KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI

RANK-BASED ADAPTIVE METHOD OF ESTIMATING BETA

BY

AKILU ATIKATU

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FUFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF M.PHIL ACTUARIAL SCIENCE

NOVEMBER 2015

DECLARATION

I hereby declare that this submission is my own work towards the award of the M.Phil degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgment had been made in the text.

Akilu Atikatu	.....	.....
Student	Signature	Date

Certified by:

Dr. G. A.Okyere	.....	.....
Supervisor	Signature	Date

Certified by:

Prof. S.K. Amponsah	.....	.....
Head of Department	Signature	Date

DEDICATION

I dedicate this work to God Almighty for being my utmost help and guide. To Mr. D. Asamoah-Owusu, words cannot express my gratitude,thank you for making this a reality I am very grateful, thank you.

ABSTRACT

Traditional methods of estimation and testing, such as the Ordinary Least Squares (OLS) method, are efficient if the normality assumption of the error distribution and other assumptions about a liner model are not violated. Adaptive tests are found to be efficient and increases the power irrespective of the condition of the observed data. In particular, stock market data comes along with some skewness, tail weights, outliers and unknown distributions that violates some underlying assumptions for which the estimates from OLS is efficient. The degree to which a security is affected by a systematic risk as compared to the effect on the market as a whole is measured by the security's beta. Beta estimates of a security on the stock market are obtained from the OLS estimates of the parameters of a linear model. In practice, however, the error distribution of the market model is not known and conclusions made solely using traditional methods may lead to invalid conclusions. Consequently, fund managers, actuaries and investment risk managers may mislead their clients based on financial decisions made based on these beta measures. This study sought to extend robust adaptive methods that considered tail weight, skewness and selector statistics, in estimating security beta with some specified lags. Further comparisons were made between the adaptive procedure and the OLS method. In line with these objectives, monthly data of three companies listed on the Ghana Stock Exchange (GSE), from January 2000 to June 2014, were used. Market models were formulated with some specific lags and estimation of model were done for both traditional and adaptive methods. The study showed that rank-based methods (Wilcoxon and Adaptive) were more robust in estimation when the distribution of the error term of the dataset was non-normal and also in the presence of outlying observations,whiles the LS method was very non-robust. Results indicated that 5% outlier-contamination was enough to cause some instability in the estimates.

ACKNOWLEDGMENT

Much appreciation goes to my supervisor Dr. G. A. Okyere for guiding me every step of the way, to Dr. Damehe and Mr. Gideon Boako of KSB, to Mr. Jerry Mensah-Jones and Mr. Richard Tawiah for your utmost support and encouragement. God richly bless you for being a help in time. To Ms. Serena Mamudu, my sweet mum, Emmanuel Kojo Amoah, Bob Kay and my project partners, thank you for always bearing me up.

CONTENTS

LIST OF ABBREVIATION

Adp .....Rank-based Adaptive estimation method CAPM .....Capital Asset Pricing Model GEER .....Generalized Estimating Equation Ranking

GSE .....Ghana Stock Exchange HFR .....Hogg Fisher Randles

IF ..... Influence Functions

JR .....Joint Ranking LAD ..... Least Absolute Deviation

LSE ..... Least Squared Error

LS ..... Least Squares MADM ..... Median Absolute Deviation about the Median

MSE ..... Mean Squared Error OLS .....Ordinary Least Squares SSR ..... Sums of Squares of the Residuals

Wil. ..... Wilcoxon scores

WLS .....Weighted Least Squares

WMW ..... Wilcoxon Mann Whitney

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background of study

Least Squares estimation of beta, as simple and applicable as it is with its extensive use in the world of business finance and actuarial science, often fails in practice raising the question of if there were a better method of estimating beta. One important limitation of the OLS model is the assumption linear regression model makes of the distribution of the error term. Statisticians have in a long time noted the loss in efficiency of OLS method when errors have nonnormal distribution. An increasing number of studies in economics has noted the difficulty in assuming errors have a normal distribution. While the theory has worked and served well for years, statisticians are increasingly concerned about the use of OLS when its assumptions are not met ie. when the probability of outlier observations in the data is much. Huber (1973) comments, that one gross outlier can damage the LS estimate. In theory, when the assumptions do not hold the standard least squares estimation for the regression coefficient will be biased and/or non-efficient. Such evidence provides a clear motivation for exploring the use of a more robust method for beta estimation. In this study a rank-based adaptive method is compared to the OLS method using monthly returns data for some stocks on the Ghana Stock Exchange(GSE)

1.2 Types of Risk

In portfolio management, risk is one important factor to be considered. The probability that observed returns could be less than expected returns creates risk in holding securities. Risk is classified into systematic and unsystematic risk.

1.2.1 Unsystematic Risk

Unsystematic risk is specific to an industry or firm and can be reduced and even eliminated by sufficient diversification of ones portfolio over a larger number of securities. Such factors as labour strikes, the capability of management, investment strategies etc.,contribute to unsystematic risk.

1.2.2 Systematic Risk

Systematic risk is associated with economic, sociological, political and other macro-level factors. Systematic risk affects the whole market and cannot be eliminated by mere diversification of ones portfolio.

Beta a proxy of systematic risk, measures the extent to which systematic risk affects a security as compared to the market.

The beta of a stock measures the volatility of the stock return an investor is exposed to in relation to the entire market. Beta measures the responsiveness of a stocks price to changes in the overall stock market. A stock market has a beta of 1.00 and the stocks beta is measured on the extent to which it deviates from the market. A stock is less volatile than the market when it has a beta less than 1.00, its price thus fluctuates less frequently as compared to the market and vice versa. Beta is used in the CAPM model to estimate an assets expected return. The Capital Asset Pricing Model (CAPM) developed in the early 1960s by William Sharpe, Jack Treynor, John Lintner and Jan Mossin was received with great enthusiasm by the finance world.

1.3 Capital Asset Pricing Model (CAPM)

CAPM is a model for measuring risk-return trade-off for all assets, including efficient and inefficient portfolios. It is a model that calculates the expected return of an asset based on its beta and the expected market returns.

$$E(R_i) = R_f + [E(R_m) - R_f]B_{im} \quad (1.1)$$

Where

R_f: the risk free interest rate (the interest rate an investor would expect to receive from a risk free investment)

B_{im}: beta of the security

R_m: is the return of the market index

R_i: is the return of security *i*

The measure of a securities Beta informs fund managers as to what securities to include in their portfolio to minimize the overall risk and maximize the return of the portfolio. According to Eugene Fama, beta as the sole variable explaining returns on stocks is dead however the fact that the average returns of stocks might not be in accordance with their beta's as predicted by the capital asset pricing model(CAPM) does not negate the usefulness of beta as a measure of a stock's risk exposure. Thus the usefulness of beta cannot be overlooked. It is very important that investors not only have a good understanding of their risk tolerance, but also know which investment match their risk preference. Using beta as a measure of risk (volatility), securities that meet these investors criteria for risk can better be chosen. Many brokerage firms calculate and publish the betas of securities they trade. Analysts, brokers and planners have used beta for decades to help them determine the risk level of an investment.

In estimating beta, econometric studies has massively employed the standard linear model. In that framework economists estimate the slope parameters, otherwise known as beta, using the ordinary least squares(OLS) technique. When errors have monotonically declining likelihood functions in the sum of squared errors, the OLS method is equivalent to the maximum likelihood(ML) method and this has increased its popularity. These estimators are efficient unbiased estimators that have the least possible asymptotic variance in a family of unbiased estimators.

The goal of OLS is to closely fit a function or model with a given dataset. It does so by minimizing the sum of squared errors from the data. The OLS method is the best

statistical method in estimating parameters of a linear model under some assumptions, Martin and Simin (2003), justifying the frequency in its use. The OLS method makes and uses some assumptions in estimating the parameters of the model as stated by Martin and Simin (2003);

- The error /residual term is normally distributed
- The variance of the residual is constant or homoscedasticity
- The mean of the error is 0

The method of OLS estimation is simplistic and has much easier computations and implementation. The estimated parameters are easy to understand and interpret.

OLS estimation of beta, as simple and applicable as it is with its extensive use in the world of business, finance and actuarial science, often fails in practice raising the question of if there were a better method of estimating beta.

One important limitation of the OLS model is the assumption of the distribution of the error terms. Statisticians are in the know that OLS loses efficiency when errors have non normal distributions and it has become increasingly difficult as pointed by economic studies to assume errors have a normal distribution. While the theory has served us well for many years, there is a growing concern about its use when underlying assumptions are violated (e.g., when the possibility of outliers is great). Huber (1973) comments, "Just a single grossly outlying observation may spoil the least squares estimate" .

Such evidence provides a clear motivation for exploring the use of a more robust estimation of beta, in this case an adaptive rank-based estimation method. We compare the adaptive rank-based estimation method with the OLS, using the measures of asymptotic relative efficiency and robustness to determine which model is most efficient in the presence of outlying observations.

This is the purpose of this work.

1.4 Statement of the Problem

Modeling the relationship between variables, in this case market returns and a specific security return by the method of least squares as is frequently done is very important, simple and easy to implement and interpret. However, more often than not, the well-known least squares regression procedure is only optimal under certain assumptions of the error term.

The error term is assumed to be normally distributed with homoscedastic variance. The assumptions of the OLS are strong Fox (2005), and there are many ways they could go wrong. The typical cases include,

- The distribution of the error terms can be skewed, heavy-tailed or non-normal
- Errors not being independent as in time series data
- Conditional variance of the error can be heteroskedastic
- The presence of outliers in the dataset
- Infrequent trading of the security as compared to the overall market, thin trading of the security

More often than not when the assumptions as stated are not met, the LS estimation of beta will be biased and/or inefficient, Hampel, Ronchetti, Rousseeuw and Stahel (1986). Several other methods of estimation have been proposed, Draper and Smith (1988), among these methods is the Adaptive method by O'Gorman (2012).

1.5 Objectives of the Study

The objectives or aims of this study are:

- To implement the use of a more robust adaptive estimation method not easily affected by outliers and non-normal data.

- Investigating and comparing beta estimation models for one which is more efficient.
- Showcase some methods of reporting adjusted Beta such as the Dimsons and Scholes-Williams method.

1.6 Methodology

In this study we report some summary results of a study undertaken to compare the properties of two alternatives to standard least squares method for simple and multiple linear regression analysis; the rank based adaptive and the Wilcoxon fits. R-codes are written to select appropriate score functions (which will inform about the distribution of the errors of the dataset) to be used in the rank based adaptive fit. In our study, we use data on stock returns from the Ghana Stock Exchange which contain outliers and some other skewed data such as uniform, exponential among others and compare the performance of all three regression methods. The sigma or tau's of the estimation methods and their asymptotic relative efficiencies are compared.

1.7 Justification of the Study

Adaptive methods of estimation have many advantages as compared to the traditional estimation methods. Adaptive methods are often times more efficient than the traditional methods when applied in estimating linear models with skewed or long-tailed error distributions. Adaptive methods are constructed carefully in order to maintain their significance level at or close to α, the probability of rejecting the null hypothesis when indeed the null hypothesis is true. Statistical properties of the adaptive methods are often superior to the traditional methods hence they are often recommended for use.

Properties of the adaptive method includes, but not limited to, the following:

- The effect of outliers on the estimate is automatically decreased.

- There is observed to be little power loss to the adaptive methods when used in estimating linear models with normal error distributions.
- For long-tailed or skewed error distributions, adaptive methods are more efficient compared to traditional methods
- As already stated, adaptive methods maintain actual significance levels at or near the nominal level of significance, α.

1.7.1 Thesis Organization

The organization of the thesis is as follows. It has five chapters, Chapter 1 presents the introduction of the study which consists of the background of the study, methodology, thesis justification and organization. Chapter 2, the literature review, discusses works done by other researchers on robust estimates of beta. Chapter 3 is the formulation of the methods used. Chapter 4 presents the data collection, analysis and formulation of the model. Chapter 5 sums up the results, summary, conclusions and recommendations of the study.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In financial economics, asset return betas needs to be estimated. Beta of the capital asset pricing model (CAPM), irrespective of academics debate on its relevance, is without doubt the best method used widely as a expected return and security's risk measure. Survey of industry users has recorded massive use of CAPM beta's. According to Graham and Harvey (2001), 73.5% of respondents use CAPM in estimating cost of equity capital. Block (1999) showed that over 30% analysts (respondents) in a financial analyst survey, viewed CAPM as an estimating model

which is very important. 50% of managers with an idea of the CAPM techniques used it as according to Gitman and Mercurio (1982). By 1997, according to Bruner, Eades, Harris and Higgins (1998), CAPM as a model for estimating cost of equity had become dominant and many practitioners use values from publications of commercial providers

2.2 Traditional Methods of Estimation

Ordinary Least Squares method is simple and it enjoys wide usage by both practitioners and academic researchers. However, factors constructed from macroeconomic data especially have large measurement error and even when measured accurately, a factor can still differ from the actual underlying factor. Example, the stock market index return is probably measured most accurately but it can be an imperfect proxy for the true market return and thus said to contain large measurement error, Meng, Gang and Bai (2007) critique.

Traditional statistical presents a unified methodology in solving problems ranging from location models to complex experimental designs, using the least squares method. First the problem is presented as a model, then the Euclidean distance between the responses and the conjectured model is minimized in a method known as the least squares method. Least squares method provides diagnostic techniques to check the models adequacy, explore the fits quality and detect outliers and other influential cases. In addition it provides inferential procedures such as confidence procedures, hypotheses tests amongs others. As Hettmansperger and McKean (2008) mentioned, the Least Squares is affected easily by outliers. One outlier can adversely affect the least squares model, inference procedures and associated diagnostics. Traditional procedures enjoy high efficiency when models are normally distributed and less efficient when errors have long tails.

2.3 Drawbacks of the Least Squares method

Mohebbi, Nourijelyani and Zeraati (2007) in a simulation study confirms the claims made above by Hettmansperger and Mckean saying, modeling data by the linear least squares method is very important and crucial. Frequently, however, the well-known least squares regression procedure is only optimal under certain distributional assumptions of errors. In practice, this assumption may not hold because of possibility of skewness or presence of outliers in a data. In theory, when the assumption of normality does not meet, the standard least squares method of estimation will be biased and/or non-efficient.

In multiple regression, the OLS method gives the best parameter estimates when its underlying assumptions are met. However, when these assumptions are not met sample estimates and inferences could be misleading. Outliers often violates the linear models assumption of normality of residuals. These outlying observations can adversely affect estimates when they are unnoticed whether they are in the direction of response or explanatory variables, adds Alma (2011)

To throw more light on what outliers are, we proceed with a discourse by Martin and Simin (2003). When market return is due to the market returns mean, outliers add no effects on the OLS fit thus has no effect on the OLS beta. If an outlier presents as a rather unusually large fall in both the market and equity returns, then this outlier is in line with OLS fit and does not affect the beta estimate. These sort of outliers improves precision in the beta estimation (ie. the outliers highly decreases the OLS beta's standard error and is termed a good outlier). If an outlier presents as a large outlying market return not an outlying asset return it equally exerts considerable influence. Neither equity returns nor market returns presents as an outlier but the pair can clearly be separated from the bulk of the data and present as a two-dimensional outlying observation. Market return outliers, are outlying observations in the one dimensional explanatory or independent variable dimension, they could

cause bias by exerting leverage on the OLS linear fit. Such outlying observations in statistics and econometrics are referred to as leverage points.

In summary Abdi and Salkind (2007) adds; Despite its popularity and versatility, Least Squares method has its problems. Probably, the most important drawback of the Least Squares is its high sensitivity to outliers (i.e. extreme observations). This is a consequence of using squares because squaring exaggerates the magnitude of differences (e.g. the difference between 20 and 10 is equal to 10 but the difference between 20^2 and 10^2 is equal to 300) and therefore gives a much stronger importance to extreme observations.

2.4 Ways of dealing with the shortfalls in Least Squares in practice.

To have an idea of how commercial providers of beta treat outliers, current standards being practiced by nine such providers is studied in a survey conducted by Martin and Simin (2003). The survey reviewed websites, published works on beta estimation methods and telephone interviews with employees with some knowledge of processes used in estimating beta. Providers of beta estimates including Barra and Ibbotson Associates, go to any extent to reduce effects of outliers. All of these commercial providers with the exception of Barra use the single factor model in reporting beta values. Ibbotson Associates deals with outliers when computing OLS betas by discarding beta estimates greater than the absolute of five. This shows that Ibbotson Associates are aware outliers can distort greatly the OLS beta. This practice by Ibbotson does not guard against distortions of estimates by outliers that are less extreme but a single outlier can shoot the beta estimates from 1.0 to 3.0. The only provider in the survey who uses statistical methods in dealing with outliers was Barra.

In its various approaches in estimating Betas and other risk models, Barra employed a unit dimensional winsorization procedure (which treated outliers by replacing them

with certain specific values) though not on returns of equity but on each exposed factor individually, Connor and Herbert (1999). From this work it was noted that commercial beta providers deal with outlying observations using the one-dimensional statistical method treatment. The treatment is termed one dimensional because outliers are separately treated in equity and market returns, then the OLS beta computation. The treatment often involves winsorizing or rejecting outlying observations. Completely removing observations seen to be outliers is termed rejecting, winsorizing involves altering outlying points by bringing them into specified values. The classical three-sigma edit rule is the best known rejection method but its said to be inferior to alternative rejection rules, Hampel et al. (1986). One common practice of dealing with outliers by winsorization is where resistant estimates of standard deviation and mean are employed to deal with outliers. Values which are more than 5.2 times the median absolute deviation about the median (MADM) distant from the sample median are reduced to a distance which is 5.2 times the MADM from the median.

This practice is employed by Barra to reduce the impact of outliers in its model. Clearly non of these one-dimensional treatments of outliers can be enough for all forms of two dimensional outliers.

According to Fox (2005), a systematic rule called the bulging rule was suggested by Mosteller and Tukey. This is used in choosing from a collection of powers and roots appropriate linearizing transformations where we replace a variable x with a power x^p. For p = 2, replace the variable with its square, x^2 ; for p = −1, replace with the inverse, x^{-1} = 1/x; for p = 1/2, replace with the square-root, $\sqrt{0.5}$ 0 $x = x$; etc. For p = 0, x =1 thus an exception is made, the log transform, log x, is used.

