymbol{Y} \qquad (3.38)$$

and the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}_{GLS}$ is

$$(\boldsymbol{X}' \Sigma^{-1} \boldsymbol{X})^{-1}. \qquad (3.39)$$

In an application Σ is often not known and is estimated. Consider that $Var(\boldsymbol{Y}) = \sigma^2\boldsymbol{A} = \Sigma$ is known up to a scalar σ^2 (Waterman, 2002). In this scenario, σ^2 is estimated by

$$\hat{\sigma}^2_{GLS} = \frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS})'\boldsymbol{A}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS})}{n-k-1}. \quad (3.40)$$

Generalized least square regression was done using R and SPSS. The R code for generalized least square regression is given in Appendix C. Robust maximum likelihood regression was employed to handle the problem of contamination in the data.

3.5 Robust Maximum Likelihood Regression

The robust maximum likelihood regression model has the form (Koller, 2007)

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad (3.41)$$

where $y_i, \boldsymbol{x}_i'$ and ε_i are the ith row of $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{\varepsilon}$, respectively. Robust maximum likelihood estimates are the answer of the regression coefficients for which

$$\sum_{i=1}^{n} \rho\left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}_s}\right) = \sum_{i=1}^{n} \rho\left(\frac{y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}}{\hat{\sigma}_s}\right) \quad (3.42)$$

is minimum, where $\rho(\mathrm{u})$ is an objective function and $\hat{\sigma}_s$ is a robust residual scale estimate (Koller, 2007; Koller and Stahel, 2011). Some popular objective functions are the Huber, Hampel and Tukey bisquare which have varying levels of resistance to influential outliers. Because Hampel and Tukey bisquare sometimes are troublesome due to low Gaussian efficiency that slows the rate of convergence, resulting in convergence to wrong roots, no convergence and multiple roots, in this study, we used Huber. Moreover, R by default uses Huber in its maximum likelihood estimation procedure. Huber's objective function is

$$\rho_H(\mathrm{u}) = \begin{Bmatrix} 0.5\mathrm{u}^2 & if|\mathrm{u}| \le c \\ c|\mathrm{u}| - 0.5c^2 & otherwise \end{Bmatrix} \quad (3.43)$$

where $c = 1.345 \times \hat{\sigma}_s$ is a tuning constant (Fox, 2002; Koller and Stahel, 2011). As Koller and Stahel (2011) argue, $\rho_H(\mathrm{u})$ is a continuous and symmetric function with a unique minimum at 0, and is a rapidly increasing function. Thus, this allows resistance of robust maximum likelihood estimates to the influence of outliers.

Since $\hat{\sigma}$ is easily affected by influential outliers it is not used for $\hat{\sigma}_s$ (Fox, 2002). Based on the iterative algorithm that converges on an estimate for residuals and scale of the residuals, two estimation methods have been proposed. The methods are re-scaled median absolute deviation (MAD) and Huber's Proposal 2. Both are also built into the maximum likelihood estimation procedure in R. Due to the fact that the re-scaled MAD leads to liberal tests and does not vary smoothly with the data, we used Huber's Proposal 2 in this dissertation. In Huber's Proposal 2 the value of $\hat{\sigma}_s$ is found by solving the equation

$$\frac{1}{n-k}\sum_{i=1}^{n}\psi^2\left(\frac{y_i-\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}}{\hat{\sigma}_s}\right) = \mathrm{E}_{\phi}(\psi^2(\hat{\boldsymbol{\varepsilon}})) \qquad (3.44)$$

where, $\psi(u)$ is derivative of $\rho_H(\mathrm{u})$ and $\mathrm{E}_{\phi}(\psi^2(\hat{\boldsymbol{\varepsilon}}))$ is the expected value of $\psi^2(\hat{\boldsymbol{\varepsilon}})$ when the residuals have a standard normal distribution (Koller and Stahel, 2011).

Differentiating (3.42) with respect to the coefficients $\widehat{\boldsymbol{\beta}}$ and setting the partial derivatives to the *(k+1) x 1* zero vector yields the following system of *k+1* estimating equations

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i-\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}}{\hat{\sigma}_s}\right) = \sum_{i=1}^{n} x_{ij}\psi\left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}_s}\right) = \mathbf{0}. \qquad (3.45)$$

Defining a weight function as

$$\mathrm{w}\,(\mathrm{u}) = \frac{\psi(u)}{u} \qquad (3.46)$$

gives $w_i = w\left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}_s}\right)$ with $w_i = 1$ if $\hat{\varepsilon}_i = 0$ (Fox, 2002). By plugging (3.46) into (3.45) we get $\sum_{i=1}^{n} x_{ij}\, w_i \frac{\hat{\varepsilon}_i}{\hat{\sigma}_s} = \sum_{i=1}^{n} x_{ij} w_i (y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}) \frac{1}{\hat{\sigma}_s} = \boldsymbol{0}$ (3.47)

provided $\hat{\sigma}_s \neq 0$. Assigning small weights to extreme values controls problematic observations by removing the effect of contamination (Mutan, 2004). This leads the maximum likelihood estimates to achieve robustness properties. Equation (3.47) implies that

$$\sum_{i=1}^{n} x_{ij} w_i (y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}) = \boldsymbol{0}$$

yielding

$$\sum_{i=1}^{n} x_{ij} w_i \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} x_{ij} w_i y_i. \tag{3.48}$$

The relation (3.48) is expressed in matrix form

$$\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\,\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{Y} \tag{3.49}$$

where $\boldsymbol{W}$ =diagonal$\{(w_i : i = 1, \dots, n)\}$. Finally, the maximum likelihood (ML) estimates are obtained as

$$\widehat{\boldsymbol{\beta}}_{ML} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{Y}. \tag{3.50}$$

Given that $\boldsymbol{W}$ depends on the residuals, which also depend on the estimated coefficients which again depend upon $\boldsymbol{W}$, maximum likelihood estimates are not calculated explicitly from data like OLS estimates (Fox, 2002). Iterative methods have been developed as a solution to this problem. In this study, the iteratively reweighted least squares (IRLS) algorithm was considered because it is used by R to calculate maximum likelihood estimates. The IRLS procedure is done in the following steps:

1) Initial estimates $\widehat{\boldsymbol{\beta}}^{(0)}$ using OLS are chosen.

2) Residuals $\hat{\varepsilon}_i^{(f-1)}$ and associated weights $w_i^{(f-1)} = w_i(\hat{\varepsilon}_i^{(f-1)})$ from previous iterations are computed at each iteration f.

3) New weighted-least-squares estimates give $\widehat{\boldsymbol{\beta}}_{ML}^{(f)} = [\boldsymbol{X}'\boldsymbol{W}^{(f-1)}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{W}^{(f-1)}\boldsymbol{Y}$ where $\boldsymbol{W}^{(f-1)} = diagonal\left\{w_i^{(f-1)}\right\}$ in the current weight matrix (Fox, 2002).

The last two steps are repeated until the estimated coefficients converge. Convergence is the point at which $\frac{\|\hat{\boldsymbol{\varepsilon}}^{(f+1)} - \hat{\boldsymbol{\varepsilon}}^{(f)}\|}{\|\hat{\boldsymbol{\varepsilon}}^{(f+1)}\|} < 0.0001$ (Fox, 2002). The fitted values at the n observed regressor locations $\boldsymbol{x}_i'$ are

$$\widehat{\boldsymbol{Y}}_{ML} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{ML} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{Y} \quad (3.51)$$

and the estimate of the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}_{ML}$ is

$$\hat{\sigma}_s{}^2(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}. \quad (3.52)$$

The R code for robust maximum likelihood regression is given in Appendix D. The last section is a summary of this chapter.

3.6 Summary

- The secondary data on the audited performance results of companies listed on the JSE at annual company financial statements for the period 2010 was used, and consisted of 278 companies in the data set for positive ROCE and 79 companies in the data set for negative ROCE.
- The dependent variable is ROCE, and the independent variables are APCE, CE, D/A, D/E, DY, EPS, EY, MC, OPM, P/E, ROA, ROE and SP.
- The procedures of classical multiple linear regression analysis were explained in which the method of estimating the model parameters; the stepwise directed search method to screen the variables; and the assumptions, statistical significance, R^2, R_{adj}^2, mutlicollinearity, influential observations, and validity to assess the model were illustrated.

- The secondary data on the performance results of various companies listed on the JSE at annual company financial statements for the period 2011 was used as a validation data set and consisted of 239 companies in the data set for positive ROCE and 83 companies in the data set for negative ROCE.
- The estimation procedures of principal components regression, generalized least squares regression, and robust maximum likelihood regression were discussed to compare in modelling ROCE with classical multiple linear regression analysis.
- Classical multiple linear regression analysis, principal components regression and generalized least squares regression were done using SPSS.
- Robust maximum likelihood regression was performed in R.

Chapter 4 is the practical implementation of the procedures using SPSS and R statistical packages.

CHAPTER 4

DATA ANALYSIS AND DISCUSSION OF RESULTS

4.1 Data Analysis

4.1.0 Introduction

The statistical techniques relating to model construction, assumption checking and model fitting discussed in Chapter 3 are applied to data on Return on Capital Employed (ROCE) of individual companies listed on the JSE. As previously mentioned in Chapter 3, we use the positive ROCE data set and negative ROCE data set in constructing the models. A preliminary data analysis was performed so as to understand the positive ROCE data and negative ROCE data, and is presented in the next section.

4.1.1 Preliminary Data Analysis

The normal Q-Q plots, normal P-P plots and frequency distribution plots of the positive ROCE data and negative ROCE data are presented in Figures 4.1 and 4.2, respectively.

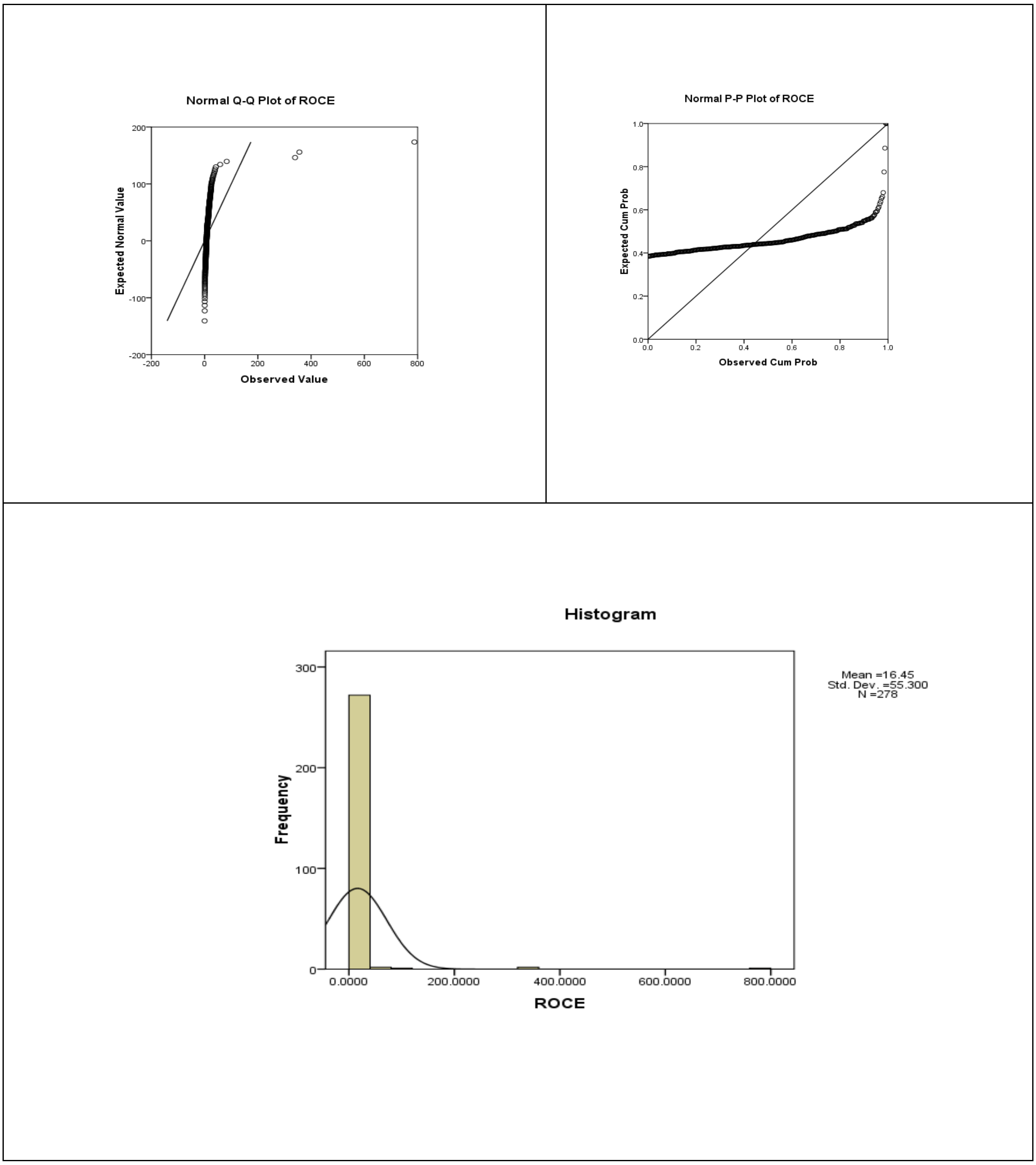

Figure 4.1. Normal Q-Q plot, normal P-P plot and frequency distribution plot of the positive ROCE data

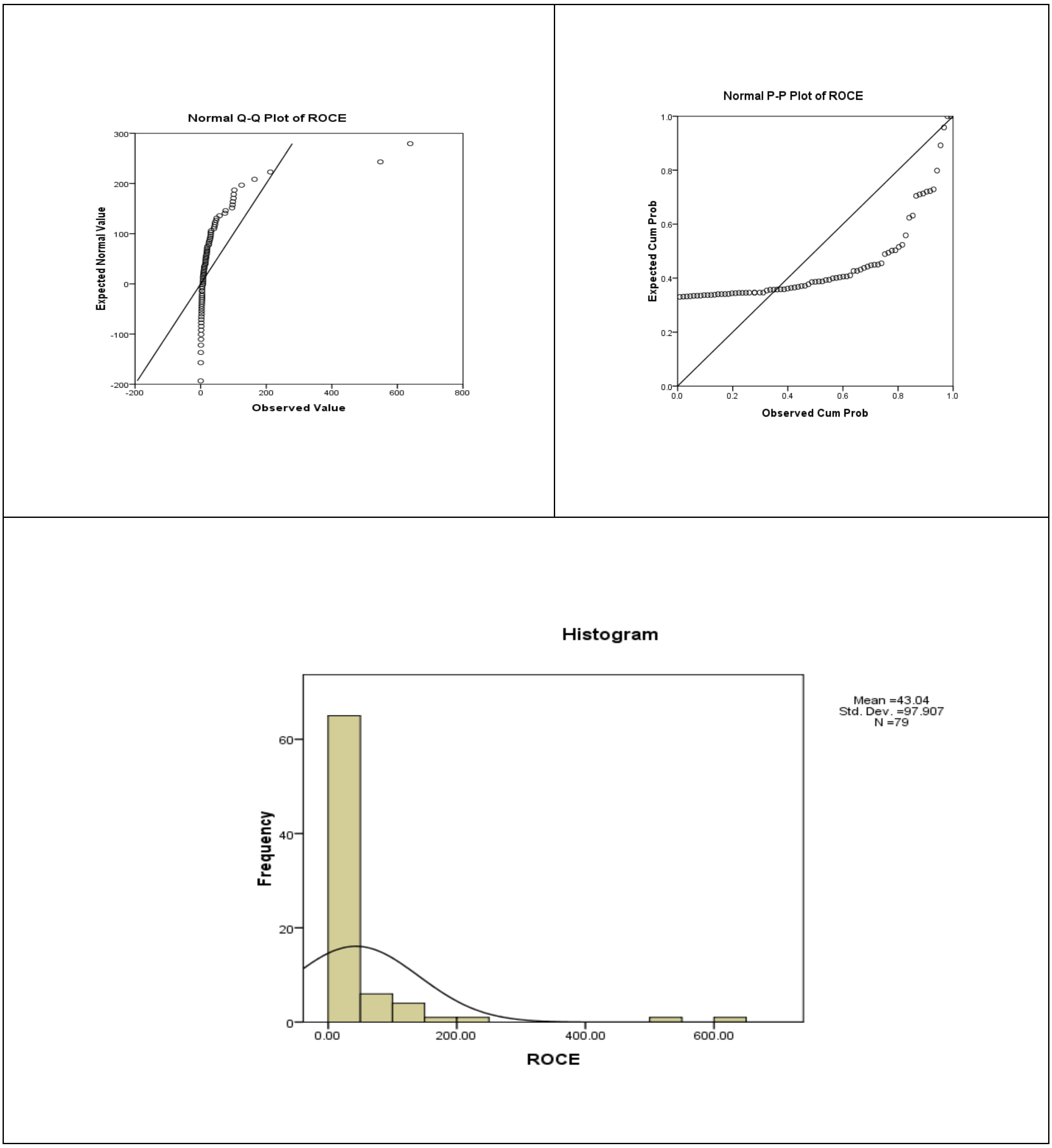

Figure 4.2. Normal Q-Q plot, normal P-P plot and frequency distribution plot of the negative ROCE data

There is a severe violation of the normality assumption based on the normal Q-Q plots and normal P-P plots in Figures 4.1 and 4.2, which show departure from a straight 45 degree line. A close inspection of the frequency distribution plots reveals that the non-normality distribution is a leptokurtic distribution, skewed to the right.

The asymmetric and extreme leptokurtic distribution nature of the raw data, in Figures 4.1 and 4.2, suggested use of test transformation methods, to verify the normality assumption. Transformation methods tested were logarithm, square root, inverse, and Box-Cox. Table 4.1 shows the results of the four transformation methods applied to data.

