UNIVERSITY FOR DEVELOPMENT STUDIES

FACULTY OF MATHEMATICAL SCIENCES

ROBUST ADAPTIVE SCHEME FOR GAUSS MARKOV MODEL

GILBERT BINEY

2021

UNIVERSITY FOR DEVELOPMENT STUDIES

ROBUST ADAPTIVE SCHEME FOR GAUSS MARKOV MODEL

BY

GILBERT BINEY

(UDS/DAS/0024/13)

THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS, FACULTY OF MATHEMATICAL SCIENCES, UNIVERSITY FOR DEVELOPMENT STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY DEGREE IN APPLIED STATISTICS

JUNE, 2021

DECLARATION

Student

I hereby declare that this thesis is the result of my own original work and that no part of it has been presented for another degree in this University or elsewhere:

Candidate's Signature: ..... Date: .....

Gilbert Biney

Supervisors

We hereby declare that the preparation and presentation of this thesis was supervised in accordance with the guidelines on supervision of thesis laid down by the University for Development Studies.

Principal Supervisor's Signature: ..... Date: .....

Prof. Gabriel Asare Okyere

Co-Supervisor's Signature: ..... Date: .....

Dr. Abukari Alhassan

ABSTRACT

The Hogg's adaptive scheme is extended to the Gauss Markov Model. The Gauss Markov model is a statistical procedure which belongs to the class of general linear model. Gauss Markov model is very sensitive to nonnormality, variance heterogeneity as well as large sample size. These assumptions may be violated as a result of departures from normality and small sample size. To overcome these problems, an Adaptive Scheme is adopted. The Adaptive Scheme is a two step procedure in which a selector statistic is used to first examine and classify given data based on measures of skewness and tailweight. Afterwards, a test statistic, independent of the selector statistic is chosen and a test conducted. A One-way Analysis of Variance and Repeated Measures Design models were considered under uncorrelated and correlated error distributions respectively. The nine winsorised scores proposed by Hettmansperger (1984) were used because they are considered the most appropriate rank scores for hypothesis testing. The Winsorised scores as well accommodate a wide range of distributions which are either symmetric or asymmetric with varying tailweights. In addition, the benchmarks for cut-off values for the measures of skewness and tailweights postulated by Al-Shomrani (2003) in his PhD dissertation were used. 10,000 simulations were conducted to compare the performance of the Adaptive Scheme and the Gauss Markov model from different continuous distributions under uncorrelated and correlated errors. Analyses of real datasets were as well performed to ascertain the efficiency of the two tests. The findings favoured the Adaptive Scheme under a broad class of continuous distributions especially for non-normal distributions. The adaptive scheme is applicable to both small and large samples. It is therefore recommended that Statisticians, Researchers and Data Analysts be encouraged to use adaptive schemes because they are applicable to a broad class of distributions.

ACKNOWLEDGMENT

A number of individuals made varied and considerable contributions to this thesis to its present form. I am particularly grateful to my supervisors, Prof. Gabriel Asare Okyere (Department of Mathematics, KNUST) and Dr. Abukari Alhassan (Former PG Coordinator, FMS, UDS, Navrongo campus) whose brotherly reception for discussions created an enabling climate for the completion of this thesis. Besides, they painstakingly undertook to correct loopholes in the thesis and their invaluable and priceless suggestions have goaded me on to complete this thesis.

I would also like to express my heartfelt gratitude and appreciation to all the Senior Members and Administrative staff of the Department of Statistics, Faculty of Mathematical Sciences of the University for Development Studies, Navrongo Campus for their immense support during the preparation of this thesis.

This acknowledgement cannot end without mentioning Prof. Suleman Nasiru, Head of Statistics Department, Dr. Salifu Katara, Dr. Solomon Sarpong and Dr. Daabo of this great University for their immense contributions towards the completion of this thesis. I appreciate you all.

DEDICATION

This thesis is dedicated to my lovely and supportive wife Mrs. Evelyn Afua Biney, to the memories of my parents and my beloved late twin sister Miss Gifty Biney.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

The Hogg's adaptive scheme is extended to the Gauss Markov model. In practice, often the Statisticians, Researchers or the Data Analysts may not have prior knowledge of the underlying distribution of a statistical model so a procedure which will give optimum power and efficiency ought to be chosen. For example, if the distribution of errors is known to be normal in a linear model then inference based on least squares which maximises power and efficiency should be chosen. On the other hand, when the assumption of normality of the error distribution is not met or violated as a result of outliers, then a more robust method can be used to analyse the problem. The rank-based method, according to Hájek and Sidák (1967), is robust to outliers and has high efficiency and power for both normal and non-normal distributions.

The Gauss-Markov model is a statistical procedure mostly used in data analysis. Its assumptions, among others, are normality of the data, large sample size as well as uncorrelated error terms. The statistical method, Analysis of Variance (ANOVA) is widely used in experimental designs whose objective is to test appropriate hypothesis about treatment means and to estimate them from three or more random samples, good sources for details can be found in (Miller, 1997; Bolton and Bon, 2009; Montgomery, 2013).

Büning (1996) indicated that, Statisticians, Data Analysts as well as Researchers often rely on the assumptions of normality for Analysis of Variance (ANOVA) models. He concluded that a lot of details on the data is usually ignored thereby violating the assumption of normality.

Generally, models are believed to be simplified representation of reality, they could be deterministic or probabilistic in nature. Statistical models should be such that the level

of uncertainty is reduced to the barest minimum. Hogg (1974) stated that a more realistic approach would be to find statistical procedures good for a broad class of underlying models, but which are not necessarily best for any of them. Such procedures are robust.

The one-way or two-way ANOVA models which could be reduced to Gauss Markov model relies mostly on normality, homogeneity of variance and large sample size for it to be modelled. However, in practise, these assumptions may be violated. ANOVA models are very sensitive to non-normality and departures from normality may originate from either skewness or outliers. Wilcox and Keselman (2003) postulated that, small departures from normality can substantially lower the power when comparing the means of two or more groups. Zar (1996) indicated though departures from normality could be remedied by using transformations such as exponential and logarithm, sometimes even after the transformation, problems with non normality data still occurs. ANOVA models which are mostly used in clinical trials may have very low enrolment at centres and hence a small sample size. This will inhibit the efficiency of the statistical procedure used.

One particular problem in which normality assumptions become inappropriate is small sample size. In most statistical modelling or techniques, sample size must be large enough for such procedure to be statistically admissible or valid. For small samples, however, some nonparanetric methods have been developed. Hao and Houser (2012) advocated for distribution-free (nonparametric) tests for differences in location problems between samples. A statistical procedure known as Adaptive Scheme would be used to handle the problems of non normality and small sample size. Adaptive Tests are categorised under Robust Statistics. Robustness theories according to Huber (1981) and Huber and Ronchetti (2009) can be described as stability theories of statistical inference and signify insensitivity to small deviations from the assumptions.

In this thesis, the nine winsorised scores proposed by Hettmansperger (1984) are considered as the most appropriate rank scores for hypothesis testing because the Winsorised scores accommodate a wide range of continuous distributions which are either symmet-

ric or asymmetric with varying tailweights. In practice what practitioners are confronted with is at what sample size will an adaptive procedure be as efficient as that of parametric procedure in a Gauss Markov model. This sample size will guide Statisticians in their modelling process.

The motivation for the study is to find a robust adaptive scheme for Gauss Markov model. In addition, we want to popularise the adaptive scheme for practising Statisticians, Researchers and Data Analysts because of the numerous advantages adaptive tests have over parametric tests. Often, parametric tests are inefficient for analysing nonnormal distributions. In addition, statistics for testing significant difference of parametric tests usually rely on large sample size among other assumptions but when sample size is small as in clinical trials and the presence of outliers in the data, then, there is the need for a more robust scheme.

1.2 Problem Statement

The Gauss Makov model according to Glen (2018) has these assumptions namely; linearity, non-collinearity, randomness, exogeneity and homoscedasticity. These assupmtions lend credence to the validity of the ordinary least squares when estimating the regression parameters. However, in practice, these assumptions are seldomly met fully. As a result, these assumptions may be violated. When for example, the assumption of exogeneity is violated, the estimators will be biased and inconsistent. Again, Gauss Markov model which is a parametric test assumes a normal distribution and large sample size. But in the event of these assumptions being violated which often happens as in clinical trials where sample size is usually small at centers, then a more robust method is employed. Similarly, outliers in data can affect the validity of data analysis and drawing of conclusions if not handled effectively (Tukey, 1960; Barnett and Lewis, 1994; Al-Shomrani, 2003). To overcome the problem of normality, large sample size, among others, the adaptive scheme is applied because it is more efficient and powerful than the parametric tests when data is generated from nonnormal distibutions, having small sample size and in the presence of outliers.

1.3 Objectives of the Study

The general objective for the study is to compare the relative efficiency of the adaptive scheme to the parametric tests.

The specific objectives are

1. To investigate and identify a more robust scheme for Gauss Markov Model.
2. To identify an adaptive scheme for ANOVA models under uncorrelated errors.
3. To identify an adaptive scheme for ANOVA models under correlated errors.
4. To investigate the relative efficiency of the adaptive scheme and Gauss Markov model to a symmetric and asymmetric distributions with varying tailweights.

1.4 Research Questions

1. What is the importance of adaptive tests in the scheme of data analysis?
2. What type of data is suitable for Adaptive test?
3. Why should adaptive test be used instead of traditional ANOVA tests?
4. At what sample size is adaptive test most appropriate?

1.5 Justification for the Study

The asymptotic properties of statistical estimates and tests are to a large extent dependent on large sample size but in reality, sample sizes as in clinical trials often has low enrolment. In this case, the use of the parametric test such as F-test in ANOVA models would not yield the optimal test because of violation of large sample assumption. In real-world testing situations reasearchers or data analysts rarely have foreknowledge of the distribution of the errors in order to apply the most powerful tests, so it is important to know just how the traditional tests compare to adaptive tests with a variety of normal and non-normal distributions. Adaptive tests of significance have been confirmed to

have significant power over traditional tests, Hogg et al. (1975), Büning (2009), Okyere (2011), O'Gorman (2012), Muhammed Di et al. (2014) and Glen (2018) are good sources of information.

The study has confirmed that adaptive tests are applicable to all research situations where the level of significance, size and power of the studies are concerned. Statisticians, Researchers and Data Analysts are therefore encouraged to use adaptive test to gain popularity over the parametric test when the underlying error distributions are both normal and non-normal with varying tailweights.

1.6 Organisation of the Study

This thesis is made up of five chapters, references and appendices. Chapter One deals with the introduction. In this chapter, the background of the study, problem statement, objectives of the study, research questions, justification and organisation of the study are discussed. Chapter Two focuses on the literature review of Gauss Markov model and some adaptive schemes. The one-way and Repeated Measures ANOVA models and hypothesis as well as the Traditional F-Test are considered. In Chapter Three, the methods and theorems used for the adaptive procedures are the focus . Chapter Four presents the results and discussions. In this chapter, simulations were conducted to confirm the underlying error distributions of some continuous distributions. It as well compared the efficiency of our adaptive scheme and the traditional test. Application of the adaptive scheme on real data were considered here as well. Chapter Five presents the summary, conclusions of the study highlighting the major findings and recommendations.

CHAPTER TWO

LITERATURE REVIEW

This chapter is focused on the review of Gauss-Markov Model and some Adaptive Procedures. The One-way ANOVA and Repeated Measures ANOVA models which are special cases of Gauss Markov Model are reviewed in this chapter.

2.1 Gauss Markov Model

The Gauss Markov model belongs to the general classes of linear models and is given by

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e} \tag{2.1}$$

where $\mathbf{Y}$ is an $n \times 1$ vector of observed responses, $\mathbf{X}$ is an $n \times p$ (design) matrix of fixed constants, β is a $p \times 1$ vector of fixed but unknown parameters, and $\mathbf{e}$ is an $n \times 1$ vector of unobserved random errors. It is assumed that $E(e) = 0$ and $Cov(e) = \sigma^2 I$ is some unknown parameter. The model (2.1) is called a linear model because the mean of the response vector $\mathbf{Y}$ is linear in the unknown parameter β. Detail information can be seen in (Christensen, 2001; Monahan, 2008).

There are several statistical models which according to Monahan (2008) are examples of the general linear model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$. These examples include but not limited to Linear Regression models, Analysis of Variance (ANOVA) models and Analysis of Covariance (ANCOVA) model. Regression models generally refer to models for which $\mathbf{X}$ is full rank, while ANOVA models refer to those for which $\mathbf{X}$ consists of zeros and ones. Brief discussions on examples of Gauss Markov model are made in the subsequent subsections

2.1.1 One - Sample Problem

Given that $Y_1, Y_2, Y_3, \ldots, Y_n$ is independent and identically distributed (iid) random variables whose mean is μ and variance $\sigma^2 > 0$. If $e_1, e_2, \ldots, e_n$ are iid random variables with mean $E(e_i) = 0$ and common variance σ^2, then model (2.1) can be written in matrix notation as

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \qquad \mathbf{X}_{n\times 1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad \beta_{1\times 1} = \mu, \qquad \mathbf{e}_{n\times 1} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

2.1.2 Simple Linear Regression

A regression model that estimates the relationship between one independent variable and one dependent variable is termed a simple linear regression. The simple linear regression model is

$$Y_i = \beta_0 + \beta_1 x_i + e_i \tag{2.2}$$

for $i = 1, 2, 3, \ldots, n$, where Y is the value of the response (dependent) variable; x the value of the independent (predictor) variable; and β_0 and β_1 are unknown regression coefficients; and e_i are uncorrelated random variables whose mean is 0 and common variance $\sigma^2 > 0$. If predictor variables $x_1, x_2, x_3, \ldots, x_n$ are fixed constants which are measured without error, then model (2.2) is considered a special case of model (2.1). Model (2.2) can be written in a matrix notation as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$Y = \quad X \quad \beta \quad + \quad e$$

2.1.3 Multiple Linear Regression

Suppose the response (dependent) variable is Y and the independent variables $X_1, X_2, X_3, \ldots, X_k$ are linearly related, then the multiple linear regression model is

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_2 x_{i3} + \cdots + \beta_k x_{ik} + e_i \tag{2.3}$$

for $i = 1, 2, 3, \ldots, n$, where $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the regression coefficients which must be estimated from sample data and e_i are uncorrelated random variables whose mean is 0 and common variance $\sigma^2 > 0$. If the independent (predictor) variables are fixed constants which are measured without error, then model (2.3) is a special case of model (2.1). In matrix presentation, model (2.3) is written as

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X}_{n\times p} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & x_{23} & \ldots & x_{2k} \\ 1 & x_{31} & x_{32} & x_{33} & \ldots & x_{3k} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \ldots & x_{nk} \end{bmatrix}, \beta_{p\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \mathbf{e}_{n\times 1} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{bmatrix}$$

where $p = k + 1$. It must be noted that $E(e) = 0$ and $cov(e) = \sigma^2 I$

2.2 Analysis of Variance

The statistical method, Analysis of Variance (ANOVA), according to Miller (1997) and Montgomery (2013) is a widely used tool for finding out which factors contribute to given measurements. This method is common in medicine, agriculture, quality control applications, commerce, education, among others. The method is based on hypothesis tests with F-distributed test variables computed from the residual quadratic sum. In the next subsection, one of the most important sampling distributions will be considered.