Such transformations are applicable for positive x values. Square-root and log transforms are undefined for negative x values. In a dataset with both negative and

positive x-values x^2 transformations will affect the order of x. To solve this problem a constant quantity c is added to all the x values before applying the power transforms, x $\longrightarrow (x + c)^p$. Negative powers like the inverse transforms, x^{-1}, reverse the x-values order, but if the original order is to be preserved , then $-x$ is used in the place of x for negative p values. Using power transforms could aid in linearizing nonlinear simple and monotone relationship. A simple relationship is smoothly curved and its curvatures direction does not change. A monotone relationship has y strictly increasing or decreasing along with x.

Relationships that are not linear, simple nor monotone are estimated using other forms of parametric regression since they cannot be linearized by power transformation. Linear least-squares regression can be used in fitting quadratic equations. Nonlinear least squares are employed in fitting a broader family of parametric models. All these methods of beating the estimates of the OLS into shape for efficient results have their limitations. For the method of trimming obviously,informative and influential parts of a dataset could be unnecessarily discarded and the transformation methods change the data entirely.

2.5 Alternative methods of the Standard Least Squares

Martin and Simin (2003) advised that practitioners needed to compute more resistant betas not easily influenced by outliers instead of relying solely on OLS betas. They further stated that a manner more effective and uniform in dealing with outliers was a straight-line fitting method more resistant and twodimensional, and this method was a weighted least squares(WLS) with weights that were data-dependent. Several alternative methods of the standard Least Squares(LS) regression have been proposed, Draper and Smith (1988). Among these, three methods are in widespread application in many branches of applied science.

According to Mohebbi et al. (2007), these methods are robust M-estimation, Least Absolute Deviation (LAD) method and nonparametric (rank based) methods. In 1973,

Huber introduced M-estimation for regression. The M in M-estimation stands for "maximum likelihood type". M-estimators are a broad class of estimators which are obtained as the minima of sums of functions of the data. Least-squares estimators are M-estimators. Robust M-estimation is an alternative to the parametric estimation when the errors have a distribution that is not necessarily normal but close to normal. The class of M-estimator models contains all models that are derived to be maximum likelihood models. The most common general method of robust regression is M-estimation, introduced by Huber (1973) that is nearly as efficient as the Ordinary Least Squares. The M-estimate as an objective minimizes the ρ of the errors instead of minimizing the sum of squares of the errors. This formula is resistant to outlying observations in the dependent variable, but turned out to be less robust to outlying observations in the independent variable (leverage points). This method has no power over the least squares when outlying observations are present in the independent variable.

In the 1980's, several alternatives to M-estimation were proposed as attempts to overcome the lack of resistance. Least trimmed squares (LTS) is a viable alternative and is the preferred choice of Rousseeuw and Leroy (2003) and Ryan (2008). In this method the largest squared residuals are excluded from the summation, which allows those outlier data points to be excluded completely. LTS method can achieve high efficiency in estimation, if the exact quantity of outliers are trimmed, the LTS method is equivalent to the OLS method computationally. However, when lesser number of outliers than is contained in the data are trimmed, the LTS can be inefficient. Conversely, when more trimming is done than there are outliers, some good data points could be eliminated. LTS in terms of breakdown is a high breakdown method with 50% breakdown. Compared to the LTS method, the Theil–Sen estimator is a low breakdown method but more efficient and popular.

The S-estimation method was next proposed. The S-estimation method estimates a line, a plane or hyperplane by minimizing a robust estimate of the scale (where the S in the name of the method is obtained) of the errors. The S-estimator though inefficient is very resistant to outlying and leverage points.

The MM-estimation method attempts to combine the high resistance of the S-estimation method and the efficiency of the M-estimator. The method does this by finding a very resistant and robust S-estimate that minimizes an M-estimate of the scale of the residuals. The scale estimated is held constant whiles locating a close-by M-estimate of the parameters.

One other method is the Least Absolute Deviation method(LAD).One optimal property of the LAD estimates of the regression coefficients is by their definition, that they are the estimates that give the smallest sum of absolute residuals. In addition, if we assume that the population of errors has a double exponential/Laplace distribution, then the LAD estimate is the maximum likelihood estimate, Birks and Dodge (1993). The strength of LAD estimation is in its robustness with respect to the distribution of response variable.

In this study by Alma (2011), four robust regression methods S-estimator, Huber M-estimator, MM-estimator and the Least Trimmed Squares (LTS) estimator, were compared to the OLS method. The MM-estimation performed best on the whole with a comprehensive presence of outliers. It however had trouble with high leverage outlying points with data size ranging from small to moderate. Its weakness was highlighted in the study. The S-estimator had a reasonable efficiency, it bounded the influence of the high outlying points. S-estimator could increase efficiency with 10% breakdown. The Huber M estimation was also efficient in the presence of outliers.

Unit weights by Wainer and Thissen (1976) is another robust method. This method is used when multiple predictors are present in a single outcome. The method was

employed by Burgess (1928) in predicting parole success. In his study, 21 positive factors were scored as absent ("prior arrest" = 0) or present (e.g. "no prior arrest" = 1), these were summed up to give a predictor score, that proved to be an efficient predictor of success on parole.

In a simulation study by Mohebbi et al. (2007), some robust estimation methods mentioned above are implemented. Four methods of linear regression; the Least Squares(LS), Huber M, Least Absolute Deviation(LAD) and nonparametric, for two important classes of distribution, symmetric and skewed, were investigated. The same sets of simulated data were used and Mean Squared Error (MSE), Mean Absolute Deviation (MAD) and Biases of these methods were compared. The Least Absolute Deviation, Huber M and nonparametric regression were shown to be more appropriate alternative to the Least Squares in heavy tailed distributions while the nonparametric and Least Absolute Deviation regression were better choices for skewed data.

The choice of symmetric distribution was so that their kurtosis was more than that of standard normal distribution (ie. heavy tailed distributions). This gave the opportunity to investigate regression methods with presence of outliers. Present results indicated that when outliers exist other alternatives of the LS are more appropriate.

The general simulation results could be summarized as follows. In almost all symmetric distributions investigated, the MSE and MAD are close for the sample sizes larger than 100 and so none of the estimation methods were superior in such circumstances. However, this has not been true for the case of the studied skewed distributions where the LS method shown to be far inferior from the other methods of estimation. In general, the bias criterion as compared with the other criteria, shown to have more fluctuation and this fluctuation persist even for large sample

size. This instability of biases created some difficulty and confusion in finding the optimum estimation in some situation.

Choosing a more efficient alternative to the LS method is closely related to the type of data and so it is advisable to use several alternative methods in data analysis, in cases of skewed distribution, the performance of LS was inferior as compared to other methods. Based on our simulated distribution in this research, the nonparametric and LAD methods were more suitable for the studied Gamma family. Mohebbi et al. concluded in this work that, in order to investigate for possibilities of more suitable methods further studies are needed.

2.6 Rank-based Procedures

Bilgic and Susmann (1999) added that, procedures based on ranks retain estimation and testing that are distribution-free. These processes are resistant to outlying observations when compared to traditional methods when random errors have non-normal distribution. Robust score functions could as an alternative be accommodated with methods based on ranks to protect analyses from influential data points in response and factor spaces. The choice of the score functions can depend on prior knowledge of the distribution of the error term. The Wilcoxon score function is quite efficient for moderate to heavy-tailed error distributions. For example, rank-based methods using Wilcoxon scores can obtain up to 95% efficiency compared to the least squares procedure with normal data and are known to be more efficient than least squares for error distributions with heavy tails. The rank-based procedures are more appealing due to these properties.

In this article Bilgic and Susmann employ the use of three methods based on ranks: Joint Ranking (JR), Generalized Rank Estimate (GR) and Generalized Estimating Equation Ranking (GEER), in estimating fixed effects.

The Joint ranking procedure in estimating nested random effects models makes use of asymptotic results of a study Kloke, McKean and Rashid (2009) did in computing standard errors and fixed effects. The asymptotic theory for the rank-based computing of fixed effects in the general mixed model was developed by Kloke et al. (2009) employing Bruner et al. (1998) general rank theory. Fixed effects in the JR method is estimated using dispersion function just as is done in independent linear models. The asymptotic distribution however has a different formula for the covariance matrix due to the model having correlated errors.

The generalized rank-based mixed model fitting is a Newton-type approximation based iterative reweighted rank method. Asymptotic properties of linearized rank estimators was developed by Hettmansperger and McKean (2011) to be used in linear model in k-step Gauss-Newton approximation with no weights. This theory was extended by Bilgic (2012) and Bilgic et al. (2013) to the k-step GR procedure used in general mixed models. After initial fitting is carried out estimates become asymptotically equivalent to the independent scenario since residuals are no more dependent because of covariance weights. This algorithm can work perfectly well for all forms of variance-covariance error structure in the general mixed models.

Abebe, McKean, Kloke and Bilgic (2013) extended the method of general estimating equations (GEE) for mixed models in norms based on rank, and derived the asymptotic normality of the rank estimators. Abebe et al. (2013) proposed that Liang and Zenger's general estimating equations expression be written in terms of the Euclidean norm.

Several papers discussed these derivations, the JR method by Kloke et al. (2009), the GR method by Bilgic et al.(2013) and the GEER method by Abebe et al. (2013). The rank-based estimators in these studies competed well with the traditional procedures including maximum likelihood (ML), restricted maximum likelihood (REML) and least squares. The rank-based methods outperformed the traditional methods when random errors had some contamination and exhibited robustness in the presence of outlying observations. Among these methods, the unweighted JR method performed poorly compared to the other methods in a Monte Carlo study carried out by Bilgic (2012). The GR and GEER methods are reported to be very similar in efficiency and empirical validity. The rank-based norm properties provided GR estimates and standard errors whiles the GEER was obtained from the rank-based norm and least squares properties combined. For a dataset with high correlation, the preferred methods will obviously be the GR or GEER method.

The Wilcoxon signed-rank test is non-parametric and used in making comparisons between matched samples, two samples that are related or a single sample with repeated measurements to find out if the ranks of their population mean differ. The Wilcoxon signed-rank test could be used instead of t-test for matched pairs, paired Student's t-test or t-test for dependent samples when the population is not normally distributed.

The Wilcoxon signed-rank test and the Wilcoxon rank-sum test are different, though they all are nonparametric and have to do with ranks summation. Both tests are named after Frank Wilcoxon (1892–1965) who proposed both tests in a single paper Wilcoxon (1945). The Wilcoxon signed-rank test was made popular by Siegel (1956) in a book on non-parametric statistics which became very influential.

Hotelling and Pabst (1936) as far back as 1936 recognized that data with small sample sizes could make use of their ranks instead to escape the assumption of normality. The Wilcoxon-Mann-Whitney (WMW) test came to being in the 1940's. It was

developed by Wilcoxon (1945) and Mann and Whitney (1947) extended it. Its simplistic nature and robustness has earned it great popularity among scientists. It turns out other simple adaptive rank tests exist that outperforms the WMW test in discovering differences in distributions. Irrespective of size, large or moderate the nonparametric adaptive procedures exhibits more power than the parametric t-test. Marking major developments in that area, we begin with Hajek and Sidak (1967), who aided in making the WMW tests better by showing that the rank test depended on the distribution of the data. When the distribution of the data is logistic the most powerful test is the WMW test, the median test is most powerful as stated by Siegel and Castellan (1988), with the Laplace/double exponential distribution. Practically the distribution of datasets are unknown, thus making it impossible to use the so called most powerful rank test. Gastwirth (1965) exhibited rank tests improvement in detecting location differences when modified by re-weighting. Gastwirth (1965) shows that to improve the power of a test one must have a fair idea of some features of the underlying distribution of the dataset.

In the mid 1950's studies showed that the rank tests could improve efficiency by discarding some parts of the dataset. The main idea was to focus the test on aspects of the distribution where location differences were revealed. For example, The data close to the hypothesized common boundaries helps a great deal to detect a location shift between two uniform distributions. Gastwirth (1965) suggested a modification in the WMW test where data was included in only the top p and the bottom r proportion of the combined sample ($0 < p, r < 1$). The underlying distribution informed the optimal values of p and r . Gastwirth (1965) proved that, rank tests that were appropriately modified were asymptotically more efficient when compared to the WMW test.

2.7 Adaptive rank tests

Gastwirth (1965) study created a path for rank-based adaptive tests. Hogg, Fisher and Randles (1975) in a paper proposed adaptive procedure simply and effectively used a dataset to choose an efficient rank test from amongs a set of alternative tests. The data in this procedure is used two times, first to select and then to perform the test nonetheless the procedure is termed as ”honest” in that the level of significance is preserved in performing the test. The strength of the HFR procedure lies in how easy it can be implemented and the great power it has compared to the WMW tests. The Hogg Fisher Randles (HFR) adaptive method has challenged a large quantity of literature.The HFR test was extended to location test(one-sample), Jones (1979) and k-sample trend tests, Buning (1999). With a more recent work by O'Gorman (1996), Xie and Priebe (2000), Xie and Priebe (2002), Kossler and Kumar (2008), Kossler (2010) and others.

A study by Hao and Houser (2012) investigated the performance of the HFR test under different sample sizes as well as optimizing some parts of the HFR algorithm. The study confirmed that, adaptive procedures are substantially more powerful than WMW tests and t-tests and almost as powerful in other cases. The study also confirmed that adaptive procedures exhibit improved power relative to t-test with moderate size samples (say $20 \leq n,m \leq 40$).

On the subject of Rank Test, O'Gorman (2012) stated that,quite a number of adaptive tests are designed to make better the performance of estimation methods and significance tests. He called a significance test adaptive, if the test procedure is altered and improved after collection and examination of the data. As an example, in using a two-sample adaptive test, data is collected and selection statistics to determine the test procedure to be applied, are calculated. If the data seems to have a normal distribution, a Wilcoxon rank-sum test is used. If there are outliers contained in the data, then we use instead a median test. Adaptive methods have more

advantages compared to traditional tests. There is observed to be little power loss to the traditional tests when adaptive methods are used in estimating linear models with normal error distributions. For long-tailed or skewed error distributions, adaptive methods are more efficient compared to traditional methods and the effect of outliers is automatically decreased when adaptive methods are used. Adaptive methods are constructed carefully in order to maintain their significance level and if that is done properly the adaptive test will have a probability at or close to α, of rejecting the null hypothesis when indeed the null hypothesis is true. Statistical properties of the adaptive methods are often superior to the traditional methods hence they are often recommended for use. The Adaptive method is above all very straightforward and practical.

Adaptive methods are said to be robust. There are two types of robustness, robustness for power and size. When a test has high power compared to other tests and the assumptions of the distributions are not met, it is said to be robust for power. On the other hand if a test maintains the actual level of significance close to the nominal level then it is robust for size. Often, traditional tests with errors not normally distributed are not robust for power but are robust for size.

2.8 Hogg's Adaptive procedure

Hogg et al. (1975) proposed the maiden practical adaptive two-sample test. Before 1975, the adaptive tests in use were not very practical but quite interesting. Tests to improve power was designed by Hajek (1962) by finding scores to produce rank test that were most powerful locally. To carry out Hajek's test, an estimate of the derivative function's first derivative and the density function was required . But the first derivative and the density function were difficult to estimate unless very large

samples were used . Hence, we do not see much of Hajek's adaptive test used in practice.

To avoid cumbersome estimations of densities and derivatives, an adaptive procedure which uses the measure of sample kurtosis to select one out of four estimators of the mean of a symmetric distribution, was proposed by Hogg (1967). In that study, symmetric distributions four in number having varying levels of kurtosis were considered. Selection statistics were used to choose a low variance estimator. One pitfall of this approach was that the sample kurtosis were highly variable, so it may fail on some occasions to select appropriate estimators for those symmetric distribution. With a sample size of 25 observations generated from the distributions in that research, the adaptive estimator had excellent performance irrespective of the limitation stated above. Hogg (1967) in urging for much more use of robust procedures stated that statisticians must take a wider view and not just choose a model before observing the sample items because the availability of excellent computing devices makes it easier for these adaptive robust procedures to be carried out. This estimator has seen various forms of modifications over the years with the most recent version designed by Hogg and Lenth (1984).

Randles and Hogg (1973) further developed Hogg's idea of modifying a statistical method using selection statistics, the methods included a one-sample and two-sample adaptive tests. In their study, based on some estimated selection statistics, rank scores were chosen and a measure of tail weight not the sample kurtosis was used as a selection statistic. The tests employed by Randles and Hogg were adaptive but less powerful as compared to traditional tests. Two years after Randles and Hogg (1973) work, an improved adaptive two-sample test was published by Hogg et al. (1975). This test was rather very practical and proved to be a more effective adaptive two-sample test. Though it attracted attention, the test seems not to be used often.

One reason for its limited use is because, being a rank-based test, it is quite difficult to generalize this method to significance tests of regression coefficients in some complex models. In a number of publications after Hogg et al. (1975) paper, many researchers constructed tests using Hogg et al. (1975) selection statistics. Buning (1996) extended Hogg's method of selecting rank scores using selection statistics, to a test of equality of medians. This test was adaptive in nature and used a one-way layout data. Buning and others made extensions of the adaptive method in Buning and Kossler (1998), Buning (1999), Buning and Thadewald (2000), Buning (2002).

2.9 Hogg's selection statistics

Buning and Hogg's tests employed the use of selector statistics to select rank scores for the tests. Their approach faced the problem of having selection statistics falling near or at the edge of a selection region. A slight altering of the data could cause the selection statistics to change which could end up in the choosing of rank scores entirely different from the previous one and this result is very undesirable due to the large change in p-value resulting from a small change in a single dataset. Ruberg (1986) in a bid to correct this condition, proposed the use of a two-sample adaptive test which is continuous with its one way layout proposed by O'Gorman (1997). Several other adaptive tests using different methods were proposed. They include tests proposed by Hall and Padmanabhan (1997) for the two-sample test, where bootstrap testing method was employed. The last 40 years has seen many studies in adaptive estimation. Yuh and Hogg (1988) as well as Hill and Padmanabhan (1991) have all proposed adaptive estimators. These estimators make use of selection statistics to select one of many regression estimators that are robust.

In 2001, a test using adaptive weighing approach was proposed by O'Gorman (2001). In this method weights are assigned observations so as to use these weighted

observations to test regression coefficients in a linear model. O'Gorman (2002) proposed an improved version of this method. Various ways that this method is applied is described in a book written by O'Gorman (2004). In that write-up he showcased computing the p-value using a method of permutation.

Other studies done by O'Gorman in this area include multivariate adaptive test O'Gorman (2006) and O'Gorman (2008) which was implemented in the analysis of repeated measure data O'Gorman (2008). Another of his work involved using the permutation of residuals which was proved to be equally effective as the permutation of independent variables.