Table 4.1. Tests of normality results for the transformed data

	Tests of normality		**Transformation methods tested**			
			Logarithm (base 10)	Square root	Inverse	Box-Cox
Positive ROCE data	Kolmogorov-Smirnov	Statistic	0.080	0.174	0.341	0.517
		D.F	278	278	278	278
		P-value	0.000	0.000	0.000	0.000
	Shapiro-Wilk	Statistic	0.946	0.546	0.281	0.034
		D.F	278	278	278	278
		P-value	0.000	0.000	0.000	0.000
Negative ROCE data	Kolmogorov-Smirnov	Statistic	0.056	0.178	0.399	0.523
		D.F	79	79	79	79
		P-value	**0.200**	0.000	0.000	0.000
	Shapiro-Wilk	Statistic	0.991	0.757	0.223	0.101
		D.F	79	79	79	79
		P-value	**0.841**	0.000	0.000	0.000

The Kolmogorov-Smirnov and Shapiro-Wilk tests were performed on the transformed data. The p-values for the Kolmogorov-Smirnov and Shapiro-Wilk test statistics for all transformations are highly significant at 5 % significance level, except logarithm for the negative ROCE data. As a result, the logarithm transformation method was selected for the negative ROCE data. For the transformed positive ROCE data, there is violation on the normality distribution. The coefficient of skewness and kurtosis values provided final assessment of normality tests for the transformed positive

ROCE data. Computation and comparison of the coefficient of skewness and kurtosis values of the positive ROCE data was conducted. Results on the four transformation methods are given in Table 4.2.

Table 4.2. Results of skewness and kurtosis for choosing transformation methods for the positive ROCE data

Transformation	Actual transformation	Coefficient of skewness	Coefficient of kurtosis	Chosen transformation method
None	ROCE	11.448	146.321	
Logarithm (base 10)	Log_{10}(ROCE)	-0.144	3.153	Logarithm
Square root	$\sqrt{ROCE}$	6.231	55.223	
Inverse	$\frac{1}{ROCE}$	10.238	131.582	
Box-Cox ($\lambda = 9.68$)	$\frac{ROCE^{9.68} - 1}{9.68}$	16.673	278.000	

The skewness and kurtosis values of the transformation methods are not zero and three, respectively. However, the rationale for choosing a particular transformation method is based on the premise that the best transformation method is the one that gives a most reduction in the skewness and kurtosis values. Explicitly, the values of the skewness and kurtosis of the transformed variable are expected to be close to zero and three, respectively. The logarithm transformation method with values of the skewness and kurtosis closest to zero and three, respectively, is chosen among the other transformation methods for this data. Figures 4.3 and 4.4, show the normal Q-Q plots, normal P-P plots and frequency distribution plots of log (ROCE) for the positive data and negative data, respectively.

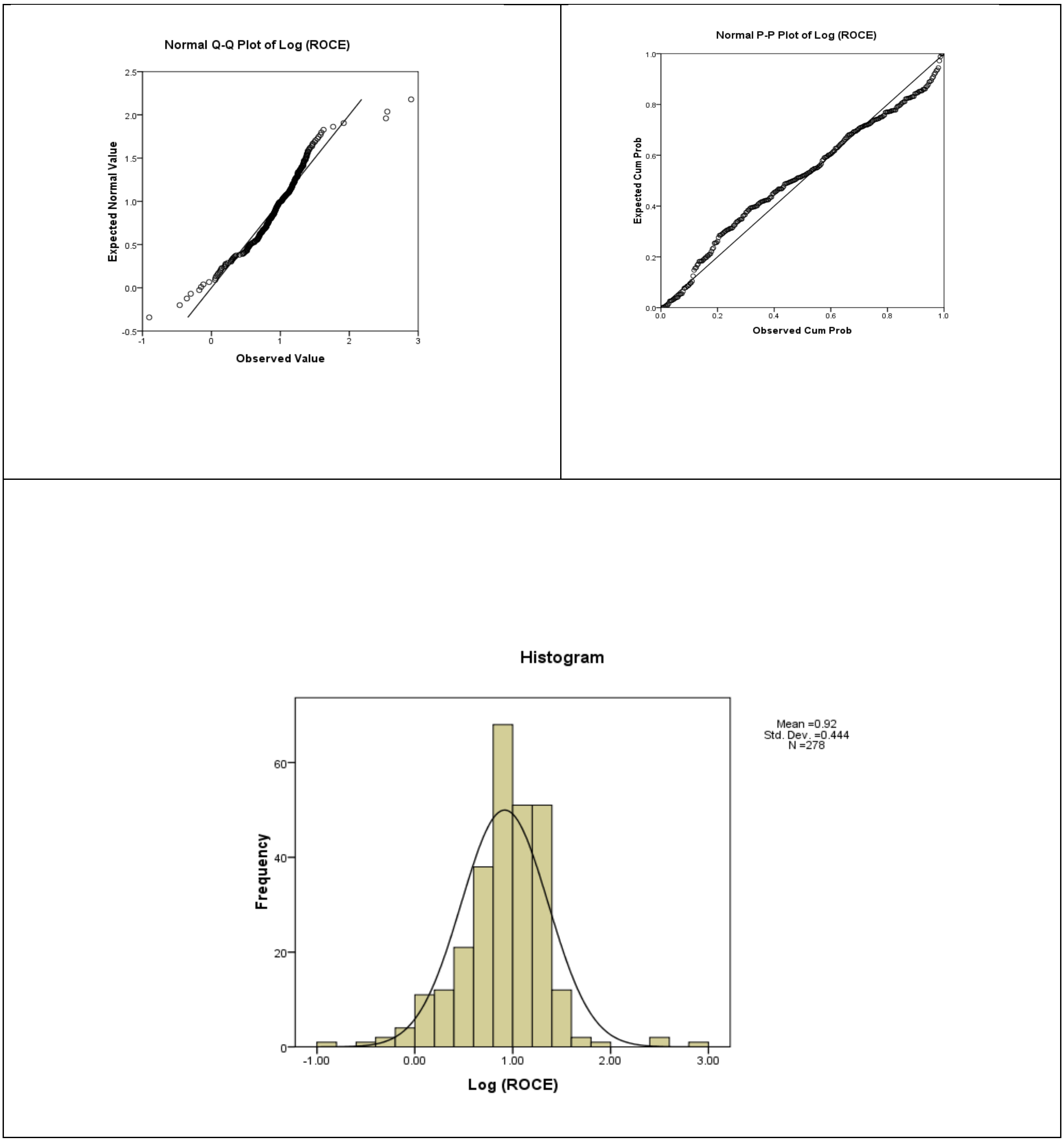

Figure 4.3. Normal Q-Q plot, normal P-P plot and frequency distribution plot of log (ROCE) for the positive data

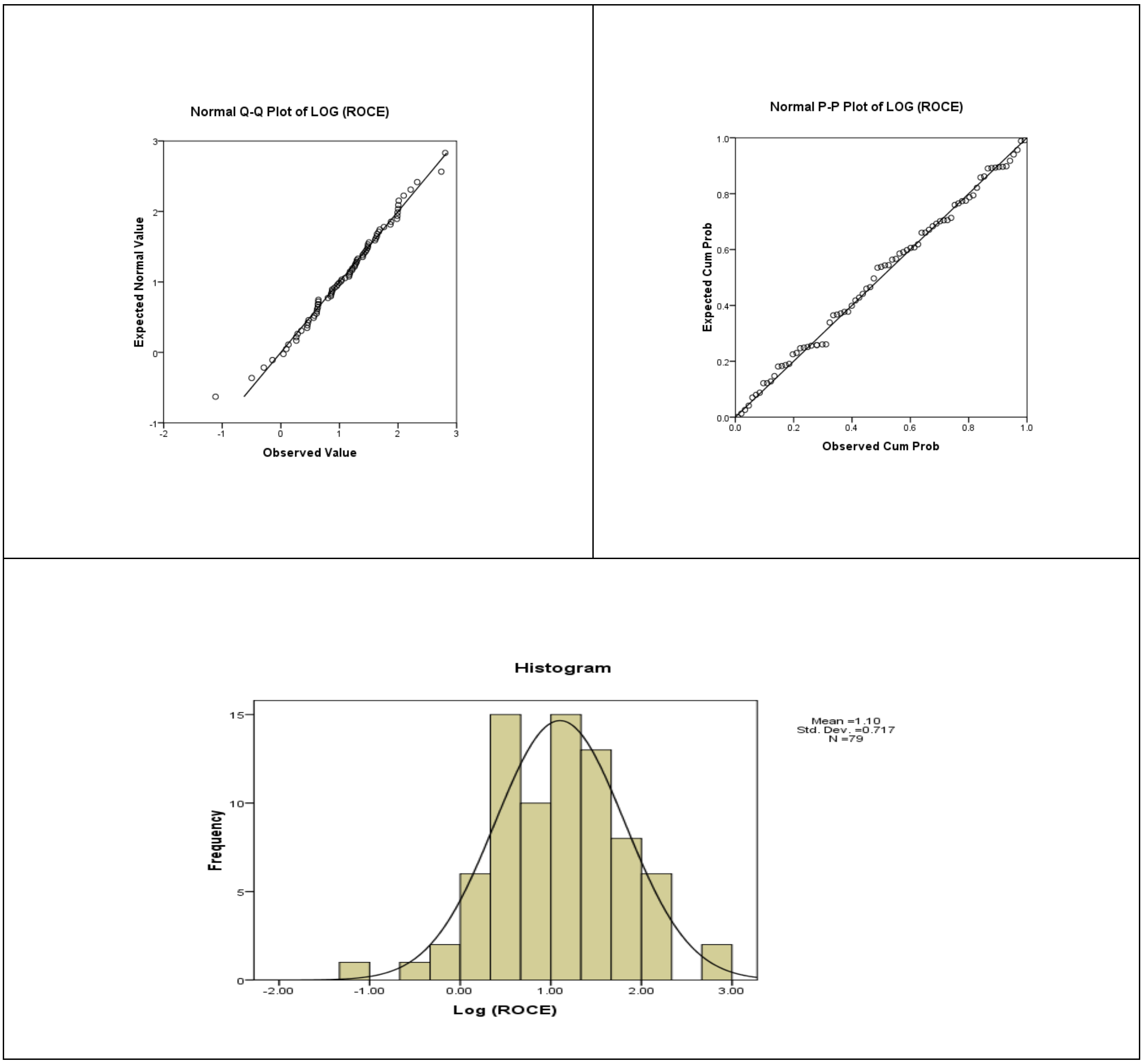

Figure 4.4. Normal Q-Q plot, normal P-P plot and frequency distribution plot of log (ROCE) for the negative data

Improved results following log transformations are presented in Figures 4.3 and 4.4. Compared to results in Figures 4.1 and 4.2, the normal Q-Q plots and normal P-P plots to the straight 45 degree line, and frequency distribution plots to the normal distribution indicate much improvement. This implies an improvement of the normality of the data. Thus, log (ROCE) is used in the ensuing analysis for the two data sets. We demonstrate the application of a classical multiple linear regression analysis to the ROCE data in the following section.

4.1.2 Classical Multiple Linear Regression Analysis

The Positive ROCE Data Set

The stepwise directed search regression procedure was implemented using the transformed positive ROCE data because the distribution of the raw/original data was leptokurtic and asymmetrical. Table 4.3 presents results of performing the stepwise directed search regression to estimate the regression coefficients.

Table 4.3. Analysis of variance for classical multiple linear regression on log (ROCE) and estimates of the regression parameter coefficients for the fitted model

Source of variation	DF	SS	MS	F-statistic	P-value
Regression	5	21.062	4.212	34.229	0.000
Residual	272	33.474	0.123		
Total	277	54.536			
	Parameters		**Estimates**	**P-values**	
	Intercept		0.817	0.000	
	ROE		0.003	0.000	
	D/E		-0.003	0.000	
	CE		-2.243×10^{-9}	0.000	
	EPS		0.000	0.000	
	DY		0.004	0.003	

Table 4.3 indicates that the p-value of the overall relationship is less than 5 % and associated p-values of the five coefficients are less than 5 %. This means that the fitted model involving the five independent variables is significant at 5 % level of significance. Table A.1 in Appendix A presents the summary and change statistics results of each variable in the fitted regression model.

The fitted regression model was assessed for statistical assumptions. Figures B.1 and B.2 in Appendix B, show normal Q-Q plot, normal P-P plot and frequency distribution plot of the standardized residuals; and scatter plot of the standardized residuals versus the standardized predicted values, respectively. Figure B.1 depicts a non-normality pattern of the residuals as evidenced by a well defined departure from a straight 45 degree line of the normal Q-Q plot and normal P-P plot. The frequency distribution plot shows that the non-normality distribution is a negatively skewed distribution (coefficient of skewness=-1.248). In Figure B.2, there is evidence of a distinguishable pattern showed in the scatter plot. Thus, the assumption of homoscedasticity is not satisfied. A logarithm transformation was applied to the independent variables in the fitted regression model so as to correct the violations in the assumptions. Figure 4.5 displays the normal Q-Q plot, normal P-P plot and frequency distribution plot of the standardized residuals for the final model.

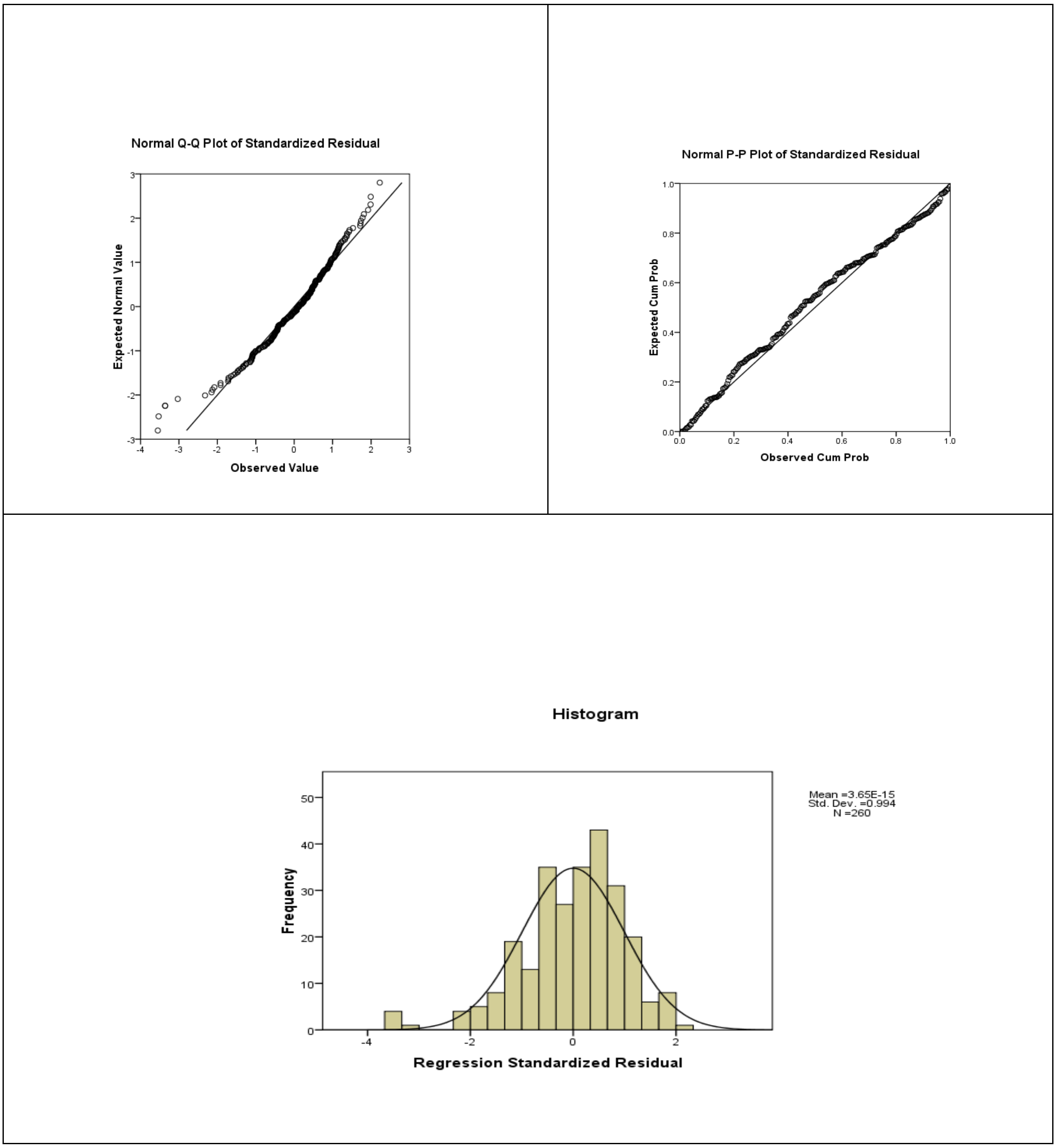

Figure 4.5. Normal Q-Q plot, normal P-P plot and frequency distribution plot of the final classical multiple linear regression standardized residuals

Results in Figure 4.5 show the normal Q-Q plot is on the straight 45 degree line and normal P-P plot does not seem to deviate much from a straight 45 degree line. The data is approximately normal as demonstrated by the histogram (coefficient of skewness=-0.773, coefficient of kurtosis=1.296) (Figure 4.5). We conclude that there is no severe non-normality of the residuals for the final model showed. The scatter plot of the standardized residuals versus standardized predicted values is shown in Figure 4.6.