2.2.1 Sampling Distributions

The sampling distribution considered here is the normal distribution because it is one of the most important distributions. If z is a normal random variable, then the probability distribution of z is

$$f(z) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{(z-\mu)^2}{2\sigma}} \tag{2.4}$$

where μ is the finite mean of the distribution and $\sigma^2 > 0$ the variance. If $z_1, z_2, z_3, \ldots, z_u$ are normally and independently distributed random variables with zero mean and variance 1, then the random variable

$$\chi^2(u) = \sum_{i=1}^{u} z_i^2 \tag{2.5}$$

follows the chi-square distribution with u degrees of freedom. If $\chi_1^2(u)$ and $\chi_2^2(v)$ are two independent chi-square random variables with u and v degrees of freedom, then the ratio

$$F(u, v) = \frac{\chi_1^2/u}{\chi_2^2/v}$$

follows the F-distribution with u numerator and v denominator degrees of freedom. The probability distribution of F is

$$h(F) = \frac{\Gamma(\frac{u+v}{2})(\frac{u}{v})^{\frac{u}{2}} F^{(\frac{u}{2})-1}}{\Gamma(\frac{u}{2})\Gamma(\frac{v}{2})[(\frac{u}{v})F + 1]^{\frac{u+v}{2}}}, \quad 0 < F < \infty. \tag{2.6}$$

Also a non-central F-distribution is defined, as $F(u, v, \delta)$, where δ is a non-centrality. If $\delta = 0$ the non-central F-distribution becomes the usual F-distribution Monahan (2008).

2.2.2 Parametric F-Test

Suppose $Y_{i1}, Y_{i2}, \ldots, Y_{in}$, where $i = 1, 2, \ldots, k$ are independent random variables then

$$Y_{ij} \sim N(\mu_i, \sigma_i^2),$$

$j = 1, 2, \ldots, n_i$ and $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$.

Hypothesis Testing

The hypothesis to be considered is

$$H_0 \colon \mu_1 = \mu_2 = \cdots = \mu_k$$

$$H_1 \colon \mu_i \neq \mu_j$$

for some $i \neq j$

Test Statistics

The test statistic for the likelihood ratio F-test is based on the statistic

$$F = \frac{(N-k)\sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2} \tag{2.7}$$

where,

$$N = \sum_{i=1}^{k} n_i$$

$$\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{k} n_i \bar{y}_i$$

The test statistic, under H_0, follows an F-distribution with $k-1$ and $N-k$ degrees of freedom.

Decision Rule

H_0 is rejected if $F \geq F_{\alpha,(k-1),(N-k)}$ where α is a pre-specified level of significance and F is the test statistic, see Montgomery (2013).

2.2.3 One -Way Analysis of Variance (ANOVA) Model

Analysis of variance problems usually concern data that arise mainly from experimental design to compare three or more ($k \geqslant 3$) treatment means. For the i^{th} treatment level, let n_i experimental units be selected at random and assigned to the i^{th} treatment.

The One-way ANOVA model is given by

$$Y_{ij} = \mu + \alpha_i + e_{ij} \qquad \begin{cases} i = 1, 2, 3, \ldots, k \\ j = 1, 2, 3, \ldots, n_i \end{cases} \tag{2.8}$$

where the Y_{ij} is the $(ij)^{th}$ observation, μ is the overall mean common to all treatments, α_i is the i^{th} treatment effect and random errors e_{ij} are uncorrelated random variables with zero mean and common variance $\sigma^2 > 0$. If the k treatment effects $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_k$ are fixed constants, then model (2.8) is a special case of model (2.1), see Monahan (2008). Thus,

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ \vdots \\ Y_{kn_k} \end{bmatrix}, \qquad \mathbf{X}_{n\times p} = \begin{bmatrix} 1_{n_1} & 1_{n_1} & 0_{n_1} & 0_{n_1} & \ldots & 0_{n_1} \\ 1_{n_2} & 0_{n_2} & 1_{n_2} & 0_{n_2} & \ldots & 0_{n_2} \\ 1_{n_3} & 0_{n_3} & 0_{n_3} & 1_{n_3} & \ldots & 0_{n_3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1_{n_k} & 0_{n_k} & 0_{n_k} & 0_{n_k} & \ldots & 1_{n_k} \end{bmatrix},$$

$$\beta_{p\times 1} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}, \qquad \mathbf{e}_{n\times 1} = \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ \vdots \\ e_{kn_k} \end{bmatrix}$$

where $p = k + 1$, 1_{ni} is an $n_i \times 1$ vector of ones and 0_{ni} is an $n_i \times 1$ vector of zeros, $E(e) = 0$, $Cov(e) = \sigma^2 I$ and

$$n = \sum_{i=1}^{k} n_i$$

2.2.4 Layout for One-Way ANOVA

The layout of the One-way ANOVA is illustrated in Table 2.1. In this layout, there are k treatments, with n_1 units receiving Treatment 1, n_2 units receiving Treatment 2, ..., n_i units receiving Treatment k. In other words, Treatment 1 has n_1 replications, Treatment 2 has n_2 replications and so on. The yield from the i^{th} unit received the j^{th} treatment is

denoted by Y_{ij}.

Table 2.1: **Data Layout for One-way ANOVA**

Treatment				
1	2	3	$\cdots$	k
Y_{11}	Y_{12}	Y_{13}	$\ldots$	Y_{1k}
Y_{21}	Y_{22}	Y_{23}	$\ldots$	Y_{2k}
Y_{31}	Y_{32}	Y_{33}	$\ldots$	Y_{3k}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$Y_{n_1 1}$	$Y_{n_2 2}$	$Y_{n_3 3}$	$\ldots$	$Y_{n_i k}$

The objectives will be to test hypothesis about the treatment effects and to estimate them. For hypothesis testing, the model errors are assumed to be normally and independently distributed random variables with mean zero and variance σ^2. The variance (σ^2) is assumed to be constant for all factor levels.

2.2.5 Two-Way Crossed ANOVA Without Interaction

Consider an experiment with two factors A and B where Factor A has a levels and Factor B has b levels. The two-way crossed model without interaction is given by

$$Y_{ijk} = \mu + \alpha_1 + \beta_j + e_{ijk}, \quad \begin{cases} i = 1, 2, 3, \ldots, a \\ j = 1, 2, 3, \ldots, b \\ k = 1, 2, 3, \ldots, n_{ij} \end{cases} \tag{2.9}$$

where the random errors e_{ijk} are uncorrelated random variables with zero mean and a constant variance $\sigma^2 > 0$. It is worth noting that model (2.9) is a special case of model (2.10) when

$$H_0 : \gamma_{11} = \gamma_{12} = \gamma_{21} = \gamma_{22} = \gamma_{31} = \gamma_{32} = 0$$

is true. That is, the no-interaction model is a reduced version of the interaction model. The assumptions of the no-interaction models is same as the interaction model. Thus, $E(e) = 0$ and $Cov(e) = \sigma^2 I$. The X matrix is not of full rank. The rank of X is $r = 4$

and there are $p = 6$ columns, see Monahan (2008) and Montgomery (2013) for details. In a matrix notation, for example, with $a = 3$, $b = 2$, and $n_{ij} = 3$, then we have

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{121} \\ Y_{122} \\ Y_{123} \\ Y_{211} \\ Y_{212} \\ Y_{213} \\ Y_{221} \\ Y_{222} \\ Y_{223} \\ Y_{311} \\ Y_{312} \\ Y_{313} \\ Y_{321} \\ Y_{322} \\ Y_{323} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{31} \\ \gamma_{32} \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_{111} \\ e_{112} \\ e_{113} \\ e_{121} \\ e_{122} \\ e_{123} \\ e_{211} \\ e_{212} \\ e_{213} \\ e_{221} \\ e_{222} \\ e_{223} \\ e_{311} \\ e_{312} \\ e_{313} \\ e_{321} \\ e_{322} \\ e_{323} \end{bmatrix}$$

2.2.6 Two-way Crossed ANOVA Model with Interaction

An experiment with two factors, say, A and B, where Factor A has a levels and Factor B has b levels is under consideration. The factors A and B are, in general, crossed if every level of A occurs in combination with every level of B. The two-factor crossed ANOVA model with interaction is given by

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \quad \begin{cases} i = 1, 2, 3, \ldots, a \\ j = 1, 2, 3, \ldots, b \\ k = 1, 2, 3, \ldots, n_{ij} \end{cases} \tag{2.10}$$

where the random errors e_{ijk} are uncorrelated random variables with zero mean and constant unknowm variance $\sigma^2 > 0$. If all the parameters are fixed, then model (2.10) is a special case of model (2.1). Without loss of generality (WLOG), supppose $a = 3$, $b = 2$ and $n_{ij} = 3$ then model (2.10) can be written in a matrix notation as

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{121} \\ Y_{122} \\ Y_{123} \\ Y_{211} \\ Y_{212} \\ Y_{213} \\ Y_{221} \\ Y_{222} \\ Y_{223} \\ Y_{311} \\ Y_{312} \\ Y_{313} \\ Y_{321} \\ Y_{322} \\ Y_{323} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{31} \\ \gamma_{32} \end{bmatrix}, \mathbf{e} = \begin{bmatrix} (e_{111} \\ e_{112} \\ e_{113} \\ e_{121} \\ e_{122} \\ e_{123} \\ e_{211} \\ e_{212} \\ e_{213} \\ e_{221} \\ e_{222} \\ e_{223} \\ e_{311} \\ e_{312} \\ e_{313} \\ e_{321} \\ e_{322} \\ e_{323} \end{bmatrix}$$

where $E(e) = 0$ and $cov(e) = \sigma^2 I$. The detailed discussion is found in (Monahan, 2008).

2.2.7 Two-Way Nested ANOVA

Consider an experiment with two factors, where one factor, say, Factor B is nested within Factor A. Thus, every level of Factor B appears with exactly one level of Factor A. The statistical model is given by

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + e_{ijk}, \quad \begin{cases} i = 1, 2, 3, \ldots, a \\ j = 1, 2, 3, \ldots, b_i \\ k = 1, 2, 3, \ldots, n_{ij} \end{cases} \tag{2.11}$$

In this model, μ denotes the overall mean, α_i represents the effect due to the i^{th} level of A, and β_{ij} represents the effect of the j^{th} level of B, nested within the i^{th} level of A. If all parameters are fixed, and the random error e_{ijk} are uncorrelated random variables with zero mean and constant unknown variance $\sigma^2 > 0$, then model (2.11) is a special case of model (2.1) where $E(e) = 0$ and $cov(e) = \sigma^2 I$). For example, with $a = 3$, $b = 2$, and $n_{ij} = 2$, then we have

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{311} \\ Y_{312} \\ Y_{321} \\ Y_{322} \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \\ \beta_{31} \\ \beta_{32} \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_{111} \\ e_{112} \\ e_{121} \\ e_{122} \\ e_{211} \\ e_{212} \\ e_{221} \\ e_{222} \\ e_{311} \\ e_{312} \\ e_{321} \\ e_{322} \end{bmatrix}$$

2.2.8 Analysis of Covariance

Conducting an experiment to compare $k \geq 2$ treatments after adjusting for the effects of a covariate x, a model for the analysis of covariance (ANCOVA) is given by

$$Y_{ij} = \mu + \alpha_i + \beta_i x_{ij} + e_{ij} \qquad \begin{cases} i = 1, 2, 3, \ldots, k \\ j = 1, 2, 3, \ldots, n_i \end{cases} \tag{2.12}$$

where the random errors e_{ij} are uncorrelated random variables with zero mean and common variance $\sigma^2 > 0$. In this model, μ represents the overall mean, α_i represent the fixed effect of receiving the i^{th} treatment (disregarding the covariates), and β_i denotes the slope of the line that relates Y to x for the i^{th} treatment. The x_{ij}'s are assumed to be fixed values measured without error so the model (2.12) is a special case of model (2.1). Suppose $k = 3$ and $n_1 = n_2 = n_3 = 3$ then in matrix notation, model (2.12) can be written as

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & x_{11} & o & 0 \\ 1 & 1 & 0 & 0 & x_{12} & 0 & 0 \\ 1 & 1 & 0 & 0 & x_{13} & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & x_{21} & 0 \\ 1 & 0 & 1 & 0 & 0 & x_{22} & 0 \\ 1 & 0 & 1 & 0 & 0 & x_{23} & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & x_{31} \\ 1 & 0 & 0 & 1 & 0 & 0 & x_{32} \\ 1 & 0 & 0 & 1 & 0 & 0 & x_{33} \end{bmatrix}, \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{31} \\ e_{32} \\ e_{33} \end{bmatrix}$$

If all the slopes are equal, say, $\beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k$, then the ANCOVA model reduces to

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + e_{ij}$$

Let $k = 3$ and $n_1 = n_2 = n_3 = 3$, then in matrix notation, model (2.12) can be written as

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & x_{11} \\ 1 & 1 & 0 & 0 & x_{12} \\ 1 & 1 & 0 & 0 & x_{13} \\ 1 & 0 & 1 & 0 & x_{21} \\ 1 & 0 & 1 & 0 & x_{22} \\ 1 & 0 & 1 & 0 & x_{23} \\ 1 & 0 & 0 & 1 & x_{31} \\ 1 & 0 & 0 & 1 & x_{32} \\ 1 & 0 & 0 & 1 & x_{33} \end{bmatrix}, \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{31} \\ e_{32} \\ e_{33} \end{bmatrix}$$

2.3 Repeated Measures Design

The experimental units or subjects in the field of social, medical and physical sciences, and business are usually people in experimental design. The differences in experience, training, or background among the subjects may to a large extent affect the responses eventhough the same treatment may be applied to the experimental situations. This variability between experimental units would become part of the experimental error if it is not controlled, and in some cases, it would significantly inflate the error mean square, making it more difficult to detect real differences between treatments. In other repeated measures the experimental units or subjects could be different stores in a marketing survey, or plants, or animals in biological survey. The control of this variability between experimental units (subjects) is overcome by employing a repeated measures design, (Montgomery, 2013).