Recent publications has seen much work on adaptive one-sample tests a shift from the norm where works before 2000 were more focused on two-sample tests. Two adaptive test for the median was proposed by Lemmer (1993). The likes of Freidlin et al. (2003) suggested adaptive test for paired data. In these study p-values from a normality test were used as the selector statistics and not measures of skewness or tail weight. These tests were quite effective with sample sizes that were moderate. When the distributions symmetry is in doubt, Baklizi (2005) suggests the use of the work by Lemmer (1993) that is a continuously adaptive test for the median. Very recently, Miao and Gastwirth (2009) suggested a test where some score functions used by Freidlin, Miao and Gastwirth (2003) are employed with a measure of tail-heaviness as selector statistics.

Neuhauser, Buning and Hothorn (2004) proposed a different method of improving and making robust the two-sample tests. In their study, first four rank scores were used to produce linear rank statistics. The maximum of the statistics was used as the representative test statistic. This test statistic together with a permutation method were then used to estimate the p-value. This test maintains its significant level and it does not make use of selector statistics. It might not be classified in the category of adaptive test but it attains the same objectives as the adaptive methods.

In considering Hogg et al. (1975) adaptive two-sample tests, the HFR test, we observed that measure of asymmetry and tail weight are used. Though Hogg (1967) suggested the use of the sample kurtosis as a selector statistics, the measures of asymmetry and tail weight assists in choosing appropriate rank scores. To calculate this measure of asymmetry in the HFR test, the observations in both samples are combined, sorted and then the selector statistics computed. HFR measure of asymmetry though robust as compared to other estimators, the average of the lower and upper 5% could be very sensitive to outliers. The most appropriate scores are selected based on the two selector statistics. After classifying data, the selected rank test is applied. For symmetric heavy-tailed distributions, HFR selected the WMW test. For light-tailed symmetric models a modified rank test that scores the top an bottom 25% of the data suggesting the extreme parts bear more information on location shifts than the central parts. For right-skewed the lower 50% of the dataset is scored, this according to Gastwirth (1965) is because lower ranks are informative about median differences since they begin near the median. Lastly, HFR selected median tests for heavy-tails since they are optimal asymptotically for Laplace distributions with heavy-tails. In a study by Hao and Houser (2012),the median test was dropped to optimize HFR's algorithm because it was believed that they performed weakly.

2.10 Power of Adaptive tests

Hogg et al. (1975) in a simulation study show that their test maintains its level of significance and their adaptive tests exhibited more power as compared to the traditional methods both parametric and non-parametric.

To demonstrate the HFR method maintains its significant level, Basu's theorem is used. The test maintains a level of significance less than or equal to α, though the data is used in obtaining the scores because the tests are distribution-free. Selector statistics and test statistic are also independent. In addition, Hogg et al. (1975) proved

the actual level of significance was approximately α, using 15 observations per group in a simulation study. To demonstrate that adaptive tests are usually more powerful compared to the traditional methods for error distributions that are not normal O'Gorman (2012) makes a power comparison between the HFR test and the pooled t-test for many error distributions using 100,000 dataset for individual distributions and for each data set 15 observations were used.

In this study by O'Gorman (2012), the tests power was in the rejection proportions obtained from the number of null hypothesis rejected. In conclusion on his study, the test obtained powers for both t-tests and HFR test with all error distribution. However, HFR test showed more power over the t-test for a greater number of the distributions. The HFR test however lost some power for the normal, uniform and bimodal error distributions. According to O'Gorman tests based on ranks, make most sense when the datasets can be ranked. That is the HFR adaptive test, Wilcoxon test amongs others. This data ranking proved to be a challenge of tests based on ranks irrespective of their significance and other benefits. As an example, if two groups need to be compared and a covariate introduced, it could be difficult to find an appropriate rank test.

2.11 O'Gorman's Adaptive test

As a solution to these problems, a non-rank based adaptive test was proposed by O'Gorman (2001). This method makes use of a weighting adaptive scheme. In recent studies, many variants of the non-rank based method are suggested to allow for increase in a tests power and allow its usage in much more diverse models. The adaptive weighted test involves two simple steps. First, observations in the model are assigned weights to generate residuals which can be said to have a normal distribution. Secondly, a p-value is computed using a method of permutation. In theory, weighted least squares ensures errors have equal variability. Weights are

assigned observations to make their errors normally distributed. In the adaptive WLS method, extreme points are assigned smaller weights to decrease the effect of outliers. A p-value is computed using a method of permutation which in this case is lower than p-values obtained from unequal variance and pooled t tests. The simulation study showcased the fact that the t test losses power to the adaptive WLS test when distributions are non-normal. Both tests, adaptive WLS tests and HFR tests have similar power for distributions that are skewed.

2.12 Current Rank-based Procedures

In this research a modification of the HFR method is used where the selector statistics are just as in the HFR method but are more adaptive to the data and the functions of differences of averages of order statistics are used. The benchmarks proposed in the dissertation of Al-Shomrani (2003) is used. In his thesis, the cutoff values for the measures of skewness and tail weight depend on the sample size n. This method was effectively used by Okyere (2011) in a study on Robust Adaptive Scheme for Linear mixed models. An R package, Rfit, was developed by Kloke and Mckean (2012) for computing these robust procedures. The Rfit package enables easy estimation and inference of the rank-based method. It employs standard syntax for linear models enabling users abreast with traditional parametric methods much ease in running robust analyses. A library of score is included in the Rfit package. The package Rfit uses the Wilcoxon (linear) scores as default, nonetheless it is easy and direct for users to create score functions, and that is exactly what we seek to do in this study. Scores based on the symmetry and tail weight of our dataset which we call the bentscores are used. Simple R codes are written to order the dataset, measure the selector statistics and select an appropriate score, which is then specified in the Rfit to be used in estimation.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

The market model index by Sharpe (1963), presents a linear relationship between the stock and market returns that can be used to decompose total risk into systematic and unsystematic risk. By nature the market model is more appropriate for the study of stock risk and easier than CAPM with no loss in its explanatory power though it has limitations. The market model estimates beta by:

$$R_{it} = \alpha_i + \gamma_i R_{mt} + \varepsilon_{it} \quad (3.1)$$

Where

R_{it}: is the realized return on asset i for the period t t: is the measurement interval and t = 1, 2, . . ., T. T: is the number of measurement intervals α_i: is the intercept term for asset i

γ_i: is the sensitivity measure of return on asset i to market R_{mt}: is the realized return on the market index for period t ε_{it}:is the residual term for asset i in period t

In practice, regression analysis has the following assumptions:

- The response variable (R_{it}) and the explanatory variable (R_{mt}) are linearly related so that

$$R_i|R_m = m(R_m) = \alpha_i + \gamma_i R_m \quad (3.2)$$

- The conditional distribution of R_i, is a normal distribution.

  $R_i \sim N(\alpha_i + \gamma_i R_m, \sigma^2)$

- Observations are independently sampled and

R_{it} and R'_{it} are independent for $it \neq it'$

We have different ways to model the conditional expectation function m(.). This work focuses on two of such models

- Parametric
- Non-parametric (Adaptive methods)

3.2 PARAMETRIC

3.2.1 Least Squares Methods

Suppose a dataset is made up of n observations $\{Y_i, X_i\}_{i=1}^{n}$. Each observation has a response variable, Y_i and a vector of p predictor variables X_i. The response variable is a linear function of the predictor variables in a linear regression model

.

$$Y_i = X_i^T \gamma + \epsilon_i \tag{3.3}$$

Where γ - is a $p \times 1$ vector of unknown parameters

$_i$ - an unobserved random variable that captures the unexplained variations in Y not caused by the $X_i's$

T - the matrix transpose

$X_i^T \gamma$ - the dot product between the vectors X and γ.

This model is written in matrix form as

$$Y = X\gamma + \epsilon \tag{3.4}$$

Where

Y - (n×1) vectors of response variables

ϵ - $(\mathrm{n}\times$1) vectors of errors

X - an (n×p) matrix of explanatory variables γ - a

(p×1)vector

3.2.2 Estimating Ordinary Least Squares

Given a linear regression model

$$Y_i = \alpha + \gamma_i X_i + \epsilon_i \tag{3.5}$$

In solving for the unknown parameter α and γ using OLS the aim is to find the best fitted model by finding that which will yield the minimum residual sum of squares (SSR). A good fit yields minimum mean and variance. Thus with

$$SSR = \sum_{i=1}^{n} \epsilon^2 = \sum_{i=1}^{n} (Y_i - (\alpha + \gamma X_i))^2 \tag{3.6}$$

Our goal is to find α and γ values that minimize the SSR. Thus *Proof* 3.2.2

$$\frac{\delta SSR}{\delta \alpha} = 0 \text{ and } \frac{\delta SSR}{\delta \gamma} = 0$$

For $\frac{\delta SSR}{\delta \alpha} = 0$

$$\frac{\delta SSR}{\delta \alpha} = \frac{\delta}{\delta \alpha}[\textstyle\sum_{i=1}^{n}(Y_i - (\alpha + \gamma X_i))^2]$$

$$= \textstyle\sum_{i=1}^{n} 2[Y_i - (\alpha + \gamma X_i)](-1)$$
$$= \textstyle\sum_{i=1}^{n} -2[Y_i - (\alpha + \gamma X_i)]$$

$$\frac{\delta SSR}{\delta \alpha} = -2\sum_{i=1}^{n}[Y_i - (\alpha + \gamma X_i)] = 0 \tag{3.7}$$

For $\frac{\delta SSR}{\delta \gamma} = 0$

$$\frac{\delta SSR}{\delta \gamma} = \frac{\delta}{\delta \gamma}[\sum_{i=1}^{n}[Y_i - (\alpha + \gamma X_i)]^2]$$

$$= \sum_{i=1}^{n} 2[Y_i - (\alpha + \gamma X_i)](-X_i)$$

$$= \sum_{i=1}^{n} -2[Y_i - (\alpha + \gamma X_i)](X_i)$$

$$\frac{\delta SSR}{\delta \gamma} = -2\sum_{i=1}^{n}[Y_i - (\alpha + \gamma X_i)](X_i) = 0 \qquad (3.8)$$

Solving equation (3.7)

$$\frac{\delta SSR}{\delta \alpha} = -2\sum_{i=1}^{n}[Y_i - (\alpha + \gamma X_i)] = 0$$

Dividing through by -2

$$P_{n_{i=1}}[Y_i - (\alpha + \gamma X_i)] = 0$$

$$P_n Y_i - P_{ni=1}\,\alpha - P_{ni=1}\,\gamma X_i = 0 \;_{i=1}$$

$$\sum_{i=1}^{n} Y_i - n\alpha - \gamma \sum_{i=1}^{n} X_i = 0$$

$$n\alpha + \gamma \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i$$

Solving equation(3.8)

$$\frac{\delta SSR}{\delta \gamma} = -2\sum_{i=1}^{n}[Y_i - (\alpha + \gamma X_i)](X_i) = 0$$

Dividing through by -2

$$\sum_{i=1}^{n}[Y_i - (\alpha + \gamma X_i)](X_i) = 0$$

$$\sum_{i=1}^{n}[X_i Y_i - (\alpha X_i + \gamma X_i^2)] = 0$$

$$\sum_{i=1}^{n} X_i Y_i - \alpha \sum_{i=1}^{n} X_i - \gamma \sum_{i=1}^{n}(X_i^2) = 0$$

$$\alpha \sum_{i=1}^{n} X_i + \gamma \sum_{i=1}^{n}(X_i^2) = \sum_{i=1}^{n} X_i Y_i$$

In matrix form in finding the inverse

$$\begin{bmatrix} \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \end{bmatrix} \quad \begin{bmatrix} n & (\mathrm{P}X_i) \\ \mathrm{P}X_i & \mathrm{P}(X_i)_2 \end{bmatrix} \quad \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} \quad \begin{bmatrix} \\ \\ \end{bmatrix} \tag{3.9}$$

$$\begin{bmatrix} \mathrm{P}y_i \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{P}X_i\mathrm{P}(X_i)_2 \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \\ \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \gamma \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \mathrm{P}X_iY_i \\ \\ \\ \end{bmatrix}$$

$$\begin{bmatrix} \\ \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \alpha \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} n & (\mathrm{P}X_i) \\ \\ \\ \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \mathrm{P}y_i \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

$$\begin{bmatrix} \\ \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

$$\begin{bmatrix} \\ \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \gamma \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \mathrm{P}X_i\mathrm{P}(X_i)_2 \\ \\ \\ \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{P}X_iY_i \\ \\ \\ \end{bmatrix}$$

of matrix M

$$\begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} n & (\mathrm{P}X_i) \\ \\ \\ \end{bmatrix}$$

let M= $\begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \\ \\ \\ \end{bmatrix}$

$$\begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \begin{bmatrix} \mathrm{P}X_i\mathrm{P}(X_i)_2 \\ \\ \\ \end{bmatrix}$$

$$detM = n\sum_{i=1}^{n}(X_i)^2 - (\sum_{i=1}^{n} X_i)^2$$

Modifying it

$$detM = n\sum_{i=1}^{n}(X_i)^2 - \frac{n}{n}(\sum_{i=1}^{n} X_i)\frac{n}{n}(\sum_{i=1}^{n} X_i)$$

$$= n\sum_{i=1}^{n}(X_i)^2 - n^2\bar{x}^2$$

$$= n^2\left[\frac{\sum_{i=1}^{n}(X_i)^2}{n} - \bar{x}^2\right]$$

$$M^{-1} = \frac{1}{detM}adjM$$

$$= \frac{1}{n\sum(X_i)^2 - \left(\sum X_i\right)^2}\begin{bmatrix}\sum(X_i)^2 & -(\sum X_i)\\ -\sum(X_i) & n\end{bmatrix}$$

$$\begin{pmatrix}\alpha\\ \gamma\end{pmatrix} = \frac{1}{n\sum(X_i)^2 - \left(\sum X_i\right)^2}\begin{bmatrix}\sum(X_i)^2 & -(\sum X_i)\\ -\sum(X_i) & n\end{bmatrix}\begin{pmatrix}\sum y_i\\ \sum X_iY_i\end{pmatrix}$$

$$= \frac{1}{n\sum(X_i)^2 - \left(\sum X_i\right)^2}\begin{bmatrix}\sum(X_i)^2\sum y_i - (\sum X_i)\sum X_iY_i\\ -\sum(X_i)\sum y_i + n\sum X_iY_i\end{bmatrix}$$

$$= \frac{1}{n\sum(X_i)^2 - \left(\sum X_i\right)^2}\begin{bmatrix}\sum(X_i)^2\sum y_i - (\sum X_i)\sum X_iY_i\\ n\sum X_iY_i - \sum(X_i)\sum y_i\end{bmatrix} \quad (3.10)$$

From 3.10

$$\alpha = \frac{1}{n\sum(X_i)^2 - \left(\sum X_i\right)^2}\left[\sum(X_i)^2\sum y_i - (\sum X_i)\sum X_iY_i\right]$$

$$\alpha = \frac{\bar{Y}\sum(X_i)^2 - \bar{X}\sum X_iY_i}{\sum(X_i)^2 - \frac{1}{n}\left(\sum X_i\right)^2}$$

$$\alpha = \frac{\bar{Y}\sum(X_i)^2 - \bar{X}\sum X_iY_i}{\sum(X_i)^2 - n\bar{X}^2}$$

$$\gamma = \frac{1}{n\sum(X_i)^2 - \left(\sum X_i\right)^2}\left[n\sum X_iY_i - \sum(X_i)\sum y_i\right]$$

$$= \frac{n}{n}\left[\frac{\sum X_iY_i - \frac{1}{n}\sum(X_i)\sum y_i}{\sum(X_i)^2 - \frac{1}{n}\left(\sum X_i\right)^2}\right]$$

$$= \frac{\sum X_iY_i - \frac{n}{n}\frac{1}{n}\sum(X_i)\sum y_i}{\sum(X_i)^2 - \frac{n}{n}\frac{1}{n}\left(\sum X_i\right)^2} \quad (3.11)$$

$$\gamma = \frac{\sum X_i Y_i - n\bar{Y}\bar{X}}{\sum (X_i)^2 - n\bar{X}^2} \tag{3.12}$$

Further

$$P(X_i - \bar{X})_2 = P(X_{i2} - 2X_i\bar{X} + \bar{X}_2)$$

$$= P(X_i)_2 - 2PX_i\bar{X} + P\bar{X}_2$$

$$= P(X_i)_2 - 2n\bar{X}_2 + n\bar{X}_2$$

$$= P(X_i)_2 - n\bar{X}_2$$

$$P(X_i - \bar{X})(Y_i - \bar{Y}) = P(X_iY_i) - X\bar{Y}_i - X_i\bar{Y} + \bar{X}\bar{Y}$$

$$= P(X_iY_i) - P(X_i\bar{Y}) - P(X\bar{Y}_i) + PX\bar{Y}$$

$$P_{(X_iY_i)} - nX\bar{Y} - nY\bar{X} + nX\bar{Y}$$

$$P_{(X_iY_i)} - nX\bar{Y}$$

Therefore 3.12 becomes

$$\gamma = \frac{\sum X_i Y_i - n\bar{Y}X}{\sum (X_i)^2 - n\bar{X}^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \tag{3.13}$$

and 3.10

$$\alpha = \bar{Y} - \gamma\bar{X} \tag{3.14}$$

In an instance where there is more than one independent variable or regressor, the matrix approach is simplest in estimation of the unknown parameters. We will suppose that the linear model is

$$Y = \gamma_o + \gamma_1 X_1 + \gamma_2 X_2 + \cdots + \gamma_k X_k + \epsilon \tag{3.15}$$

In matrix representation

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ 1 & x_{13} & x_{23} & \cdots & x_{k3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \quad \hat{\gamma} = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_k \end{bmatrix}$$

Where Y is an (n×1) vector of observations with X of order (n×p) and γ a (p×1) vector parameter where p is (k+1) . To understand how the matrix approach works we look at a replica of the problem.