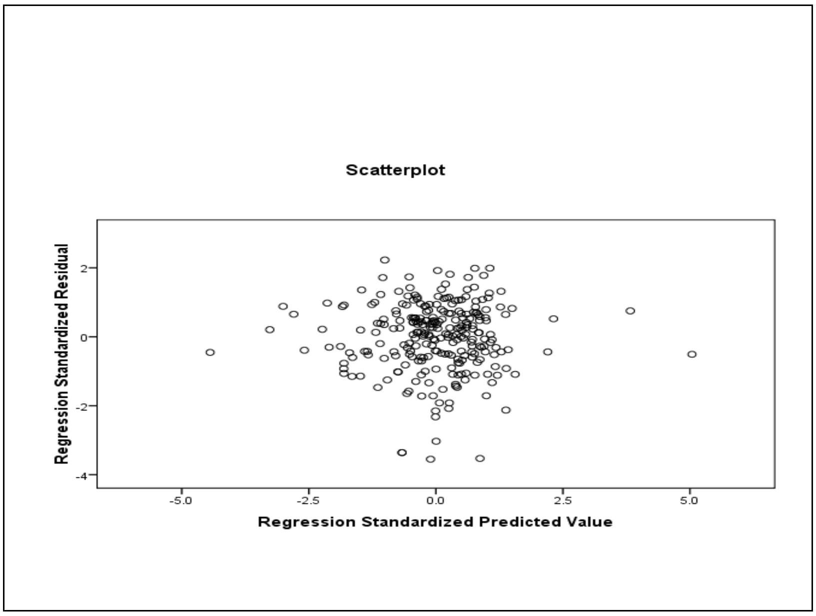

Figure 4.6. Scatter plot of standardized residuals versus standardized predicted values of the final classical multiple linear regression model

The random scattering of points with no clear pattern in Figure 4.6, indicates that the homoscedasticity assumption of the residuals for the final model is not violated. To test for autocorrelation in our final model (model with an intercept), the Durbin-Watson test was conducted. The results are:

Durbin-Watson	Durbin-Watson number of observations
1.597	260

For a two-tailed test at 5 % significance level, the Durbin-Watson critical values for n=200 and k=3 are $d_L = 1.738$, $4 - d_L = 2.262$, $d_U = 1.799$, and $4 - d_U = 2.201$ (Gujarati, 2004). This implies that for positive or negative autocorrelations, the null hypothesis that the residuals are not autocorrelated is not rejected. We conclude that the independence assumption of the residuals is satisfied. To check the existence of influential observations in the data we examined standardized residuals, leverages and Cook's Distances for the prediction model. Observations 54, 113, 135, and 162 were identified as outliers ($|\text{standardized residuals}| > 2$), high leverages points ($\text{leverages} > \frac{2\times k+2}{n} = \frac{2\times 3+2}{278} = 0.0288$) and influential points (Cook's Distances $> \frac{4}{n} = \frac{4}{278} = 0.0144$). Observations 3, 67, 77,104,137,145,176, and 225 were classified as both outliers ($|\text{standardized residuals}| > 2$) and influential points (Cook's Distances > 0.0144), and observations 163, 206, and 277 were classified as both high leverages points ($\text{leverages} > 0.0288$) and influential points (Cook's Distances > 0.0144). Observations 65, 174, and 203 were identified as just influential points (Cook's Distances > 0.0144). Observations 1, 75, 89, 99, 102, 141, 149, 151,193, 209, 212, 220, 242, 250, and 275 identified as high leverages points ($\text{leverages} > 0.0288$) were retained in the data because they were not influential points (Cook's Distances < 0.0144). All the identified outliers were influential points. Overall, eighteen observations had influence on the prediction model. These were removed from the data. We present the Cook's Distance plot in Figure 4.7.

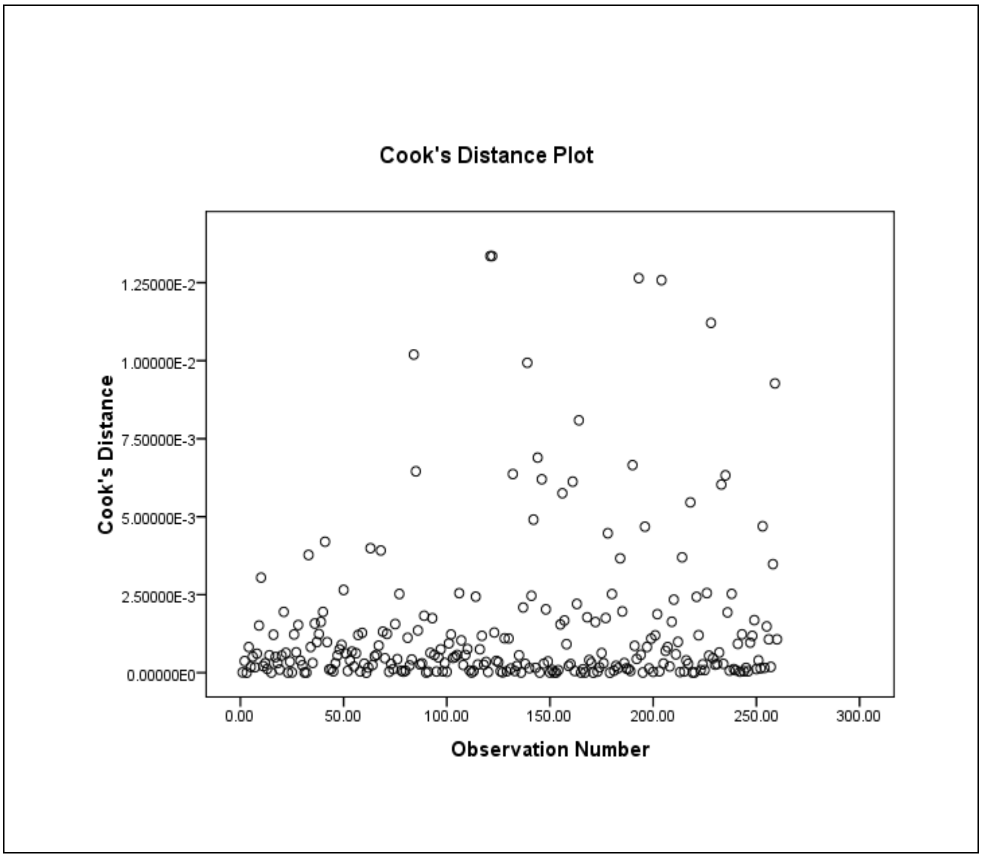

Figure 4.7. Cook's Distance plot for the final classical multiple linear regression model

Under the final model, the Cook's Distance values of the observations in the data set are less than 0.0144 (Figure 4.7). The data contained no observations with serious influence on the parameter estimates of the final model. The final and validation classical multiple linear regression models for predicting log (ROCE) are presented in Table 4.4.

Table 4.4. Classical multiple linear regression model for log (ROCE) with standard error of estimates in parentheses

Parameters	Final		Validation	
	Estimates	**P-values**	**Estimates**	**P-values**
Intercept	0.175 (0.052)	0.001	0.274 (0.100)	0.007
Log (ROE)	1.020 (0.018)	0.000	0.960 (0.032)	0.000
Log (D/E)	-0.193 (0.015)	0.000	-0.230 (0.024)	0.000
Log (CE)	-0.063 (0.008)	0.000	-0.074 (0.015)	0.000
Regression Equation Statistics				
	R^2	0.928	R^2	0.805
	R^2_{adj}	0.927	R^2_{adj}	0.803
	MSE	0.013	MSEP	0.043
	VIF_{max}	1.036	VIF_{max}	1.027

The values of R^2, R^2_{adj}, MSE, maximum VIF for any independent variable, and p-values for the parameters in the validation model are close to those in the final model (Table 4.4). The results in Table 4.4 reflect consistency in magnitude and sign of parameter estimates, and standard error of parameter estimates values; between the final model and validation model. The percentage differences between elasticities of ROE, D/E, and CE are 5.88 %, 19.17 %, and 17.46 %, respectively. These results imply that the statistics of the validation model are in general agreement with those of the final model. Thus, we conclude that the prediction model is satisfactory.

The Negative ROCE Data Set

Table A.2 in Appendix A presents correlation matrix both transformed dependent variable and independent variables on the negative ROCE data set. Results in Table A.2 show significant correlations at 5 % significance level, of log (ROCE) with independent variables APCE, D/A, EY, ROA, CE, MC and SP. There are four pairs of these independent variables where EY has significant correlations at 5 % significance level (Table A.2). As a result, EY was removed from the data set. A pair of CE and MC has high significant correlation at 5 % significance level (Table A.2).

The variable CE was discarded from the data set because it has lower correlation with log (ROCE) than MC (Table A.2). We considered independent variables APCE, D/A, ROA, MC and SP as possible candidates for inclusion in the model. The stepwise directed search regression procedure was applied to the transformed negative ROCE data. The results for the fitted model are displayed in Table 4.5.

Table 4.5. Analysis of variance for classical multiple linear regression on log (ROCE) and estimates of the regression parameter coefficients for the fitted model

Source of variation	**DF**	**SS**	**MS**	**F-statistic**	**P-value**
Regression	3	23.093	7.698	34.052	0.000
Residual	75	16.954	0.226		
Total	78	40.048			
	Parameters		**Estimates**	**P-values**	
	Intercept		0.595	0.000	
	ROA		-0.009	0.000	
	APCE		0.248	0.000	
	MC		-2.001×10^{-11}	0.000	

Table 4.5 shows that the p-value for the overall model is less than 5 % and p-values of the three coefficients are less than 5 %. This implies that the fitted model involving the three independent variables is significant at 5 % level of significance. The summary and change statistics results of each variable in the fitted regression model are presented in Table A.3 in Appendix A. The fitted classical multiple linear regression model was assessed for statistical assumptions. Figure 4.8 presents the normal Q-Q plot, normal P-P plot and frequency distribution plot of the standardized residuals for the final model.

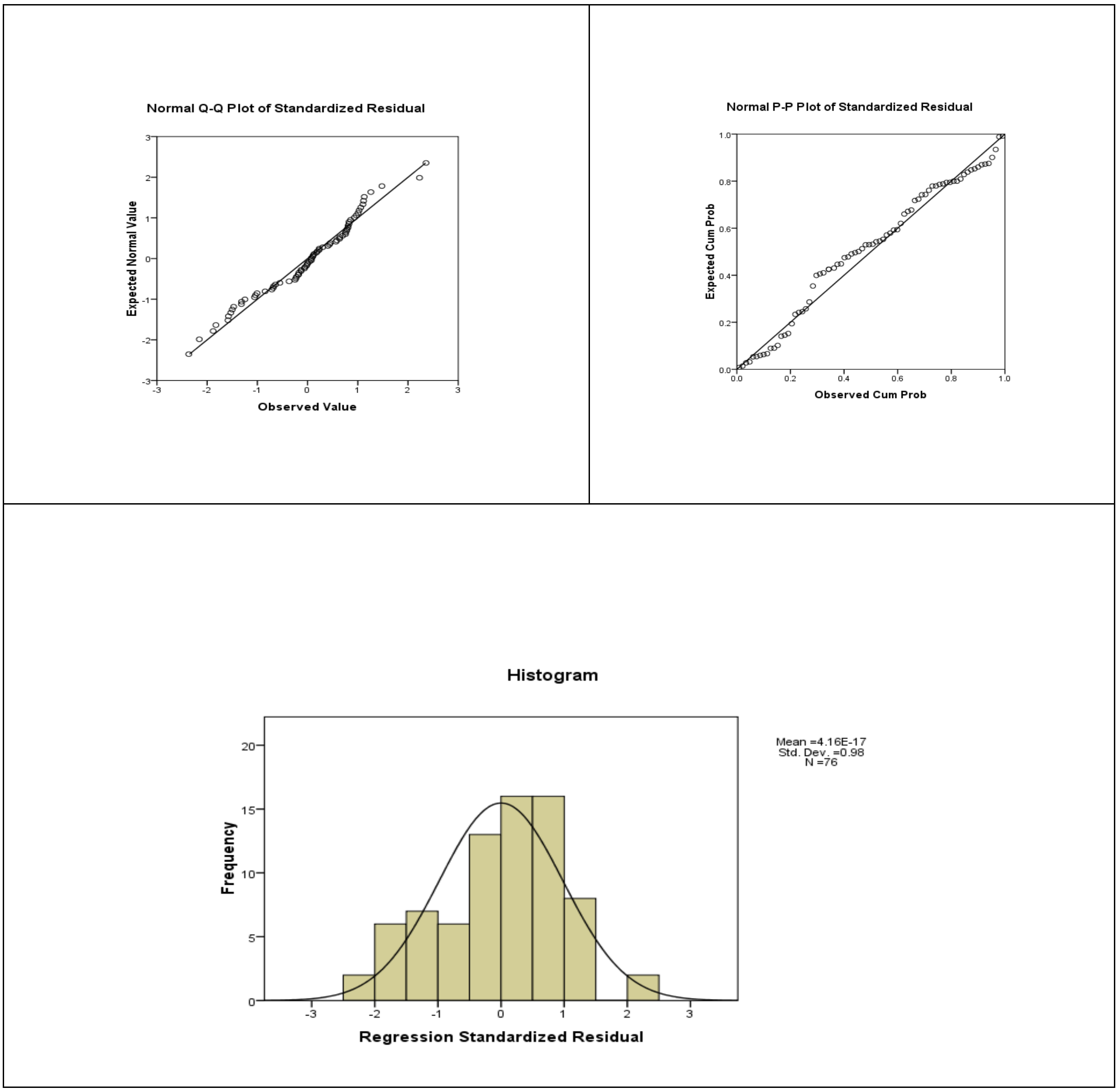

Figure 4.8. Normal Q-Q plot, normal P-P plot and frequency distribution plot of the final classical multiple linear regression standardized residuals

The normal Q-Q plot and normal P-P plot in Figure 4.8 does not seem to deviate much from a straight 45 degree line suggesting that the normality assumption is not violated. The data is almost normal as shown by the histogram (coefficient of skewness=-0.309, coefficient of kurtosis=-0.071) (Figure 4.8). Figure 4.9 presents the scatter plot of the standardized residuals versus the standardized predicted values.

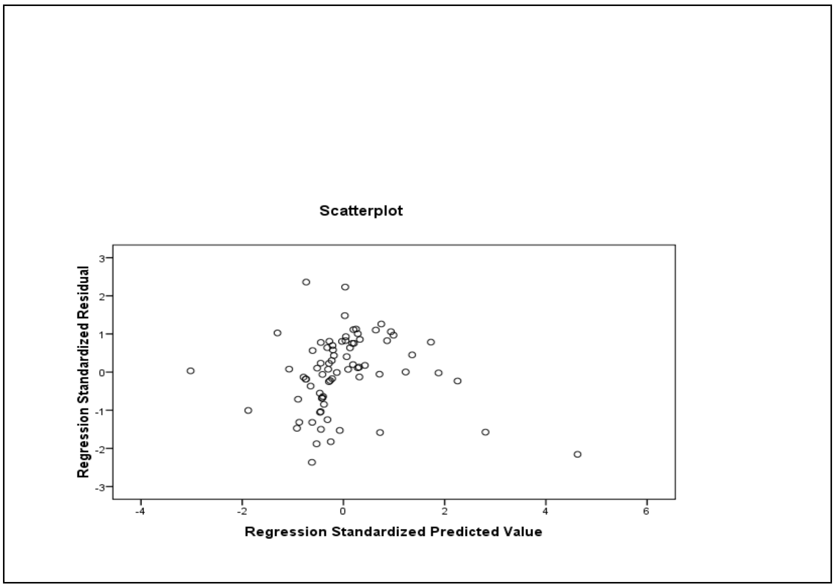

Figure 4.9. Scatter plot of standardized residuals versus standardized predicted values of the final classical multiple linear regression model

The systematic pattern in Figure 4.9 is mild. Thus, the assumption of homoscedasticity is fulfilled. The Durbin-Watson test results for autocorrelation in the final model (model with an intercept) are as follows:

Durbin-Watson	Durbin-Watson number of observations
2.391	76

For a two-tailed test at 5 % significance level, the Durbin-Watson critical values for n=75 and k=3 are $d_L = 1.543$, $4 - d_L = 2.457$, $d_U = 1.709$, and $4 - d_U = 2.291$ (Gujarati, 2004). This suggests that the data does not contain serious autocorrelation problems. We conclude that the independence assumption of the residuals is satisfied. Standardized residuals, leverages and Cook's Distances were used to assess influential observations in the data for the prediction model. Observations 3, 22, 37, and 63 identified as high leverage points ($\text{leverages} > \frac{2\times k+2}{n} = \frac{2\times 3+2}{79} = 0.1013$) were not influential (Cook's Distances $< \frac{4}{n} = \frac{4}{79} = 0.0506$)

on the prediction model. These were retained in the data. Observation 8 was identified as an outlier ($|\text{standardized residual}| > 2$), a high leverage point ($\text{leverage} > 0.1013$) and an influential point (Cook's Distance > 0.0506). Observation 49 identified as an outlier ($|\text{standardized residual}| > 2$) and observation 28 identified as a point of high leverage ($\text{leverage} > 0.1013$) were influential points (Cook's Distances > 0.0506). All the identified outliers were influential points. The three influential observations were deleted. The Cook's distance plot is presented in Figure 4.10 under the final classical multiple linear regression model.

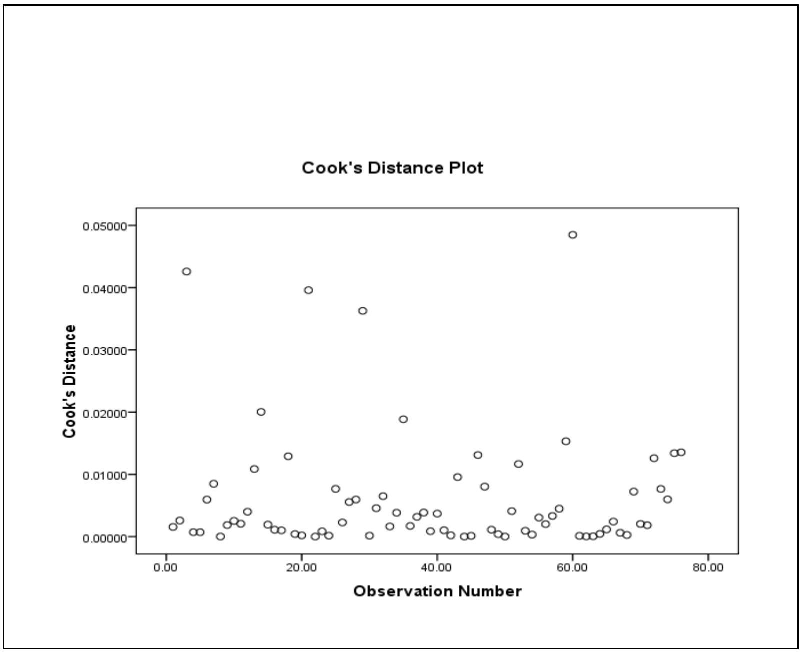

Figure 4.10. Cook's Distance plot for the final classical multiple linear regression model

Figure 4.10 shows that Cook's distance value of each of the observations is less than 0.0506. This confirms that none of the observations were influential on the parameter estimates in the final classical multiple linear regression model. Table 4.6

shows the final and validation classical multiple linear regression models for predicting log (ROCE).

Table 4.6. Classical multiple linear regression model for log (ROCE) with standard error of estimates in parentheses

Parameters	Final			Validation		
	Estimates		P-values	Estimates		P-values
Intercept	0.450 (0.113)		0.000	0.529 (0.136)		0.000
APCE	0.368 (0.079)		0.000	0.284 (0.069)		0.000
ROA	-0.011 (0.001)		0.000	-0.008 (0.001)		0.000
MC	-1.979$\times 10^{-11}$ (0.000)		0.000	-3.145$\times 10^{-11}$ (0.000)		0.061
Regression Equation Statistics						
	R^2	0.666		R^2	0.399	
	R^2_{adj}	0.652		R^2_{adj}	0.377	
	MSE	0.149		MSEP	0.429	
	VIF_{max}	1.025		VIF_{max}	1.041	

The signs of parameter estimates and values of parameter estimates, standard error of parameter estimates, and maximum VIF for any independent variable; between the final model and validation model are generally concordant (Table 4.6). The percentage differences between estimates of APCE, ROA, and MC are 22.83 %, 27.27 %, and 58.92 %, respectively. However, there is a substantial change in the values of R^2, R^2_{adj} and MSE between the final model and validation model; and insignificance of the coefficient for MC in the validation model (Table 4.6). These statistics showed poor performance of the final model. The application of principal components regression to the ROCE data is illustrated in the next section.