A repeated measures design is one in which multiple measurements of the response variable are obtained from each experimental unit or subject. The measurements might be taken serially in time such as hourly, daily, weekly or monthly. The dependency, or correlation, among responses measured under the same experimental unit is the defining feature of a repeated measures design.

The repeated measures design in its simplest case is a generalisation of the paired t test. A repeated measures within-subjects design can be thought of as an extension of the paired t test that involves $n \geq 3$ assessments in the same experimental unit. Repeated measures experiments can as well be viewed as a

type of factorial experiment, with group and time as the two factors.

Repeated Measures ANOVA model for one sample and multiple samples are examples of Two-way Crossed ANOVA model where time and subjects are crossed for the one sample case. In the case of multiple samples, Group and Time are crossed. The two factors, Factor A and Factor B could be subjects and time or group and time depending on the type of the sample.

2.3.1 Repeated Measures ANOVA Model for One Sample

Suppose there are n independent experimental units (subjects) and t treatments or time points. Each treatment or time point is to be used exactly once on each of n esprimental units. The observation Y_{ij} is the response of subject i at time j. The model for this design is

$$Y_{ij} = \mu + \pi_i + \tau_j + e_{ij} \qquad \begin{cases} i = 1, 2, 3, \ldots, n \\ j = 1, 2, 3, \ldots, t \end{cases} \tag{2.13}$$

where μ is the overall mean, π_i is a random effect for subject i which is constant over all occasions, τ_j is the fixed effect of time j and e_{ij} is a random error component specific to subject i at time j.

Assumptions

1 The fixed effects τ_j are assumed to satisfy the sum-to-zero constraints

$$\sum_{j=1}^{t} \tau_j = 0$$

2 The random effects π_i are independent $N(0, \sigma_\pi^2)$

3 the random errors e_{ij} are independent $N(0, \sigma_e^2)$

4 The term π_i is common to all n measurements on the same subject so, the covariance between Y_{ij} and $Y_{i'j}$ is not, in general zero.

5 The covariance between Y_{ij} and $Y_{i'j}$ is constant across all treatments and subjects.

It can be said that all the random variables in model (2.13) are independent, the repeated observations from a subject are correlated. Details are found in Montgomery (2013) and Davis (2002). The repeated measures ANOVA for One-Sample case can be viewved as a special case of the randomised block design in which the the block is the individual experimental unit. If all the parameters are fixed, then model (2.13) is a special case of model (2.1).

2.3.2 Layout for the One-Sample Case

The layout of a repeated measures design for one sample is shown in Table 2.2.

Table 2.2: **Data Layout of a Repeated Measures Design for One Sample Case**

	Time Points						
Subject	1	2	3	...	j	...	t
1	Y_{11}	Y_{12}	Y_{13}	...	Y_{1j}	...	Y_{1t}
2	Y_{21}	Y_{22}	Y_{32}	...	Y_{2j}	...	Y_{2t}
3	Y_{31}	Y_{32}	Y_{33}	...	Y_{3j}	...	Y_{3t}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
i	Y_{i1}	Y_{i2}	Y_{i3}	...	Y_{ij}	...	Y_{it}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
n	Y_{n1}	Y_{n2}	Y_{n3}	$\cdots$	Y_{nj}	...	Y_{nt}

The objective of the analysis is to determine whether the distribution of the response variable is changing over time. For hypothesis testing, the model errors are assumed to be normally and independently distributed random variables with mean zero and variance σ^2. The variance (σ^2) is assumed to be constant for all factor levels. The details can be seen in Davis (2002) and Montgomery (2013).

2.3.3 Repeated Measures ANOVA Model for Multiple Samples

Let us assume that repeated measurements at t time points are obtained from s groups of subjects. Let n_h denote the number of subjects in group h, and let

$$n = \sum_{h=1}^{s} n_h.$$

Let Y_{hij} denote the response at time j from the i^{th} subject in group h then the model for repeated measures ANOVA for multiple samples follows the two-way ANOVA and is given by

$$Y_{hij} = \mu + \pi_h + \tau_j + \gamma_{hj} + e_{hij} \qquad \begin{cases} h = 1, 2, 3, \ldots, s \\ i = 1, 2, 3, \ldots, n_h \\ j = 1, 2, 3, \ldots, t \end{cases} \tag{2.14}$$

where μ is the overall mean effect, π_h is the fixed effect of group h, τ_j is the fixed effect of time j , γ_{hj} is the interaction of the time and group and e_{hij} is a random error component. Both factors are assumed

to be fixed, and the treatment group effects are defined as deviations from the overall mean, so

$$\sum_{h=1}^{s} \pi_h = 0$$

and

$$\sum_{j=1}^{t} \tau_j = 0.$$

Similarly, the interaction effects are fixed and are defined such that

$$\sum_{h=1}^{s} \gamma_{hj} = \sum_{j=1}^{t} \gamma_{hj} = 0.$$

Let the random errors e_{hij} be uncorrelated random variables with zero mean and constant unknown variance $\sigma^2 > 0$. If $E(e) = 0$ and $cov(e) = \sigma^2 I$, then model (2.14) is a special case of model (2.1).

In the multiple samples Repeated Measures ANOVA design, both row and column effects are of equal interest. As such, test of hypotheses of interest are:

- Treatment (Group) Effects

  $H_0 : \pi_1 = \pi_2 = \cdots = \pi_s = 0$

  $H_1 : \pi_h \neq 0$ for at least one π_h

- Time Effects

  $H_0 : \tau_1 = \tau_2 = \cdots = \tau_t = 0$

  $H_1 : \tau_j \neq 0$ for at least one τ_j

- Interaction Effects

  $H_0 : \gamma_{hj} = 0$ for all h, j

  $H_1 : \gamma_{hj} \neq 0$ for at least one γ_{hj}

  The group and time effects could be referred to as row and column treatment effects respectively.

2.3.4 Repeated Measures ANOVA Layout for Multiple Samples

The layout for repeated measures ANOVA for the multiple samples which is an equivalent of a two-way crossed factorial design is presented in Table 2.3. The two factors Group and Time Point have levels s and t respectively.

Table 2.3: **Data Layout for Repeated Measures Design for Multiple Samples**

		Time Points					
Group	Subject	1	2	...	j	...	t
	1	Y_{111}	Y_{112}	...	Y_{11j}	...	Y_{11t}
	2	Y_{121}	Y_{122}	...	Y_{12j}	...	Y_{12t}
1	⋮	⋮	⋮	⋱	⋮	⋱	⋮
	i	Y_{1i1}	Y_{1i2}	...	Y_{1ij}	...	Y_{1it}
	⋮	⋮	⋮	⋱	⋮	⋱	⋮
	n_1	Y_{1n_11}	Y_{1n_12}	...	Y_{1n_1j}	...	Y_{1n_it}
	1	Y_{211}	Y_{212}	...	Y_{21j}	...	Y_{21t}
	2	Y_{221}	Y_{222}	...	Y_{22j}	...	Y_{21t}
2	⋮	⋮	⋮	⋱	⋮	⋱	⋮
	i	Y_{2i1}	Y_{2i2}	...	Y_{2ij}	...	Y_{2it}
	⋮	⋮	⋮	⋱	⋮	⋱	⋮
	n_2	Y_{2n_21}	Y_{2n_22}	...	Y_{2n_2j}	...	Y_{2n_2t}
	1	Y_{h11}	Y_{h12}	...	Y_{h1j}	...	Y_{h1t}
	2	Y_{h21}	Y_{h22}	...	Y_{h2j}	...	Y_{h2t}
h	⋮	⋮	⋮	⋱	⋮	⋱	⋮
	i	Y_{hi1}	Y_{hi2}	...	Y_{hij}	...	Y_{hit}
	⋮	⋮	⋮	⋱	⋮	⋱	⋮
	n_h	Y_{hn_h1}	Y_{hn_h2}	...	Y_{hn_hj}	...	Y_{hn_ht}
	1	Y_{s11}	Y_{s12}	...	Y_{s1j}	...	Y_{s1t}
	2	Y_{s21}	Y_{s22}	...	Y_{s2j}	...	Y_{s2t}
s	⋮	⋮	⋮	⋱	⋮	⋱	⋮
	i	Y_{si1}	Y_{si2}	...	Y_{sij}	⋱	Y_{sit}
	⋮	⋮	⋮	⋱	⋮	⋱	⋮
	n_s	Y_{sn_s1}	Y_{sn_s2}	...	Y_{sn_sj}	...	Y_{sn_st}

2.4 Covariance Structure

The correlation between time points of repeated measures are assumed to be constant and such covariance structure is called compound symmetry. However, the correlation between time points is higher when times of measurements are closer than measures that are far apart in time. Variances of repeated measures change with time. This structure of the covariance is referred to as First Order Autoregressive (AR(1)) covariance.

The potential patterns of correlation and variance may be combined to produce a complicated covariance structure of repeated measures, see Davis (2002). The covariance structures discussed here include the compound symmetry (CS), autoregressive of order 1 AR(1), autoregressive with heterogenuous variance ARH(1), unstructured (un) and Toeplitz (toep).

The compound symmetry assumes that the correlations between all pairs of measures are the same and the variances are homogeneous. The AR(1) structure has homogeneous variances and correlations

that decline exponential with distance. Thus, the correlations between adjacent pairs are greater than the correlations between distant pairs. The unstructured covariance structure assumes that each pair of measurements has its own correlation. The Toeplitz structure is similar to the AR(1) in that all measurements next to each other have the same correlation. However, the corelations do not necessarily have the same pattern as in the AR(1).

2.4.1 Compound Symmetry

The compound symmetry (CS) covariance structure is given by

$$\sum = \begin{bmatrix} \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 + \sigma_\pi^2 & \sigma_\pi \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma^\pi + \sigma_e^2 \end{bmatrix}$$

$$\sum = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

where

$\rho = \dfrac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_e^2} = Corr(y_{ij}, y_{ij'})$ and $\sigma^2 = \sigma_\pi^2 + \sigma_e^2$.

The intraclass correlation coefficient ρ ranges from 0 to 1 as $\dfrac{\sigma_\pi^2}{\sigma_e^2}$ ranges from 0 to ∞. The variance $\sigma_\pi^2 + \sigma_e^2$ and covariances σ_π^2 are both homogeneous across time.

2.4.2 First Order Autoregressive (AR1)

The first Order Autoregressive AR(1) covariance structure is

$$\sum = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

where the $Cov(Y_{ij}, Y_{ij'}) = \sigma^2 \rho^{|i-j|}$ for $i, j = 1, 2, \ldots, t$.

2.4.3 Unstructured

The unstructured covariance structure is

$$\sum = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^3 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^4 \end{bmatrix}$$

where the $Cov(Y_{ij}, Y_{ij'}) = \sigma_{ij}$ for $i, j = 1, 2, \ldots, t$

2.4.4 Toeplitz

The Toeplitz covariance structure is given as

$$\sum = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix},$$

where the covariance $Cov(Y_{ij}, Y_{ij'}) = \sigma_{|i-j|+1}$, for $i, j = 1, 2, \ldots, t$.

2.4.5 Autoregressive with Heterogeneous Variance

The covariance structure for the autoregressive with heterogeneous variance ARH(1) is

$$\sum = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix},$$

where the covariance $Cov(Y_{ij}, Y_{ij'}) = \sigma_i\sigma_j\rho^{|i-j|}$, for $i, j = 1, 2, \ldots, t$.

2.5 Estimators and Outliers

The presence of outliers in a dataset is one of the most important topics in statistical inference. An outlier can be defined as observations which appear to be inconsistent with the remaining set of data. Outliers can be contaminants, i.e. arising from other distributions or can be typical observations generated from the assumed model, see Barnett and Lewis (1994). Therefore, outliers need very special attention because a small departure from the assumed model can have strong negative effects on the efficiency of classical estimators for location and scale, see Tukey (1960).

2.6 Adaptive Statistical Methods

Adaptive procedures, according to O'Gorman (2004) use the given data to ascertain which statistical method or technique is the most appropriate and efficient. Generally, it is a two stage procedure. In the first stage, a selection statistic is computed from the estimate of skewness and tailweight, that is, the shape of the error distribution of the data. The selector statistic is used to determine the appropriate statistical procedure for the analysis in the second stage. This procedure has been proven to increase the power of the test if the error distribution is skewed and makes narrow confidence intervals, are robust for both validity and efficiency and automatically downweight outliers, which has the effect of making the results less sensitive to observations that do not agree with the model.

Husková (1985) and Hájek et al. (1999) distinguished between non-restrictive and restrictive adaptive schemes. In the case of non-restrictive procedures the optimal scores $a_{opt}(k)$ for the locally most powerful rank test, which depend on the (unknown) underlying distribution function F and its density f, are estimated directly from the data.

However, in the case of restrictive procedures, a 'reasonable' family of distributions and a corresponding class of 'suitable' tests are chosen. At the first stage, the unknown distribution function is classified with respect to some measures like tailweight and skewness. At the second stage, an appropriate test for that classified type of distribution is selected and then carried out. Hogg states 'so adapting the test to the data provides a new dimension to parametric tests which usually improves power of the overall test'. This two-staged adaptive test maintains the level α for all continuous distribution functions. The adaptive tests automatically reduce the influence of outliers. They are sometimes said to be robust; but to be clear about robustness, we should describe the two kinds of robustness.

A test is said to be robust for size if its actual significance level is quite close to the nominal significance level, even when the usual assumptions are not met. For example, a test that is derived by

assuming normality of the error distribution would be robust for size if it maintains its level of significance with non-normal errors. A test is said to be robust for power if it has high power relative to other tests when the usual distributional assumptions are not met. Many traditional tests are robust for size with non-normal errors but are not robust for power. The objective is to develop adaptive tests that are robust for size and robust for power O'Gorman (2012).

In the parametric case of testing hypotheses the efficiency of a test statistic strongly depends on the assumption of the underlying distribution of the data, for example, if we assume normality then optimal tests are available for the one- two- and c-sample location or scale problem such as t-tests, F-tests and Chi-square-tests Büning (1994).