The matrix procedure for expressing a system of k simultaneous equations in k unknowns. If the equations are written in the orderly pattern *Example* 3.2.1

$$a_{11}v_1 + a_{12}v_2 + \cdots + a_{1k}v_k = g_1 \; a_{21}v_1 +$$

$$a_{22}v_2 + \cdots + a_{2k}v_k = g_2$$

...

$$a_{k1}v_1 + a_{k2}v_2 + \cdots + a_{kk}v_k = g_k$$

then the set of simultaneous linear equations can be expressed as the matrix equation,

$$AV = G \tag{3.16}$$

where

$$A=V=G=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \quad \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix} \quad \begin{matrix} g_1 \\ g_2 \\ \dots \\ g_k \end{matrix}$$

Now let us solve this system of simultaneous equations. If they are uniquely solvable, it can be shown that (A^{-1}) exists. Multiplying both sides of the matrix equation by (A^{-1}) ,we have *Solution*

$$(A^{-1})AV = (A^{-1}G) \tag{3.17}$$

But since $A^{-1}A = I$,we have

$$I(V) = A^{-1}G$$

$$V = A^{-1}G \tag{3.18}$$

That is if we know A^{-1} ,we can find the solution to the set of simultaneous linear equation by obtaining the product $A^{-1}G$. For I, an identity matrix with elements being 1 as the diagonal entries and 0 elsewhere, any element multiplied by *I* results in that element thus *I(V)=V*. To implement this in our matrix representation of the OLS method, where

$$Y = \gamma X$$
$$0 \qquad 0$$

$$(X X)\hat{\gamma} = X Y$$

$$(X X)^{-1}(X_0 X)\hat{\gamma} = (X_0 X)^{-1} X_0 Y \tag{3.19}$$

where $(X X)^{-1}(X_0 X) = I$

$$I\hat{\gamma} = (X X)^{-1} X_0 Y$$

$$\hat{\gamma} = (X X)^{-1} X_0 Y \tag{3.20}$$

This yields a (k+1)×1 vector of the γ elements. The matrix formulas for the SSE are

$$SSE = Y_0 Y - \gamma_0 (X_0 Y) \tag{3.21}$$

Example 3.2.2: Using the method of matrix differentiation

Let $\beta=(\beta_1,\cdots,\beta_k)_0$ be a k×1 vector and let $f(\beta)=F(\beta_1,\cdots,\beta_k)$ be a real-valued function that depends on β. i.e. $f(\cdot) : R^k \rightarrow R$ maps the vector β into a single number, $f(\beta)$. Then the derivative of $f(\cdot)$ with respect to β is defined as

$$\frac{\delta f(\beta)}{\delta \beta} = \begin{pmatrix} \frac{\delta f(\beta)}{\delta \beta_1} \\ \vdots \\ \frac{\delta f(\beta)}{\delta \beta_k} \end{pmatrix} \tag{3.22}$$

This is a k×1 column vector with typical elements given by the partial derivative $\frac{\delta f(\beta)}{\delta \beta_i}$ which is also known as the gradient. To illustrate the use of matrix differentiation consider the linear regression model in matrix notation,

$$Y = X\beta + \epsilon \tag{3.23}$$

where Y is an (n×1) vector of observation, X is a (n× p) matrix of explanatory variables, β is a (p×1)vector of parameters to be estimated, and is a (n×1) vector of error terms. One way to motivate the ordinary least squares (OLS) principle is to choose the estimator, $\hat{\beta}_{OLS}$ of β, as the value that minimizes the sum of squared residuals, i.e.

$$\hat{\beta}_{OLS} = argmin \sum_{t=1}^{T} \hat{\epsilon}_t^2 = argmin \hat{\epsilon}' \hat{\epsilon}$$ (3.24) *Solution* Looking at the function to be minimized, we find that

$$\hat{\epsilon}'\hat{\epsilon} = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$= (Y' - \hat{\beta}'X')(Y - X\hat{\beta})$$

$$= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

$$= Y'Y - 2Y'X\hat{\beta} + \hat{\beta}X'X\hat{\beta} \quad (3.25)$$

Where the last line uses the fact that $Y'X\hat{\beta}$ and $\hat{\beta}'X'Y$ are identical scalar variables. Taking the first derivative with respect to $\hat{\beta}$ yields the (n×1) vector

$$\frac{\delta(\hat{\epsilon}'\hat{\epsilon})}{\delta\hat{\beta}} = \frac{\delta(Y'Y - 2Y'X\hat{\beta} + \hat{\beta}X'X\hat{\beta})}{\delta\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} \quad (3.26)$$

Solving the k equations, $\frac{\delta(\hat{\epsilon}'\hat{\epsilon})}{\delta\hat{\beta}} = 0$, yields the OLS estimator

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \quad (3.27)$$

Provided that $X'X$ is non-singular. From the above it is observed that Least squares offers a method of model fitting that minimizes the Euclidean distance between the response variable and its mean given the explanatory variables.

3.3 NORM

A norm is a function that assigns a strictly positive length or size to each vector in a vector space other than the zero vector which has zero length assigned to it. As defined by Wikipedia, the free encyclopedia a norm is a nonnegative function $\|.\|$ defined on R^n which follows the following properties

- $\|y\| \geq 0$ for all y
- $\|y\| = 0$, iff (if and only if) $y = 0$

- kαyk = $|\alpha|$kyk for all real z, Positive homogeneity
- ky + zk ≤ kyk + kzk, Triangle inequality

A consequence of the last two conditions is that a norm only assumes nonnegative values, and that it is convex.

3.3.1 L_p - norm

Popular norms include the so called l_p-norms, where p=1, 2 or p=∞:

$$\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p} \tag{3.28}$$

with the convention that when p=∞,

$$\text{k}x\text{k}_\infty = max_{1\leq i\leq n}|x_i| \text{ or } max|x_i| \tag{3.29}$$

i=1, 2, ..., n.

Note that when p=1 we get the taxicab norm and p=2 is the Euclidean norm. For $0 < p < 1$ the resulting function does not define a norm because the triangle inequality is violates .

3.3.2 Derivative of the L_p - norm.

Proof: The derivative of the L_p-norm is given by

$$\frac{\delta}{\delta x_k}\|x\|_p = \frac{x_k|x_k|^{p-2}}{\|x\|_p^{p-1}} \tag{3.30}$$

For p=2,

$$\frac{\delta}{\delta x_k}\|x\|_2 = \frac{x_k|x_k|^{2-2}}{\|x\|_2^{2-1}} = \frac{x_k}{\|x\|_2}$$

or

$$\frac{\delta}{\delta x_k}\|x\|_2 = \frac{x}{\|x\|_2} \tag{3.31}$$

3.3.3 Euclidean norm

On an n-dimensional Euclidean space R^n, the intuitive notion of length of the vector $x = (x_1,x_2,....,x_n)$ is captured by the formula

$$\|x\| = (\sum_{i=1}^{n} |x_i|^2)^{1/2} = \sqrt{x_1^2 + ... + x_n^2} \tag{3.32}$$

this gives the ordinary distance from the origin to the point x, a consequence of the Pythagorean theorem. The Euclidean norm is by far the most commonly used norm on R^n. The norm can also be expressed as the square root of the inner product of the vector and itself

$$\text{k}x\text{k} = \sqrt{x^*x} \tag{3.33}$$

Where x is represented as a column vector$(x_1;x_2;...;x_n)$ and x^* denotes its conjugate transpose. This formula is valid for any product space, including Euclidean and complex spaces. For Euclidean spaces, the inner product is equivalent to the dot product. Hence, in this specific case the formula can also be written with the notation:

$$\text{k}x\text{k} = \sqrt{x.x} \tag{3.34}$$

The Euclidean norm is also called the Euclidean length,L_2 distance,L_2 norm.

3.3.4 Manhattan or Taxicab norm

$$\|x\|_1 = (\sum_{i=1}^{n} |x_i|^1)^{1/1} = \sum_{i=1}^{n} |x_i| \tag{3.35}$$

The taxicab norm also known as the L_1 norm gets its name from the distance a taxi has to drive in a rectangular street grid to get from the origin to the point x. The distance derived from this norm is called the Manhattan distance or the L_1 distance.

The L_1 -norm is simply the sum of the absolute values of the columns. In contrast $\sum_{i=1}^{n} x_1$ is not a norm because it may yield negative results

3.3.5 Dispersion Function

The distance between two vectors x and y, is represented and expressed in norms as;

$$d(x,y) = \mathrm{k}x - y\mathrm{k} \tag{3.36}$$

Given a linear model, and a specific norm $\mathrm{k}\cdot\mathrm{k}$, the estimate of γ is given as

$$\hat{\gamma} = argmin\mathrm{k}Y - \gamma X\mathrm{k} \tag{3.37}$$

That is a value that minimizes the distance between Y and γX. The minimum distance, known as the dispersion function, between Y and γX is D(γ). The dispersion function is expressed in terms of a norm is

$$D(\gamma) = \mathrm{k}Y - \gamma X\mathrm{k} \tag{3.38}$$

the dispersion function induced by the norm. $D(\gamma)$ is differentiable.

The gradient process is defined by the function

$$S(\gamma) = -\frac{\delta}{\delta\gamma} D(\gamma) \tag{3.39}$$

$S(\gamma)$ is a non increasing. At the points where $D(\gamma)$ is non differentiable, the function is discontinuities. From the derivative function of an L_2 - norm above, we have the gradient process,

$$S(\gamma) = \frac{\delta}{\delta\gamma}\|Y - \gamma X\|_2 = \frac{Y_k - \gamma X_k|Y_k - \gamma X_k|^{2-2}}{\|Y - \gamma X\|_2^{2-1}} = \frac{Y_k - \gamma X_k}{\|Y - \gamma X_k\|_2} \tag{3.40}$$

The value for which $S(\gamma)$ is 0 is the minimizing value.

That is γ is the solution to

$$S(\hat{\gamma}) = 0$$

3.3.6 Pseudo-norm

An operation k.k is called a Pseudo-norm if it satisfies the following four conditions

- $\|u+v\|_\varphi \leq \|u\|_\varphi + \|v\|_\varphi \ \forall$ u,v $\epsilon\ R^n$
- $\|\ \alpha u\|_\varphi \leq |\alpha|\|u\| \ \forall\ \alpha \epsilon$R, u$\epsilon\ R^n$
- $\|u\|_\varphi \geq 0 \ \forall$ u $\epsilon\ R^n$
- kuk$_\phi$ = 0, if and only if $u_1 = u_2 = u_3 = \ldots = u_n$

The rank-bases score is estimated using the following Pseudo-norm,

$$\|v\|_\varphi = \sum_{i=1}^{n} \alpha(R(v_i))v_i, v\epsilon R^n$$

3.4 RANK-BASED ANALYSIS

Let Y be an n×1 vector of responses which follows a linear model given by

$$Y = 1\gamma_o + X\gamma + \varepsilon \tag{3.41}$$

where

1 is a vector of n ones X is an n × p

design matrix α is an intercept

parameter γ is a p × 1 vector of

parameters ε is an n × 1 vector of

errors

Assume the ε are iid and it had pdf f(x) and an unknown cdf F(x). $\hat{\gamma}_{LS}$ is the estimator that minimizes the Euclidean distance between Y and X satisfying

$$\hat{\gamma}_{LS} = Argmin\|y - X\gamma\|_2^2 \tag{3.42}$$

where $\|v\|_2^2 = \sum_{i=1}^{n} v_i^2$ is an L_2 norm. For rank-based estimates replace the L_2 norm $\|.\|_2^2$ with another norm, the pseudo-norm

$$\|v\|_{\varphi} = \sum_{i=1}^{n} a(R(v_i))v_i, \quad v \epsilon R^n \tag{3.43}$$

$R(v_i)$ represents the rank of v_i for $v_1,\ldots,v_n$ and *a(i)* are scores generated as $a[i] = \varphi(\frac{i}{n+1})$, for a nondecreasing bounded square-integrable function $\phi(u)$,satisfying, without loss of generality, the standardizing conditions $\int \phi(u)du = 0$ and $\int \phi^2(u)du = 1$. Then the rank-based estimator minimizes the $\|\cdot\|_{\phi}$ distance between Y and the column space of X; ie.,

$$\hat{\gamma}_{\phi} = Argmin\|y - X\gamma\|_{\phi} \tag{3.44}$$

The Wilcoxon score function $\phi[u] = \sqrt{12}[u - (1/2)]$, is used. The rank-based estimator of the intercept parameter is a location estimate based on the residuals. For the LS method, the arithmetic mean is used while for the rank-based estimates, generally, the median is used. i.e.,

$$\hat{\alpha} = med_i Y_i - x^T \hat{\gamma}_{\phi} \tag{3.45}$$

The R package Rfit can compute rank-based analysis. Kloke and Mckean (2012) developed this package and it can be downloaded at CRAN. Note that closed form solutions exist for least squares, however, this is not the case for rank estimation. The R estimates are obtained by minimizing a convex optimization problem. It can be shown, see for example Hettmansperger and McKean (2011), that the solution to $\hat{\gamma}_{\phi} = Argmin\|y - X\gamma\|_{\phi}$ is consistent with the asymptotically normal distribution given by

$$\hat{\gamma}_{\varphi} \sim N(\gamma, \tau_{\varphi}^2(X^T X)^{-1}) \tag{3.46}$$

where τ_{ϕ} is the scale parameter which depends on the pdf f(t) and the score function $\phi(u)$. In Rfit, the Koul, Sievers, and McKean (1987) consistent estimator of τ_{ϕ} is computed.

3.4.1 Robustness

In this section, we briefly discuss the robustness properties of the rank-based estimators. Three of the main concepts in robustness are efficiency, influence, and breakdown.

3.4.2 Influence Functions

The finite sample version of the influence function of an estimator is its sensitivity curve. It measures the change in an estimator when an outlier is added to the sample. More formally, let the vector $x_n = (x_1,x_2,\ldots,x_n)$ denote a sample of size n. Let $\hat{\theta}_n = \hat{\theta}_n(x_n)$ denote an estimator. Suppose we add a value x to the sample to form the new sample $x_{n+1} = (x_1,x_2,\ldots,x_n,x)$ of size n + 1. Then the sensitivity curve for the estimator is defined by

$$S(x;\hat{\theta}) = \frac{\hat{\theta}_{n+1} - \hat{\theta}_n}{1/n+1} \tag{3.47}$$

The value $S(x;\hat{\theta})$ measures the rate of change of the estimator at the outlier x. While intuitive, a sensitivity curve depends on the sample items. Its theoretical analog is the influence function which measures rate of change of the functional of the estimator at the probability distribution, F(t), of the random errors of the location model. We say an estimator is robust if its influence function is bounded. Note that the sensitivity curve for the mean is unbounded; i.e., as the outlier becomes large the rate in change of the mean becomes large; i.e., the curve is unbounded. Influence functions describes the approximate and standardized effect of an additional observation in any point x on a statistic T, given a (large) sample with distribution F. The effect of one outlier on the estimator can be described by the influence function.

Let T be a statistical functional defined on a space of distribution function, F be a distribution function in the domain of T. T is gateaux differentiable at F if for any

distribution function G, such that the distribution function $(1 - s)F + sG$, which is a contaminated distribution, lie in the domain of T. The Gateaux derivative of T at F in the direction G is defined by:

$$L_F(T, G) = lim_{\epsilon \longrightarrow 0} \left[\frac{T(1-\epsilon)F + \epsilon G - T(F)}{\epsilon} \right] \tag{3.48}$$

An equivalent way of stating the definition is to define D = G - F and the above becomes

$$L_F(T, G) = lim_{\epsilon \longrightarrow 0} \left[\frac{T(F + \epsilon D - T(F)}{\epsilon} \right] \tag{3.49}$$

From a statistical perspective, it represents the rate of change in a statistical functional upon a small amount of contamination by another by another distribution G

As an example;

Suppose F is a continuous CDF, and G is the distribution that places all of its mass at the point x_o. The Gateaux derivative of $T(F) = f(x_o)$ is

$$L_F(T, G) = lim_{\epsilon \to 0} \left[\frac{\frac{\delta}{\delta x}((1-\epsilon)F(x) + \epsilon G(x))_{x=x_0} - \frac{\delta}{\delta x} F(x)|_{x=x_o}}{\epsilon} \right]$$

$$L_F(T, G) = lim_{\epsilon \to 0} \left[\frac{(1-\epsilon)f(x_o) + \epsilon g(x_o) - f(x)}{\epsilon} \right] = \infty \tag{3.50}$$

A useful condition is that if the functional is bounded, then the plug-in estimate will converge to the true value. Statisticians usually do not work with the general Gateaux derivative but a special case of it called the influence function, in which G places a point mass of 1 at x:

$$0 \; \delta_x(u) = \begin{cases} \text{if } u < x \text{ if} \\ 1 \qquad u \geq x \end{cases}$$

The influence function is usually written as a function of x, and defined as

$$L(x) = lim_{\epsilon \longrightarrow 0} \left[\frac{T(1-\epsilon)F + \epsilon \delta_x - T(F)}{\epsilon} \right]$$

A closely related concept is that of the empirical influence function:

$$\hat{L}(x) = lim_{\epsilon \longrightarrow 0}\left[\frac{T(1-\epsilon)\hat{F} + \epsilon\delta_x - T(\hat{F})}{\epsilon}\right] \tag{3.51}$$

If T(F) can be written in the form $T(F) = \alpha\{T_1(F), T_2(F), \ldots.\}$ then

$(x) = \sum_j \frac{\delta a}{\delta T_j}|_F L_j(x)$ where $L_j(x)$ is the influence function of $T_j(F)$. Next we derive the Influence function of $\hat{\gamma_\phi}$, that is the estimate of gamma for a specified score function $\phi(u)$ Let H be the joint distribution function of X and Y. Let the px1 vector T(H) denote the functional corresponding to $\hat{\beta}_\phi$ assume without loss of generality that the true $\gamma = 0, \alpha = 0$and that E(x)=0. Hence the distribution function of Y is F(y) and Y, X are independent meaning H(x,y) = M(x)F(y).

Recall that the R-estimate satisfies the equations

$$\sum x_i a(R(Y_i - X_i'\gamma)) = 0 \tag{3.52}$$

Let $\hat{G}_n^*$denote the empirical distribution function of$Y_i - X_i'\gamma$.Then we can rewrite the above equation as;

$$n\sum x_i \varphi(\frac{n}{n+1}\hat{G}_n^*(Y_i - X_i'\gamma))\frac{1}{n} = 0 \tag{3.53}$$

Let G^*denote the distribution function of $Y - X_0 T(H)$ then the functional T(H) satisfies

$$\int \phi(G^*(Y - x_0 T(H))xdH(x,y) = 0 \tag{3.54}$$

We can show that $G^*(t) = s_{u \le vT(H)+t}\, dH(v,u)$ Let $H_s = (1-s)H + sW$ for an arbitrary distribution function W. Then the functional T(H) evaluated at H_s satisfies the equation.

$$(1-s)\int \phi(G^*_s(Y - x\,T(H))xdH(x,y) + s\int \phi(G^*_s(Y - x\,T(H_s))xdW(x,y) = 0 \tag{3.55}$$

WhereG_s^* is the distribution function of $Y - x_0 T(H_s)$. We obtain $\frac{\delta T}{\delta s}$ by implicit

differentiation. Then upon substituting $M_{x_o y_o}$ for W the influence function is given by $\frac{\delta T}{\delta s}|_{s=0}$ which we will denote by T. Implicit differentiation leads to

$$0 = -\int \varphi(G_s^*(Y - x'T(H_s))xdH(x,y) - (1-s)\int \varphi(G_s^*(Y - x'T(H_s)) + \frac{\delta G_s^*}{\delta s}xdH(x,y) + \int \varphi(G_s^*(Y - x'T(H_s))xdW(x,y) + s\gamma_1$$

Since s will be set to 0, γ_1 is irrelevant.