4.1.3 Principal Components Regression

The Positive ROCE Data Set

Due to large variation of the independent variables in the positive ROCE data set, a logarithm (base 10) transformation was applied to the independent variables. The Kaiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity results for the transformed independent variables are presented in Table 4.7.

Table 4.7. KMO and Bartlett's test

KMO		0.579
Bartlett's Test of Sphericity	Approx. Chi-Square	4.236×10^3
	D.F	78
	P-value	$\leq$0.000

The highly significant Bartlett's test (Table 4.7) confirms existence of sufficient correlations between the independent variables. The KMO value in Table 4.7 is greater than 0.5. Thus need to apply principal components analysis. Principal components analysis on the correlations matrix of transformed independent variables on the positive ROCE data set was conducted. The eigenvalues of each principal component before extraction, after extraction and after varimax rotation are displayed in Table 4.8.

Table 4.8. Total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.508	26.988	26.988	3.508	26.988	26.988	3.291	25.314	25.314
2	2.668	20.524	47.512	2.668	20.524	47.512	2.441	18.777	44.091
3	2.218	17.059	64.571	2.218	17.059	64.571	2.335	17.960	62.051
4	1.484	11.412	75.982	1.484	11.412	75.982	1.811	13.932	75.982
5	.903	6.946	82.929						
6	.776	5.972	88.901						
7	.649	4.989	93.890						
8	.338	2.601	96.491						
9	.258	1.988	98.479						
10	.084	.647	99.126						
11	.060	.462	99.588						
12	.053	.408	99.996						
13	.001	.004	100.000						

Table 4.8 shows that the data has 13 principal components before extraction, with principal component 1 explaining 26.988 % of total variance. By Kaiser-Guttman criterion (Andreica, 2009), there are four principal components with eigenvalues greater than 1 and these principal components accounted for about 76 % of the total variance. The eigenvalues pertaining to these principal components are given in the middle column of Table 4.8 without the values of the excluded principal components. The last column of the table shows the eigenvalues of the four principal components after varimax rotation. Rotation has the effect of equalising the relative importance of the four principal components (Ul-Saufie et al., 2011). Before rotation, principal component 1 accounted for 26.988 % of the total variance compared to 20.524 %,

17.059 % and 11.412 %. However, after rotation it accounted for 25.314 % of the total variance compared to 18.777 %, 17.960 % and 13.932 %. The scree plot for the principal components solution before extraction is displayed in Figure 4.11.

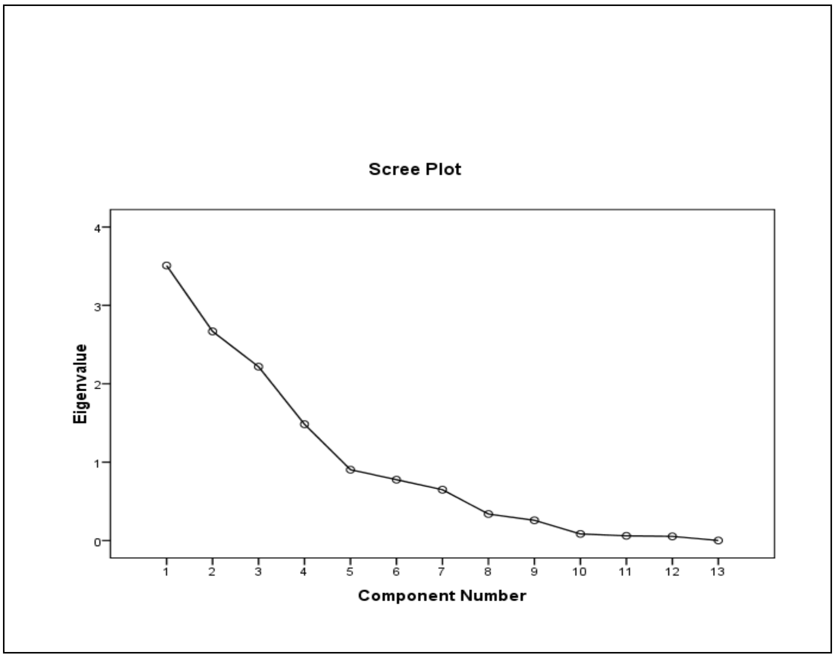

Figure 4.11. Scree plot for the principal components

The plot levels off from component 5 to component 6 (Figure 4.11), refuting the previous result from Table 4.8 that the optimal principal components solution is described by four components. Table 4.9 presents the rotated matrix using varimax rotation with Kaiser normalization.

Table 4.9. Rotated component matrix

	Component			
	1	2	3	4
Log (MC)	**.918**	-.182	.046	.061
Log (EPS)	**.897**	.258	-.024	.095
Log (CE)	**.887**	-.147	.032	-.049
Log (SP)	**.848**	-.238	-.059	.166
Log (EY)	-.065	**.964**	.082	.054
Log (P/E)	.064	**-.964**	-.080	-.054
Log (DY)	-.105	.498	-.124	.102
Log (D/E)	.064	-.100	**.922**	.117
Log (D/A)	-.127	-.071	**.878**	.202
Log (APCE)	.072	.113	**.706**	-.148
Log (ROA)	-.092	.137	.177	**.834**
Log (ROE)	.211	.281	.133	**.798**
Log (OPM)	.199	-.176	-.362	**.585**

The rotated matrix has the loading of each variable on each component for all values. The variables dominating the principal components are as follows (Table 4.9):

Principal component 1 (PC1) has log (MC), log (EPS), log (CE), and log (SP).

Principal component 2 (PC2) has log (EY) and log (P/E).

Principal component 3 (PC3) has log (D/E), log (D/A) and log (APCE).

Principal component 4 (PC4) has log (ROA), log (ROE) and log (OPM).

A principal components regression model which considered the transformed positive ROCE data was constructed. The regression results for the fitted model are shown in Table 4.10.

Table 4.10. Analysis of variance for principal components regression on log (ROCE) and estimates of the regression parameter coefficients for the fitted model

Source of variation	**DF**	**SS**	**MS**	**F-statistic**	**P-value**
Regression	4	30.524	7.631	86.761	0.000
Residual	273	24.012	0.088		
Total	277	54.536			
	Parameters		**Estimates**	**P-values**	
	Intercept		0.016	0.892	
	PC1		-0.007	0.207	
	PC2		0.115	0.000	
	PC3		-0.050	0.004	
	PC4		0.309	0.000	

The p-value for the overall model is less than 5 % and p-values of three coefficients are less than 5 % (Table 4.10). This suggests that the fitted model is significant at 5 % level of significance. The fitted principal component regression model was assessed for statistical assumptions. The normal Q-Q plot, normal P-P plot and frequency distribution plot of the standardized residuals are shown in Figure 4.12.

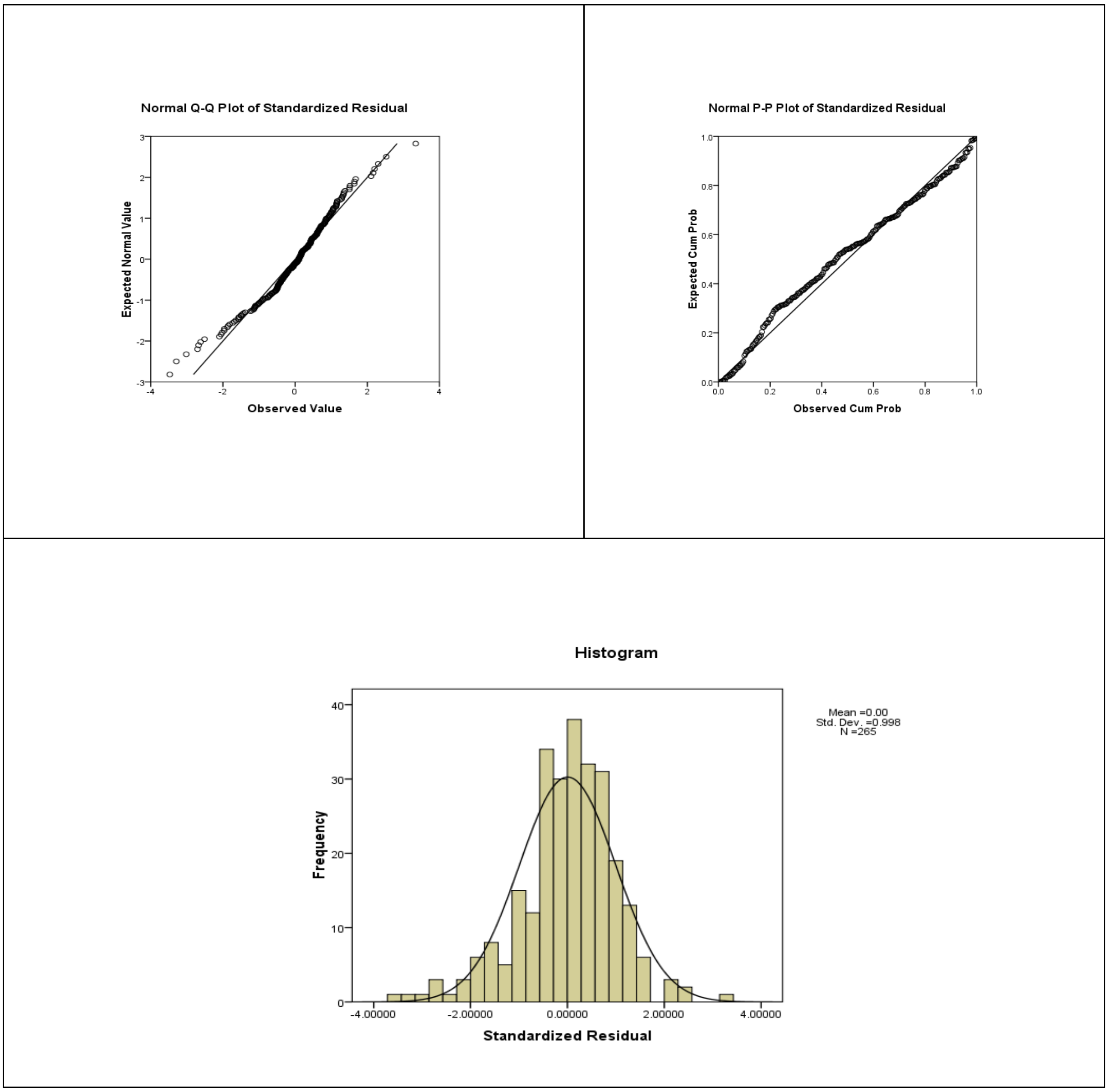

Figure 4.12. Normal Q-Q plot, normal P-P plot and frequency distribution plot of the final principal regression standardized residuals

Most parts of the normal Q-Q plot and normal P-P plot in Figure 4.12 are on the straight 45 degree line, suggesting that the normality assumption of the residuals is satisfied. The frequency plot resembles a normal distribution (coefficient of skewness=-0.501, coefficient of kurtosis=1.284). The scatterplot of the standardized residuals versus the standardized predicted values is presented in Figure 4.13.

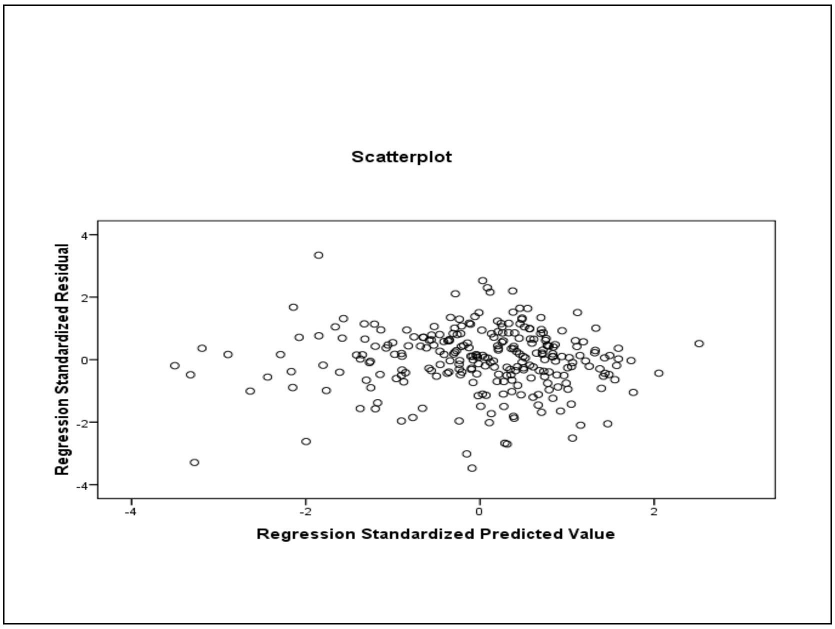

Figure 4.13. Scatter plot of standardized residuals versus standardized predicted values of the final principal component regression model

The plot in Figure 4.13 shows that the residuals have a random scatter. Thus, we accept the model as fitting the homoscedasticity assumption. The results for the Durbin-Watson test for positive or negative autocorrelation in the final principal component regression model (model with no intercept) are as follows:

Durbin-Watson	Durbin-Watson number of observations
1.967	265

For a two-tailed test at 5 % significance level, the Durbin-Watson critical values for n=200 and k=2 are $d_L = 1.644$, $4 - d_L = 2.356$, $d_U = 1.738$, and $4 - d_U = 2.262$ (Gujarati, 2004). This shows that the data has no autocorrelation, suggesting that the independence assumption of the residuals is fulfilled. We used standardized residuals, leverages and Cook's Distances to identify points that do not fit with the regression model. Observations 5, 77, 113, 135, and 163 were classified as both outliers $(|\text{standardized residuals}| > 2)$ and influential points (Cook's Distances $> \frac{4}{n} =$

$\frac{4}{278} = 0.0144$), and observations 202, 209, 212, 234, and 262 were classified as both high leverages points ($\text{leverages} > \frac{2 \times k+2}{n} = \frac{2 \times 2+2}{278} = 0.0216$) and influential points (Cook's Distances $> \ 0.0144$). Observations 101,193 and 250 had Cook's Distances greater than 0.0144. This shows that these observations are influential points. Observations 24, 184 and 189 identified as high leverages points ($\text{leverages} > 0.0216$) and observations 1, 54,128, 129, 145, 149, 162, 198, 211, and 252 identified as outliers ($|\text{standardized residuals}| > 2$) were retained in the data because they were not influential points (Cook's Distances $< \ 0.0144$). Overall, 13 observations were influential on the prediction model. The thirteen influential observations were deleted. Figure 4.14 presents the Cook's Distance plot under the final principal component regression model.

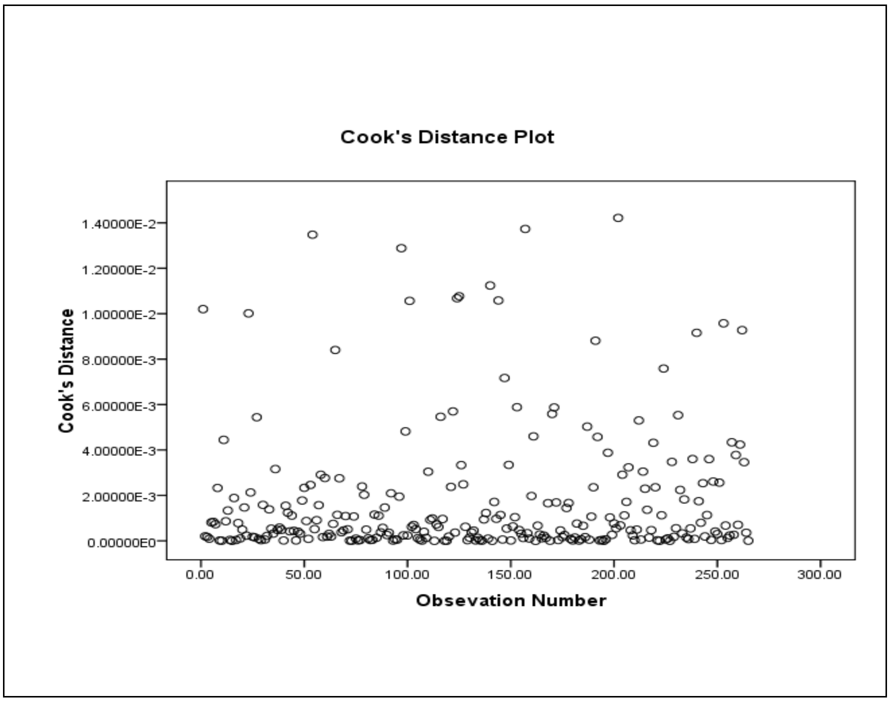

Figure 4.14. Cook's Distance plot for the final principal components regression model

Figure 4.14 reveals that Cook's Distance values are less than 0.0144, implying that there were no observations which were influential on the parameter estimates in the final model. The final and validation principal components regression models for predicting log (ROCE) appear in Table 4.11.

Table 4.11. Principal components regression model for log (ROCE) with standard errors of estimates in parentheses

Parameters	**Final**		**Validation**	
	Estimates	**P-values**	**Estimates**	**P-values**
PC2	0.142 (0.024)	0.000	0.174 (0.032)	0.000
PC4	0.281 (0.005)	0.000	0.280 (0.007)	0.000
Regression Equation Statistics				
	R^2 0.929		R^2 0.893	
	R^2_{adj} 0.929		R^2_{adj} 0.892	
	MSE 0.069		MSEP 0.111	
	VIF_{max} 1.064		VIF_{max} 1.037	

There is general resemblance in values of R^2, R^2_{adj}, MSE, maximum VIF for principal components, and significance of the parameters; between the final model and validation model (Table 4.11). There is conformity in sign and size of parameter estimates, and values of standard error of parameter estimates; between the final model and validation model (Table 4.11). The percentage differences between estimates of PC2 and PC4 are 22.54 % and 0.36 %, respectively. Thus, the final model has resulted in a satisfactory validation performance.

The Negative ROCE Data Set

The KMO and Bartlett's test of sphericity results for the negative ROCE data set are presented in Table 4.12.

Table 4.12. KMO and Bartlett's test

KMO		0.319
Bartlett's Test of Sphericity	Approx. Chi-Square	381.940
	D.F	78
	P-value	≤0.000

The KMO value in Table 4.12 is less than 0.5, thus the data is not adequate for principal components analysis. However, Bartlett's Test is highly significant (Table 4.12) and sixteen pairs of independent variables have significant correlations at 5 % significance level (Table A.2). Thus principal components analysis is suitable for this data. We proceed with the analysis. Table 4.13 shows the eigenvalues of each principal component before extraction, after extraction and after varimax rotation.