In the non-parametric nonadaptive case the distribution of the test statistic is not based on a special distribution of the data like the normal, only the assumption of continuity of the distribution is needed in general. It is well known, however, that the efficiency of non-parametric tests depends on the underlying distribution, too, for example, the Kruskal-Wallis test in the c-sample location problem has high power for symmetric and medium- up to long-tailed distributions in comparison to its parametric and non-parametric competitors whereas the Kruskal-Wallis test can be poor for asymmetric distributions.

Büning (1994) indicated that for the practising statistician it is more the rule rather than the exception that he has no foreknowledge of the underlying distribution of his data. Consequently, one should apply an adaptive test which takes into account the given data set. A power comparison by means of Monte Carlo simulation shows that the adaptive test is very efficient over a broad class of distributions in contrary to its parametric and non-parametric nonadaptive competitors .

2.6.1 Single Location Adaptive Procedures

Bandyopadhyay and Dutta (2007) proposed two adaptive tests for a single location problem without making assumptions about the symmetry of the continuous distribution of the data. Whereas one is based on a measure of symmetry, used as a standard of deciding between the Wilcoxon signed rank test (W^+) and the signed test (S^+) (the deterministic approach), the other (the probabilistic approach) is a combination of the signed test and the Wilcoxon signed rank test based on evidence of asymmetry provided by the p-value from the triples test defined as

$$\begin{aligned}\hat{\lambda} &= \frac{1}{\binom{n}{3}} \\ &= \sum_{i<j<k} g(X_i, X_j, X_k)\end{aligned} \tag{2.15}$$

and

$$g(x_1, x_2, x_3) = \frac{1}{3}\left[sign(x_1 + x_2 - 2x_3) + sign(x_1 + x_3 - 2x_2) + sign(x_2 + x_3 - 2x_1)\right]$$

where $sign(X) = 1, 0, -1$ according as $x >, =, < 0$ based on equation (2.15).
For the probabilistic approach, Bandyopadhyay and Dutta (2007) used p to denote the p-value associated with the observed value of $\hat{\lambda}$ in equation (2.15), using the p-value to denote the amount of symmetry of the distribution present in the data. For any value of p, a Bernoulli trial with probability of success p is performed. If a success was realized, the Wilcoxon signed rank test was used otherwise, the sign test was used. The adaptive test rule was: Reject H_0 with probability p if $W^+ > w^+$ and with probability $(1-p)$ if $S^+ > s^+$ are the upper -critical values of W^+ and S^+.

However with the deterministic approach, a sample measure of symmetry on which a preliminary test was based was used. The proposed measure of symmetry was given as

$$Q = \frac{X_{(n)} - 2\tilde{X} + X_{(1)}}{X_{(n)} - X_{(1)}}$$

where $-1 \leq Q \leq 1$, $\tilde{X}$ is the median of the data and $X_{(i)}$ is the order statistics of the data. The median was equidistant from both extremes if the distribution of the data was symmetric, closer to the minimum value for a positively skewed distribution and closer to a maximum value for a negatively skewed distribution. The test statistic was then proposed as

$$T = S^+ I(|Q| > c) + W^+(|Q| \leq c)$$

where $I(y)$ is an indicator function assuming value 1 or 0 depending on whether y is true or false. For all values of c considered, $c = 0.075$ was regarded the best in terms of robustness of the test Bandyopadhyay and Dutta (2007).

Consequently, from simulation studies, when the two adaptive methods were compared, Bandyopadhyay and Dutta (2007) concluded that the probabilistic approach was in general found to be very robust and had high power over the deterministic approach, thus concluding that when nothing was known about the skewness of the distribution, the probabilistic approach should be used.

2.6.2 Two-Sample Rank Test Statistics

Hao and Houser (2012) presented a seven decade advances of adaptive procedures for non-parametric test and extensively narrated the progress made in this area. This section employs their material as useful

reference for this review. For a given $f(.)$, as the probability function of the cumulative distribution function $F(.)$, let R_i represent the rank of observation $Y_i (i = 1, 2, \ldots, n_2)$ in the order statistics of the combined sample $N = n_1 + n_2$ observations with $1 \leq R_i \leq N$. Hájek and Sidák (1967) as found in Hao and Houser (2012) showed that, in general the asymptotically most powerful rank test statistic S depends on the inverse c.d.f F^{-1}

$$S = \sum_{i=1}^{n_2} a(R_i) \tag{2.16}$$

and

$$a(R_i) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \tag{2.17}$$

where $u = \dfrac{R_i}{N+1}$ is the $Y_i^{'}s$ rank normalised in the combined sample, $\dfrac{R_i}{N+1} \in (0, 1)$ and $a(R_i)$ is defined as the scores or the ranks, since it maps the observation Y_i to the rank of Y_i in the combined sample. As $N \to \infty$, $F^{-1}(u)$ shows the corresponding observation using its rank R_i and the inverse c.d.f of the data, that is $a(u)$ provides the information in the ranks. In later development, Hájek et al. (1999) established that for any particular distribution of interest models (2.16) and (2.17) provides the most powerful rank test. Thus as $n_1, n_2 \to \infty$, $\dfrac{S - E(S)}{\sqrt{Var(S)}} \sim N(0, 1)$. Hence they provided three examples:

1. **Normal Score Test**

   The most powerful rank test for normal distributions, also known as the normal score test is defined by

   $$S_{nor} = \sum_{i=1}^{n_2} \Phi^{-1}\left[\frac{R_i}{N+1}\right] \tag{2.18}$$

   where Φ is the c.d.f. of standard normal distribution.

   The sampling distribution of S_{nor} is symmetric with mean equal to

   $$E(S_{nor}) = 0$$

   and variance equal to

   $$Var(S_{nor}) = \frac{n_1 n_2}{N(N-1)} \sum_{i=1}^{N} \left[\Phi^{-1}\left(\frac{i}{N+1}\right)\right]^2$$

   The hypotheses of interest are

   $$H_0 : \Delta = 0$$

   versus

   $$H_1 : \Delta \neq 0,$$

where the parameter Δ denotes a shift in location between the two distributions. Other alternatives can be used. At an asymptotic level of α, reject H_0 in favour of H_1 if

$$|z_{nor}| \geq z_{\frac{\alpha}{2}}$$

where

$$z_{nor} = \frac{S_{nor}}{\sqrt{Var_{H_0}(S_{nor})}}$$

2. **Mann-Whitney-Wilcoxon (MWW)**

The MWW test was selected as the most powerful rank test once the data was known to have been drawn from a logistic distribution, with the test statistic given by

$$S_{log} = \frac{2n_2}{N+1}\sum_{i=1}^{n_2} R_i - n_2. \tag{2.19}$$

The linear transformation of equation (2.19) test statistic is given by

$$S_{MWW} = \sum_{i=1}^{n_2} R_i$$

. The sampling distribution of S_{MWW} is symmetric with mean equal to

$$E(S_{MWW}) = \frac{1}{2}n_2(N+1)$$

and the variance is equal to

$$Var(S_{MWW}) = \frac{1}{12}n_1n_2(N+1)$$

.

3. **Median Test**

The median test which was adjudged the most powerful test when the data was from a Laplace (double exponential) distribution, with the test statistic defined as

$$S_{lap} = \sum_{i}^{n_2} sign\left(R_i - \frac{N+1}{2}\right) \tag{2.20}$$

where

$$sign(X) = \begin{cases} 1, & if \quad x > 0 \\ 0, & if \quad x = 0 \\ -1 & if \quad x < 0 \end{cases}$$

Equation (2.20) is practically the same as the test that counts the number of $Y_i^{'}s$ above the median

of the combined sample and increases by $\frac{1}{2}$ when the median falls in the sample of $Y_i^{'}s$. Therefore,

$$\begin{aligned} S_{median} &= \sum_{i=1}^{n_2} \frac{1}{2}\left[sign\left(R_i - \frac{n+1}{2}\right)+1\right] \\ &= \frac{1}{2}S_{lap} + n_2 \end{aligned}$$

with the mean and variance of the median test given as

$$E[S_{median}] = \frac{n_2}{2}$$

and

$$Var[S_{median}] = \frac{n_1 n_2}{4(N-1)}$$

if N is even, and

$$Var[S_{median}] = \frac{n_1 n_2}{4N}$$

if N is odd (Hao and Houser, 2012).

To ascertain the comparative strength of efficiency of these statistics to their parametric counterparts, the Asymptotic Relative Efficiencies (ARE) (that is, for two consistent test statistics, A and B under H_0, ARE is the reciprocal of the ratio of sample sizes needed to derive similar power against the same alternative hypothesis H_1, taking the limit as the sample size $N \to \infty$ as $H_1 \to H_0$). For instance, Pitman (1949) computed the asymptotic relative efficiency (ARE) of the Mann-Whitney-Wilcoxon (MWW) test relative to the t-test as

$$A.R.E_{w,t} = 12\sigma^2\left[\int f^2(x)dx\right]^2$$

where σ is the standard deviation of the underlying distribution of $f(x)$.

2.6.3 Hogg's Adaptive Scheme

In this section we review the general theory of Hogg-type adaptation. In adaptation, selector statistics are sought for. These statistics assist us in adapting to some features of an unknown distribution of the given data. Hogg's adaptive scheme is a two-stage procedure,(Hogg, 1974). Suppose sampling is done from an unknown distribution F(t). First the unknown distribution with its regression score is classified by the skewness and tail weight. This is done through the selector statistics. Second, by the classification, a test statistic which is independent of the selector statistic is selected and a test performed. This two-stage adaptive test maintains the level α for all continuous distributions. The main theorem behind adaptation is stated below.

Theorem 2.6.1. ***Lemma***

1. *Let K denote the class of distribution functions under consideration. Suppose that each of the m tests based on the statistics $T_1, T_2, ..., T_m$ is distribution-free over the class K i.e. $P_{H_0}(T_h \in C_h \mid f) = \alpha$ for each $F \in K$, $h = 1, ..., m$.*

2. *Let S be some statistic that is statistically independent of $T_1, T_2, ..., T_m$ under H_0 for each $F \in K$. Suppose S is used to decide which test T_h to conduct. (S is called a selector statistic.) Specially, let Q denotes the set of all values of S with the following decomposition: $Q = D_1 \cup D_2 \cup D_3 ... \cup D_m$ and $D_h \cap D_k = \emptyset$ for $h \neq k$, so that $S \in D_h$ corresponds to the decision to use the test T_h. The overall testing procedure is then defined by:*
   If $S \in D_h$ then reject H_0 if $T_h \in C_h$.

This two-staged adaptive test is, under H_0, distribution-free over the class K, i.e. it maintains the level α for each $F \in K$.

That is

$$\begin{aligned}
P_{H_0}(reject H_0 \mid F) =& P_{H_0}[\bigcup_{h=1}^{m}(S \in D_h \bigwedge T_h \in C_h \mid F)] \\
=& \sum_{h=1}^{m} P_{H_0}(S \in D_h \bigwedge T_h \in C_h \mid F) \\
=& \sum_{h=1}^{m} P_{H_0}(S \in D_h \mid F).P_{H_0}(T_h \in C_h \mid F) \\
=& \alpha.\sum_{h=1}^{m} P_{H_0}(S \in D_h \mid F) \\
=& \alpha \cdot (1) \\
P_{H_0}(reject H_0 \mid F) =& \alpha
\end{aligned}$$

So the procedure of selecting T_h using an independent statistics S and then constructing a test of significance level α with test statistic T_h has an overall significance level α. Hence the overall testing procedure is defined by, if $S \in D_m$ then reject H_0 if $T_h \in C_h$.

2.6.4 Two Sample Location Problem and Adaptation

The concept of adaptation within the context of two sample location problem is reviewed in this sub subsection.

Let $X_1, X_2, \ldots, X_{n_1}$ be an independent and identically distributed (iid) random sample with cumulative distribution function $F_X = F(x)$ and density function $f_X = f(x)$. Also, let $Y_1, Y_2, \ldots, Y_{n_2}$ be another random sample, independent and identically distributed from the cumulative distribution function $F_Y = F(X - \Delta)$ and density function $f_Y = f(X - \Delta)$, where $\Delta = \mu_y - \mu_x$ represents a shift in location between the two distributions, μ_y and μ_x are the means of $F(X - \Delta)$ and $F(X)$ respectively, and F is unknown.

The objective is to test the hypothesis $H_0 : \Delta = 0$ against $H_1 : \Delta \neq 0$. Hence let $n = n_1 + n_2$ represent the combined sample of both $X_i^{'}s$ and $Y_i^{'}s$. Let $Z^{'} = (X_1, X_2, \ldots, X_{n_1}; Y_1, Y_2, \ldots, Y_{n_2})$ denote the vector of observations; let $n = n_1 + n_2$ denote the total sample. Then the location model can be written as

$$Z_i = \Delta c_i + e_i, \quad 1 \leq i \leq n \tag{2.21}$$

where $e_1, e_2, \ldots, e_n$ are iid with distribution function $F(x)$,

$$c_i = \begin{cases} 0 \;\; if \;\; 1 \leq i \leq n_1 \\ 1, \;\; if \;\; n_1 + 1 \leq i \leq n \end{cases} \tag{2.22}$$

and

$$Z_i = \begin{cases} X_i, \quad for \;\; 1 \leq i \leq n_1 \\ Y_{i-n_1}, \;\; for \;\; n_1 + 1 \leq i \leq n_1 + n_2 = n \end{cases} \tag{2.23}$$

where $Z_{(i)}$ is the order statistics for Z_i. Then under the null hypothesis H_0, for $Z_{(i)}$, the conditional distribution of Z_i is discrete with probability $\frac{1}{n!}$ for all $n!$ permutations of the vector Z_i. This implies that the conditional distribution does not depend on $F(x)$. Hence by the definition of sufficiency the order statistics are sufficient for F, (Hogg et al., 2013). Next the completeness of nonparametric family as found in Bhattacharyya et al. (1977) is considered.

Theorem 2.6.2. *Let K be a family of probability distributions for a real or vector random variable Z and let $W = h(Z)$ be a measurable function with K_W denoting the induced family of distributions of W. If Z is complete with respect to the family K, then W is complete with respect to K_W*

Under H_0, $Z_{(1)} < Z_{(2)} < \cdots < Z_{(n)}$, the ordered observations for the combined sample has a common distribution $F \in K$ where K is the family of all distributions absolutely continuous with respect to Lebesgue measure, a good source of information is (Bhattacharyya et al., 1977). If the function h such that $h(Z_1, Z_2, \ldots, Z_n) = (Z_{(i)}, Z_{(2)}, \ldots, Z_{(n))}$ is considered, then by Theorem 2.6.2, the combined ordered sample $Z_{(1)} < Z_2 < \cdots < Z_{(n)}$ is complete for K. It is worth mentioning that the order statistics are complete with respect to the corresponding induced family of distributions

Corollary 2.6.1. *Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ and $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$ be the order statistics for each sample with distributions F_x and F_y respectively, then $X_{(1)} < X_{(2)} < \cdots < X_{(n)}; Y_{(i)} < Y_{(1)} < \cdots < Y_{(n)}$ are complete with respect to the family distributions induced by F_x and F_Y, for $F_x, F_Y \in K$.*

The next theorem is based on the above corollary. Since under $H_0 : F_x = F_y = F$ the ordered statistics are complete and that given Z_i, the conditional distribution of Z_i does not depend on F, it implies the order statistics are complete and sufficient.