First get the partial derivative of G_s^* with respect to s

$$G^*_s(Y - x^0T(H_s)) = s_{u \le y - T(Hs)^0(x-v)}\, dH_s(v,u) =$$

$$(1-s)^R F[y - T(H_s)^0(x-v)]dM(v) + ss_{u \le y - T(Hs)^0(x-v)}\, dW(v,u)$$ Thus:

$$\frac{\delta G_s^*(Y - X'T(H_s))}{\delta s} = -\int F[y - T(H_s)'(x-v)]dM(v) + (1-s)\int F'[y - T(H_s)'(-v)](v-x)'\frac{\delta T}{\delta s}dM(v) + \iint_{u \le y - T(H_s)'(x-v)} dW(v,u) + s\gamma_2$$ Since s is set to 0, γ_2 is irrelevant as well. Therefore using the independence between Y and X at H, we get

$$\frac{\delta G_s^*(Y - XT(H_s))}{\delta s}|_{s=0} = -F(y) - f(y)x'T + W_y(y)$$

$$0 = -\int x\varphi(F(y))dH(x,y) + \int x\varphi'(F(y))[-F(y) - f(y)x'T + W_y(y)]dH(x,y) + \int x\varphi(F(y))dW(x,y)$$

$$= -\int \varphi'(F(y))f(y)xx'TdH(x,y) + \int x\varphi(F(y))dW(x,y) \quad (3.56)$$

Substituting $M_{x_o y_o}$ in for W we get

$$0 = \tau^P T + x_o\phi(F(y_o))$$

Solving for T, the influence function of $\hat{\gamma_\phi}$ is given by

$$(x_o, y_o, \hat{\gamma}_\varphi) = \tau \sum^{-1} \varphi(F(y_o))x_o \quad (3.57)$$

Thus the proof of robustness of $\hat{\gamma_\phi}$

3.4.3 Breakdown Point

The idea of contaminating a distribution with a small amount of additional data has a long history in statistics and the investigation of robust estimators.The breakdown value is defined as how much contaminated data an estimator can tolerate before it becomes useless. Breakdown point measures the ability of a statistic to resist the outliers contained in the data set.

Let $X_n = (x_1,x_2,....,x_n)$ represent a realization of a sample wherex_i's are n independent and identically distributed observations from the distribution F_x.

Assume that $m < n$ and replace $x_1,x_2,...,x_m$ with $x_1^*, x_2^*, ..., x_m^*$ let $X^{(m)} = (x_1^*, x_2^*, ..., x_m^*, x_{m+1}, x_{m+2}, ..., x_n)'$ represent the corruption of any m of the n observations with$Q(X_n), Q(X_n^*)$ being estimators or test statistics of θ .

Now the maximum bias

$$maxBias = maxBias(m, X_n, Q) = sup_{x_1^*,x_2^*,...,x_m^*} d(Q(X_n), Q(X_n^*)) \quad (3.58)$$

Where d (.,.) denotes some distance function (eg. The Euclidean distance). The finite sample breakdown point is now given by $BP(Q,n)$ =

$min_m\{\frac{m}{n}|maxBias(m, x_n, Q) = \infty\}$

And the asymptotic breakdown point is

$$BP(Q) = lim_{n\to\infty} BP(Q,n) \quad (3.59)$$

Often there exists an integer m such that $x_{(m)} \le \hat{\theta} \le x_{(n-m+1)}$ and either $\hat{\theta}$, an estimated parameter, tends to $-\infty$ as $x_{(m)}$ tends to $-\infty$ or $\hat{\theta}$ tends to $+\infty$ as $x_{(n-m+1)}$ tends to $+\infty$.

If m^* is the smallest such integer then ϵ_n^*, breakdown point, is $\frac{m^*}{n}$. For breakdown points values close to 0.5 are desirable.

3.4.4 Breakdown value of the L_1 and L_2 estimates

The L_1 estimate is the sample median. If the sample size is n = 2k; then when $x_{(k)}$ tends to $-\infty$, the median also tends to $-\infty$ hence the breakdown value of the sample

median is $\frac{k}{n}$ which tends to 0.5. By a similar argument, when the sample size is n = 2k+1, the breakdown value is $\frac{(k+1)}{n}$ and it also tends to 0.5 as the sample size increases. Hence we say the sample median is a 50% breakdown estimate thus the sample median can tolerate almost half of the data being contaminated. The L_2 estimate is the sample mean. Notice for the sample mean that one point of contamination suffices to make the mean meaningless $(asx_1^* \longrightarrow \infty, x \longrightarrow \infty)$. The breakdown value is $\frac{1}{n}$, hence the breakdown point of the mean is 0.

3.4.5 Asymptotic Relative Efficiency

For any two test statistics that are consistent, P and Q, of any hypothesis H_o, the asymptotic relative efficiency is the ratio of sample sizes needed to get identical power against the same alternative H_1, taking the limit as the sample size n tends to infinity and as H_1 tends to H_o, according to Hao and Houser (2012). This implies that the asymptotic relative efficiency(ARE) lies in the interval (0,1) when the tests are positive ie. ARE (P,Q)$^{\epsilon}$ $(0, \infty)$.

When ARE (P,Q) (0,1) then the test statistic P is regarded less efficient than Q, the test P is however considered efficient as the test Q when the ARE (P,Q) = 1, lastly the test P is more efficient than the test Q when the ARE (P,Q) (1, +∞), Hao and Houser (2012). Alternatively, let T_P and T_Q be two linear rank statistics based on the score generating functions P and Q. Then the asymptotic relative efficiency (ARE) is given by

$$ARE(T_P, T_Q/f) = \frac{AE(T_P/f)}{AE(T_Q/f)} \tag{3.60}$$

where $AE(T_P/f)$ and $AE(T_Q/f)$ are the asymptotic efficacies of P and Q respectively, Kossler (2010)

Definition: The asymptotic relative efficiency between two tests or estimates based on the score functions $\phi_1(u)$ and $\phi_2(u)$ or one score function relative to other score function is defined by;

$$e(\varphi_1, \varphi_2) = \frac{C_{\varphi_1}^2}{C_{\varphi_2}^2} = \frac{\tau_{\varphi_2}^2}{\tau_{\varphi_1}^2} \tag{3.61}$$

where $C_{\phi 1}$ and $C_{\phi 2}$ are respectively the efficacies of the two estimates and $\tau_{\phi i}$, i = 1, 2 are the scale parameters of the two score functions.

3.4.6 Optimal Scores

Rank-based analysis require the use of score functions $\phi(u)$. If there is an idea on the form of the errors underlying distribution, one can get optimal score function which minimizes the estimators variance. From $\tau_{\varphi}^{-1} = \int \varphi(u)\varphi_f(u)du$ and $\varphi_f(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}$ we can rewrite $1/\tau_{\phi}$

$$\begin{aligned} \tau_{\varphi}^{-1} &= \int_0^1 \varphi(u)\varphi_f(u)du \\ &= \frac{\int_0^1 \varphi(u)\varphi_f(u)du}{(\int_0^1 \varphi_f^2(u)du)^{1/2}} (\int_0^1 \varphi_f^2(u)du)^{1/2} \\ &= \rho(\int_0^1 \varphi_f^2(u)du)^{1/2} \\ &= \rho\sqrt{I(f)} \end{aligned} \tag{3.62}$$

where ρ is a correlation coefficient and $sqrtI(f)$ is Fisher Information. Therefore, minimizing τ_{ϕ} is equivalent to maximizing the above identity. By the last equality, this is accomplished by making ρ = 1; i.e., by taking $\phi(u)$ to be $\phi_f(u)$. So $\varphi_f(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}$ we can rewrite $1/\tau_{\phi}$ is the score function which optimizes the rank-based analysis. Since $\hat{\gamma}_{\phi}$ is location and scale equivariant, only the form of f(x) is needed. Furthermore, since in this case $\tau_{\phi} = 1/\sqrt{I(f)}$, the rank based estimator $\hat{\gamma}_{\phi}$ is asymptotically fully efficient, i.e., $\hat{\gamma}_{\phi}$ has the same asymptotic distribution as the maximum likelihood estimator (mle).

3.4.7 Estimates of the scale parameter τ_ϕ

The estimators of τ_ϕ that we discuss are based on the R-residuals formed after estimating γ. In particular, the estimators do not depend on the estimate of intercept parameter α. Suppose then we have fit a model based on a score function ϕ which is bounded, and is standardized so that $^{\mathrm{R}}\phi = 0$ and $^{\mathrm{R}}\phi^2 = 1$. Let $\hat{\gamma}_\phi$ denote the R-estimate of γ and let $\hat{e}_R = Y - X\hat{\gamma}_\phi$ denote the residuals based on the R-fit.

3.5 ADAPTIVE PROCEDURES

Adaptive methods of estimation and testing have several advantages over traditional methods, OLS. Adaptive methods make use of rank-based estimates for linear regression models.

3.5.1 The Adaptive Procedure of Hogg Fisherand Randles(HFR)

There is a pool of score functions ,ϕ, from which the most appropriate score is chosen to be implemented in estimating parameters. Hogg Fisher Randles (1975) propose a two step procedure for choosing an appropriate score function. In summary, the HFR adaptive procedure is:

- Selection statistics Q_1(Skewness) and Q_2(tail weight) are computed.
- These selector statistics Q_1 and Q_2 depending on the selection region they fall, informs the choice and use of the most appropriate rank scores. These selection regions are seen in fig 3.1

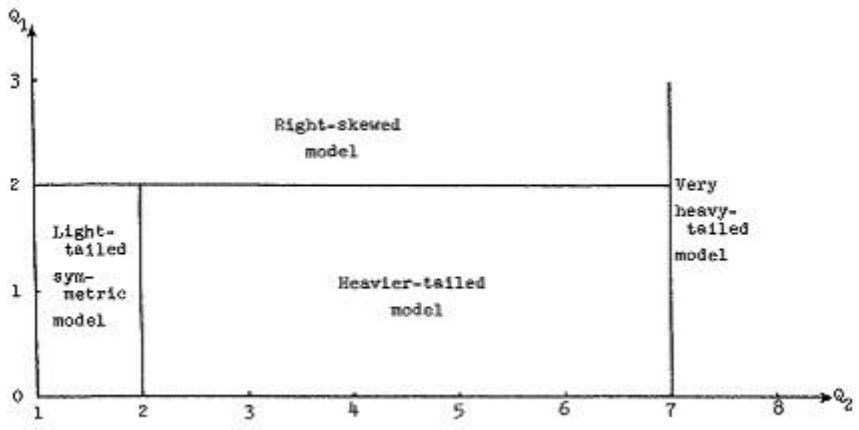

Figure 3.1: Selection criteria for the HFR procedure

The selected scores are then used in estimating and testing.

As an example to how the selection method works, a data set is obtained, the measure of asymmetry is obtained to be $Q_1 = 1.5$, this is an indication that the dataset has a nearly symmetric distribution. The measure of tail weight is obtained next as $Q_2 = 1.5$. These two values are then located on the selection criteria in 3.1 to determine the distribution for the dataset and appropriate scores for that distribution, in this case a light tailed symmetric model. If we had $Q_1 = 1.5$ and $Q_2 = 3.5$ as selection statistics, meaning the dataset had a slightly skewed distribution, then we would have chosen the Wilcoxon scores as the appropriate.

3.5.2 Significance Level of the HFR method

Hogg et al. (1975) demonstrates that their two step procedure maintains its level of significance. With a simulation study, they proved that this method was more powerful than traditional methods.Hogg et al. (1975) designed this adaptive procedure to maintain its significance level. This fact is adequately proven with the Basu's theorem.

3.5.3 Basu's Theorem

Theorem 3.5.1 *Any boundedly complete sufficient statistic is independent of any ancillary statistic. Often used in statistics as a tool to prove independence of two statistics, by first demonstrating one is complete sufficient and the other is ancillary.*

Definition 3.5.3 Let *F* denote the class of distribution functions under consideration. Suppose that each of the r tests based on the statistics $T_1,\ldots,T_r$ is distribution-free over the class *F* : ie $P_{H_o}(T_t \epsilon C_t|F) = \alpha$ for each $F \epsilon F, i = 1,\ldots,r$. C_i is the critical region of T_i Let Q be some statistic that is independent of $T_1,\ldots,T_r$ under H_o for each *FF*. Suppose we use Q to decide which test T_i to conduct. Specifically, let S denote the set of all values of Q with the following decomposition:

$$S = D_1 \cup D_2 \cup \ldots \cup D_r, D_i \cap D_j = \varphi \text{ for } i \neq j$$

So that $Q \epsilon D_i$ corresponds to the decision to use the test T_i.Then a test based on is distribution-free

Proof.. That is,

$$\begin{aligned} P_{H_o}(reject H_o|F) &= \sum P_{H_o}(Q \epsilon D_t, T_t \epsilon C_t|F) \\ &= \sum P_{H_o}(Q \epsilon D_t|F).P_{H_o}(T_t \epsilon C_t|F) \\ &= \sum P_{H_o}(Q \epsilon D_t|F)\alpha \\ &= \alpha \end{aligned} \tag{3.63}$$

3.5.4 Estimating Selector Statistics, Cutoff points and Scores

Hogg (1974) used a pair of selector statistics, Q_1 and Q_2, which are measures of skewness and tail weight respectively. The measure of skewness Q_1 is

$$Q_1 = \frac{\bar{U}_{.05} - \bar{M}_{.5}}{\bar{M}_{.5} - \bar{L}_{.05}} \tag{3.64}$$ Where $\bar{U}_{.05}$, $\bar{M}_{.5}$ and $\bar{L}_{.05}$ are the averages of the largest 5 percent of the ordered data, the middle 50% and the smallest 5% of the ordered data, respectively. The measure of tail weight Q_2 is

$$Q_2 = \frac{\bar{U}_{.05} - \bar{L}_{.05}}{\bar{U}_{.5} - \bar{L}_{.5}} \tag{3.65}$$

It is important to note that these are functions of differences of averages of order statistics of the form

$$\bar{A}_{\alpha 1} - \bar{B}_{\alpha 2}$$

Where α_1 and α_2 are some fraction to be trimmed from the combined ordered data. Let

$$m(\alpha_1, \alpha_2) = \frac{1}{l} \sum_{i=t_1+1}^{n-t_2} Z_{(t)} \tag{3.66}$$

Where $Z_{(t)}$'s are ordered combined sample $t_1 = [n\alpha_1]$, $t_2 = [n\alpha_2]$,$[x]$ denotes the smallest integer greater than x,$l = n - t_1 - t_2$ and redefine a measure of skewness Q_1^* and tail weight Q_2^* by

$$Q_1^* = \frac{m(0.95, 0) - m(0.25, 0.25)}{m(0.25, 0.25) - m(0, 0.95)} \tag{3.67}$$

$$Q_2^* = \frac{m(0.95, 0) - m(0, 0.95)}{m(0.5, 0) - m(0, 0.5)} \tag{3.68}$$

Suppose we want to adapt on residuals, then we need the ordered residuals from an initial fit. The measures of tail weight and skewness of the residuals are obtained by using Q_1^* and Q_2^* respectively. In this research, the benchmarks proposed in the dissertation of Al-Shomrani (2003) are used and the cutoff values for the measures of skewness and tail weight depend on the sample size n. This is a modified version of Hogg's (1975). However it converges to Hogg's (1975) as

$n \longrightarrow \infty$

For Q_1^* we have,

lower cutoff=0.$36 + (\frac{0.68}{n})$

$$uppercutoff = 2.73 - (\frac{3.72}{n}) \tag{3.69}$$

And for Q_2^*, if sample size is less than 25,

lower cutoff=2.$17 + (\frac{3.01}{n})$

$$uppercutoff = 2.63 - (\frac{3.94}{n}) \tag{3.70}$$

however, if sample size is equal or greater than 25 then,

lower cutoff=2.$24 - (\frac{4.68}{n})$

$$uppercutoff = 2.63 - (\frac{9.37}{n}) \tag{3.71}$$

These cutoff points are used to select a rank test which is based on a rank score function corresponding to an unknown distribution. Different scores based on tail

weight and/or skewness have been proposed in literature. Most of these scores are selected depending on tail weight and/or skewness. The rank tests that we consider in this thesis are of the form,

$$T_{\varphi} = \sum_{j=i}^{n} \varphi[\frac{R(Z_j)}{n+1}]R(Z_j) \tag{3.72}$$

where ϕ satisfies the following conditions;

- ϕ nondecreasing function and square-integrable on (0,1)
- ϕ is differentiable on (0,1)

Since ϕ is square integrable,we assume without loss o generality that,

$$\int_0^1 \varphi(u)du = 0 \text{ and} \int_0^1 \varphi^2(u)du = 1$$

Note that a test statistic is synonymous with score function. We use the two interchangeably. We may also write $a_{\varphi}(t) = \varphi[\frac{t}{n+1}]$ and think of $a_{\phi}(1)$,....,$a_{\phi}(n)$ as scores. As discussed by Hettmansperger and Mckean(1998), for model

$$Z = \quad C_i + e_i \tag{3.73}$$

where e_i has density f and distribution F, the optimal score, $\varphi = \frac{f'(F^{-1}(u))}{f(F^{-1}(u))}$ these are optimal in the sense that the corresponding test statistics are asymptotically efficient. For example, Gastwirth (1965), Buning (1996) proposed rank test based on scores corresponding to some selected distributions. They showed that the scores below with the type of distribution in parenthesis have high power see Buning (2005) over their targeted area of distribution.