Table 4.13. Total variance explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	2.599	19.992	19.992	2.599	19.992	19.992	2.046	15.735	15.735
2	1.975	15.194	35.186	1.975	15.194	35.186	1.866	14.351	30.086
3	1.564	12.032	47.218	1.564	12.032	47.218	1.690	12.999	43.084
4	1.195	9.196	56.414	1.195	9.196	56.414	1.404	10.802	53.886
5	1.129	8.686	65.100	1.129	8.686	65.100	1.319	10.148	64.034
6	1.049	8.072	73.172	1.049	8.072	73.172	1.188	9.138	73.172
7	.820	6.309	79.481						
8	.770	5.924	85.405						
9	.657	5.055	90.460						
10	.542	4.168	94.628						
11	.428	3.294	97.923						
12	.251	1.930	99.852						
13	.019	.148	100.000						

The initial solution has 13 principal components, with principal component 1 explaining 19.992 % of total variance (Table 4.13). There are six principal components with eigenvalues greater than 1. These six principal components accounted for about 73 % of the total variance using the Kaiser-Guttman criterion (Andreica, 2009). Before rotation, principal component 1 accounted for 19.992 % of the total variance compared to 15.194 %, 12.032 %, 9.196 %, 8.686 %, and 8.072 %. After rotation it accounted for 15.735 % of the total variance compared to 14.351 %, 12.999 %, 10.802 %, 10.148 %, and 9.138 %. The scree plot for the initial solution is shown in Figure 4.15.

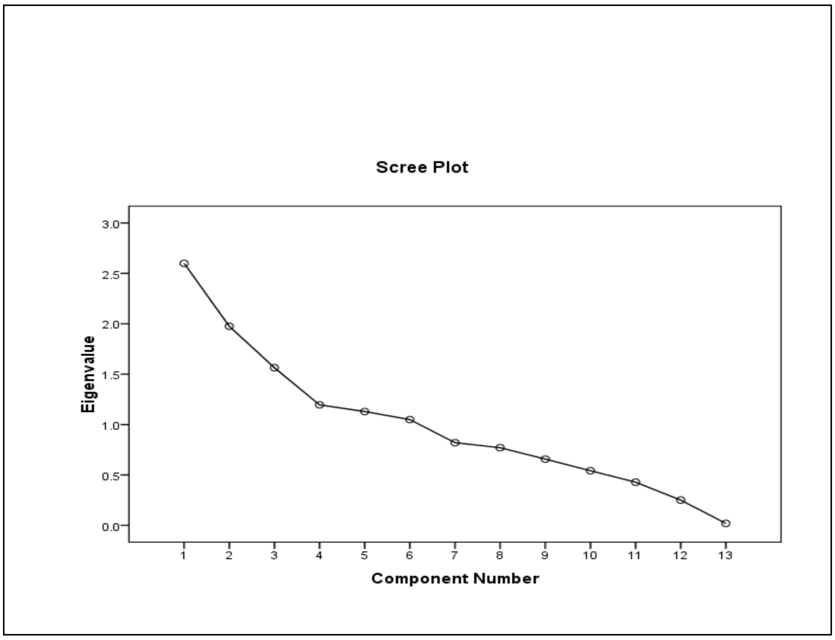

Figure 4.15. Scree plot for the principal components

The plot levels off from component 5 to component 6 (Figure 4.15), validating the result from Table 4.13 that the optimal principal components solution is described by six components. The rotated matrix using varimax rotation with Kaiser normalization is presented in Table 4.14.

Table 4.14. Rotated component matrix

	Component					
	1	2	3	4	5	6
MC	**.968**	.017	.072	-.112	-.101	-.093
CE	**.943**	.028	.003	.082	-.162	.046
ROE	.018	**-.887**	-.056	-.017	-.025	.043
D/E	.022	**.795**	-.063	.067	-.030	-.015
APCE	.089	.071	**-.748**	.030	.098	.218
EY	.113	.570	**.649**	-.056	.032	.028
ROA	.178	-.078	**.634**	.026	.114	.380
D/A	-.046	-.325	-.476	.131	.154	.467
EPS	.189	.038	.115	**.822**	-.059	-.065
SP	.276	-.036	.216	**-.739**	.000	-.076
P/E	-.217	-.024	-.083	.212	**.794**	.011
DY	-.046	.020	.037	-.300	**.768**	-.061
OPM	-.050	.020	-.011	-.036	-.097	**.866**

The principal components are as follows (Table 4.14):

Principal component 1 (PC1) has MC and CE.

Principal component 2 (PC2) has D/E and ROE.

Principal component 3 (PC3) has ROA, EY and APCE.

Principal component 4 (PC4) has EPS and SP.

Principal component 5 (PC5) has DY and P/E.

Principal component 6 (PC6) has OPM.

A principal components regression model which considered the transformed negative ROCE data was constructed. The regression results are displayed in Table 4.15.

Table 4.15. Analysis of variance for principal components regression on log (ROCE) and estimates of the regression parameter coefficients for the fitted model

Source of variation	DF	SS	MS	F-statistic	P-value
Regression	6	24.152	4.025	18.234	0.000
Residual	72	15.995	0.221		
Total	78	40.048			
	Parameters		**Estimates**	**P-values**	
	Intercept		0.777	0.000	
	PC1		-1.474$\times 10^{-11}$	0.017	
	PC2		0.000	0.002	
	PC3		-0.007	0.000	
	PC4		-1.217$\times 10^{-5}$	0.745	
	PC5		0.001	0.201	
	PC6		1.172$\times 10^{-6}$	0.972	

The p-value for the overall model is less than 5 % and p-values of the three coefficients are less than 5 % (Table 4.15). This shows that the fitted model involving the six principal components is significant at 5 % level of significance.

The fitted regression model was assessed for statistical assumptions. Figure 4.16 shows normal Q-Q plot, normal P-P plot and frequency distribution plot of the standardized residuals of the final principal components regression model.

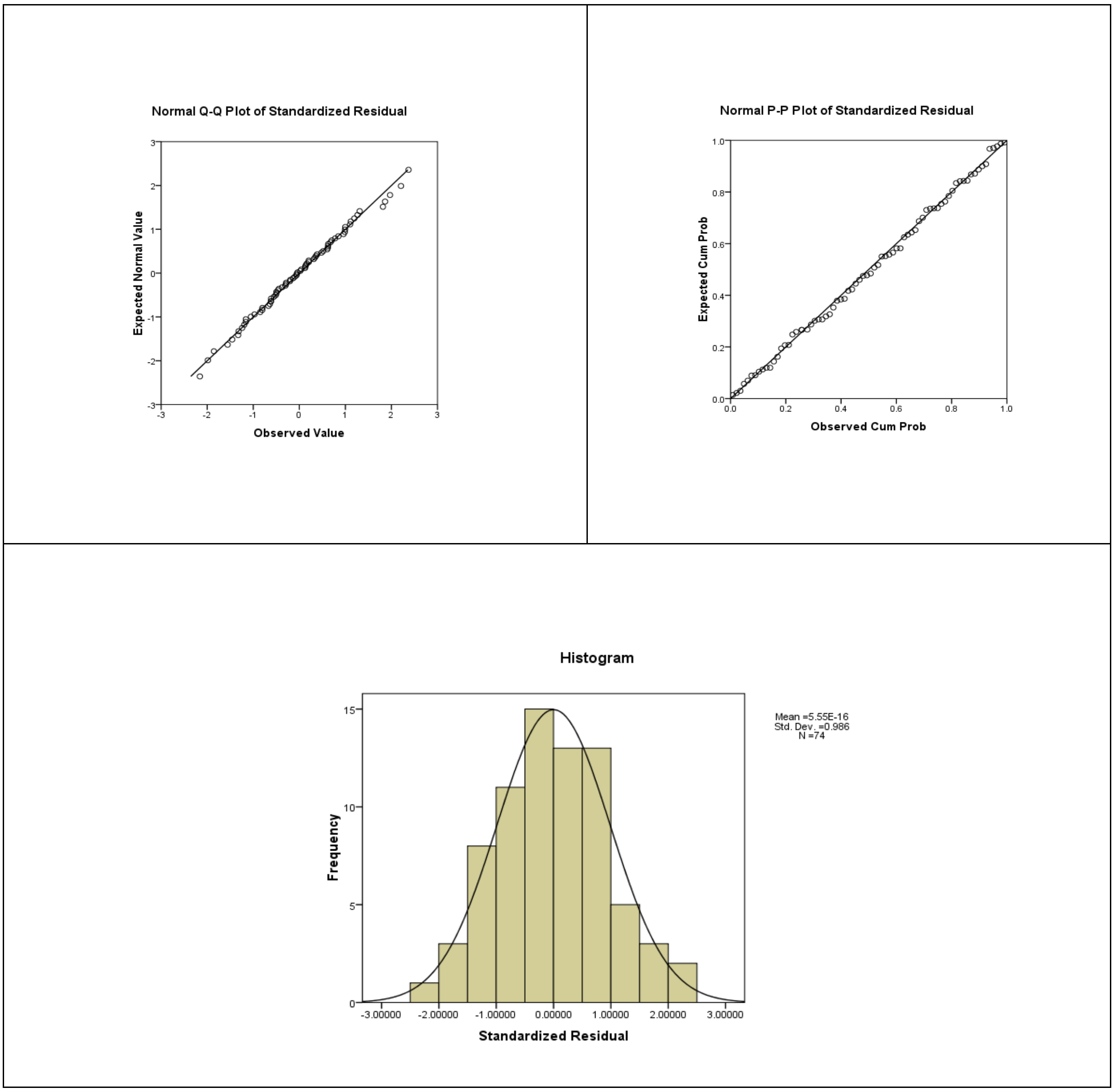

Figure 4.16. Normal Q-Q plot, normal P-P plot and frequency distribution plot of the final principal regression standardized residuals

Figure 4.16 portrays a normality pattern of the residuals as evidenced by a normal Q-Q plot and normal P-P plot which are on a straight 45 degree line. The frequency distribution plot shows the normality distribution (coefficient of skewness=0.168, coefficient of kurtosis=-0.170). The scatter plot of the standardized residuals versus the standardized predicted values is displayed in Figure 4.17.

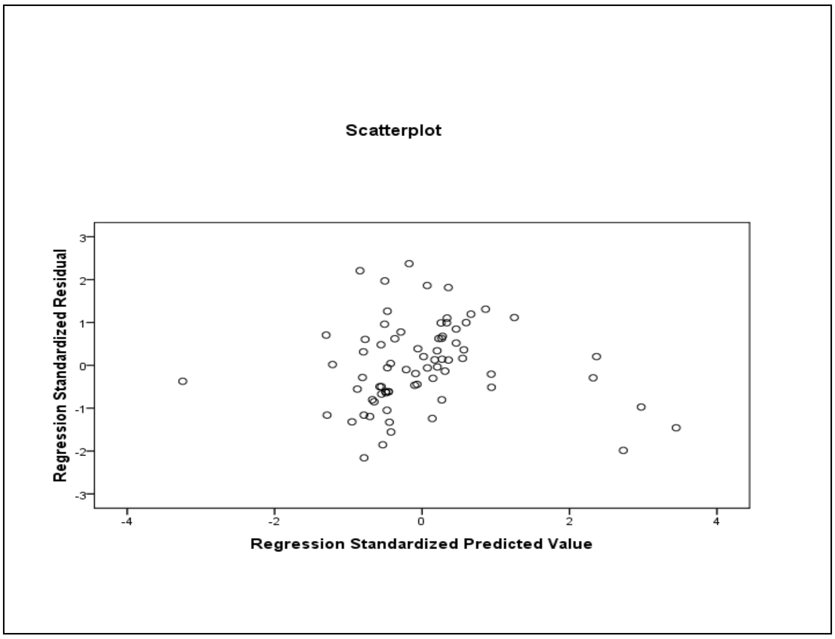

Figure 4.17. Scatter plot of standardized residuals versus standardized predicted values of the final principal component regression model

There is a weak distinguishable pattern showed in the scatter plot (Figure 4.17). Thus, the assumption of homoscedasticity is satisfied. The Durbin-Watson test for autocorrelation in the final model (model with an intercept) was conducted. The results are:

Durbin-Watson	Durbin-Watson number of observations
2.108	74

For a two-tailed test at 5 % significance level, the Durbin-Watson critical values for n=70 and k=2 are $d_L = 1.554$, $4 - d_L = 2.446$, $d_U = 1.672$, and $4 - d_U = 2.328$ (Gujarati, 2004). This shows that the data has no autocorrelation, suggesting that the independence assumption of the residuals is satisfied. Standardized residuals, leverages and Cook's Distances were used to check the existence of influential observations in the data for the fitted model. Observation 75 was identified as an outlier ($|\text{standardized residual}| > 2$), a high leverage point ($\text{leverage } > \frac{2\times k+2}{n} = \frac{2\times 2+2}{79} = 0.0759$) and an influential point (Cook's Distance $> \frac{4}{n} = \frac{4}{79} = 0.0506$).

Observation 49 identified as an outlier ($|\text{standardized residual}| > 2$) and observations 3 and 22 identified as points of high leverages ($\text{leverages} > 0.0759$) were influential points (Cook's Distances > 0.0506). Observation 28 was classified as just an influential point (Cook's Distance > 0.0506). All the outliers and points of high leverages were influential points. The five influential observations were removed from the data. A Cook's Distance plot is presented in Figure 4.18 for the final principal components regression model.

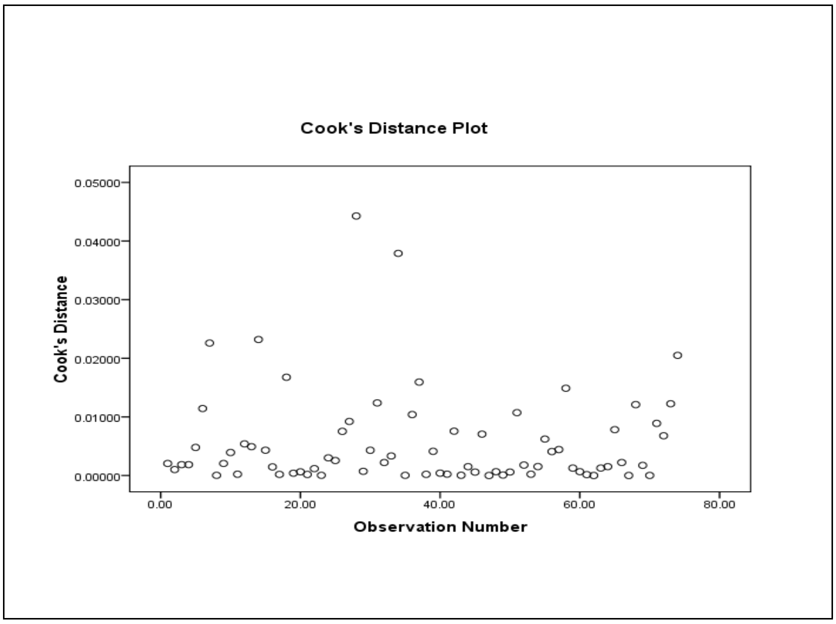

Figure 4.18. Cook's Distance plot for the final principal components regression model

The Cook's Distance values of the observations in the data set are less than 0.0506 (Figure 4.18). This indicates that none of the observations were influential on the parameter estimates of the final model. Table 4.16 presents the final and validation principal components regression models for predicting log (ROCE).

Table 4.16. Principal components regression model for log (ROCE) with standard errors of estimates in parentheses

Parameters	**Final**		**Validation**	
	Estimates	**P-values**	**Estimates**	**P-values**
Intercept	0.794 (0.066)	0.000	0.802 (0.084)	0.000
PC1	-1.443$\times 10^{-11}$ (0.000)	0.007	-2.764$\times 10^{-11}$ (0.000)	0.075
PC3	-0.007 (0.001)	0.000	-0.005(0.001)	0.000
Regression Equation Statistics				
	R^2	0.544	R^2	0.478
	R^2_{adj}	0.532	R^2_{adj}	0.465
	MSE	0.167	MSEP	0.368
	VIF_{max}	1.045	VIF_{max}	1.035

The signs of parameter estimates and values of R^2, R^2_{adj}, MSE, parameter estimates, standard error of parameter estimates, and maximum VIF for principal components; between the final model and validation model are generally similar (Table 4.16). However, there is insignificance of the coefficient for PC1 in the validation model (Table 4.16). The percentage differences between estimates of PC1 and PC3 are 91.55 % and 28.57 %, respectively. Overall, the prediction model is accurate. The next section presents the application of generalized least squares regression to the ROCE data.

4.1.4 Generalized Least Squares Regression

The Positive ROCE Data Set

The stepwise directed search regression procedure was applied to the raw/original positive ROCE data; and the results were used to construct classical multiple linear regression and generalized least squares regression models. Table A.4 in Appendix A presents the summary and change statistics results of each variable in the fitted regression models. The results of fitting classical multiple linear regression and generalized least squares regression models are displayed in Table A.5 in Appendix A. The Durbin-Watson results for testing autocorrelation in the fitted classical multiple linear regression model (model with an intercept) are:

Durbin-Watson	Durbin-Watson number of observations
1.933	278

For a two-tailed test at 5 % significance level, the Durbin-Watson critical values for k=4 and n=200 are $d_L = 1.728$, $4 - d_L = 2.272$, $d_U = 1.809$, and $4 - d_U = 2.191$ (Gujarati, 2004). This implies that the data has no autocorrelation. Thus, the data satisfies the independence assumption of the residuals. Figure B.3 in Appendix B, presents the scatter plot of the standardized residuals versus the standardized predicted values of the fitted classical multiple linear regression model. The scatter plot in Figure B.3 shows signs of a systematic pattern which implies that the data violated the assumption of homogeneity of variance. However, the models for the classical multiple linear regression and generalized least squares regression fits are in perfect agreement (Table A.5). Thus, generalized least squares regression is not sensitive to heteroscedasticity. To correct the anomaly, weighted generalized least squares regression model was fitted. The results of fitting classical multiple linear regression model and weighted generalized least squares regression model are displayed in Table 4.17.