Theorem 2.6.3. *Under $H_0 : F_x = F_y$, the order statistics for the combined sample, $Z_{(1)} < Z_{(2)} < \cdots < Z_{(n)}$ are sufficient and complete for F.*

This means that, under H_0, the order statistics for the combined sample exhausts all information of F.

Let $T = T[R(\mathbf{X},\mathbf{Y})])$ be such a statistic whose distribution is free under H_0. Then from Basu's theorem, T is a test statistic and ancillary for F. Furthermore, if the ordered combined sample, $Z_{(1)}, Z_{(2)}, \ldots, Z_{(n)}$ is complete and sufficient for F, then it follows that for all measurable functions G, $G(Z_{(1)}, Z_{(2)}, \ldots, Z_{(n)})$ is also complete and sufficient for F.

Theorem 2.6.4. ***Basu's Theorem***

Let $T = T[R(\boldsymbol{X}, \boldsymbol{Y})]$ be a statistic whose distribution is free of F, then under H_0, T and $G(Z_{(1)}, Z_{(2)}, \cdots, Z_{(n)})$ are independent, for all (measurable) functions G.

Here the $T's$ are rank tests, thus, $T = T[R(\mathbf{X},\mathbf{Y})] = \{T_1, T_2, T_3, \cdots, T_r\}$. This means that the test statistics depend on the joint ranks of the combined ordered sample $Z_i's$. Applying Theorem 2.6.1 to the two sample problem, K is the class of all continuous distribution functions of F and $T_1, T_2, \ldots, T_r$ are rank statistics. So T_i is distribution-free over K, for $i = 1, 2, \ldots, r$.

Under H_0, the order statistics are complete ad sufficient for the common, but unknown distribution F. This implies for an adaptive scheme, if distribution free statistics are used and the selector is based on the combined order statistics then the adaptive scheme maintains level.

Definition 2.6.1. ***(Sufficiency)***

Consider a random variable $X \in R^n$ on some measurable space Ω. Let K denote a family of distribution of X such that

$$K = F(X, \theta), \theta \in \Theta$$

A statistic $T \equiv T(X)$ is said to be sufficient for θ if and only if the conditional distribution of X given T does not depend on θ for every $F \in K$

2.6.5 The Adaptive Procedure of Hogg, Fisher and 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 et al. (1975) proposed a two step procedure for choosing an appropriate score function. In summary, the HFR adaptive procedures are:

1. Selector statistics Q_1 for Skewness and Q_2 for tailweight are computed.
2. 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.

The selector statistics is computed as

$$Q_1 = \frac{\bar{U}_{0.05} - \bar{M}_{0.5}}{\bar{M}_{0.5} - \bar{L}_{0.05}} \tag{2.24}$$

$$Q_2 = \frac{\bar{U}_{0.05} - \bar{L}_{0.05}}{\bar{U}_{0.5} - \bar{L}_{0.5}} \tag{2.25}$$

where $\bar{U}_{0.05}, \bar{M}_{0.5}$ and $\bar{L}_{0.05}$ are the averages of the largest 5%, the middle 50% and the smallest 5% of the ordered data respectively.

If Q_1 is large (say 2 or more) then there is an indication that the distribution is skewed to the right. On the other hand, if $Q_1 < \frac{1}{2}$, the sample indicates left skewed distribution. Large values of Q_2 shows heavy tailed distribution while small values indicates light tailed distribution.

An illustration cited in Hogg et al. (2013), pages 571 - 575 is used.

Let C_{1lw}, C_{1up}, C_{2lw}, and C_{2up} denote some prespecified emperical cutoffs of the measures of skewness and tailweight for a class of distributions such that the following rules are needed for some scores:

- $Q_1 < C_{1lw}$ indicates left-skewed distributions
- $Q_1 \geq C_{1up}$ indicates right-skewed distributions
- $C_{1lw} < Q_1 < C_{1up}$ indicates symmetric distributions.
- $Q_2 < C_{2lw}$ indicates light tailed distributions
- $Q_2 \geq C_{2up}$ indicated heavy tailed distributions.
- $C_{2lw} < Q_2 < C_{2up}$ indicates medium tailed distributions.

The benchmarks proposed by Hogg et al. (1975) is used in this illustration. Some test statistics which depends on the ranks of combined sample are considered here. Thus, let $R(X_1), R(X_2), R(X_3), \ldots, R(X_{n1}), R(Y_1), R(Y_2),$ denote the combine ranks of $X_1, X_2, X_3, \ldots, X_{n1}, Y_1, Y_2, Y_3, \ldots, Y_{n2}$ of random samples with cdf $F(x)$ and $F(x - \Delta)$ respectively. Define the elements of a set of test statistics by

$$W_i = \sum_{j=1}^{n2} a_i[R(Y_j)], \qquad i = 1, 2, 3, 4 \tag{2.26}$$

where $a_i(j) = \varphi_i\left[\frac{j}{n+1}\right]$, φ_i is an associated score optimal for a specified distribution and $n = n_1 + n_2$. Suppose $F(x)$ can vary from light to heavy tailed distribution, one may consider the Wilcoxon test

$$W_1 = \sum_{j=1}^{n2} a_1[R(Y_j)] \tag{2.27}$$

where $a_1(j) = \varphi_1\left[\frac{j}{n+1}\right]$ and $\varphi_1(u) = \sqrt{12}(u - 0.5)$ is an associated optimal score for the logistic distribution function which is slightly heavier than normal distribution.

As a second example, we consider the Mood's median test which depends on the ranks

$$W_2 = \sum_{j=1}^{n2} a_2[R(Y_j)] \tag{2.28}$$

where $a_2(j) = \varphi_2\left[\frac{j}{n+1}\right]$ and $\varphi_2(u) = sgn(u - 0.5)$ is an associated optimal score for the laplace distribution.

Suppose the right-skewed distribution is considered then

$$W_3 = \sum_{j=1}^{n2} a_3[R(Y_j)] \tag{2.29}$$

where $a_3(j) = \varphi_3\left[\frac{j}{n+1}\right]$ and $\varphi_3(u)$ is an associated optimal score for such a distribution.
For light-tailed distribution,

$$W_4 = \sum_{j=1}^{n2} a_4[R(Y_j)] \tag{2.30}$$

where $a_4(j) = \varphi_4\left[\frac{j}{n+1}\right]$ and $\varphi_4(u)$ is an associated optimal score.
Using the benchmarks proposed by Hogg et al. (1975) and the computed selector statistic, the following scores will be selected.

Example 2.6.1. ***Median Test***

$$a(j) = \begin{cases} 1, & j > \frac{n+1}{2}, \\ 0, & otherwise \end{cases}$$

Example 2.6.2. ***Right-skewed***

$$a(j) = \begin{cases} j - \frac{n+1}{2} - 1, & j \leq \frac{n+1}{2}, \\ 0, & otherwise \end{cases}$$

Example 2.6.3. ***Light-tailed Symmetric***

$$a(j) = \begin{cases} j - \frac{n+1}{2} - \frac{1}{2}, & j \leq \frac{n+1}{4}, \\ j - n + \frac{n+1}{4} - \frac{1}{2}, & j \geq n - \frac{n+1}{4} + 1, \\ 0, & otherwise \end{cases}$$

Example 2.6.4. ***Moderate Heavy-tailed***

$$a(j) = \begin{cases} j, & 1 \leq j \leq n, \\ 0, & otherwise \end{cases}$$

Table 2.4: **Type of Distribution, Benchmark along with their Selected Score**

Benchmark	Distribution	Score Selected
$Q_2 > 7$	Heavy-tailed symmetric	$\varphi_2(u)$ score (1)
$Q_1 > 2$ and $Q_2 < 7$	Right-skewed	$\varphi_4(u)$ score(2)
$Q_1 \leq 2$ and $Q_2 \leq 2$	Light-tailed symmetric	$\varphi_3(u)$ score (3)
Elsewhere	Moderate heavy-tailed	$\varphi_1(u)$ score(4)

The order statistics of the combined sample of all n values is used to compute the selector statistics (Q_1, Q_2) and is displayed in Table 2.4. For example, if $Q_1 > 2$ and $Q_2 < 7$, the right-skewed distribution is selected. In all other cases (elsewhere), a model for moderate heavy-tailed distributions are chosen. These selection regions are shown in fig 2.1

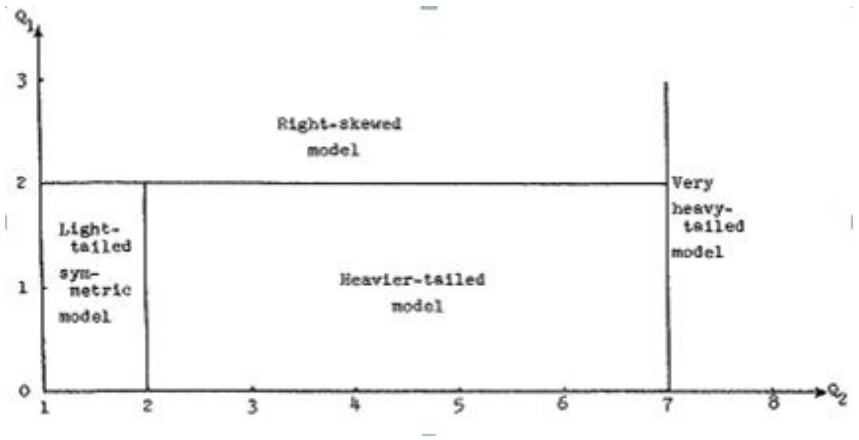

Figure 2.1: Selection Criteria for the HFR Procedure

The selected scores are then used in estimating and testing. As an example to show how the selection method works, a data set is obtained, the measure of skewness is obtained to be $Q_1 = 1.5$, this is an indication that the dataset has a nearly symmetric distribution. The measure of tailweight is obtained next as $Q_2 = 1.5$. These two values are then located on the selection criteria in figure 2.1 to determine the distribution for the dataset and appropriate scores for that distribution, in this case a light tailed symmetric model. If $Q_1 = 1.5$ and $Q_2 = 3.5$, then the dataset had indicated heavier-tailed model, so we would chose the Wilcoxon scores.

2.6.6 Adaptive Rank Tests

Gastwirth (1965) in his study created a path for rank-based adaptive tests. Hogg et al. (1975) in a paper proposed adaptive procedure simply and effectively used a dataset to choose an efficient rank test from among 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 MWW 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) by Jones (1979) and c-sample trend tests Büning (1996). With more recent works on adaptive rank test by Xie and Priebe (2000), Xie and Priebe (2002), Kössler and Kumar (2008), Kössler (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 MWW 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 instead a median test is used. 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.

2.6.7 O'Gorman's Adaptive Test

As a solution to an adaptive rank test 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 to 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.

2.6.8 Büning's Adaptive Test

For model

$$Z = \Delta C_i + e_i$$

where e_i has density f and distribution F, the optimal score, $\varphi_f(u)$ is given by

$$\varphi_f(u) = \frac{f^{'}[F^{-1}(u)]}{f[F^{-1}(u)]}$$

These are optimal in the sense that the corresponding test statistics are asymptotically efficient (Hettmansperger and McKean, 1998). For example, Gastwirth (1965), Randels and Wolfe (1979), Büning (1994), and Büning (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 over their targeted area of distribution (Büning, 2009).

Example 2.6.5 (Gastwirth test (Short Tails))**.**

$$a(k) = \begin{cases} k - \dfrac{N+1}{4}, & k \leq \dfrac{N+1}{4} \\ 0, & \dfrac{N+1}{4} \leq k \leq \dfrac{3(N+1)}{4} \\ k - \dfrac{3(N+1)}{4}, & k > \dfrac{3(N+1)}{4} \end{cases}$$

Example 2.6.6 (Kruskal-Wallis (Medium Tails)).

$$a(k) = k$$

Example 2.6.7 (Hogg Fisher Randles Test (Right Skewed)).

$$a(k) = \begin{cases} k - \dfrac{N+1}{2}, & k \leq \dfrac{N+1}{2} \\ 0, & k > \dfrac{N+1}{2}. \end{cases}$$

For left-skewed distributions interchange the terms $k - \dfrac{N+1}{2}$ and 0 in the definition of the score above.

As an efficient test for long tails, Büning (1996) proposed the Long tail (LT)-test with scores chosen analogously to Huber's Ψ-function referring to M-estimates.

Example 2.6.8 (LT-test (Long Tails)).

$$a(k) = \begin{cases} -\dfrac{N+1}{4}, & k < \dfrac{N+1}{4} \\ k - \dfrac{N+1}{2}, & \dfrac{N+1}{4} \leq k \leq \dfrac{3(N+1)}{4} \\ \dfrac{N+1}{4}, & k > \dfrac{3(N+1)}{4} \end{cases}$$

A selector statistics $S = (\hat{Q}_1, \hat{Q}_2)$ where $\hat{Q}_1$ and $\hat{Q}_2$ are Hogg's measures of skewness and tailweight defined by

$$\hat{Q}_1 = \frac{\bar{U}_{0.05} - \bar{M}_{0.50}}{\bar{M}_{0.50} - \bar{L}_{0.05}}$$

and

$$\bar{Q}_2 = \frac{\bar{U}_{0.05} - \bar{L}_{0.05}}{\bar{U}_{0.50} - \bar{L}_{0.50}}$$

where $\bar{U}_\lambda, (\bar{M}_\lambda, \bar{L}_\lambda)$ denote the average of the largest (middle, smallest) λN order statistics in the combined sample $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \cdots \leq X_{(N)}$. Obviously, $\hat{Q}_1 \geq 0$; $\hat{Q}_1 < 1$ if the data are skewed to the left, $\hat{Q}_1 = 1$ if the data are symmetric and $\hat{Q}_1 > 1$ if the data are skewed to the right. $\hat{Q}_2 \geq 1$, the longer the tails the greater $\hat{Q}_2$, (Hogg, 1974).