(Short Tails)

$$a(t) = \begin{cases} t - \left[\frac{n+1}{4}\right], & t \leq \frac{n+1}{4} \\ 0, & \frac{n+1}{4} \leq t \leq \frac{3(n+1)}{4} \\ t - \left[\frac{3(n+1)}{4}\right], & t > \frac{3(n+1)}{4} \end{cases}$$

56 (3.74)

Wilcoxon (Medium Tails)

$$a(t) = t \quad (3.75)$$

Hogg Fisher Randles Test(Right Skewed)

$$a(t) = \begin{cases} t - \left[\frac{n+1}{2}\right], & t \leq \frac{n+1}{2} \\ 0, & t > \frac{n+1}{2} \end{cases} \quad (3.76)$$

Note that these scores are not standardizes. We make use of nine Winsorised scores. These could be classified into four generic scores. Thus

1. $\phi_I(u) =$ [illegible] s_3, $u > s_1$; $s_3 + \frac{s_3 - s_2}{s_1}(u - s_1)$, otherwise

2. $$\varphi_{II}(u) = \begin{cases} -\frac{s_3}{s_1}, & u < s_1 \\ -\frac{s_4}{s_2 - 1}(u - 1) + s_4, & u > s_2 \\ 0, & \text{otherwise} \end{cases}$$

3. $\phi_{III}(u) =$ [illegible] s_2, $u < s_1$; $s_3 + \frac{s_2 - s_3}{s_1 - 1}(u - 1)$, otherwise

4.

$$\phi_{IV}(u) = \begin{cases} s_3, & u < s_1 \\ s_4, & u > s_2 \\ s_3 + \frac{s_4 - s_3}{s_2 - s_1}(u - s_1) & \text{otherwise} \end{cases},$$

Where s_1, s_2, s_3, s_4 are parameters and $a_i(t) = \phi_i(t/(n+1))$. Table (3.1) shows distributions and scores with their corresponding parameters.

Table 3.1: Winsorised Scores

Skewness	Tail Weight	Score Function
Left	Light	$\phi_{LL} = \phi_{II}$,with parameter$(s_1 = .1, s_2 = -1, s_3 = 2.0)$
Left	Medium	$\phi_{LM} = \phi_{III}$,with parameter$(s_1 = .3, s_2 = -1, s_3 = 2.0)$
Left	Heavy	$\phi_{LH} = \phi_{III}$,with parameter$(s_1 = .5, s_2 = -1, s_3 = 2.0)$
Symmetric	Light	$\phi_{SL} = \phi_{II}$,with parameter$(s_1 = .25, s_2 = .75, s_3 = -1, s_4 = 1)$
Symmetric	Medium	Wilcoxon Scores,$\varphi_{SM} = \sqrt{12}[u - \frac{1}{2}]$
Symmetric	Heavy	$\phi_{SH} = \phi_{IV}$,with parameter$(s_1 = .25, s_2 = .75, s_3 = -1, s_4 = 1)$
Right	Light	$\phi_{RL} = \phi_{II}$,with parameter$(s_1 = .9, s_2 = -2, s_3 = 1)$
Right	Medium	$\phi_{RM} = \phi_{II}$,with parameter$(s_1 = .7, s_2 = -2, s_3 = 1)$
Right	Heavy	$\phi_{RH} = \phi_{I}$,with parameter$(s_1 = .5, s_2 = -2 and s_3 = 1)$

Adapting on both samples and residuals can be done. In adapting on residuals, an initial fit is done then the residuals are used for the adaptation.

3.5.5 The Nine winsorised scores

Nine regions which depends on the selector statistics $S = \{Q_1^*, Q_2^*\}$ are defined by,

LH= $\left\{Q_1^* < \hat{Q}_{1l}^*, Q_2^* > \hat{Q}_{2u}^*\right\}$ SH= $\left\{\hat{Q}_{1l}^* < Q_1^* < \hat{Q}_{1u}^*, Q_2^* > \hat{Q}_{2u}^*\right\}$

RH= $\left\{Q_1^* > \hat{Q}_{1u}^*, Q_2^* > \hat{Q}_{2u}^*\right\}$ LM= $\left\{Q_1^* > \hat{Q}_{1l}^*, \hat{Q}_{2l}^* < Q_2^* < \hat{Q}_{2u}^*\right\}$ SM= $\left\{\hat{Q}_{1l}^* < Q_1^* < \hat{Q}_{1u}^*, \hat{Q}_{2l}^* < Q_2^* < \hat{Q}_{2u}^*\right\}$

RM= $\left\{Q_1^* > \hat{Q}_{1u}^*, \hat{Q}_{2l}^* < Q_2^* < \hat{Q}_{2u}^*\right\}$

LL= $\left\{Q_1^* < \hat{Q}_{1l}^*, Q_2^* < \hat{Q}_{2l}^*\right\}$ SL= $\left\{\hat{Q}_{1l}^* < Q_1^* < \hat{Q}_{1u}^*, Q_2^* < \hat{Q}_{2l}^*\right\}$

RL= $\left\{Q_1^* > \hat{Q}_{1u}^*, Q_2^* < \hat{Q}_{2l}^*\right\}$

Where $\hat{Q}_{1l}^*, \hat{Q}_{1u}^*, \hat{Q}_{2l}^*, \hat{Q}_{2u}^*$ are benchmarks from the ordered samples or residuals (Al-Shomrani,2003). Each region identifies a type of score with their corresponding parameters for distributions with their classifications shown in Table 3.1.

Figure (3.2) gives the graphical representation of the nine winsorised scores.

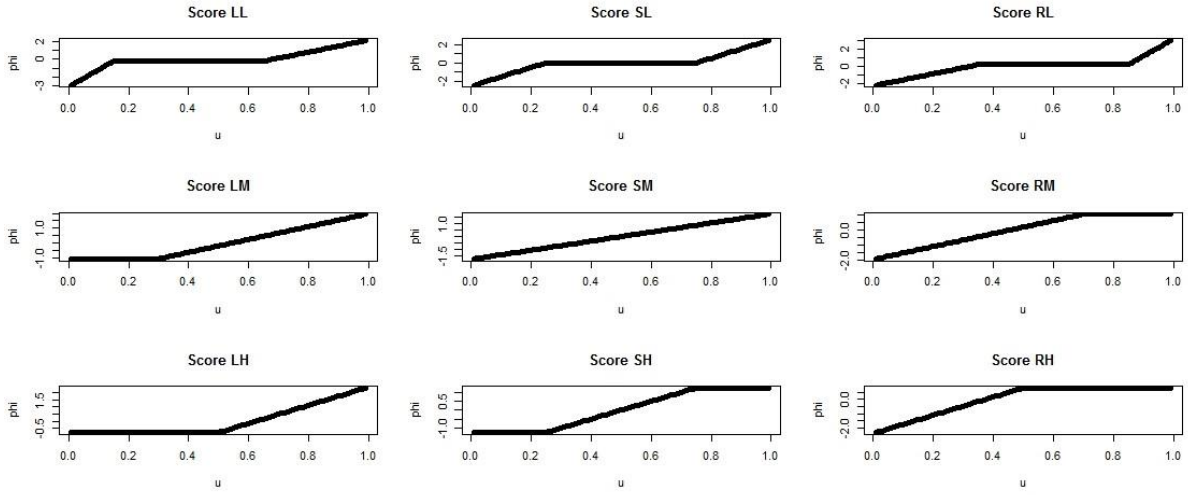

Figure 3.2: Plots of the nine winsorised scores

Figure (3.3) below is an example of the benchmarks of the nine scores based on a sample of size 50.

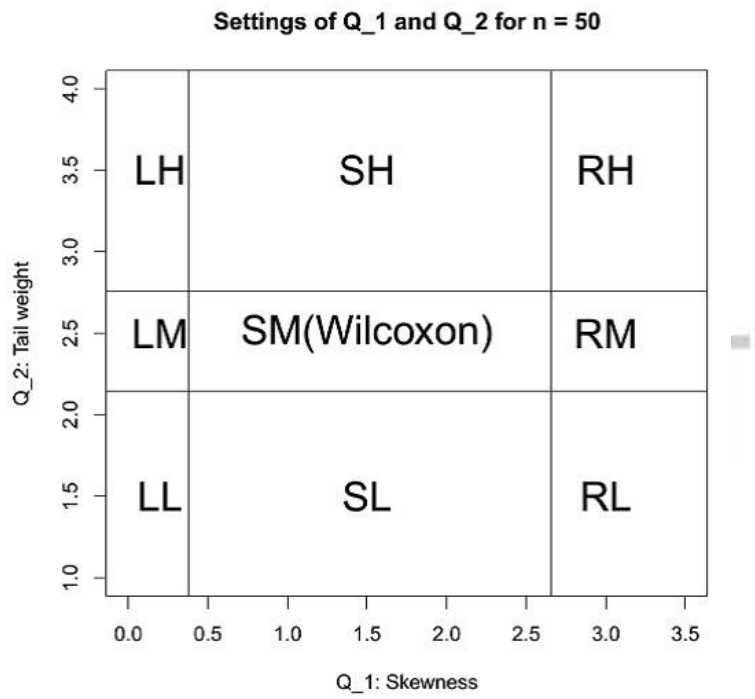

Figure 3.3: Plot of scores with n=50

Let D_k and ϕ_k be a region and score selected respectively, with k = {1,2, ..., 9}. Then the adaptive test, (S,ϕ), is

$$AD(S,\varphi) = T_{\varphi k}, S \epsilon D_k \tag{3.77}$$

where

$$T_{\phi k}(\Delta) = \sum_{i=1}^{n_2} a_{\phi k}(R(y_i - \Delta)) \tag{3.78}$$

is a test statistics based on the ranks and score, ϕ_k associated with region D_k and hence distribution-free. Under H_o, the mean of $T_{\phi k}(\Delta)$ is zero. Thus

$$E_{H_o}[T_{\varphi k}] = \sum_{i=1}^{n_2} E_{H_o}[a_{\varphi k}(R(y_i))]$$

$$= \sum_{i=1}^{n_2}\sum_{j=1}^{n} a_{\varphi k}(j)\frac{1}{n} = 0 \tag{3.79}$$

because the ranks of $y_i's$ are uniform on the integers 1, 2,..., n and $\sum_{j=1}^{n} a_{\varphi k}(j) = 0$.

Since $E_{H_o}[T_{\phi k}] = 0$. From literature, $AD(S,\phi)$ is asymptotically distribution-free. This is because the selector statistic S is based on the order statistics only, the $T_{\phi k}$-statistics is based on the ranks only, and asymptotic critical values are used.

CHAPTER 4

ANALYSIS

4.1 Introduction

In this study, the market model was used in estimating beta using the OLS and Rank based estimation (Wilcoxon and Adaptive) methods. This study was based on observed daily share prices on 37 Ghana Stock Exchange (GSE) listed companies from January 2000 to June 2014, making 173 months.

In the model, historical stock returns was regressed on historical market returns using the GSE Composite Index as proxy for the market index. The sample size was reduced to 10, this is because only these 10 companies had consistent data. Thus in the presence of non-synchronous trading, beta estimation techniques such as Scholes-Williams' beta and Dimsons' beta are employed. According to Dimson (1979), infrequently traded securities have a beta estimate which is biased downwards whiles frequently traded securities are upward biased. The regression was run on discrete monthly returns.

4.2 Fitting the market model

In estimating Beta of stocks and fitting models for predictions we investigate whether there is a need for multiple or simple linear regression and what number of lags and leads is accurate.

No Lags

In the Scenario of no lags, a regression to estimate the market model involves only one explanatory variable, thus the returns on a given asset is regressed on the market returns. Figure 4.1, a boxplot of all ten stocks and the market index, shows little variations in the values of each individual factor in the plot and a rather obviously large proportion of outlying observations in individual stocks and this could be as a result of wrong data (asset price) entries. Since all the stocks contain outliers a random use of three stocks is representative enough of the behavior of all the stocks, thus the fits done in this study is done using stock returns of PZ, Fanmilk (FML) and Ghana Commercial Bank (GCB).

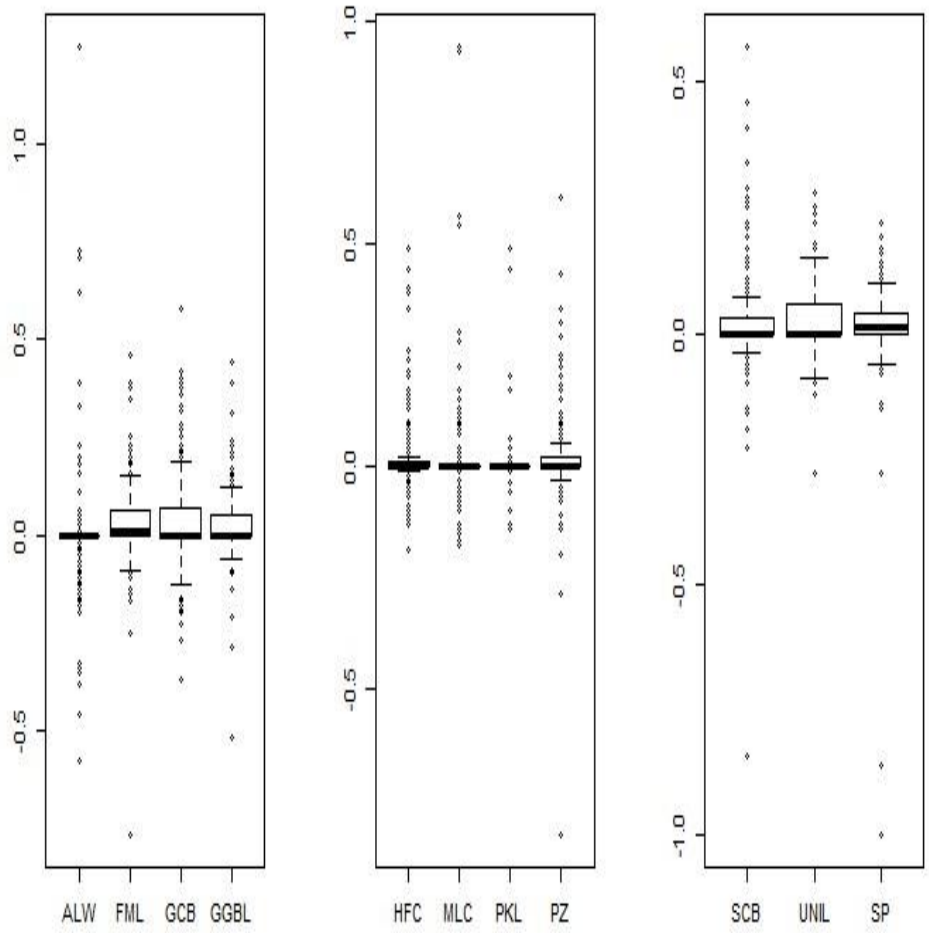

Figure 4.1: Boxplot of 10 stocks on the GSE

The Least-Squares, Wilcoxon R-fit and Adaptive fits are carried out and the respective Beta's, p-value and the Sigma/tau of these fits are recorded for all three stocks. The scores used for each adaptive fit is displayed as well. The results in Table 4.1, shows a positive and significant relation between stock returns and market returns for all three assets and all three fitting methods including the adaptive method, for which a score for symmetric heavy tailed data was selected for all stocks. The rank-based adaptive method gives the minimum tau's for all three stocks giving the best fit and a better model for predicting expected Share returns.

Table 4.1: Estimated parameters with the market model

Stocks	Fits	γ	p-value	σ or τ	Scores
PZ	LS	0.0975	0.1760	0.1118	-
	Wil.	0.0294	0.0126 *	0.0182	wscores
	Adp.	0.0294	0.0005519 ***	0.0130	SH
FML	LS	0.1650	0.02161*	0.1109	-
	Wil.	0.1667	9.931e-11***	0.0377	wscores
	Adp	0.2	9.217e-14***	0.0384	SH
GCB	LS	0.2244	0.00618**	0.1261	-
	Wil.	3.3333e-01	5.032e-14***	0.0632	wscores
	Adp	5.0000e-01	<2e-16***	0.0388	SH

Lag one

Multiple regression is employed in estimating the Beta of a stock with one lag and one lead term. From Table 4.2 the estimates of the coefficient parameters (γ) are significant for FML and GCB but that is not the case in the estimated coefficients for PZ. The adaptive fits uses scores for symmetric heavy-tailed data (SH). For all stocks the adaptive fit had the least standard error.

Table 4.2: Estimated parameters with one lag market model

Stocks	Fit	γ_1	γ_2	γ_3	σ orτ	Score
PZ	LS	0.0005	0.1203	0.0501	0.1128	-
	p-value	0.9959	0.2278	0.4976		

	Wil. p-value	0.008	0.0123	0.0228	0.01816	wscores
		0.6149	0.4419	0.0566.		
	Adp p-value	3.9839e-15	7.4791e-15	3.467e-15	0.0133	SH
		1	1	1		
FML	LS p-value	0.0486	0.1212	0.46070	0.0972	-
		0.5669	0.1584	1.45e-11 ***		
	Wil. p-value	0.0686	0.1310	0.3102	0.0458	wscores
		0.0875 .	0.00139 **	< 2.2e-16 ***		
	Adp p-value	0.06401	0.1256	0.2588	0.0369	SH
		0.0481 *	0.000157 ***	< 2.2e-16 ***		
GCB	LS p-value	0.1755	0.3792	0.0017	0.1233	-
		0.1050	0.0007 ***	0.983159		
	Wil. p-value	0.1411	0.6017	0.0998	0.0594	wscores
		0.0072 **	< 2e-16 ***	0.01112 *		
	Adp p-value	0.1348	0.6325	0.2210	0.0459	SH
		0.00096 ***	< 2.2e-16 ***	7.895e-12 ***		

Lag Two

Stock beta is estimated using two lead and lag terms. These lead and lag terms of the market returns are regressed against the stock returns. The coefficients of these lagged terms are significant for GCB and in the two leads in FML but none whatsoever in PZ. For GCB two different scores are selected after initial fit with LS and Wilcoxon with the LS initial fit giving the best fit with the least standard error.