Table 4.17. Classical multiple linear regression and weighted generalized least squares regression models for ROCE with standard errors of estimates in parentheses

Fitted				
Parameter	**Classical multiple linear regression**		**Weighted generalized least squares regression**	
	Estimate	**P-value**	**Estimate**	**P-value**
ROE	0.857 (0.008)	0.000	0.729 (0.014)	0.000
D/E	-0.091 (0.020)	0.000	-0.075 (0.014)	0.000
DY	-0.120 (0.034)	0.001	0.025 (0.012)	0.032
CE	-4.184$\times 10^{-8}$ (0.000)	0.001	-1.946$\times 10^{-8}$ (0.000)	0.000
Regression Equation Statistics				
	R^2	0.976	R^2	0.920
	R^2_{adj}	0.976	R^2_{adj}	0.919
	MSE	79.506	MSE	0.044
	VIF_{max}	1.054	VIF_{max}	Not applicable
Validation				
	R^2	0.914	R^2	0.921
	R^2_{adj}	0.912	R^2_{adj}	0.920
	MSEP	107.995	MSEP	0.045

The Durbin-Watson results for testing autocorrelation in the fitted classical multiple linear regression model (model with no intercept) are:

Durbin-Watson	Durbin-Watson number of observations
1.841	278

For a two-tailed test at 5 % significance level, the Durbin-Watson critical values for k=4 and n=200 are $d_L = 1.624$, $4 - d_L = 2.376$, $d_U = 1.718$, and $4 - d_U = 2.282$ (Gujarati, 2004). This implies that the data has no autocorrelation. Thus, the data satisfies the independence assumption of the residuals. Figure 4.19 shows the scatter plot of the standardized residuals versus the standardized predicted values of the fitted classical multiple linear regression model.

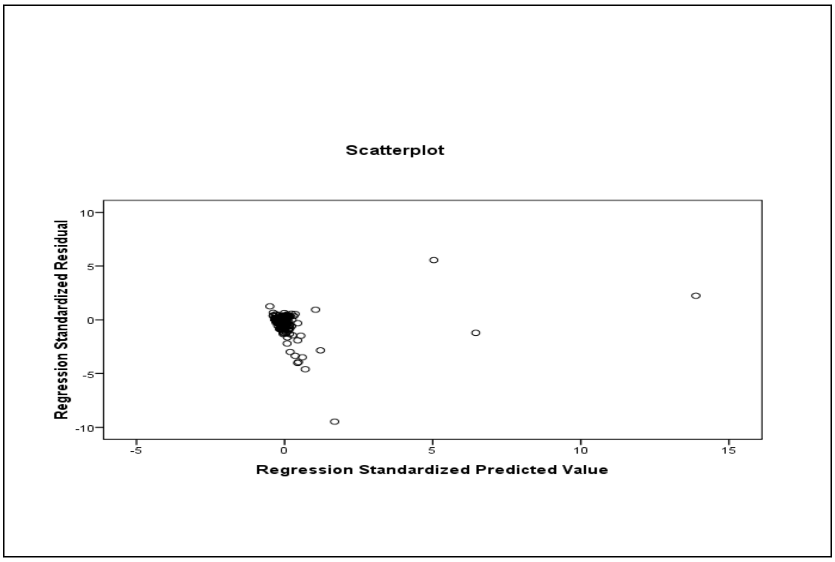

Figure 4.19. Scatter plot of standardized residuals versus standardized predicted values of the fitted classical multiple linear regression model

The plot in Figure 4.19 shows that there is increasing residual with increasing predicted value. This testifies that the homoscedasticity assumption is invalid in the fitted classical multiple linear regression model.

The percentage differences in OLS estimates and weighted generalized least square estimates of ROE, D/E, DY, and CE are 14.94 %, 17.58 %, 116 %, and 53.49 %, respectively (Table 4.17). The standard error values of weighted generalized least square estimates are generally less than those for OLS estimates (Table 4.17). Overall, the weighted generalized least square regression prediction model appears to fit the data considerably better than the classical multiple linear regression prediction model, with the largest difference noticed in the MSE statistic (Table 4.17). The weighted generalized least square regression validation model is considerably better than the classical multiple linear regression validation model, with the largest difference noticed in the MSEP statistic (Table 4.17). The differences between the models for the classical multiple linear regression and weighted generalized least square regression fits for the positive ROCE data set are due to the heteroscedasticity in the data (Figure 4.19).

The Negative ROCE Data Set

The stepwise directed search regression procedure was applied to the raw/original negative ROCE data. The modulus values of the negative ROCE were used. The summary and change statistics results of each variable in the fitted regression models are presented in Table A.6 in Appendix A. The stepwise directed search results were used to construct classical multiple linear regression and generalized least squares regression models. The results of fitting classical multiple linear regression and generalized least squares regression models are shown in Table A.7 in Appendix A. The Durbin-Watson results for testing autocorrelation in the fitted classical multiple linear regression model (model with an intercept) are:

Durbin-Watson	Durbin-Watson number of observations
1.913	79

For a two-tailed test at 5 % significance level, the Durbin-Watson critical values for k=2 and n=75 are $d_L = 1.571$, $4 - d_L = 2.429$, $d_U = 1.680$, and $4 - d_U = 2.320$ (Gujarati, 2004). This indicates that the data has no autocorrelation, suggesting that the independence assumption of the residuals is fulfilled. Figure 4.20 presents the scatter plot of the standardized residuals versus the standardized predicted values for the fitted classical multiple linear regression model.

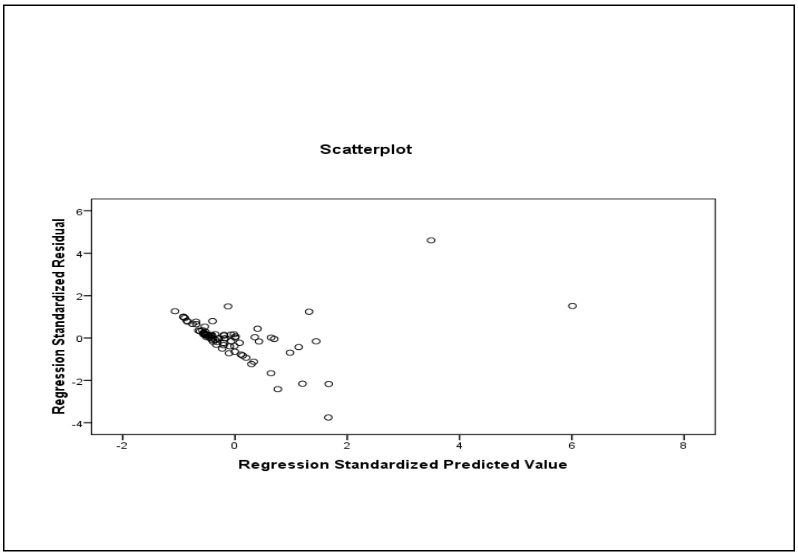

Figure 4.20. Scatter plot of standardized residuals versus standardized predicted values of the fitted classical multiple linear regression model

The plot in Figure 4.20 shows that there is increasing variance with increasing predicted value. This confirms that the homoscedasticity assumption is not satisfied. However, the models for the classical multiple linear regression and generalized least squares regression fits are identical (Table A.7), suggesting that generalized least squares regression is not sensitive to heteroscedasticity. The weighted generalized least squares regression model was fitted so as to correct the anomaly. The results of fitting weighted generalized least squares regression model are shown in Table 4.18.

Table 4.18. Weighted generalized least squares regression model for ROCE with standard errors of estimates in parentheses

Parameter	Fitted: Weighted generalized least squares regression	
	Estimate	P-value
Intercept	-20.144 (4.966)	0.000
APCE	17.827 (3.538)	0.000
ROA	1.472 (0.175)	0.000
Regression Equation Statistics		
	R^2	0.559
	R^2_{adj}	0.548
	MSE	57.125
	VIF_{max}	Not applicable
Validation		
	R^2	0.078
	R^2_{adj}	0.055
	MSEP	387.985

The percentage differences in OLS estimates and weighted generalized least squares estimates of APCE and ROA are 73.37 % and 214.02 %, respectively (Tables A.7 and 4.18). The standard error values of weighted generalized least squares estimates are generally less than that for OLS estimates (Tables A.7 and 4.18). The weighted generalized least squares regression prediction model is considerably better than the classical multiple linear regression prediction model, with the largest difference noticed in the MSE statistic (Tables A.7 and 4.18). The weighted generalized least squares regression validation model is considerably better than the classical multiple linear regression validation model, with the largest difference noticed in the MSEP statistic (Tables A.7 and 4.18). The differences between the models for the classical multiple linear regression and weighted generalized least squares regression fits for the negative ROCE data set are due to the heteroscedasticity in the data (Figure 4.20). The application of robust maximum likelihood regression to the ROCE data is illustrated in the next section.

4.1.5 Robust Maximum Likelihood Regression

The Positive ROCE Data Set

The stepwise directed search regression results of the raw/original positive ROCE data were used to construct robust maximum likelihood regression model. Huber objective function was used. The results of fitting robust maximum likelihood regression model are displayed in Table 4.19.

Table 4.19. Robust maximum likelihood regression model for ROCE with standard errors of estimates in parentheses

Fitted		
Parameter	**Robust maximum likelihood regression**	
	Estimate	**P-value**
Intercept	-0.816 (0.214)	0.000
ROE	0.857 (0.003)	0.000
D/E	-0.081 (0.007)	0.000
DY	-0.082 (0.012)	0.000
CE	0.000 (0.000)	0.000
Regression Equation Statistics		
	R^2	0.998
	R^2_{adj}	0.997
	MSE	6.739
	VIF_{max}	Not applicable
Validation		
	R^2	0.995
	R^2_{adj}	0.993
	MSEP	5.499

The percentage differences in OLS estimates and maximum likelihood estimates of ROE, D/E, DY, and CE are 1.039 %, 2.410 %, 5.750 %, and 0 %, respectively (Tables A.5 and 4.19). The standard error values of maximum likelihood estimates are slightly less than those for OLS estimates (Tables A.5 and 4.19). The robust maximum likelihood regression prediction model is slightly better than the classical multiple linear regression prediction model, with the largest difference noticed in the MSE statistic (Tables A.5 and 4.19). The robust maximum likelihood regression validation model is slightly better than the classical multiple linear regression validation model, with the largest difference noticed in the MSEP statistic (Tables A.5 and 4.19). The differences between the models for the classical multiple linear regression and robust maximum likelihood regression fits for the positive ROCE data set are due to the non-normality distribution of ROCE (Figure 4.1) and heteroscedasticity characteristics of the data (Figure B.3).

The Negative ROCE Data Set

The stepwise directed search regression results of the raw/original modulus values of the negative ROCE were used to construct robust maximum likelihood regression model. Huber objective function was used. The results of fitting robust maximum likelihood regression model are shown in Table 4.20.

Table 4.20. Robust maximum likelihood regression model for ROCE with standard errors of estimates in parentheses

	Fitted	
Parameter	**Robust maximum likelihood regression**	
	Estimate	**P-value**
Intercept	-35.061 (3.016)	0.000
APCE	32.142 (1.850)	0.000
ROA	-1.171 (0.039)	0.000
Regression Equation Statistics		
	R^2	0.990
	R_{adj}^2	0.984
	MSE	98.883
	VIF_{max}	Not applicable
Validation		
	R^2	0.999
	R_{adj}^2	0.993
	MSEP	121.661

The percentage differences in OLS estimates and maximum likelihood estimates of APCE and ROA are 51.99 % and 9.30 %, respectively (Tables A.7 and 4.20). The standard error values of maximum likelihood estimates are far less than those for OLS estimates (Tables A.7 and 4.20). The robust maximum likelihood regression prediction model is considerably better than the classical multiple linear regression prediction model, with the largest difference noticed in the MSE statistic (Tables Tables A.7 and 4.20). The robust maximum likelihood regression validation model is considerably better than the classical multiple linear regression validation model with largest differences noticed in the R^2, R^2_{adj} and MSEP statistics (Tables A.7 and 4.20). The appreciable large differences between the models for the classical multiple linear regression and robust maximum likelihood regression fits for the negative ROCE data set are due to the non-normality distribution of ROCE (Figure 4.2) and heteroscedasticity characteristics of the data (Figure 4.20). The next section gives a discussion of the results.

4.2 Discussion of Results

4.2.1 Classical Multiple Linear Regression Analysis

The Positive ROCE Data Set

The results are:

1) There is positive elasticity of ROCE with respect to change in ROE, and are negative elasticities of ROCE with respect to changes in D/E and CE.

2) The lead key determinant of ROCE is ROE, followed by D/E and CE.

This indicates that practitioners should maximize ROE, and minimize D/E and CE in order to generate maximum profit. Given that ROE has a positive elasticity and thus more powerful in maximizing profit, practitioners should consider prioritizing ROE in crafting marketing strategies.

The Negative ROCE Data Set

The results are:

1) There is positive effect on ROCE with respect to change in APCE, and are negative effects on ROCE with respect to changes in ROA and MC.

2) The lead key determinant of ROCE is APCE, followed by ROA and MC comes far much later.

This shows that practitioners should minimize ROA and MC, and maximize APCE in order to minimize risk of making loss. Given that ROA and MC have negative effects and thus more influential in minimizing risk, practitioners should prioritize ROA and MC in crafting marketing strategies.

Comparing with Previous Studies

The results are more and less precise than those found by previous studies (Aanu et al., 2014; Azhagaih and Gavoury, 2011; Mohamad and Saad, 2010; Nimalathasan and Brabete, 2010; Ray, 2011; Sulait, 2010), indicating that ROCE is quite difficult to predict. This suggests that the key determinants are also important sources of variability in ROCE of individual companies that the practitioners need to work with.

4.2.2 Principal Components Regression

The Positive ROCE Data Set

The results are:

1) There are positive effects on ROCE with respect to changes in all key determinants.

2) The lead key determinant of ROCE is PC4 and followed closely by PC2.

Practitioners should maximize and prioritize all key determinants in order to generate maximum profit and when crafting marketing strategies, respectively.

The Negative ROCE Data Set

The results are:

1) There are negative effects on ROCE with respect to changes in all key determinants.

2) The lead key determinant of ROCE is PC3 and followed by PC1.

Practitioners should minimize and prioritize all key determinants in order to minimize risk of making loss and when crafting marketing strategies, respectively.

Comparing with Classical Multiple Linear Regression Analysis

In the discussion, the results indicate less precision of the principal components regression models, reflecting that using principal components as independent variables do not improve classical multiple linear regression model prediction for the two data sets. This suggests that the principal components regression models have led to accounting less of the total variation, highlighting that the key determinants are less important sources of variability in ROCE of individual companies.

4.2.3 Generalized Least Squares Regression

In the discussion, the results indicate more precision of the weighted generalized least squares regression model than the classical multiple linear regression model, signifying that improvements in modelling ROCE using weighted generalized least squares regression is possible for the positive and negative data sets. This suggests that the independent variables in the weighted generalized least squares regression models are important sources of variability in ROCE of individual companies that the practitioners need to work with.

The Positive ROCE Data Set

The results are:

1) There is positive effect on ROCE with respect to change in ROE and DY, and are negative effects on ROCE with respect to changes in D/E and CE.

2) The lead key determinant of ROCE is ROE, followed by DE, DY, and CE.

This indicates that practitioners should maximize ROE and DY, and minimize D/E and CE in order to generate maximum profit. Thus practitioners should prioritize ROE and DY in crafting marketing strategies.

The Negative ROCE Data Set

The results are:

1) There is positive effect on ROCE with respect to changes in APCE and ROA.

2) The lead key determinant of ROCE is APCE and ROA comes far much later.

Practitioners should maximize all key determinants in order to minimize risk of making loss. Thus practitioners should not prioritize key determinants when crafting marketing strategies.

4.2.4 Robust Maximum Likelihood Regression

The differences between the classical multiple linear regression model and robust maximum likelihood regression model for the positive ROCE data set and negative ROCE data set, point to the danger of using classical multiple linear regression model when some of the assumptions are not satisfied. This leads to wrong decision making about the effects of variables upon ROCE resulting in incorrect levels of company performance. Maximum likelihood estimates have a higher breakdown point; and are more consistent, sufficient and efficient than OLS estimates; making robust maximum likelihood regression more suitable for modelling ROCE than classical multiple linear regression. In the discussion, the results show that improving classical multiple linear regression using robust maximum likelihood regression in modelling ROCE is very possible for the two data sets. This reflects that the independent variables in the robust maximum likelihood regression models are major sources of variability of individual companies contributing to ROCE, that the practitioners need to work with.

The Positive ROCE Data Set

The results are:

1) There is positive effect on ROCE with respect to change in ROE, and are negative effects on ROCE with respect to changes in D/E and DY.

2) The lead key determinant of ROCE is ROE, followed by DY and D/E comes very closely.

This indicates that practitioners should maximize ROE, and minimize D/E and DY in order to generate maximum profit. Thus practitioners should prioritize ROE in crafting marketing strategies.

The Negative ROCE Data Set

The results are:

1) There is positive effect on ROCE with respect to change in APCE and negative effect on ROCE with respect to change in ROA.

2) The lead key determinant of ROCE is APCE and ROA comes far much later.

Practitioners should minimize ROA and maximize APCE in order to minimize risk of making loss. Thus practitioners should prioritize ROA in crafting marketing strategies.

We provide a summary of this chapter in the next section.

4.3 Summary

Classical multiple linear regression analysis reflected that:

- The final model has a R^2 of 0.928, an R^2_{adj} of 0.927, a MSE of 0.013, and maximum VIF of 1.036 for the positive data set.
- There is positive elasticity of ROCE with respect to change in ROE, and are negative elasticities of ROCE with respect to changes in D/E and CE for the positive data set.
- The lead key determinant of ROCE is ROE, followed by D/E and CE for the positive data set.
- The final model has a R^2 of 0.666, an R^2_{adj} of 0.652, a MSE of 0.149, and maximum VIF of 1.025 for the negative data set.
- There is positive effect on ROCE with respect to change in APCE, and are negative effects on ROCE with respect to changes in ROA and MC for the negative data set.
- The lead key determinant of ROCE is APCE, followed by ROA and MC for the negative data set.
- ROCE is difficult to predict.

Principal components regression revealed that:

- The final model has a R^2 of 0.929, an R^2_{adj} of 0.929, a MSE of 0.069, and maximum VIF of 1.064 for the positive data set.
- There are positive effects on ROCE with respect to changes in all key determinants for the positive data set.

- The lead key determinant of ROCE is PC4 and followed by PC2 for the positive data set.
- The final model has a R^2 of 0.544, an R^2_{adj} of 0.532, a MSE of 0.167, and maximum VIF of 1.045 for the negative data set.
- There are negative effects on ROCE with respect to changes in all key determinants for the negative data set.
- The lead key determinant of ROCE is PC3 and followed by PC1 for the negative data set.
- Using principal components as independent variables do not improve classical multiple linear regression model prediction for the two data sets.