Now, the follwing four categories which are based on S are defines as follows:

$$D_1 = \left\{ S \mid 0 \leq \hat{Q}_1 \leq 2; \hat{Q}_2 \leq 2 \right\}$$

$$D_2 = \left\{ S \mid 0 \leq \hat{Q}_1 2; 2 < \hat{Q}_2 < 3 \right\}$$

$$D_3 = \left\{ S \mid \hat{Q}_1 \geq 0; \hat{Q}_2 > 3 \right\}$$

$$D_4 = \left\{ S \mid \hat{Q}_1 > 2; 1 \leq \hat{Q}_2 \leq 3 \right\}$$

Büning (1996) proposed the following adaptive test A:

$$A = \begin{cases} G & if \quad S \in D_1 \\ KW & if \quad S \in D_2 \\ LT & if \quad S \in D_3 \\ HFR & if \quad S \in D_4 \end{cases},$$

where

G, KW, LT and HFR denote Gastwirth, Kruskal Wallis, Long Tail and Hogg, Fisher and Randles tests respectively.

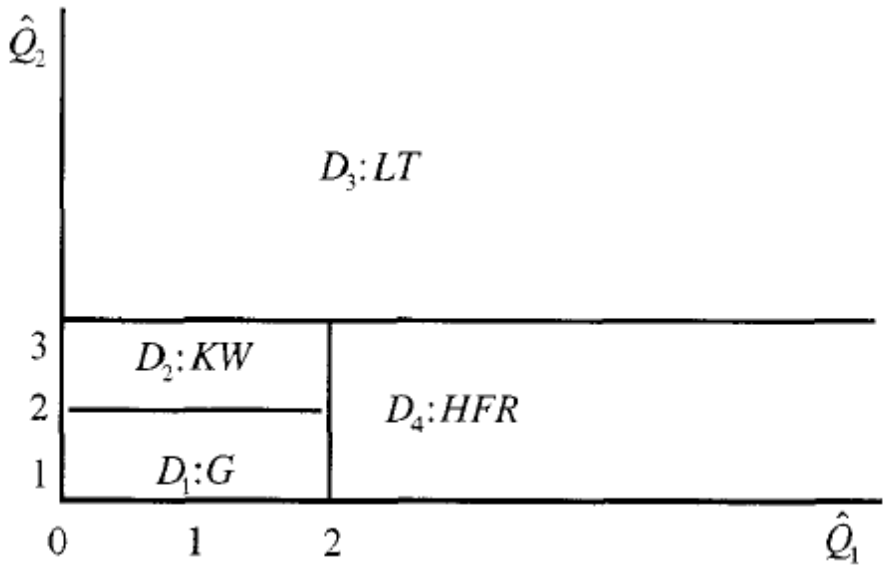

Figure 2.2: Büning's Adaptive Scheme

For example if $\hat{Q}_1 = 1.235$ and $\hat{Q}_2 = 3.3$ then we have to apply LT-test.

The adaptive test above is based on the measures $\hat{Q}_1$ and $\hat{Q}_2$ calculated from the combined sample $X_{(1)}, X_{(2)}, X_{(3)} \ldots, X_{(N)}$ in order to guarantee that the resulting test is distribution-free in the sense of the Lemma. These scores are not standard so we will make use of the nine winsorised scores proposed by Hettmansperger (1984) in this thesis.

2.6.9 Power of Adaptive Tests

Hogg et al. (1975) in a simulation study showed 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 nonadaptive. To demonstrate that the HFR method maintains its significant level, Lemma is used. The test maintains a level of significance less than or equal to α, though the dataset 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 datasets 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 of 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 to the normal, uniform and bimodal error distributions. According to O'Gorman (2012), tests based on ranks, make most sense when the datasets can be ranked. That is the HFR adaptive test, Wilcoxon test among 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.

2.6.10 Okyere's Adaptive Test

Okyere (2011) extended Hogg's adaptive test to Linear Mixed Models. A comparison was made among four adaptive procedures, one parametric procedure and one nonparametric nonadaptive procedure. The parametric procedure considered here is the maximum likelihood (ML) or restricted maximum likelihood estimates (REML). The adaptive procedures such as Hogg-McKean adaptation on sample (HMS), Hogg-McKean adaptation on residuals (HMR), Many ranking (MR) and Hogg-McKean-Ignoring-Center (HMIC) adaptation on sample. For HMS procedure, adaptation is done on the sample from center to center. Meta analysis is applied to formulate an overall test and estimate of the fixed effect parameter. The HMR procedure adapts on residual from the robust Wilcoxon fit. Further analyses are similar to HMS. The HMIC method ignores the centers and adaptation is done on the combined sample. The theory, analysis of testing and estimating, and formal algorithm for HMR, HMS and HMIC were discussed. The fourth adaptive scheme is the rank-based many rankings (MR) fitting procedure. Under this scheme, adaptation is obtained on the center-residuals of the initial fit of the linear mixed model. The nonparametric nonadaptive procedure, called many rankings Wilcoxon (MRW) is similar to the many ranking (MR) except that Wilcoxon score is used for each center.

In his work, he demonstrated that an adaptive procedure is applicable to mixed models under exchangeable errors. He indicated that besides the exchangeability assumption and the classification of skewness and tailweight, no other condition is required. It was discovered that the adaptive scheme he proposed are well behaved over a broad class of distributions ranging from heavy-tailed to light-tailed

distributions and from symmetric distributions to skewed distributions. At normal distribution, based on the simulations and numerical examples, the HMS is as efficient as the ML method. When the distribution of random errors are not normal, his adaptive schemes outperforms ML method. In particular HMS and HMR are highly efficient for symmetric light tailed, right skewed heavy tailed and left skewed heavy tailed distributions. The MR procedure and MRW seems to be conservative. The HMIC method did not perform well among the adaptive procedures. The estimation procedures that have been developed paved way for diagnostics checking of the fitted model.

2.6.11 Some Current Works on Adaptation

Afrifa-Yamoah et al. (2016) presented work on a robust procedure to fit Oneway ANOVA model under adaptation on the observed samples. In their work, the data generated are assumed to have been derived from unknown continuous distributions. The data were used to reveal the underlying distribution by assessing the associated values for skewness and tailweight proposed by Hogg et al. (1975). This initial classifications will determine the specific scores functions on which inferences will be based. Simple adaptive procedures established by Hogg et al. (1975) were used in the estimation of the scores and were classified according to the family of winsorised Wilcoxon scores. Four procedures were considered in their work. Three adaptive procedures and one parametric procedure. The adaptive procedures are the Pure-Hogg where adaptation is done on the samples. They considered adaptation on residuals from the ordinary least square (OLS) and Wilcoxon t. The F-test is the parametric procedure considered. Simulation studies were performed to prove the dominance or otherwise of the adaptive procedures over the parametric procedures over a wide range of continuous distributions.

The findings of the study revealed that although, the F-test displayed superiority in efficiency in symmetric, medium and light tailed distributions, the adaptive test was more efficient in more broader class of continuous distributions. The performance of these test at small sample sizes was of much importance because most sensitive areas of the application of oneway ANOVA model often has low sample usage.

Another study on adaptive robust profile analysis of a longitudinal data has been done by (Okyere et al., 2018). In that study, a statistical test of significance concerning comparison of location of two independent samples situated in a longitudinal data setting was considered to construct adaptive test for testing group and time interaction in profile analysis. The study focused on the two dimensional selector statistics $S = (Q_1^*, Q_2^*)$ where Q_1^* and Q_2^* are respective measures for skewness and tailweight of the unknown distribution function. The nine winsorised scores were considered as the most appropriate set of rank scores for testing group and time interaction. From the simulations and real data analysis, the adaptive appeared to be more efficient than the parametric ANOVA-F test for a class of nonnormal dis-

tributions. The study and that of Büning (2009) among others consistently demonstrate the advantages of adaptive tests over the traditional parametric tests nevertheless, adaptive tests are not frequently used compared to the parametric ones.

Saleem and Sherani (2020) presented paper on Selecting and estimating rank score functions based on residuals for linear mixed models. They indicated in their paper that the rank-based method is a robust estimation method in the presence of outliers and performs as an alternative to Ordinary Least Squares and Restricted Maximum Likelihood Estimate developed for linear models. This estimation method is based on a pseudo norm established on score functions. The rank-based fit could be sufficiently improved by selecting the accurate score function according to the underlying distribution of the error term. In situation where the distribution of the error term is not known, the performance of two selection criteria developed for linear models for a class of error distribution, thus, symmetric, asymmetric, light-tailed to heavytailed distributions was investigated by their paper. The two selection schemes for random intercept multilevel models with cluster-correlated error terms were evaluated. The selection of appropriate score functions is made from a class of suitable score functions. The efficiency of each score function is compared with other score functions by following the recommended shapes of error distributions. All the score functions performed well when group size is 30 or more and the individual sample size is 5, 10, 30 and 50. Some of the score functions such as Bentscores1 and Bentscores3 show minimum standard error among all other score functions even for the smallest sample size and its magnitude reduces as sample size increases. Another criterion for choosing an appropriate score function is Hogg type adaptive scheme. A simulation study is conducted based on the Hogg's adaptive scheme was applied for several shapes of distributions on the multilevel model. The efficiency of the rank-based fit with the selected score function is compared with the Wilcoxon score based on minimum standard error. For the case of right-skewed, moderately heavy-tailed and light-tailed distribution, selected fit from the adaptive scheme is more precise than Wilcoxon fit. For contaminated normal distribution selected fit is more precise in small sample sizes only. In group size 30 or more, the selection of score function does not make a significant change in standard error.

In sample size more than 900, almost for every score function, precision tends to reach 1. The results indicate the significance of sample size at each level. Generally speaking, when the total sample size is around 1000, rank-based fit through selected score function and by Wilcoxon fit produces quite similar standard error of fixed effect estimates. The application of both selection schemes is illustrated through an example of block design with cluster-correlated errors. All the score functions provided through both selection procedures give good results in terms of lower standard error and unbiased estimates for multilevel models compared with restricted maximum likelihood estimate.

The review of adaptive designs have generally indicated the efficiency of the adaptive schemes over

their parametric and nonparametric non-adaptive tests. However, to the best of our knowledge much work has not been done by extending adaptive scheme to the Gauss Markove model. In this study, an adaptive scheme for estimating and testing fixed effects for Gauss Markov model is presented.

CHAPTER THREE

METHODOLOGY

This chapter focuses on the methods of the adaptive procedures for the work. The rank-based test, excchangeable random variables, estimation of the scale parameter, asymptotic relative efficiency and adaptive statistical methods were discussed.

3.1 Rank Based Test

Consider the function

$$\|v\| = \sum_{j=1}^{n} a[R(v_i)]v_i, \qquad v_i \in R^n \tag{3.1}$$

where $a(j)'s$ are the scores such that $a(1) \leq a(2) \leq \cdots \leq a(n))$ and $\sum a(j) = 0$ and $R(v_i)$ denotes the rank of v_i among the $v_1, v_2, v_3, \cdots, v_n$ and the scores at each observed data point is generated by

$$a(i) = \varphi\left(\frac{i}{N+1}\right) \tag{3.2}$$

It is assumed that $a(j) = -a(n+1-j)$.

The rank-based procedures are used due to the fact that they are robust and the overall dispersion function denoted $D(\Delta)$ is convex. It is worth noting that the adaptation is performed at each observed data points, since it has been established that at each observed data point, under H_0 for the model

$$Z_i = v_i\delta + e_i, \quad 1 \leq i \leq n$$

the error measurements are exchangeable.

3.1.1 Norms

A norm is a nongegative function $\|.\|$ defined on R^n such that

- $\|y\| \geq 0$ for all y
- $\|y\| = 0$ if and only if $y = 0$
- $\|ay\| = |a|\|y\|$ for all real a and
- $\|y + z\| \leq \|y\| + \|z\|$

Theorem 3.1.1. *Suppose* $a_j(1) \leq a_j(2) \leq a_j(3) \leq \cdots \leq a_j(n)$, *and* $a(j) = -a(n+1-j)$, *then the function* $\|.\|_\varphi$ *is a pseudo norm*

Theorem 3.1.2 (Pseudo-norm)**.** *Suppose that* $a(1) \leq a(2) \leq \cdots \leq a(n)$, $\sum a(j) = 0$ *and* $a(j) = -a(n+1-j)$. *Then the function* $\|.\|_\varphi$ *is a pseudo-norm if it satisfies the following four conditions:*

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

3.1.2 General Rank Scores

A set of rank scores, equation (3.2), are selected for a nondecreasing score function φ which is standardised as

$$\int_0^1 \varphi(u)du = 0$$

and

$$\int_0^1 \varphi^2(u)du = 1$$

The rank-based estimator of β is given by

$$\hat{\beta}_\varphi = Argmin\|Y - X\beta\| \tag{3.3}$$

where

$$\|v\|_\varphi = \sum_{t=1}^{N} a(R(v_t))v_t, v \in R^N \tag{3.4}$$

A set of rank-based scores is generated by a function $\varphi(u)$ defined on the interval $(0,1)$. It is assumed that $\varphi(u)$ is a square-integrable and, without loss of generality standardised as

$$\int_0^1 \varphi(u)du = 0$$

and

$$\int_0^1 \varphi^2(u)du = 1$$

The generated scores are then

$$a_\varphi(i) = \varphi\left(\frac{i}{n+1}\right) \tag{3.5}$$

Because

$$\int_0^1 \varphi(u)du = 0,$$

we may assume that

$$\sum_{i=1}^{n} a[i] = 0$$

satisfying the pitman regularity.

For example, the Wilcoxon pseudo-norm is generated by the score function

$$\varphi(u) = \sqrt{12}\left(u - \tfrac{1}{2}\right) \tag{3.6}$$

and the sign score is generated by

$$\varphi(u)) = sgn\left(u - \tfrac{1}{2}\right). \tag{3.7}$$

Hájek and Sidák (1967) showed that, in general, the optimal score function is

$$\varphi(u) = \varphi_f(u) = -\frac{f'[F^{-1}(u)]}{f[F^1(u)]} \tag{3.8}$$

The rank test considered in this thesis is of the form

$$T_\varphi = \sum_{i=1}^{n} \varphi\left[\frac{R(Z_i)}{n+1}\right] I(Z_i = Y_i) \tag{3.9}$$

where

$$a_\varphi(i) = \varphi\left(\frac{i}{n+1}\right),$$

$a_\varphi(1), a_\varphi(2), \cdots, a_\varphi(n)$ are scores and φ satisfies the following conditions

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

Since φ is square integrable, we assume

$$\int_0^1 \varphi^2(u)du = 1,$$

see (Hettmansperger and McKean, 1998).