Table 4.3: Estimated parameters with Lag two market model

Fit	γ_1	γ_2	γ_3	γ_4	γ_5	σ or τ	Score
PZ							
LS	0.132	-0.034	0.115	0.032	0.037	0.113	-
p-value	0.185	0.741	0.261	0.752	0.619		
Wil.	0.024	-0.014	0.030	-0.011	0.034	0.019	wscores
p-value	0.165	0.433	0.084.	0.511	0.008**		
Adp	0.013	-0.002	0.019	-0.009	0.014	0.014	SH
p-value	0.282	0.912	0.126	0.489	0.119		

FML							
LS	0.016	0.055	0.051	0.757	-0.049	0.090	-
p-value	0.838	0.500	0.531	< 2e-16***	0.410		
Wil.	0.042	0.046	0.037	0.630	-0.046	0.052	wscores
p-value	0.355	0.320	0.425	<2e-16***	0.183		
Adp	0.047	0.047	0.026	0.581	-0.080	0.039	SH
p-value	0.173	0.182	0.456	< 2.2e-16***	0.002**		
GCB							
LS	0.018	0.180	0.335	0.159	-0.008	0.122	-
p-value	0.870	0.102	0.0025**	0.145	0.923		
Wil.	-0.025	0.157	0.482	0.350	-0.002	0.055	wscores
p-value	0.598	0.002**	< 2.2e-16***	2.24e-11***	0.967		
Adp	0.047	0.047	0.026	0.581	-0.080	0.039	SH
p-value	0.173	0.182	0.456	< 2.2e-16***	0.002**		
Adp 2	0.039	0.036	-0.027	0.710	-0.015	0.046	RH
p-value	0.345	0.389	0.515	<2e-16***	0.630		

Lag three

In estimating beta with three lags and leads, GCB shows much significant coefficient parameters but not FML, thus for FML estimating up to lag two with an adaptive fit using score SH gives the best estimate of its beta value.

FML	γ_1	0.0030	0.0397	0.0451	-
	p-value	0.9707	0.3424	0.1520	-
	γ_2	0.0260	0.0398	0.0437	-
	p-value	0.7515	0.3522	0.1742	-
	p-value γ_3	0.0566	0.0503	0.0503	-
	γ_4	0.4888	0.2381	0.1166	-
	p-value	0.0448	0.0179	0.0124	-
		0.5869	0.6767	0.6993	-

	γ_5	0.7741	0.6803	0.6053	-
	p-value	< 2e-16***	< 2e-16***	< 2.2e-16***	-
		-0.1178	-0.1304	-0.1335	-
		0.1509	0.00251**	4.604e-05***	-
		0.0134	-0.0072	-0.0094	-
		0.8242	0.8175	0.6883	-
		0.0909	0.04727731	0.0355	-
		-	wscores	SH	-

		-0.0610	-0.0465	-0.0428	-0.0639
		0.5732	0.3400	0.2626	0.1833
		0.03658	-0.0162	-0.0215	-0.0355
		0.7410	0.7443	0.5822	0.4687
		0.1767	0.1578	0.1436	0.1397
		0.1100	0.0017**	0.00029***	0.004647**
		0.3266	0.4324	0.4754	0.4636
	p-value	0.0037**	4.520e-15***	< 2.2e-16***	< 2.2e-16***
	γ_5	0.1649	0.3902	0.4128	0.3895
	p-value	0.1414	8.365e-13***	< 2.2e-16***	4.599e-13***
	γ_6	-0.0586	-0.0450	-0.0402	0.0432
	p-value	0.1414	0.3640	0.3010	0.3753
	γ_7	0.1125	0.0572	0.0534	0.0511
	p-value	0.1414	0.1173	0.0629.	0.1549
GCB	σ or τ	0.1224	0.0550	0.0432	0.0542

	Scores	-	wscores	SH	RH

Table 4.4:

Estimated parameters with Lag three market model

Stocks	Fit	LS	Wil	Adp	Adp1

γ_6
p-value
γ_7
p-value
σ or τ
Scores
γ_1
p-value
γ_2
p-value
γ_3
p-value
γ_4

Lag four

Table 4.5: Estimated parameters with Lag four market model

Stocks	Fit	LS	Wil	Adp
FML	γ_1	-0.1027	-0.1406	-0.1285
	p-value	0.2056	0.0015**	0.00037***
	γ_2	0.0290	0.05175	0.0565
	p-value	0.7258	0.2445	0.1154
	γ_3	0.0303	0.0451	0.0544
	p-value	0.7128	0.3084	0.1288
	γ_4	0.0545	0.0488	0.0430
	p-value	0.5120	0.2755	0.2325
	γ_5	0.0554	0.0182	0.0038
	p-value	0.5067	0.6852	0.9169
	γ_6	0.7803	0.7609	0.7005
	p-value	< 2e-16***	2.2e-16***	< 2.2e-16***
	γ_7	-0.1184	-0.1199	-0.1193
	p-value	0.1562	0.0080**	0.0011**
	γ_8	0.0751	0.0425	0.0349
	p-value	0.3628	0.3379	0.3281
	γ_9	-0.0690	-0.0661	-0.0625
	p-value	0.2545	0.0431*	0.0180*
	σ or τ	0.0911	0.0489	0.0394

From Tables 4.5 and 4.6 below, its observed that the standard errors for the Wilcoxon and Adp 2 are close with the Adp1 having the minimum value.

Table 4.6: continuation of Table 4.5

Stocks	Fit	LS	Wil	Adp	Adp 2
GCB	γ_1	-0.1359	-0.1036	-0.0922	-0.1232
	p-value	0.2098	0.0545.	0.0207*	0.0275*
	γ_2	-0.0506	-0.0394	-0.0328	-0.0657
	p-value	0.6463	0.4713	0.4163	0.2467
	γ_3	0.0304	-0.0035	-0.0179	-0.0127
	p-value	0.7820	0.9493	0.6562	0.8214

γ_4	0.1811	0.1540	0.1399	0.1242
p-value	0.1042	0.0057**	0.00072***	0.0304*
γ_5	0.3263	0.4252	0.4934	0.4638
p-value	0.0038**	1.307e-12***	< 2.2e-16***	1.246e-13***
γ_6	0.1580	0.4048	0.4420	0.4229
p-value	0.1569	1.008e-11***	< 2.2e-16***	7.114e-12***
γ_7	-0.0817	-0.0756	-0.0747	-0.0266
p-value	0.4627	0.1707	0.0675.	0.6406
γ_8	0.1684	0.0846	0.0774	0.0654
p-value	0.1272	0.1231	0.0561.	0.2477
γ_9	0.1480	0.1231	0.0444	0.1549
p-value	0.0682.	0.0024**	0.1338	0.00025***
σ or τ	0.1216	0.06025	0.0445	0.0624

4.3 Asymptotic Relative Efficiency (ARE)

The asymptotic relative efficiency between two tests or estimates based on the score functions $\phi_1(u)$ and $\phi_2(u)$ or one score function relative to other score function is defined by;

$$e(\varphi_1, \varphi_2) = \frac{C_{\varphi_1}^2}{C_{\varphi_2}^2} = \frac{\tau_{\varphi_2}^2}{\tau_{\varphi_1}^2} \tag{4.1}$$

where C_{ϕ_1} and C_{ϕ_2} are respectively the efficacies of the two estimates and τ_{ϕ_i}, with i=1, 2, are the scale parameters of the two score functions. The efficiency of the LS, Wilcoxon and adaptive fits are compared for all five datasets up to lag four.

Table 4.7: Asymptotic Relative Efficiency for no lag

ARE	PZ	FML	GCB
ARE(Wil.,LS)	37.8743	8.6741	3.978859
ARE(Adp,LS)	74.0575	8.338419	10.55278
ARE(Adp,Wil.)	1.95535	0.9613057	2.652213

Table 4.8: Asymptotic Relative Efficiency for lag one and two

Lags	ARE	PZ	FML	GCB
LAG 1	ARE(Wil.,LS)	38.5940	4.5075	4.3027
	ARE(Adp,LS)	71.7643	6.9421	7.2118

	ARE(Adp,Wil.)	1.8595	1.5401	1.6761
LAG 2	ARE(Wil.,LS)	34.6175	3.0705	4.96367
	ARE(Adp,LS)	67.1154	5.3505	9.7518
	ARE(Adp,Wil.)	1.9388	1.7425	1.9646
	ARE(Adp2,LS)	-	-	6.8821
	ARE(Adp2,Wil.)	-	-	1.3865
	ARE(Adp2,Adp.)	-	-	0.7057

Table 4.9: Asymptotic Relative Efficiency for lag three and four

Lag	ARE	FML	GCB
LAG 3	ARE(Wil.,LS)	3.0705	4.9527
	ARE(Adp,LS)	5.3505	8.0413
	ARE(Adp.,Wil)	1.7425	5.1052
	ARE(Adp2,LS)	-	1.6236
	ARE(Adp2,Wil.)	-	1.0308
	ARE(Adp2,Adp.)	-	0.6349
LAG 4	ARE(Wil.,LS)	3.46544	4.0740
	ARE(Adp,LS)	5.3354	7.4659
	ARE(Adp.,Wil)	1.5396	3.7968
	ARE(Adp2,LS)	-	1.8326
	ARE(Adp2,Wil.)	-	0.9320
	ARE(Adp2,Adp.)	-	0.5086

4.4 Adjusted Beta

The study on the impact of non-synchronous trading for the measurement of beta was originally done by Scholes-Williams (1977). Their estimator of beta is

$$\hat{\beta}_{SW} = (\gamma_j + \gamma_{j+n} + \gamma_{j-n})/(1 + 2\rho) \qquad (4.2)$$

where $\pm n$ represents the number of lag and lead terms and ρ is the correlation of the market. The beta estimator derived by Dimson is given by

$$\hat{\beta}_{DIM} = \gamma_j + \sum_{n=1}^{N} \gamma_{j+n} + \sum_{n=1}^{N} \gamma_{j-n} \qquad (4.3)$$

These methods estimate $\hat{\beta}$ which adjusts for non-synchronous trading, such as we have with stocks on the Ghana Stock Exchange (GSE). Dimson suggests that the coefficients γ_{j-n} and γ_{j+n} all be simultaneously estimated using multiple regression as opposed to independently estimated as suggested by ScholesWilliams.

Table 4.10: Dimson and Scholes-Williams Beta estimates

Stock	Method	Fits	β	β	β	β
			Lag 1	Lag 2	Lag 3	Lag 4
PZ	Dimson	LS	0.1709	0.2824	-	-
		Wil	0.0431	0.0629	-	-
		Adp	0.0000046	0.0366	-	-
	Scholes-Williams	LS	0.2017	0.3332	-	-
		Wil	0.0509	0.07419	-	-
		Adp	5.429e-6	0.0432	-	-
FML	Dimson	LS	0.6306	0.8299	0.8001	0.7346
		Wil	0.5098	0.7096	0.6905	0.6406
		Adp	0.4484	0.6212	0.6139	0.5828
	Scholes-Williams	LS	0.7442	0.9795	0.9442	0.8669
		Wil	0.6016	0.8375	0.8149	0.7560
		Adp	0.5292	0.7332	0.7245	0.6879
GCB	Dimson	LS	0.5564	0.6837	0.6976	0.7441
		Wil	0.8425	0.9615	0.9299	0.9697
		Adp	0.9883	0.6212	0.9807	0.9794
		Adp2	-	0.7425	0.9878	1.0030
	Scholes-Williams	LS	0.6567	0.8069	0.8233	0.8781
		Wil	0.9943	1.1347	1.0975	1.1444
		Adp	1.1663	0.7332	1.1575	1.1559
		Adp2	-	0.8763	1.1657	1.1837

4.5 Some distributions and scores the Adaptive method selects

This section demonstrates the effectiveness of the rank-based adaptive method in selecting the most appropriate scores for some specific distribution; the normal, lognormal, exponential, uniform, cauchy and gamma distributions. Random numbers from these distributions have their selector statistics and cutoffs as well as scores

estimated. For some distributions more than one score is selected and the score with the largest number of occurrence in a trial of 500 fits is selected as the appropriate score for that distribution. The scores selected as well as a description of the form of the error distributions, after an initial fit is done by the LS and wilcoxon method, are shown in Table 4.11.

Table 4.11: Scores for various distributions

Distribution	Scores selected	appropriate score	Descriptive
Normal	SM (500)	SM	Symmetric, medium-tails
lognormal	RH (500)	RH	Right skewed, heavy-tails
Exponential	RH(100) RM(400)	RM	Right skewed, medium-tails
Uniform	SL (500)	SL	Symmetric, light-tails
Cauchy	LH(74) RH(90) SH(336)	SH	Symmetric, heavy-tails
Gamma	RH(111) RM(389)	RM	Right skewed, medium-tails

In a trial adaptive fit, we assume initial fits have been carried out and the errors are from these distributions thus their scores are specified and used in the adaptive fits. The relative efficiency amongs the least-squares, wilcoxon and adaptive fits are compared.

Table 4.12: ARE for various distributions

Fits	Normal	Lognormal	Exponential	Cauchy	Uniform	Gamma
ARE(LS,Wil)	1.1517	0.1774	0.712	0.4687	1.0347	0.3468
ARE(LS.Adp)	1.1517	0.0942	0.5044	0.39036	0.7066	0.2724
ARE(Wil,Adp)	1	0.5309	0.7085	0.8328	0.6829	0.7852

The LS gains over 115% efficiency over the adaptive and wilcoxon estimators when the errors are normally distributed and enjoys a loss in efficiency when used for other data types other than the normal distribution, ranging from heavy-tailed to light-tailed, symmetric to right skewed. The adaptive method, when errors are normally distributed, is as efficient as the wilcoxon . In general, the adaptive estimator has a substantial gain in efficiency over the LS and wilcoxon estimator for non-symmetric,

heavy-tailed error distributions with the wilcoxon method having greater efficiency than the LS method.

4.6 The Contaminated Normal Distribution

In this section, we consider the contaminated normally distributed random error terms generated at each point where the dataset is contaminated in the response variable with 5%, 10%, 15% and 20% contamination and its Beta values and selected scores recorded.

Table 4.13: Contaminated Normal distribution

Fits	Normal	10% cont.	5% cont.	10% cont.	15% cont.	20% cont.
LS	0.033	-0.318	-0.015	0.408	0.888	0.047
Wil	0.067	-0.207	0.089	0.202	0.146	0.203
Adp	0.067	-0.214	0.103	0.124	0.036	0.230
Scores	SM	LH	SH	RH	RH	SH

Its noted that 5% contamination causes the data to have heavy tails and causes the LS fit to lose almost 30% efficiency to the Wilcoxon and about 40% to the adaptive method and over 93% efficiency to the wilcoxon rank-based method when there is 15% contamination.

Table 4.14: Asymptotic Relative Efficiency

Fits	Normal	10% cont.	5% cont.	10% cont.	15% cont.	20% cont.
ARE(LS,Wil)	1.078	0.418	0.708	0.231	0.067	0.304
ARE(LS,Adp)	1.078	0.401	0.606	0.268	0.046	0.259
ARE(Wil,Adp)	1	0.959	0.857	1.160	0.697	0.853

4.6.1 Example on Contamination using Baseball Salaries Data

This example considers salaries of 176 professional baseball players for the 1987 season, a data in the Rfit package. The dataset initially has no outliers and some outliers are introduced to reflect the response of the least squares, wilcoxon and adaptive fits.

Table 4.15: Contamination of Baseball Data

Fits	No Contamination	5% cont.	10% cont.	15% cont.	20% cont.
LS	0.8527	0.6626	0.0622	-0.1828	-0.0731
Wil	0.9171	0.8853	0.8431	0.8735	0.9032
Adp	0.9389	0.8734	0.8647	0.8968	0.9106
Scores	SH	RH	RH	RH	RH

The LS fit shows much variation in its estimated values but the wilcoxon and adaptive fits exhibit much robustness with the introduction of the increasing percentages of outliers, 5 %, 10%, 15% and 20% outliers.

Table 4.16: Asymptotic Relative Efficiency

Fits	No Contamination	5% cont.	10% cont.	15% cont.	20% cont.
ARE(LS,Wil)	0.8065	0.073	0.0233	0.0197	0.0204
ARE(LS,Adp)	0.9521	0.0845	0.0216	0.0118	0.00635
ARE(Wil,Adp)	1.1806	1.1577	0.9292	0.5974	0.3107

As expected the ARE of the LS method show some loss in efficiency when the LS method is used to estimate the parameters instead of the other methods. The LS sees up to 98.03% loss in efficiency when there is 15% contamination with outliers. When there is 5% contamination and no contamination at all the wilcoxon gains 115.77% and 118% efficiency respectively over the Adaptive method whiles the Adaptive method shows a greater efficiency in the other percentages of contamination.

4.7 Robustness

In this section, diagnostics based on both highly efficient and high breakdown robust fits are explored to confirm the adequacy of the model and check the quality of fit. These diagnostics are primarily concerned with the determination of robust estimators.

For motivation, we consider a simple dataset with two predictors and n = 100 data points. The values of the x's are drawn from the uniform distribution. The first set of responses is drawn from the model

$$y_i = 5x_{1i} + 5x_{2i} + e_i \quad (4.4)$$

where $e_1,\ldots,e_n$ were drawn independently from a N(0,1) distribution. In the second set of response the last 25 data points representing 25% of the data is replaced with 20 to 44 standard deviation's of the original dataset which we call "1st" dataset. This second set is labeled the "2nd" dataset. The LS, wilcoxon and adaptive fits of the two datasets are obtained, summarizing them in Table 4.17.

Table 4.17: Estimates of fit for normal and contaminated data

Dataset	Fits	Intercept	x_1	x_2	τ or σ
1st	LS	-0.0590	4.8631	5.3505	0.9552
	Wil	-0.1393	4.9308	5.2694	1.0621
	Adp	-0.1393	4.9308	5.2694	1.06208
2nd	LS	24.175	11.348	-11.932	33.04
	Wil	0.5678	4.7795	4.9694	7.4167
	Adp (RM)	0.4645	4.6993	5.1706	4.9539
	Adp2 (RH)	0.5025	4.6354	5.1588	4.8772

On the "1st" dataset all three fits agree estimating quite accurately the coefficient and intercept parameters close to their true values in the original model,with the LS having the least error. On the "2nd" dataset the LS is impaired greatly while the Wilcoxon and adaptive fits exhibited robustness. Two adaptive fits, adp and adp 2, are seen here because two different scores were selected for residuals from LS and wilcoxon initial fits respectively. The adaptive fit after wilcoxon initial fits has the least tau thus giving the best fit.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

A parametric estimation method, like the Least-Squares method, is a method based on the assumptions that the random errors in the data have a particular type of distribution,in this case a normal distribution. The rank-based fitting, a non-parametric method, is an alternative to the parametric estimation when the errors have a distribution that is not necessarily normal. Rank-based adaptive methods which specifies scores to be used in fitting are carried out in this study. In this research performance of three estimating methods are investigated and the fitting of lagged beta's are carried out for some stocks on the Ghana Stock Exchange (GSE). Some methods of reporting adjusted beta values are showcased. The performance and robustness of the three methods are investigated in the presence of varying percentages of outliers. An investigation of the accuracy of selector statistics, in selecting best scores for different types of distribution for the adaptive fit, are carried out.