Generalized least squares regression showed that:

- Improving classical multiple linear regression using weighted generalized least squares regression in modelling ROCE is possible for the positive and negative data sets.
- The weighted generalized least squares regression model has a R^2 of 0.920, an R^2_{adj} of 0.919 and a MSE of 0.044 for the positive data set.
- There is positive effect on ROCE with respect to changes in ROE and DY, and are negative effects on ROCE with respect to changes in D/E and CE for the positive data set.
- The lead key determinant of ROCE is ROE, followed by DE, DY and lastly by CE for the positive data set.
- The weighted generalized least squares regression model has a R^2 of 0.559, an R^2_{adj} of 0.548 and a MSE of 57.125 for the negative data set.

- There is positive effect on ROCE with respect to changes in APCE and ROA for the negative data set.
- The lead key determinant of ROCE is APCE and followed by ROA for the negative data set.

Robust maximum likelihood regression indicated that:

- Improving classical multiple linear regression using robust maximum likelihood regression in modelling ROCE is possible for the positive and negative data sets.
- The robust maximum likelihood regression model has a R^2 of 0.998, an R^2_{adj} of 0.997 and a MSE of 6.739 for the positive data set.
- There is positive effect on ROCE with respect to change in ROE, and are negative effects on ROCE with respect to changes in D/E and DY for the positive data set.
- The lead key determinant of ROCE is ROE, followed by DY and lastly by D/E for the positive data set.
- The robust maximum likelihood regression model has a R^2 of 0.990, an R^2_{adj} of 0.984 and a MSE of 98.883 for the negative data set.
- There is positive effect on ROCE with respect to change in APCE and negative effect on ROCE with respect to change in ROA for the negative data set.
- The lead key determinant of ROCE is APCE and followed by ROA for the negative data set.

Chapter 5 is the overall conclusion of the dissertation.

CHAPTER 5

CONCLUDING REMARKS

This dissertation centred on identifying and quantifying key determinants of Return on Capital Employed (ROCE) of individual companies listed on the Johannesburg Stock Exchange (JSE), by comparing classical multiple linear regression analysis, principal components regression, generalized least squares regression, and robust maximum likelihood regression approaches. The procedures of the different regression approaches were discussed.

For the positive ROCE data set, the classical multiple linear regression model had a R^2 of 0.928, an R^2_{adj} of 0.927, a MSE of 0.013, maximum VIF of 1.036, and the lead key determinant was ROE with positive elasticity, followed by D/E and CE, both with negative elasticities. For the negative ROCE data set, the classical multiple linear regression model had a R^2 of 0.666, an R^2_{adj} of 0.652, a MSE of 0.149, maximum VIF of 1.025, and the lead key determinant was APCE with positive effect, followed by ROA and MC, both with negative effects.

For the positive ROCE data set, the principal components regression model had a R^2 of 0.929, an R^2_{adj} of 0.929, a MSE of 0.069, maximum VIF of 1.064, and the lead key determinant was PC4 (log (ROA), log (ROE), log (OPM)) and followed by PC2 (log (EY), log (P/E)), both with positive effects. For the negative ROCE data set, the principal components regression model had a R^2 of 0.544, an R^2_{adj} of 0.532, a MSE of 0.167, maximum VIF of 1.045, and the lead key determinant was PC3 (ROA, EY, APCE) and followed by PC1 (MC, CE), both with negative effects. In this dissertation, we worked with only two principal components selection methods. It is possible to use diverse principal components selection methods in order to explore the accuracy and reliability of each method (Fekedulegn et al., 2002; Maitra and Yan, 2008). Further investigation of the sources of variation on ROCE could be pursued by exploring other shrinkage regression methods. A comparative analysis of principal components regression, partial least squares regression, ridge regression, least absolute shrinkage and selection operator regression, and elastic net regression methods can be performed to ascertain their efficacy. The superior method is adopted based on R^2, R_{adj}^2 and MSE.

For the positive ROCE data set, the weighted generalized least squares regression model had a R^2 of 0.920, an $\mathrm{R}_{\mathrm{adj}}^2$ of 0.919, a MSE of 0.044, and the lead key determinant was ROE with positive effect, followed by DE with negative effect, DY with positive effect and lastly CE with negative effect. For the negative ROCE data set, the weighted generalized least squares regression model had a R^2 of 0.559, an R_{adj}^2 of 0.548, a MSE of 57.125, and the lead key determinant was APCE and followed by ROA, both with positive effects. Thorough data analyses are essential in order to investigate the validity of these findings.

For the positive ROCE data set, the robust maximum likelihood regression model had a R^2 of 0.998, an $\mathrm{R}_{\mathrm{adj}}^2$ of 0.997, a MSE of 6.739, and the lead key determinant was ROE with positive effect, followed by DY and lastly D/E, both with negative effects. For the negative ROCE data set, the robust maximum likelihood regression model had a R^2 of 0.990, an R_{adj}^2 of 0.984, a MSE of 98.883, and the lead key determinant was APCE with positive effect and followed by ROA with negative effect. The work presented here was restricted to the general approach of robust regression method. Future research work could focus on higher breakdown point, more efficient and more equivariant robust; and nonparametric regression methods, such as, least absolute deviation, least trimmed squares, least median of squares, modified maximum likelihood, Theil, weighted Theil, and Winsorized least squares as described in Koller and Stahel (2011), and Mutan (2004). Simulation studies can be carried out in order to establish the behaviours of these regression estimators.

The findings showed that the use of robust maximum likelihood regression produced more accurate/precise results compared to those obtained using the three competing approaches, because it has a higher breakdown point and is more consistent, sufficient and efficient. Besides that, the use of robust maximum likelihood regression was considered less complex due to no or minimal conditions. However, assessment of models performance suggested that the constructed models can also be used for identification and quantification of key determinants of ROCE. This study's findings suggest that companies management can establish and control proper marketing strategies using the key determinants and results of these strategies can see an improvement in ROCE.

BIBLIOGRAPHY

Aanu, O.S., Odianonsen, I.F. and Foyeke, O.I. (2014). Effectiveness of Audit Committee and Firm Financial Performance in Nigeria: An Empirical Analysis. *Journal of Accounting and Auditing: Research and Practice*, Vol. 2014: No. 301176.

Andreica, M.E. (2009). Predicting Romanian Financial Distressed Companies. *Academy of Economic Studies, Bucharest.*

Arafat, N.M. and Shahimi, M. (2013). An Exploratory of the Intangible Assets: Methods of Measuring Intellectual Capital. *International Journal of Humanities and Applied Sciences,* Vol. 2: No. 2.

Azhagaiah, R. and Gavoury, C. (2011). The Impact of Capital Structure on Profitability with Special Reference to IT Industry in India. *Managing Global Transitions*, 9:371-392.

Bowerman, B.L. and O'Connel, R.T. (1990). Linear Statistical Models: An Applied Approach, Second Edition. *PWS-Kent Publishing Company*.

Cameron, C.P. (2011). An Investigation into Ownership Concentration and Financial Performance of Listed South African Industrial Companies. *Gordon Institute of Business Science, University of Pretoria. Pretoria.*

Chen, C. (2002). Robust Regression and Outlier Detection with the ROBUSTREG Procedure. SUGI Paper 265-27. *SAS Institute Inc. Cary, N.C.*

Comincioli, N., Poddi, L. and Vergalli, S. (2012). Does Corporate Social Responsibility Affect the Performance of Firms? *Universities of Brescia and Ferrara.*

Cook, R.D. (2007). Fisher Lecture: Dimension Reduction in Regression. *Statistical Science,* 22: 1-26.

Damodaran, A. (2007). Return on Capital, Return on Invested Capital and Return on Equity: Measurement and Implications. *Stern School of Business.*

Draper, N.R. and Smith, H. (1998). Applied Regression Analysis, Third Edition. *John Wiley and Sons, Inc. New York.*

Econex and Quantec Research (2010). The Contribution of the South African Breweries to the South African Economy. *Working for South Africa.*

Eljelly, M.A.A. (2009). Ownership and firm performance: The experience of Saudi Arabia's emerging economy. *International Business and Economics Research Journal*, Vol. 8: No. 8.

Erasmus, N. (2013). Dividend Yield Investment Strategies in the South African Stock Market. *University of the Witwatersrand. Johannesburg*.

Faraway, J. J. (2002). Practical regression and ANOVAs using R.

Fekedulegn, B.D., Colbert, J.J., Hicks, R.R. and Schuckers, M.E. (2002). Coping with Multicollinearity: An Example on Application of Principal Components Regression in Dendroecology. *USDA Forest Service.*

Fielding, J.L. and Gilbert, G.N. (2000). Understanding social statistics. *Sage Publications Ltd*.

Fox, J. (2002). Robust Regression. *Appendix to an R and S-PLUS Companion to Applied Regression.*

Fox, J. and Weisberg, S. (2010). Time-Series Regression and Generalized Least Squares in R. An *Appendix to an R Companion to Applied Regression*.

Frost, J. (2013). How to Interpret Regression Analysis Results. *The Minitab Blog*.

Grosswindhager, S. (2009). Using Penalized Logistic Regression Models for Predicting the Effects of Advertising Material. *Vienna University of Technology*.

Gujarati, D.N. (2004). Basic Econometrics, Fourth Edition. *The McGraw-Hill Companies.*

Hair, Jr. J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010). Instructor's

Manual for Multivariate Data Analysis: A Global Perspective, Seventh Edition. *Pearson.*

Hillestad, O.C. (2007). An analysis of financial ratios for the Oslo Stock Exchange. *Economic Bulletin* No.3.

Hobarth, L.L. (2006). Modeling the relationship between financial indicators and company performance. An empirical study for US-listed companies. *WU Vienna University of Economics and Business.*

Holsclaw, T.N. (2007). Investigation of Repeated Measures Linear Regression Methodologies. *San Jose State University.*

Howell, E. (2013). Robust regression procedures to handle outliers. *Southern Methodist University.*

Howarth, R.J. (2001). A history of regression and related model-fitting in the earth sciences. *Natural Resources Research*,10: 241-286.

Hussien, Y. (2010). Modeling the Factors Affecting Cereal Crop Yields in the Amhara National Regional State of Ethiopia. *University of KwaZulu-Natal. Pietermaritzburg.*

Johannesburg Stock Exchange (2010). JSE Limited Annual Report 2010, viewed 10 October 2014, http://phx.corporate-ir.net/phoenix.html?c=1981&p=irol-reportsannual.

Johannesburg Stock Exchange (2011). JSE Limited Integrated Annual Report 2011, viewed 17 October 2014, http://financialresults.co.za/2012/jse_ar2011/downloads/shareholder_report_2011.pdf.

Kan, R. and Robotti, C. (2009). A Note on the Estimation of Asset Pricing Models Using Simple Regression Betas. *Federal Reserve Bank of Atlanta: Working Paper Series.*

Karlsson, L. (2002). Garch-Modelling. Theoretical Survey, Model Implementation and Robustness Analysis. *Royal Institute of Technology.*

Kariya, T. and Kurata, H. (2004). Generalized Least Squares. *John Wiley and Sons.*

Koller, M. (2007). Robust Statistics: Tests for Robust Linear Regression. *Swiss Federal Institute of Technology. Zurich.*

Koller, M. and Stahel, W.A. (2011). Sharpening Wald-type Inference in Robust Regression for Small Samples. *Swiss Federal Institute of Technology. Zurich*.

Li, Y. (2010). A Comparison Study of Principle Component Regression, Partial Least Squares Regression and Ridge Regression with Application to Fourier Transform Infrared (FTIR) Data. *Uppsala University.*

Lokshin, M. and Sajaia, Z. (2004). Maximum Likelihood Estimation of Endogenous Switching Regression Models. *The Stata Journal* 4: 282-289.

MacLeod, R.H. (2008). A Guide to Listing Companies on the JSE Limited. *Business Blue-Book of South Africa.*

Ma, S. (2007). Principal Component Analysis in Linear Regression Survival Model with Microarray Data. *Journal of Data Science* 5: 183-198.

Maitra, S. and Yan, J. (2008). Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression. *Casualty Actuarial Society.*

Mandal, N., Mahavidyalaya, B.N. and Goswami, S. (2010). Impact of Working Capital Management on Liquidity, Profitability and Non-Insurable Risk and Uncertainty Bearing: A Case Study of Oil and Natural Gas Commission. *Great lakes Herald,* Vol.4: No.2.

Mkhize, H. and Mbanga, P.M. (2006). A critical review of the restructuring of the South African market. *International Review of Business Research Papers,* 2: 80-91.

Mohamad, N.E.A.B. and Saad, N.B.M. (2010). Working Capital Management: The Effect of Market Valuation and Profitability in Malaysia. *International Journal of Business and Management,* Vol.5: No.11.

Motyka, M.S. (2003). Risk Measurement of Mortgage-Backed Security Portfolios via Principal Components and Regression Analyses. *Worcester Polytechnic Institute.*

Mutan, O.C. (2004). Comparison of Regression Techniques via Monte Carlo. *Middle East Technical University.*

Myung, I.J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology,* 47: 90-100.

Nimalathasan, B. and Brabete, V. (2010). Capital structure and its impact on profitability: A study of listed manufacturing companies in Sri Lanka. *The Young Economists Journal*, 1: 7-16.

Orsini, N., Bellocco, R. and Greenland, S. (2006). Generalized Least Squares for Trend Estimation of Summarized Dose-Response Data. *The Stata Journal*, 6: 40-57.

Pattabiraman, S. (2013). Growth vs Profitability: The Importance of ROCE. *CFO Connect*.

Person, C. (2011). Predicting Recycling Efficiency-Multiple Regression Modelling of the Recycling Rates of Tetra Pak's Beverage Cartons. *University of Lund.*

Pierce, M.W., Thornton, C.I. and Abt, S.R. (2010). Predicting Peak Outflow From Breached Embankment Dams. *Colorado State University. Fort Collins. Colorado*.

Ray, S. (2011). Financial Performance of Paper and Paper Product Companies in India in Post-Liberalization Period: An Exploratory Study. *Research Journal of Finance and Accounting,* Vol 2: No.5.

Ryan, T.P. (2008). Modern Regression Methods, Second Edition. *John Wiley and Sons, Inc.*

Sabbagh, Y. (2003). Maximum Likelihood Estimation of Hammerstein Models. *Linkoping University.*

Shaffer, L. B. (2007). Examining Regression Analysis Beyond the Mean of the Distribution using Quantile Regression. *University of Tennessee. Knoxville*.

Sharma, S. (2011). Determinants of Equity Share Prices in India. *International Refereed Research Journal,* Vol 2: Issue 4.

Shirgaokar, A. (2009). Principal Component Regression Models for Thermo-Mechanical Reliability of Plastic Ball Grid Arrays on Cu-Core and No Cu-Core PCB Assemblies in Harsh Environments. *Auburn University. Alabama.*

Shobhana, V.K. and Karpagavalli, R. (2011). Determinants of Market Price of Shares of the Select Banking Companies Listed at Bombay Stock Exchange. *Academia: South Asian Academic Research Journals*, Vol 1: Issue 3.

Shourvarzi, M.R. and Sadeddin, S. (2011). Analysis of the predictive ability of the components of economic value added in predicting next period's operating profit: Evidence from Iran. *African Journal of Business Management*, 5:3251-3258.

Singh, J. and Yadav, P. (2013). Return on Capital Employed-A Tool for Analyzing Profitability of Companies. *International Journal of Techno-Management Research*, Vol 1: Issue 1.

Smadi, A.A. and Abu-Afouna, N.H. (2012). On Least Squares Estimation in a Simple Linear Regression Model with Periodically Correlated Errors: A Cautionary Note. *Austrian Journal of Statistics*, 41: 211-226.

Soosova, K. (2005). A Predictive Model for Event Driven Hedge Fund Returns. *University of Zurich*.

Sopipan, N., Kanjanavajee, W. and Sattayatham, P. (2012). Forecasting SET50 Index with Multiple Regression based on Principal Component Analysis. *Journal of Applied Finance and Banking,* Vol 2: No.3.

Steyn, J. P. (2012). Using Capital Intensity and Return on Capital Employed as filters for security selection. *Stellenbosch University*.

Sulait, T. (2007). Relational Capital and Firm Performance: A Case of Manufacturing Tea Firms in Uganda. *Makerere University*.

Tran, N.M. (2008). An Introduction to Theoretical Properties of Functional Principal Component Analysis. *The University of Melbourne.*

Tudor, C. (2009). Price Ratios and the Cross-Section of Common Stock Returns on Bucharest Stock Exchange: Empirical Evidence. *Romanian Journal of Economic Forecasting, Issue 2.*

Ul-Saufie, A.Z., Yahya, A.S. and Ramli, N.A. (2011). Improving Multiple Linear Regression Model using Principal Component Analysis for Predicting Particulate Matter Concentration in Seberang Prai, Pulau Pinang. *International Journal of Environment Sciences*, Vol. 2: No.2.

Waterman, M. J. T. (2002). Linear Mixed Model Robust Regression. *Virginia State University. Blacksburg.*

Wilcox, S.E. (2007). The Adjusted Earnings Yield. *Financial Analysts Journal,* Vol. 63: No. 5.

World Federation of Exchanges (2012). Annual Report and Statistics, viewed 7 May 2014, http://www.world-exchanges.org/world_federation_of_exchanges/2012_annual report and statistics.htm.

http://www.mcgregorbfa.com, viewed 5 February 2012.