3.1.3 Jaeckel's Dispersion Function

The geometry of rank-based estimation is similar to that of least squares. In rank based regression however, the Euclidean distance is replaced with another measure of distance, the Jaeckel's dispersion

function defined by the rank based estimator of the shift parameter Δ denoted by $\hat{\Delta}$ is given by

$$\hat{\Delta}_\varphi = Argmin\|Z - C\Delta\|_\varphi \tag{3.10}$$

Denoting the negative of the gradient of $\|Z - C\Delta\|$ by $S_\varphi(\Delta)$ then based on equation (3.1)

$$S_\varphi(\Delta) = \sum a_\varphi[R(X_j - \Delta)] \tag{3.11}$$

where $\hat{\Delta}_\varphi$ approximately solves the equation $S_\varphi(\hat{\Delta}_\varphi) = 0$, see (Hettmansperger and McKean, 2011). Thus, for each observed data, under the null hypothesis, the gradient of the rank test statistic is

$$S_\varphi = \sum_{j=1}^{n_2} a_\varphi[R(X_j)] \tag{3.12}$$

Since the test statistic only depends on the ranks of the combined sample it is distribution free under the null hypothesis. Thus

$$E_0(S_\varphi) = 0 \tag{3.13}$$

and

$$\begin{aligned} \sigma_\varphi^2 &= V_0(S_\varphi) \\ \sigma_\varphi^2 &= \frac{n_1 n_2}{n(n-1)} \sum_{i=1}^{n} a^2(i) \end{aligned} \tag{3.14}$$

where the variance can be expressed as

$$\begin{aligned} \sigma_\varphi^2 &= \frac{n_1 n_2}{n(n-1)} \left\{ \tfrac{1}{n} \textstyle\sum_{i=1}^{n} a^2(i) \right\} \\ \sigma_\varphi^2 &= \frac{n_1 n_2}{n(n-1)} \end{aligned} \tag{3.15}$$

the approximation is due to the fact that the term in braces is a Riemann sum of

$$\int_0^1 \varphi^2(u)du = 1$$

and hence converges to 1, a good source of information is found in (Hettmansperger and McKean, 2011).

3.1.4 Asymptotic Distribution and Efficacy of $\hat{\Delta}_\varphi$

To obtain the asymptotic null distribution of S_φ using equation (3.12), it follows then that from equations (2.21) and (2.22) the linear rank statistic is

$$\begin{aligned} S_\varphi &= \sum_{i=1}^{n} c_i a(R(Z_i)) \\ S_\varphi &= \sum_{i=1}^{n} (c_i - \bar{c}) a(\frac{n}{n+1} F_n(Z_i)) \end{aligned} \tag{3.16}$$

where F_n is the empirical distribution function of $Z_1, Z_2, \ldots, Z_n$. The score function is monotonic and square integrable. Now, let T_φ be the random variable defined by

$$T_\varphi = \sum_{i}^{n} (c_i - \bar{c})_\varphi (F(Z_i)). \tag{3.17}$$

Hence, comparing equations (3.16) and (3.17), it implies that T_φ is an approximate of S_φ. Consequently, under H_0 the distribution of T_φ is approximately normal and has the same distribution as S_φ on condition that the second moment of their difference goes to 0, a good source of information is (Hettmansperger and McKean, 2011). That is

$$Var\left[\frac{T_\varphi - S_\varphi}{\sigma_\varphi}\right] \longrightarrow 0;$$

hence, S_φ is asymptotically normal with mean and variance given by the equations (3.13) and (3.14) respectively. Hence an asymptotic level α test of the $H_0 : \Delta = 0$ versus $H_1 : \Delta > 0$ is reject H_0 in favour of H_1 if

$$S_\varphi \geq z_\alpha \sigma_\varphi$$

where σ_φ is defined by (3.14).

It is also assumed that $f(x)$ has a finite Fisher information, This means that f is absolutely continuous, non decreasing and square integrable $\varphi(u)$ such that

$$0 \leq I(f) = \int_0^1 \varphi_f^2(u) du < \infty$$

and

$$lim\left(\frac{n!}{n}\right) = \lambda_i$$

, $0 < \lambda_i < 1, i = 1, 2$ and $\lambda_1 + \lambda_2 = 1$, see page 109 of (Hettmansperger and McKean, 2011). The square integrable is defined as

$$\varphi_f(u) = \frac{-f'(F^{-1}(u))}{f(F^{-1}(u))}$$

where f is uniformly bounded.

Defining the scale parameter τ_φ as

$$\begin{aligned}\tau_\varphi^{-1} &= \int_0^1 \varphi(u)\varphi_f(u)du \\ \tau_\varphi^{-1} &= \int_0^1 \varphi(u)\left[\frac{-f'(F^{-1}(u))}{f(F^{-1}(u))}\right]du\end{aligned} \tag{3.18}$$

Since the error measurements are independently distributed, then for the asymptotic representation of $\hat{\Delta}_\varphi$, the gradient $S_\varphi(\Delta)$ should satisfy the four conditions under subsection 3.1.4 for the observed data, see Hettmansperger and McKean (2011). Hence $S_\varphi(\Delta)$ is non-increasing which satisfies the the first condition. Thus, from equation (3.11)

$$S_\varphi(\Delta) = \sum_{i=1}^{2} \varphi\left[\frac{n_1}{n+1}F_{n_1}(Y_i-\Delta) + \frac{n_2}{n+1}F_{n_2}(Y_i)\right] \tag{3.19}$$

where F_{n_1} and F_{n_2} are the empirical distribution functions of $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ respectively, see (Hettmansperger and McKean, 2011). The second condition of subsection 3.1.3 is satisfied if $E[\frac{1}{n}S_\varphi(\Delta)] = \mu(\Delta)$ and the $\mu^{'} > 0$. Now, from equation (3.19),

$$\begin{aligned}E\left[\frac{1}{n}S_\varphi(\Delta)\right] &\longrightarrow \lambda_2\int_{-\infty}^{\infty}\varphi[\lambda_1F(x)+\lambda_2F(x-\Delta)]f(x-\Delta)dx \\ &= \lambda_2\int_{-\infty}^{\infty}\varphi[\lambda_1F(x+\Delta)+\lambda_2F(x)]f(x)dx \\ E\left[\frac{1}{n}S_\varphi(\Delta)\right] &= \mu_\varphi(\Delta) > 0\end{aligned}$$

Differentiating $\mu_\varphi(\Delta)$ and evaluating at $\Delta = 0$, to obtain an asymptotic efficacy results. This is illustrated as follows;

$$\begin{aligned}\mu_\varphi^{'}(0)|_{\Delta=0} &= \lambda_1\lambda_2\int_{-\infty}^{\infty}\varphi^{'}[F(t)]f^2(t)dt \\ &= \lambda_1\lambda_2\int_{-\infty}^{\infty}\varphi[F(t)]\left[\frac{-f'(t)}{f(t)}\right]f(t)dt \\ &\lambda_1\lambda_1\int_0^1\varphi(u)\varphi_f(u)du \\ \mu_\varphi^{'}(0)|_{\Delta=0} &= \lambda_1\lambda_2\tau_\varphi^{-1} > 0\end{aligned}$$

The second condition of Pitman regular is thus satisfied.

For the third condtion to be met, the asymptotic linearity of $S_\varphi(\Delta)$ is given by

$$\frac{1}{\sqrt{n}}S_\varphi\left(\frac{\delta}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}}S_\varphi(0) - \tau_\varphi^{-1}\lambda_1\lambda_2\delta + O_p(1) \tag{3.20}$$

uniformly for $|\delta| \leq B$, where $B > 0$ and τ_φ is as defined in equation (3.18). Finally, from condition 4 of

subsection 3.1.3,

$$\frac{1}{\sqrt{n}}\frac{S_\varphi(0)}{\sqrt{\lambda_1\lambda_2}} \sim N(0.1).$$

Thus, the efficacy of the test based on S_φ found on page 111 of Hettmansperger and McKean (2011) is given by

$$\begin{aligned} c_\varphi &= \frac{\tau_\varphi^{-1}\lambda_1\lambda_2}{\sqrt{\lambda_1\lambda_2}} \\ c_\varphi &= \tau_\varphi^{-1}\sqrt{\lambda_1\lambda_2} \end{aligned} \tag{3.21}$$

since the asymptotic efficiency is given by $c_\varphi = \dfrac{\mu^{'}(\Delta)}{\sigma(0)}$.

Theorem 3.1.3. *Suppose $S_\varphi(\Delta)$ is a Pitman regular with efficacy c_φ, then $\sqrt{n}(\hat{\Delta}-\Delta)$ converges in distribution to $Z \sim N\left(0, \dfrac{1}{c_\varphi^2}\right)$.*

Thus, since the estimate $\hat{\Delta}_\varphi$ solves the equation $S_\varphi(\hat{\Delta}) = 0$, then, based on the Pitman regulrity and theorem 3.1.4, the asymptotic distribution of $\hat{\Delta}$ is given by

$$\sqrt{n}(\hat{\Delta}-\Delta) \xrightarrow{D} N(0, \tau_\varphi^2(\lambda_1\lambda_2)^{-1}) \tag{3.22}$$

By using equation (3.20) and $T_\varphi = 0$ to approximate $S_\varphi = 0$, we have the following result:

$$\sqrt{n}\hat{\Delta} = \frac{\tau_\varphi}{\lambda_1\lambda_2}\frac{1}{\sqrt{n}}T_\varphi(0) + O_\rho(1) \tag{3.23}$$

We want to select scores such that the efficacy c_φ (3.21), is as large as possible, or equivalently such that the asymptotic variance of $\hat{\Delta}$ is as small as possible.

3.2 Exchangeable Random Variables

Error measurements or observations are said to be exchangeable if they are considered independent and identically distributed (i.i.d), or if they are jointly normal with identical covariances, see Good (2002). Suppose the two-sample location problem is considered, where $X_i^{'}s$ are random sample, i.i.d with continuous distribution function $F(x)$ and $Y_i^{'}s$ being random sample, i.i.d with distribution function $F(x-\Delta)$ with the hypothesis defined as $H_0 : \Delta = 0$ versus $H_1 : \Delta \neq 0$, then the exact test for H can be obtained by transforming the variable by subtracting 0 from each of the $X_i^{'}s$ and Δ from each of the $Y_i^{'}s$. Thus, "a set of random variables X will be said to be transformably exchangeable if there exists a transformation(measurable transformation) T, such that $T\mathbf{X}$ is exchangeable" see Good (2002).

Theorem 3.2.1. *An infinite sequence of random variables $(Y_1, Y_2, \ldots, Y_n, \ldots)$ is said to be infinitely exchangeable under probability measure P, if the joint probability of every finite subsequence $(Y_{n1}, Y_{n2}, \ldots, Y_{nk})$*

satisfies $(Y_{n1}, Y_{n2}, \ldots, Y_{nk}) \stackrel{d}{=} (Y_{\tau(ni)}, Y_{\tau(n2)}, \ldots Y_{\tau(nk)})$ *for all permutations* τ *defined on the set* $\{1, 2, 3, \ldots, k\}$

An infinite sequence of random variables is said to be infinitely exchangeable, if every finite sequence of its variables (events) is exchangeable, see Mahmoud (2008).

Definition 3.2.1

The random variables $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n, \ldots$ are exchangeable if the $n!$ permutations $\varepsilon_{k1}, \varepsilon_{k2}, \ldots, \varepsilon_{k_{ni}}$ have the same n_i-dimensional probability distribution. The variables of of an infinite sequence ε_{n_i} are exchangeable if $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n_i}$ are exchangeable for each n_i.

Let's consider the white-blue Pólya's urn. An urn containing b blue balls and w white balls. A ball drawn at random is replaced together with k balls of the same colour. This process is repeated infinitely such that

$$W_i = \begin{cases} 1, & if\ the\ ball\ in\ the\ i^{th}\ draw\ is\ white \\ 0, & otherwise. \end{cases}$$

Thus, for the first three picks, the probabilities of all sequences with only 1 blue colour is given by;

$$P(W_1 = 1, W_2 = 1, W_3 = 0) = \frac{1}{2} \times \frac{2}{3} \times \frac{1}{4} = \frac{1}{12}$$
$$P(W_1 = 1, W_2 = 0, W_3 = 1) = \frac{1}{2} \times \frac{1}{3} \times \frac{2}{4} = \frac{1}{12}$$
$$P(W_1 = 0, W_2 = 1, W_3 = 1) = \frac{1}{2} \times \frac{1}{3} \times \frac{2}{4} = \frac{1}{12}$$

Similarly, the probabilities of all sequences of obtaining only one white ball in the

first three picks is given by

$$\begin{aligned} P(W_1 = 1, W_2 = 0, W_3 = 0) &= P(W_1 = 0, W_2 = 1, W_3 = 0) \\ &= P(W_1 = 0; W_2 = 0; W_3 = 1) \\ &= \frac{1}{12} \end{aligned}$$

Finally the probabilities of all sequences of the same colour in all three picks is

$$\begin{aligned} P(W_1 = 1, W_2 = 1, W_3 = 1) &= P(W_1 = 0, W_2 = 0, W_3 = 0) \\ &= \frac{1}{4} \end{aligned}$$

Suppose a ball drawn at random is replaced together with k balls of the same colour and this process is repeated infinitely, then for the first five picks, we have $P(1, 0, 1, 1, 0) = P(1, 1, 0, 1, 0) = P(1, 1, 1, 0, 0)$

$$P(1, 0, 1, 1, 0) = \frac{w}{w+b} \times \frac{b}{w+b+k} \times \frac{w+k}{w+b+2k} \times \frac{w+2k}{w+b+3k} \times \frac{b+k}{w+b+4k}$$
$$P(1, 1, 0, 1, 0) = \frac{w}{w+b} \times \frac{w+k}{w+b+k} \times \frac{b}{w+b+2k} \times \frac{w+2k}{w+b+3k} \times \frac{b+2k}{w+b+4k}$$

$$P(1,1,1,0,0) = \frac{w}{w+b} \times \frac{w+k}{w+b+k} \times \frac{w+2k}{w+b+2k} \times \frac{b}{w+b+3k} \times \frac{b+k}{w+b+4k}$$

It can be seen that for all 5! permutations, the same distribution is obtained.