5.2 Conclusion

In this study, first its observed that the rank-based methods (Wilcoxon and Adaptive) are more robust(for power) in estimation when the distribution of the error term of the dataset is non-normal and also in the presence of outlying observations,whiles the LS method is less robust. Contamination with as little as 5 outlying observations is enough to cause some disturbance in the LS estimates. The standard error of the LS method, as expected, were the smallest with the rank-based fit close in estimation and losing just a little efficiency for errors with normal distribution.

Secondly, the school of thought that debunks the need for lagged market returns in estimating stock returns is challenged by the study results which shows significant lagged coefficients for FML and GCB up to lag four.

Thirdly the adaptive fit done after initial fit with the LS often shows relatively higher efficiency than an initial fit with the rank-based Wilcoxon method. In a large

proportion of time the adaptive method proved more efficient than the Wilcoxon fit with the Adaptive method (after LS initial fit) losing efficiency to the Wilcoxon fit on very few occasions.

The adaptive procedure effectively selects appropriate scores for distributions.

This study confirms that the least-Squares method as according to Mohebbi et al. (2007) is very important and crucial, however it is only optimal under certain distributional assumptions including the error being normally distributed. The adaptive fit is more robust in the presence of outliers and consistently gives the best fit with a minimum standard error .

5.3 Recommendations

When the researcher has no idea of the error distributions of the dataset (as is the case in practice), the adaptive fit gives an idea of what the error distribution looks like and an appropriate score selected in fitting the model. Indeed technology has made model fitting easy so why not go the extra mile to ensure an efficient and more accurate fit in the absence of knowledge of the error distribution of the data. The study also shows some stocks on the GSE are indeed thinly traded and require lagged beta estimates. Though adaptive fits with LS initial fit exhibits more efficiency than a Wilcoxon initial fit,both initial fits need to be carried out to give the optimum score for the adaptive fit.

5.4 Further Studies

In this study adaptive methods of estimation were carried out after initial fits had been carried out with both LS and Wilcoxon. Scores selected after initial fits with LS were more efficient.

Further studies needs to be done to ascertain which of these initial fits should be carried out for more precision as to what exactly the procedure in an adaptive fit is. Complete diagnostics like the ANOVA for the rank-based adaptive method needs to be studied.

REFERENCES

Abdi, H. and Salkind, N. (2007). The method of least squares. *Encyclopedia of Measurement and Statistics.Sage*.

Abebe, A., McKean, J. W., Kloke, J. D., and Bilgic, Y. K. (2013). Iterated reweighted rank-based estimates for gee models. *Under revision*, pages 73, 74.

Al-Shomrani, A. A. (2003). *A Comparison of different schemes for selecting and estimating score functions based on residuals*. PhD thesis, Western Michigan University.

Alma, O. G. (2011). Comparison of robust regression methods in linear regression. *Int. J. Contemp. Math. Sciences*, vol. 6,no.9:409–421.

Baklizi, A. (2005). A continuously adaptive rank test for shift in location. *Australian and New Zealand Journal of Statistics*, 47:203–209.

Bilgic, Y. K. (2012.). *Rank-based estimation and prediction for mixed effects models in nested designs*. PhD thesis, Western Michigan University, Department of Statistics.

Bilgic, Y. K., McKean, J. W., Kloke, J. D., and Abebe, A. (2013). Iteratively reweighted generalized rank-based method in mixed models. *Manuscript*, pages p73,74.

Bilgic, Y. K. and Susmann, H. (December, 1999). Rfit: An r package for rankbased estimation and prediction in random effects nested models. *The R Journal*, 5/2:19.

Birks, D. and Dodge, Y. (1993). Alternative methods of regression. *Wiley, New York.*

Block, S. B. ((July/August)1999). A study of financial analysts:practice and theory. *Financial Analysts Journal*, vol. 55, no. 4:86–95.

Bruner, R. F., Eades, K. M., Harris, R. S., and Higgins, R. C. (1998). Best practices in estimating the cost of capital:survey and synthesis. *Financial Practice and Education*, vol. 27(Spring/Summer):13–28.

Buning, H. (1996). Adaptive tests for the c-sample location problem-the case of two-sided alternatives. *Communications in Statistics-Theory and Methods*, 25:1569–1582.

Buning, H. (1999). Adaptive johckheere-type tests for ordered alternatives. *Journal of Applied Statistics*, 26:541–551.

Buning, H. (2002). An adaptive distribution-free test for the general two-sample problem. *Computational Statistics*, 17:297–313.

Buning, H. and Kossler, W. (1998). Adaptive tests for umbrella alternatives. *Biometrical Journal*, 40:573–587.

Buning, H. and Thadewald, T. (2000). An adaptive two-sample location-scale test of lepage type for symmetric distributions. *Journal of Statistical Computation and Simulation*, 65:287–310.

Burgess, E. W. (1928). Factors determining success or failure on parole. in a. a. bruce (ed.), the workings of the indeterminate sentence law and parole in illinois. *Springfield, Illinois: Illinois State Parole Board. Google books*, pages 205–249.

Connor, G. and Herbert, N. (1999). Estimation of the european equity model. *Horizon:The Barra Newsletter*, no. 169(Winter).

Draper, N. and Smith, H. (1988). Applied regression analysis. *3rd Edn. Wiley, New York*.

Fox, J. (2005). Introduction to nonparametric regression. *Economic & Social Research Council (ESRC)*.

Freidlin, B., Miao, W., and Gastwirth, J. L. (2003). On the use of the shapirowilk test in twostage adaptive inference for paired data from moderate to very heavy tailed distributions. *Biometrical Journal*, 45:442–448.

Gastwirth, J. L. (1965). Percentile modifications of two sample rank tests. *Journal of the American Statistical Association*, 60:1127–1141.

Gitman, L. J. and Mercurio, V. A. (1982). Cost of capital techniques ysed by major u.s firms:survey and analysis of fortune's 1000. *Financial Management*, vol. 14,no.4 (Winter)::21–29.

Graham, J. R. and Harvey, C. R. (2001). The theory and practice of corporate finance:evidence from the field. *Journal of Financial Economics*, vol.60,nos. 2-3(May):187–243.

Hajek, J. (1962). Asymptotically most powerful rank-order tests. *The Annals of Mathematical Statistics*, 33:1124–1147.

Hajek, J. and Sidak, Z. S. (1967). Theory of rank tests. *Academic Press, New York, London.*

Hall, P. and Padmanabhan, A. (1997). Adaptive inference for the two-sample scale problem. *Technometrics*, 39:412–422.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). Robust statistics. *Wiley,New York*.

Hao, L. and Houser, D. (2012). Adaptive procedures for wilcoxon-mann-whitney test: Seven decades of advances.

Hettmansperger, T. P. and McKean, J. W. (2008). *Robust Nonparametric Statistical Methods*. Chapman Hall, New York.

Hettmansperger, T. P. and McKean, J. W. (2011). Robust nonparametric statistical methods. *Chapman Hall, New York*, 2nd edition:p57–64.

Hill, N. J. and Padmanabhan, A. R. (1991). Some adaptive robust estimators which work with real data. *Biometrical Journal*, 33:81–101.

Hogg, R. V. (1967). Some observations on robust estimation. *Journal of the American Statistical Association*, 62:1179–1186.

Hogg, R. V., Fisher, D. M., and Randles, R. H. (1975). A two-sample adaptive distribution-free test. *Journal of the American Statistical Association*, 70:656– 661.

Hogg, R. V. and Lenth, R. V. (1984). A review of some adaptive statistical techniques. *Communications in Statistics-Theory and Methods*, 13:1551–1579.

Hotelling, H. and Pabst, M. (1936). Rank correlation and tests of significance involving no assumption of normality. *Annals of Mathematical Statistics*, 7:29– 43.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *Annals of Statistics*, 1:799–821.

Jones, D. (1979). An efficient adaptive distribution-free test for location. *Journal of the American Statistical Association*, 74(368):822–828.

Kloke, J. and Mckean, J. W. (December 2012). Rfit: Rank-based estimation for linear models. *The R Journal*, Vol. 4/2.

Kloke, J. D., McKean, J. W., and Rashid, M. (2009). Rank-based estimation and associated inferences for linear models with cluster correlated errors. *Journal of the American Statistical Association*, 104:384–390.

Kossler, W. (2010). Max-type rank tests, u- tests and adaptive tests for the twosample location problem- an asymptotic power study. *Journal of Computational Statistics and Data Analysis*, 54(9):2053–2065.

Kossler, W. and Kumar, N. (2008). An adaptive test for the two-sample location problem based on u-statistics.

Lemmer, H. H. (1993). Adaptive tests for the median. *IEEE Transactions on*

Reliability, 42:442–448.

Mann, H. B. and Whitney, D. R. (1947). On a test whether one of two samples is stochstically larger than the other. *Annals of Mathematical Statistics*, 18:50–60.

Martin, R. D. and Simin, T. (September,2003.). Outlier resistant estimates of beta. *Financial Analysts Journal*, 2:202–203.

Meng, J. G., Gang, H., and Bai, J. (2007). A simple method for estimating betas when factors are measured with error.

Miao, W. and Gastwirth, J. (2009). A new two stage adaptive nonparametric test for paired differences. . *Statistics and its Interface*, 2:213–221.

Mohebbi, M., Nourijelyani, K., and Zeraati, H. (2007). A simulation study on robust aalternatives of least squares regression. *Journal of Applied Sciences*, 7(22):3469–3476.

O'Gorman, T. W. (1996). An adaptive two-sample test based on modified wilcoxon scores. *Communications in Statistics-Simulation and Computation*, 25:459–479.

O'Gorman, T. W. (1997). An adaptive test for the one-way layout. *The Canadian Journal of Statistics*, 25:269–279.

O'Gorman, T. W. (2001). An adaptive permutation test procedure for several common tests of significance. *Computational Statistics and Data Analysis*, 35:335–350.

O'Gorman, T. W. (2002). An adaptive test of significance for a subset of regression coefficients. *Statistics in Medicine*, 21:3527–3542.

O'Gorman, T. W. (2004). Applied adaptive statistical methods. *Society for Industrial and Applied Mathematics,Philadelphia*.

O'Gorman, T. W. (2006). An adaptive test for a subset of coefficients in a multivariate regression model. *Biometrical Journal*, 48:849–859.

O'Gorman, T. W. (2008). Using adaptive tests for the analysis of repeated measurements. , *Journal of Biopharmaceutical Statistics*, 18:595–610.

O'Gorman, T. W. (2012). *Adaptive Tests of Significance Using Permutations of Residuals with Rand SAS*. John Wiley and Sons, Inc.

Okyere, G. A. (2011). *Robust adaptive schemes for linear mixed models*. PhD thesis, Western Michigan University, Department of Statistics.

Randles, R. H. and Hogg, R. V. (1973). Adaptive distribution-free tests. *Communications in Statistics*, 2:337–356.

Rousseeuw, P. J. and Leroy, A. M. (2003). Robust regression and outlier detection. wiley.

Ruberg, S. J. (1986). A continuously adaptive nonparametric two-sample test. *Communications in Statistics-Theory and Methods*, 15:2899–2920.

Ryan, T. P. (2008). Modern regression methods. wiley.

Siegel, S. (1956). Nonparametric statistics, new york: Mcgraw-hill.

Siegel, S. and Castellan, N. J. J. (1988). Nonparametric statistics for the behavioral sciences. 2nd Edition, McGraw-Hill International Editions.

Wainer, H. and Thissen, D. (1976). Three steps toward robust regression. *Psychometrika*, vol. 41(1):9–34.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, 1:80–83.

Xie, J. and Priebe, C. E. (2000). Generalizing the mann-whitney-wilcoxon statistic. nonparametric statistics. 12:661–682.

Xie, J. and Priebe, C. E. (2002). A weighted ggeneralization of the mann-whitneywilcoxon statistic. *Journal of Statistical Planning and Inference*, 102:441–466.

Yuh, L. and Hogg, R. V. (1988). On adaptive m-regression. *Biometrics*, 44:433– 445.

APPENDIX

Appendix I - R-codes for computing cutoffs, selector statistics and selecting scores

```
upmean = function(alp1,alp2,xs){ n =
    length(xs) ind1 = floor(alp1*n + 0.5)
    ind2 = floor(alp2*n + 0.5) if(ind2 >
    n){ind2 = n} if(ind1 > ind2){ind1 =
    ind2}

    s = sum(xs[ind1:ind2]) upmean =
    s/(ind2 - ind1 + 1)
    upmean
  }

  lomean = function(alp1,alp2,xs){ n =
    length(xs) ind1 = floor(alp1*n + .5)
    ind2 = floor(alp2*n + .5) if(ind1 <
    1){ind1 = 1} if(ind1 > ind2){ind2 =
    ind1} s = sum(xs[ind1:ind2])

    lomean = s/(ind2 - ind1 + 1)
    lomean
}

  mimean = function(alp,xs){ n = length(xs) ind1 =
    floor(((1.0-alp)/2.0)*n + 0.5) if(ind1 < 1){ind1 =
    1}
```

```
    ind2 = n - ind1

    s = sum(xs[(ind1+1):ind2]) mimean =
    s/(ind2 - ind1) mimean

}

selstat = function(x,par=c(0.95,0.5,0.05,0.5,0.5)){

    n = length(x) #
    resids are x

    clq1 = 0.36 + (0.68/n) cuq1 =
    2.73 - (3.72/n) if(n < 25){ clq2
    = 2.17 - (3.01/n) cuq2 = 2.63 -
    (3.94/n)
    } else { clq2 = 2.24 - (4.68/n)
        cuq2 = 2.95 - (9.37/n)
    } cus = c(clq1,cuq1,clq2,cuq2) iord
    = order(x) xs = x[iord]

    um1 = upmean(par[1],1.0,xs) lm1
    = lomean(0.0,par[3],xs) mm1 =
    mimean(par[2],xs) um2 =
    upmean(par[4],1.0,xs) lm2 =
    lomean(0.0,par[5],xs)

    ulmeans = c(um1,lm1,mm1,um2,lm2)
```

```
q1 = (um1 - mm1)/(mm1 - lm1) q2 =
(um1 - lm1)/(um2 - lm2)


qs = c(q1,q2)


if(q1 <= clq1){ if(q2 <= clq2){score = 'LL'} if((q2 > clq2) &&
   (q2 <= cuq2)){score = 'LM'} if(q2 > cuq2){score = 'LH'}
} else if(q1 <= clq2){ if(q2 <= clq2){score = 'SL'} if((q2 >
   clq2) && (q2 <= cuq2)){score ='SM'} if(q2 > cuq2){score
   = 'SH'}
} else { if(q2 <= clq2){score = 'RL'} if((q2 > clq2) && (q2 <=
   cuq2)){score = 'RM'} if(q2 > cuq2){score = 'RH'}
}          if          (score=="SM"){
selscores='bentscoressm'  }  else
if(score=="SH"){
selscores='bentscoressh'  }  else
if(score=="SL"){
selscores='bentscoressl'  }  else
if(score=="RL"){
selscores='bentscoresrl'  }  else
if(score=="RM"){
selscores='bentscoresrm'  }  else
if(score=="RH"){
selscores='bentscoresrh'  }  else
if(score=="LL"){
selscores='bentscoresll'  }  else
if(score=="LM"){
selscores='bentscoreslm'  }  else
```

```
    if(score=="LH"){
    selscores='bentscoreslh'
    }
                    list(score=score,qs = qs,cus=cus,ulmeans=ulmeans,selscores=selscores)
}
```


Appendix II - R-codes for defining Scores

```
bent.phi<-function(u,...) ifelse(u<.1,-20*u-.1,0) bent.Dphi<-
function(u,...) ifelse(u<0.1,-20,0) bentscoresll<-
new("scores",phi=bent.phi,Dphi=bent.Dphi)

bent.phi<-function(u,...) ifelse(u<.3,-1,2+(3/0.7)*(u-1)) bent.Dphi<-
 function(u,...) ifelse(u<0.3,0,3/0.7) bentscoreslm<-
 new("scores",phi=bent.phi,Dphi=bent.Dphi)


 bent.phi<-function(u,...) ifelse(u<.5,-1,2+6*(u-1))
 bent.Dphi<-function(u,...) ifelse(u<0.5,0,6)

 bentscoreslh<-new("scores",phi=bent.phi,Dphi=bent.Dphi)


 bent.phi<-function(u,...) ifelse(u>.7,1,1+(3/0.7)*(u-.7))
 bent.Dphi<-function(u,...) ifelse(u>0.7,0,3/0.7)
 bentscoresrm<-
 new("scores",phi=bent.phi,Dphi=bent.Dphi)
```

```
bent.phi<-function(u,...) ifelse(u>.5,1,1+6*(u-.5)) bent.Dphi<-
function(u,...) ifelse(u>0.5,0,6) bentscoresrh<-
new("scores",phi=bent.phi,Dphi=bent.Dphi)


bent.phi<-function(u,param){ s1=param[1] s2=param[2] s3=param[3]
   s4=param[4] ifelse(u<s1,-s3/s1*(u-s1),ifelse(u>s2,-s4/(s2-1)*(u-1)+s4,0))
} bent.Dphi<-function(u,param){ s1=param[1] s2=param[2]
s3=param[3] s4=param[4] ifelse(u<s1,-s3/s1,ifelse(u>s2,-
s4/(s2-1),0))
} bent.param<-c(0.25,0.75,-1,1) bentscoressl<-
new("scores",phi=bent.phi,Dphi=bent.Dphi,param=bent.param)

bent.phi<-function(u) sqrt(12)*(u-0.5)
   bent.Dphi<-function(u)
   rep(sqrt(12),length(u)) bentscoressm<-
   new("scores",phi=bent.phi,Dphi=bent.Dphi)


bent.phi<-function(u,param){ s1=param[1] s2=param[2] s3=param[3]
  s4=param[4] ifelse(u<s1,s3,ifelse(u>s2,s4,s3+((s4-s3)/(s2-s1))*(u-s1)))
} bent.Dphi<-function(u,param){ s1=param[1] s2=param[2]
s3=param[3] s4=param[4] ifelse(u<s1,0,ifelse(u>s2,0,(s4-
s3)/(s2-s1)))
} bent.param<-c(0.25,0.75,-1,1) bentscoressh<-
new("scores",phi=bent.phi,Dphi=bent.Dphi,param=bent.param)
```