APPENDIX A

ADDITIONAL TABLES

Table A.1. Stepwise directed search regression model summary and change statistics for the transformed positive ROCE data set

Variables as retained by stepwise directed search	R^2	R^2_{adj}	MSE	R^2 Change	F Change	DF1	DF2	P-value Change
ROE	0.255	0.253	0.147	0.255	94.620	1	276	0.000
D/E	0.297	0.292	0.139	0.042	36.605	2	275	0.000
CE	0.324	0.317	0.135	0.027	14.247	3	274	0.000
EPS	0.366	0.357	0.127	0.042	4.398	4	273	0.000
DY	0.386	0.375	0.123	0.020	5.141	5	272	0.000

Table A.2. Correlation coefficients and the associated p-values in parentheses for the transformed negative ROCE data set

	Log ROCE	APCE	D/A	D/E	DY	EPS	EY	OPM	P/E	ROA	ROE	CE	MC	SP
Log ROCE	1	0.407 (0.00)	0.310 (0.01)	0.045 (0.70)	0.104 (0.36)	-0.150 (0.19)	-0.492 (0.00)	-0.018 (0.88)	0.218 (0.054)	-0.633 (0.00)	0.001 (0.995)	-0.336 (0.00)	-0.369 (0.00)	-0.223 (0.049)
APCE		1	0.317 (0.00)	-0.018 (0.89)	-0.024 (0.83)	-0.010 (0.93)	-0.356 (0.00)	0.065 (0.57)	0.080 (0.49)	-0.145 (0.20)	-0.021 (0.86)	-0.006 (0.96)	-0.046 (0.69)	-0.129 (0.26)
D/A			1	-0.216 (0.06)	-0.041 (0.72)	0.014 (0.91)	-0.379 (0.00)	0.214 (0.06)	0.186 (0.10)	-0.099 (0.38)	0.245 (0.03)	-0.078 (0.49)	-0.159 (0.16)	-0.126 (0.27)
D/E				1	-0.008 (0.94)	0.021 (0.86)	0.261 (0.02)	-0.023 (0.84)	-0.030 (0.79)	-0.007 (0.95)	-0.501 (0.00)	0.060 (0.60)	0.042 (0.72)	-0.124 (0.28)
DY					1	-0.191 (0.09)	0.014 (0.90)	0.000 (0.996)	0.302 (0.01)	-0.008 (0.95)	-0.001 (0.991)	-0.244 (0.03)	-0.033 (0.78)	0.104 (0.37)
EPS						1	0.096 (0.40)	-0.023 (0.84)	-0.017 (0.88)	-0.002 (0.99)	-0.080 (0.48)	0.158 (0.16)	0.091 (0.42)	-0.313 (0.01)
EY							1	-0.017 (0.89)	-0.062 (0.59)	0.295 (0.01)	-0.549 (0.00)	0.104 (0.36)	0.151 (0.18)	0.239 (0.03)
OPM								1	-0.045 (0.70)	0.064 (0.57)	0.06 (0.60)	0.049 (0.67)	-0.092 (0.42)	-0.077 (0.50)
P/E									1	-0.040 (0.73)	0.001 (0.996)	-0.229 (0.04)	-0.351 (0.00)	-0.156 (0.17)

ROA										1	0.047 (0.68)	0.111 (0.33)	0.142 (0.211)	0.098 (0.391)
ROE											1	0.004 (0.97)	0.006 (0.96)	-0.014 (0.90)
CE												1	0.917 (0.00)	0.131 (0.25)
MC													1	0.317 (0.00)
SP														1

Table A.3. Stepwise directed search regression model summary and change statistics for the transformed created negative ROCE data set

Variables as retained by stepwise directed search	R^2	R^2_{adj}	MSE	R^2 Change	F Change	DF1	DF2	P-value Change
ROA	0.400	0.393	0.312	0.400	54.420	1	77	0.000
APCE	0.502	0.489	0.262	0.102	13.131	2	76	0.000
MC	0.577	0.560	0.226	0.075	4.237	3	75	0.000

Table A.4. Stepwise directed search regression model summary and change statistics for the raw positive ROCE data set

Variables as retained by stepwise directed search	R^2	R^2_{adj}	MSE	R^2 Change	F Change	DF1	DF2	P-value Change
ROE	0.973	0.973	82.992	0.973	9.931×10^3	1	276	0.000
D/E	0.975	0.974	78.245	0.002	4.655×10^3	2	275	0.000
DY	0.975	0.975	76.701	0.000	1.686×10^3	3	274	0.000
CE	0.976	0.975	75.419	0.001	0.85×10^3	4	273	0.000

Table A.5. Classical multiple linear regression and generalized least squares regression models for ROCE with standard errors of estimates in parentheses for the raw positive ROCE data set

Fitted				
Parameter	**Classical multiple linear regression**		**Generalized least squares regression**	
	Estimate	**P-value**	**Estimate**	**P-value**
Intercept	-2.389 (0.600)	0.000	-2.389 (0.600)	0.000
ROE	0.866 (0.008)	0.000	0.866 (0.008)	0.000
D/E	-0.083 (0.019)	0.000	-0.083 (0.019)	0.000
DY	-0.087 (0.035)	0.013	-0.087 (0.035)	0.013
CE	-2.898$\times 10^{-8}$ (0.000)	0.018	-2.898$\times 10^{-8}$ (0.000)	0.018
Regression Equation Statistics				
	R^2	0.976	R^2	0.976
	R^2_{adj}	0.975	R^2_{adj}	0.975
	MSE	75.419	MSE	75.419
	VIF_{max}	1.013	VIF_{max}	Not applicable
Validation				
	R^2	0.897	R^2	0.897
	R^2_{adj}	0.895	R^2_{adj}	0.895
	MSEP	108.028	MSEP	108.028

Table A.6. Stepwise directed search regression model summary and change statistics for the raw negative ROCE data set

Variables as retained by stepwise directed	R^2	R^2_{adj}	**MSE**	R^2 **Change**	**F Change**	**DF1**	**DF2**	**P-value Change**
APCE	0.496	0.490	4892.493	0.496	75.823	1	77	0.000
ROA	0.815	0.810	1817.788	0.319	91.835	2	76	0.000

Table A.7. Classical multiple linear regression and generalized least squares regression models for ROCE with standard errors of estimates in parentheses for the raw negative ROCE data set

Fitted				
Parameter	**Classical multiple linear regression**		**Generalized least squares regression**	
	Estimate	**P-value**	**Estimate**	**P-value**
Intercept	-79.401 (8.751)	0.000	-79.401 (8.751)	0.000
APCE	66.947 (5.367)	0.000	66.947 (5.367)	0.000
ROA	-1.291 (0.113)	0.000	-1.291 (0.113)	0.000
Regression Equation Statistics				
	R^2	0.815	R^2	0.815
	R^2_{adj}	0.810	R^2_{adj}	0.810
	MSE	1817.788	MSE	1817.788
	VIF_{max}	1.021	VIF_{max}	Not applicable
Validation				
	R^2	0.007	R^2	0.007
	R^2_{adj}	-0.018	R^2_{adj}	-0.018
	MSEP	192492.427	MSEP	192492.427

APPENDIX B

ADDITIONAL FIGURES

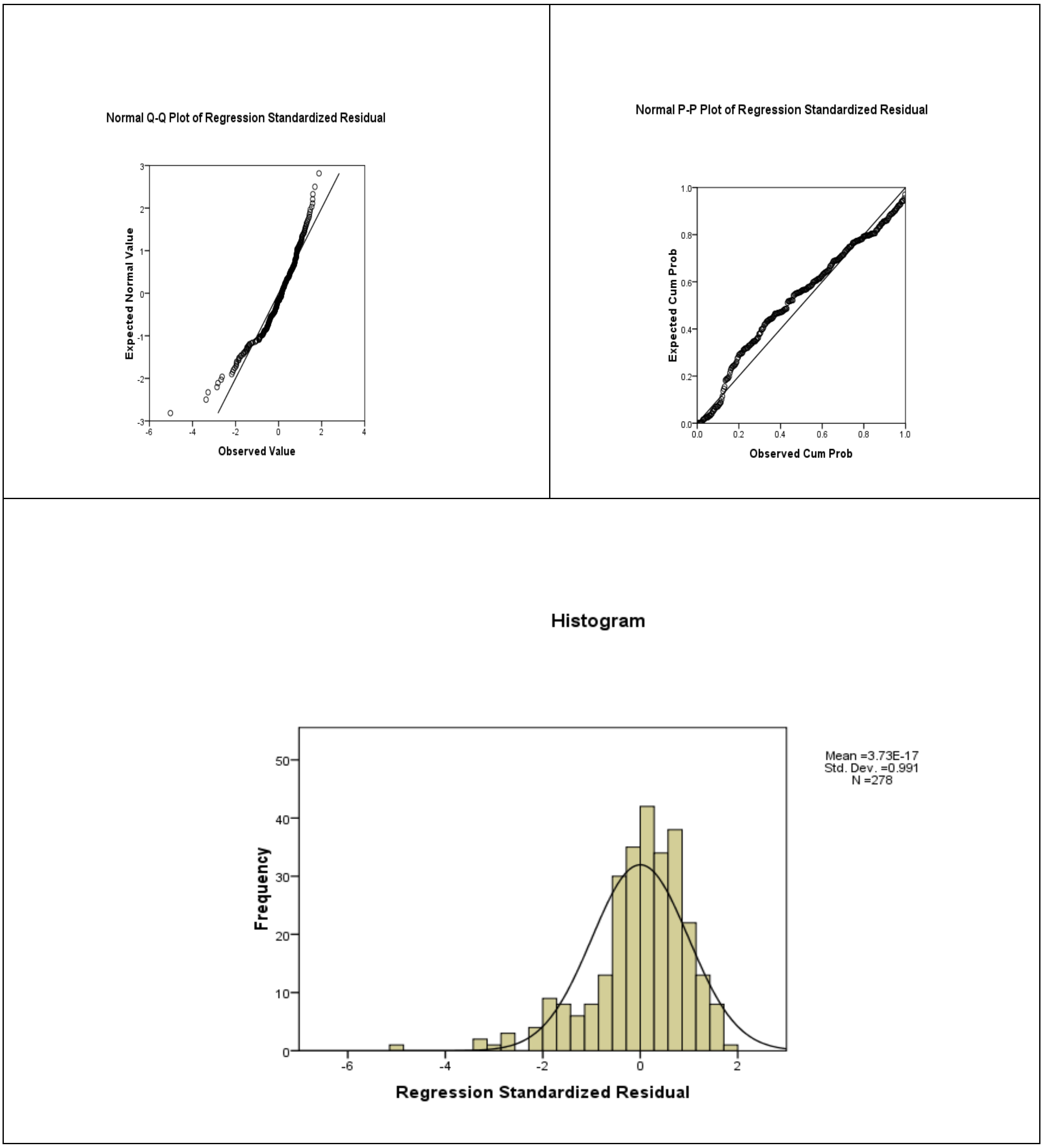

Figure B.1. Normal Q-Q plot, normal P-P plot and frequency distribution plot of the classical multiple linear regression standardized residuals for the transformed positive ROCE data set

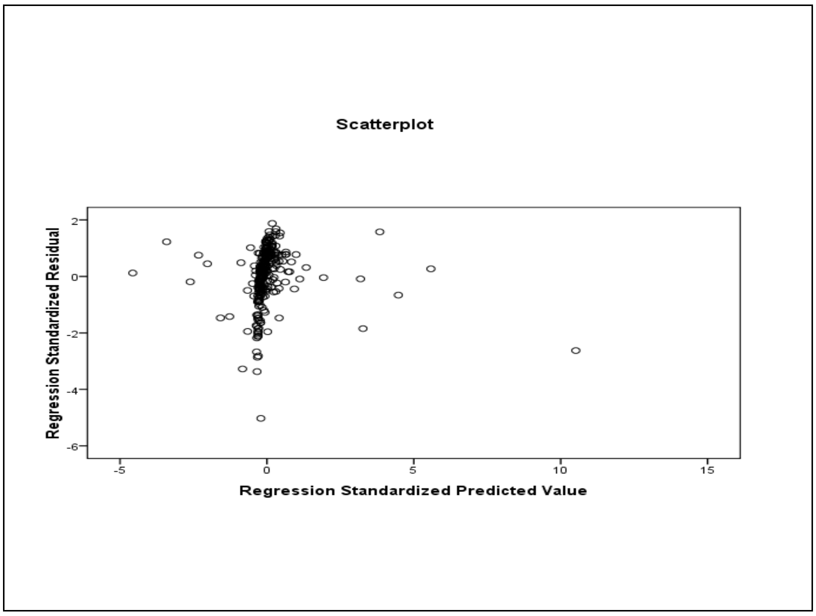

Figure B2. Scatter plot of standardized residuals versus standardized predicted values of classical multiple linear regression for the transformed positive ROCE data set

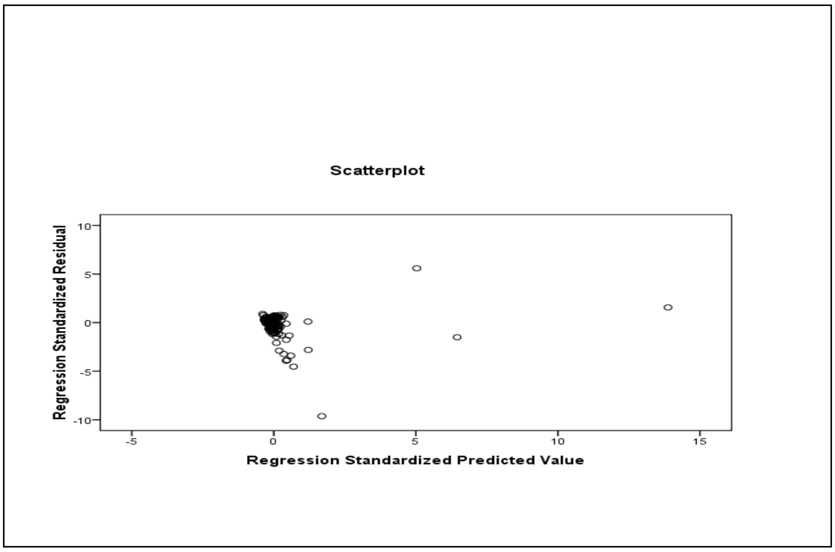

Figure B.3. Scatter plot of standardized residuals versus standardized predicted values of the fitted classical multiple linear regression model for the raw positive ROCE data set

APPENDIX C

R CODE FOR GENERALIZED LEAST SQUARES REGRESSION

Positive ROCE Data Set

```
## First we load the “foreign” package in order to read the data files since the data files were generated using SPSS
        >require(foreign)

## Here we import the SPSS file into R, and assign it to the variable PositiveData
        >PositiveData<-read.spss("E:/POSITIVE DATA.sav")

## We load the nlme package which enables us to perform the Generalized Least Squares Regression analysis
        >Library(nlme)

## We perform the Generalized Least Squares Regression analysis. ROCE is the dependent variable
        >g<-gls(ROCE~APCE+ROA, PositiveData)

Finally we get the summary statistics for the Generalized Least Squares Regression analysis
        >Summary(g)
```

Positive ROCE Validation Data Set

```
## First we load the “foreign” package in order to read the data files since the data files were generated using SPSS
        require(foreign)

## Here we import the SPSS file into R, and assign it to the variable PositiveValidationData
        PositiveValidationData<-read.spss("E:/POSITIVE VALIDATION DATA.sav")

## We load the nlme package which enables us to perform the Generalized Least Squares Regression analysis
        Library(nlme)

## We perform the Generalized Least Squares Regression analysis. ROCE is the dependent variable
        g<-gls(ROCE~APCE+ROA, PositiveValidationData)

Finally we get the summary statistics for the Generalized Least Squares Regression analysis
        Summary(g)
```

Negative ROCE Data Set

```
## First we load the "foreign" package in order to read the data files since the data files were generated using SPSS
        >require(foreign)

## Here we import the SPSS file into R, and assign it to the variable NegativeData
        >NegativeData<-read.spss("E:/NEGATIVE DATA.sav")

## We load the nlme package which enables us to perform the Generalized Least Squares Regression analysis
        >Library(nlme)

## We perform the Generalized Least Squares Regression analysis. ROCE is the dependent variable
        >g<-gls(ROCE~APCE+ROA, NegativeData)

Finally we get the summary statistics for the Generalized Least Squares Regression analysis
        >Summary(g)
```

Negative ROCE Validation Data

```
## First we load the "foreign" package in order to read the data files since the data files were generated using SPSS
        require(foreign)

## Here we import the SPSS file into R, and assign it to the variable NegativeValidationData
        NegativeValidationData<-read.spss("E:/NEGATIVE VALIDATION DATA.sav")

## We load the nlme package which enables us to perform the Generalized Least Squares Regression analysis
        Library(nlme)

## We perform the Generalized Least Squares Regression analysis. ROCE is the dependent variable
        g<-gls(ROCE~APCE+ROA, NegativeValidationData)

Finally we get the summary statistics for the Generalized Least Squares Regression analysis
        Summary(g)
```

APPENDIX D

R CODE FOR ROBUST MAXIMUM LIKELIHOOD REGRESSION

Positive ROCE Data Set

```
## First we load the "foreign" package in order to read the data files since the data files were generated using SPSS
        >require(foreign)

## Here we import the SPSS file into R, and assign it to the variable PositiveData
        >PositiveData<-read.spss("E:/POSITIVE DATA.sav")

## We load the MASS package which enables us to perform Robust Maximum Regression analysis
        >Library(MASS)

## We perform the Robust Maximum Regression analysis. ROCE is the dependent variable
        >g<-rlm(ROCE~APCE+ROA, PositiveData)

Finally we get the summary statistics for the Robust Maximum regression analysis
        >Summary(g)
```

Positive ROCE Validation Data Set

```
## First we load the "foreign" package in order to read the data files since the data files were generated using SPSS
        > require(foreign)

## Here we import the SPSS file into R, and assign it to the variable PositiveValidationData
        >PositiveValidationData<-read.spss("E:/ POSITIVE VALIDATION DATA.sav")

## We load the MASS package which enables us to perform Robust Maximum Regression analysis
        >Library(MASS)

## We perform the Robust Maximum Regression analysis. ROCE is the dependent variable
        >g<-rlm(ROCE~APCE+ROA, PositiveValidationData)

Finally we get the summary statistics for the Robust Maximum regression analysis
        >Summary(g)
```

Negative ROCE Data Set

```
## First we load the "foreign" package in order to read the data files since the data files were generated using SPSS
>require(foreign)

## Here we import the SPSS file into R, and assign it to the variable NegativeData
>NegativeData<-read.spss("E:/NEGATIVE DATA.sav")


## We load the MASS package which enables us to perform Robust Maximum Regression analysis
>Library(MASS)

## We perform the Robust Maximum Regression analysis. ROCE is the dependent variable
>g<-rlm(ROCE~APCE+ROA, NegativeData)

Finally we get the summary statistics for the Robust Maximum regression analysis
>Summary(g)
```

Negative ROCE Validation Data Set

```
## First we load the "foreign" package in order to read the data files since the data files were generated using SPSS
> require(foreign)

## Here we import the SPSS file into R, and assign it to the variable NegativeValidationData
>NegativeValidationData<-read.spss("E:/NEGATIVE VALIDATION DATA.sav")


## We load the MASS package which enables us to perform Robust Maximum Regression analysis
>Library(MASS)

## We perform the Robust Maximum Regression analysis. ROCE is the dependent variable
>g<-rlm(ROCE~APCE+ROA, NegativeValidationData)

Finally we get the summary statistics for the Robust Maximum regression analysis
>Summary(g)
```