Theorem 3.2.2 (Lemma). *If $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n_i}, \ldots$ are independent and identically distributed, then they are exchangeable, but not conversely.*

In the example of Pölya's Urn, the $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_5, \ldots$ are not independent but are exchangeable.

Generally, for $W_1, W_2, \ldots, W_n$ being an arbitrarily large but fixed $n \geq 2$ and there is a total of k among the indicators that are 1, and the rest 0 occurring at positions $1 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_k \leq n$. Then, the probability of this event is

$$\begin{aligned} P(W_1 = 0, \ldots, W_{\tau 1-1} = 0, W_{\tau 1} = 1, W_{\tau 1+1} = 0, \ldots, W_{\tau k-1} = 0, W_{\tau k} = 1, W_n = 0) \\ = \frac{1}{2} \times \frac{2}{3} \times \cdots \times \frac{\tau_1 - 1}{\tau} \times \frac{1}{\tau+1} \times \frac{\tau_1}{\tau_1+2} \times \frac{\tau_1+1}{\tau_1+3} \times \cdots \times \frac{\tau_2-2}{\tau_2} \times \\ \frac{2}{\tau_2+1} \times \frac{\tau_2-1}{\tau_2+2} \times \cdots \times \frac{\tau_k-k}{\tau_k} \times \frac{k}{\tau_{k+1}} \times \frac{\tau_k-k+1}{\tau_k+2} \times \cdots \times \frac{n-k}{n+1} \\ = \frac{k!(n-k)!}{(n+1)!} \end{aligned} \tag{3.24}$$

Conversely,

$$\begin{aligned} P[W_1 = 1, W_2 = 1, \ldots, W_k = 1, W_{k+1} = 0, W_{k+2} = 0, \ldots, W_n = 0] \\ = \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \cdots \times \frac{k}{k+1} \times \frac{1}{k+2} \times \frac{2}{k+3} \times \cdots \times \frac{n-k}{n+1} \\ = \frac{k!(n-k)!}{(n+1)!} \end{aligned} \tag{3.25}$$

Comparing models (3.24) and (3.25), the probability of drawing k white balls in n draws is independent of where in the sequence the white balls were drawn, see Mahmoud (2008). The characteristics of interest is that there are k white balls and that all the sequences with the same number of balls have the same probability.

Theorem 3.2.3 (De Finetti's). *Let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n, \ldots$ be an infinite sequence of random variables. Suppose that for any n, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ is exchangeable: $P(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = P(\varepsilon_{\tau_1}, \varepsilon_{\tau_2}, \ldots, \varepsilon_{\tau_n})$ for permutations τ of $(1, 2, 3, \ldots, n)$. Then*

$$P(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \int \{\Pi_{i=1}^n P(\varepsilon_{i/\theta})\} P(\theta) d\theta.$$

for some parameter θ, some prior distribution of θ and some sampling model $P(\varepsilon\theta)$. The prior and sampling model depend on te form of the belief model $P(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$. θ is the parameter that describes te conditions under which the random variables are generated.

The implication here is that any probability measure describing an exchangeable sequence that is infinite can be expressed as a mixture of independent and identically distributed (iid) probability measures.

3.3 Estimation of the Scale Parameters

For a specific distribution, the optimum scores is selected such that the asymptotic efficacy C_φ is "as large as possible" or equivalently the asymptotic variance $\hat{\Delta}_\varphi$ "is small as possible", a good source of information is Hettmansperger and McKean (2011).

The scale parameter τ_φ is defined as

$$\begin{aligned}\tau_\varphi^{-1} &= \int_0^1 \varphi(u)\varphi_f(u)du \\ &= \int_0^1 \varphi(u)\left\{-\frac{f^{'}F[^{-1}(u)]}{f[F^{-1}(u)]}\right\}du\end{aligned} \tag{3.26}$$

$\varphi_f(u)$ is referred to as the optimal score function. If $\hat{\Delta}$ is an estimator whose variance achieves the Cramer-Rao lower bound ($\forall\ \hat{\Delta}$), it is called efficient.

That is;

$$var(\hat{\Delta}) \geq \frac{\left[\frac{d}{d\Delta}E(\hat{\Delta})\right]^2}{nI(\Delta)} \tag{3.27}$$

Thus for the jth observation in the k^{th} sample, select scores with efficacies as large as possible or with asymptotic variance τ_φ as small as possible, see Hettmansperger and McKean (2011).

The proof of equation (3.26) is shown below

$$\begin{aligned}\tau_\varphi^{-1} &= \int_0^1 \varphi(u)\varphi_f(u)du \\ &= \int_0^1 \varphi(u)\left\{-\frac{f^{'}F[^{-1}(u)]}{f[F^{-1}(u)]}\right\}du \\ &= \frac{\int_0^1 \varphi(u)\varphi_f(u)du}{\sqrt{\int_0^1 \varphi_f^2(u)du}\sqrt{\int_0^1 \varphi(u)du}}\sqrt{\int_0^1 \varphi_f(u)du} \\ &= \left\{\frac{\int_0^1 \varphi(u)\varphi_f(u)du}{\sqrt{\int_0^1 \varphi_f^2(u)du \times 1}}\right\}\sqrt{\int_0^1 \varphi_f(u)du} \\ &= \rho\sqrt{\int_0^1 \varphi_f^2(u)du} \\ \tau_\varphi^{-1} &= \rho\sqrt{I(f)}\end{aligned}$$

where ρ is the correlation coefficient and $\int_0^1 \varphi_f^2(u)du$ is the Fisher Information denoted by $I(f)$. Hence, by the Cramér-Rao lowr bound, the smallest asymptotic variance obtainable is asymptotically efficient. Thus, to maximise τ_φ the score function is chosen such that $\rho = 1$ and $\varphi(u) = \varphi_f(u)$, see (Hettmansperger and McKean, 2011). Since $\hat{\Delta_\varphi}$ is location and scale equivalent, only the form $f(x)$ is needed. Therefore

$$\tau_\varphi = \frac{1}{\sqrt{I(f)}}.$$

The resulting estimate $\hat{\Delta_\varphi}$ is asymptotically efficient, implying that τ_i is a consistent estimator for τ. Hence for an estimator τ, the average of these estimators of tthe data is evaluated resulting in

$$\tau = \frac{1}{j}\sum_{i=1}^{j} \tau_i$$

which is consistent for τ, see (Rashid et al., 2012)

3.4 Asymptotic Relative Efficiency

For any two test statistics that are consistent, P and Q, of any hypothesis H_0, 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_0, 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. $ARE(P,Q) \in (0,\infty)$. When $ARE(P,Q) \in (0,1)$ 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 $ARE(P,Q) \in (1,+\infty)$, (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

$$ARE(T_P, T_Q/f) = \frac{AE(T_P, T_Q/f)}{AE(T_Q/f)} \tag{3.28}$$

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

The asymptotic relative efficiency between two tests or estimates based on the score functions $\varphi_1(u)$ and $\varphi_2(u)$ or one function relative to other score function is defined by

$$e(\varphi_1, \varphi_2) = \frac{C_{\varphi_1}^2}{C_{\varphi_2}^2} = \frac{\tau_{\varphi_2}^2}{\tau_{\varphi_1}^2} \tag{3.29}$$

where C_{φ_1} and C_{φ_2} are respectively the efficacies of the two estimates and $\tau_{\varphi_i}, i = 1, 2$ are scale parameters of the two score functions.

3.5 Adaptive Scheme

The Adaptive scheme is a two step procedure. The data is first examined and classified based on skewness and tailweight from a class of continuous distributions. This is done through a selector statistic. Second, by the classification, a test statistic which is independent of the selector statistic is selected and a test conducted. This two-staged adaptive procedure maintains the level α for all continuous distribution functions.

The procedures are as follows:

1 Let $Y_{(1)}, Y_{(2)} \cdots, Y_{(N)}$ be the combined ordered residuals of independent random samples $Y_{11}, Y_{12}, \cdots, Y_{1n_1}, Y_{21},$ from continuous distribution function $f(t)$ with some amount of variations denoted by Δ among the samples, that is $f(t - \Delta)$

2 Adaptation is based on the residuals after an initial fit of the winsorised Wilcoxon scores on the observed samples has been done.

3 The nine winsorised scores are incorporated and extended to the HFR test in the context of Gauss Markov model.

4 Residuals will be obtained from an initial R-fit using Wilcoxon scores.

5 The residuals will be ordered and their distribution will be classified by using both Q_1^* and Q_2^*.

6 Once the distribution of these residuals is classified, a corresponding score function will be selected based on the scores presented in table 3.1.

7 After the selection of the score function, the model will be refit using this selected score function and an inference such as estimates of parameters can be obtained.

3.5.1 Selector Statistics

The selector statistics aids in selecting score function, $S = (Q_1^*, Q_2^*)$ where Q_1^* and Q_2^* are the respective measures of skewness and tailweight. The measures of skewness and tail weight are defined respectively by

$$Q_1^* = \frac{(m(0.95, 0) - m(0.25, 0.25))}{((m(0.25, 0.25) - m(0, 0.95))} \tag{3.30}$$

$$Q_2^* = \frac{((m(0.95, 0) - m(0, 95))}{((m(0.5, 0) - m(0, 0.5))} \tag{3.31}$$

where $m(\alpha_1, \alpha_2) = \frac{1}{h}\sum_{i=t_1+1}^{n-t_2} Z_{(i)}$ and

$Z_i's$ are ordered residuals from an initial fit

$t_1 = [n\alpha_1]$

$t_2 = [n\alpha_2]$

$[x]$ denotes the smallest integer greater than x

$h = n - t_1 - t_2$

We want to adapt on residuals, so the combined ordered residuals from an initial fit is used. The measures of tail weight and skewness of the residuals are obtained by using Q_1^* and Q_2^* respectively.

In this thesis, the benchmarks proposed by Al-Shomrani (2003) as cited in Okyere (2011) for the cut off values were used. These benchmarks depend on the sample size n.This is a modified version of Hogg et al. (1975). However, as $n \to \infty$, the measures converge to those proposed by Hogg et al. (1975).

For Q_1^*,

$$\begin{aligned} \text{Lower cut off} &= 0.36 + \frac{0.68}{n} \\ \text{Upper cut off} &= 2.73 - \frac{3.72}{n} \end{aligned} \tag{3.32}$$

For Q_2^*,

if $n < 25$,

$$\begin{aligned} \text{Lower cut off} &= 2.17 - \frac{3.01}{n} \\ \text{Upper cut off} &= 2.63 - \frac{3.94}{n} \end{aligned} \tag{3.33}$$

If $n \geqslant 25$, then

$$\begin{aligned} \text{Lower cut off} &= 2.24 - \frac{4.68}{n} \\ \text{Upper cut off} &= 2.95 - \frac{9.37}{n} \end{aligned} \tag{3.34}$$

3.5.2 Winsorisation

Winsorisation is the transformation of statistics by limiting extreme values in the statistical data to reduce the effect of possibly spurious outliers (Winsor et al., 1947). The distribution of many statistics can be heavily influenced by outliers. A typical strategy is to set all outliers to a specified percentile of the data; for example, a 90% winsorisation would see all data below the 5th percentile set to the 5th percentile, and data above the 95th percentile set to the 95th percentile. Winsorised estimators are usually more robust to outliers than their more standard forms.

Consider the data set consisting of

$92, 19, \mathbf{101}, 58, \mathbf{1053}, 91, 26, 78, 10, 13, \mathbf{-40}, \mathbf{101}, 86, 85, 15, 89, 89, 28, -5, 41$

$(N = 20, mean = 101.5)$ The data is ordered as follows:

-40, $-5, 10, 13, 15, 19, 26, 28, 41, 58, 78, 85, 86, 89, 89, 91, 92,$ **101**, **101**, **1053**

$(N = 20, mean = 101.5)$

For the given data the 5th percentile lies between -40 and -5, while the data above the 95th percentile lies between 101 and 1053. Then a 90% winsorisation would result in the following:

-5, **-5**, $10, 13, 15, 19, 26, 28, 41, 58, 78, 85, 86, 89, 89, 91, 92,$ **101**, **101**, **101**

(N = 20, mean = 55.65)

3.5.3 Distinction between Winsorisation and Trimming

Winsorising is not equivalent to simply excluding data which is a simpler procedure, called trimming or truncation. In a trimmed estimator, the extreme values are discarded. In a winsorised estimator, the extreme values are instead replaced by certain percentiles. Thus a winsorized mean is not the the same a trimmed mean. For instance,10% trimmed mean of the data will be obtained from :

$-5, 10, 13, 15, 19, 26, 28, 41, 58, 78, 85, 86, 89, 89, 91, 92, 101, 101$

(N = 18, mean = 56.5)

The nine Winsorised scores proposed by Hettmansperger (1984) were considered the most appropriate set of rank scores for testing hypothesis. These could be classified into four generic scores. Thus,

$$\varphi_I(u)) = \begin{cases} s_3, & u > s_1 \\ s_3 + \frac{s_3 - s_2}{s_1}(u - s_1), & otherwise \end{cases}$$

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

$$\varphi_{III}(u) = \begin{cases} s_2, & u < s_1 \\ s_3 + \frac{s_2 - s_3}{s_1 - 1}(u - 1), & otherwise \end{cases}$$

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

where s_1, s_2, s_3,s_4 and s_5 are parameters and

$$a_i(t) = \varphi_i\left(\frac{t}{n+1}\right).$$

Table 3.1 shows the distributions and scores with their corresponding parameters.

Table 3.1: **Winsorised Scores**

Skewness	Tail Weight	Score Function
Left	Light	$\varphi_{LL} = \varphi_{II}$, with parameters $(s_1 = .15, s_2 = 0.65, s_3 = -1), s_4 = 2$
Left	Medium	$\varphi_{LM} = \varphi_{III}$, with parameters $(s_1 = .3, s_2 = -1, s_3 = 2.0)$
Left	Heavy	$\varphi_{LH} = \varphi_{III}$, with parameters $(s_1 = .5, s_2 = -1, s_3 = 2.0)$
Symmetric	Light	$\varphi_{SL} = \varphi_{II}$, with parameters $(s_1 = .25, s_2 = .75, s_3 = -1, s_4 = 1)$
Symmetric	Medium	Wilcoxon Scores, $\varphi_{SM} = \sqrt{12}[u - \frac{1}{2}]$
Symmetric	Heavy	$\varphi_{SH} = \varphi_{IV}$, with parameters $(s_1 = .25, s_2 = .75, s_3 - 1, s_4 = 1)$
Right	Light	$\varphi_{RL} = \varphi_{II}$, with parameters $(s_1 = .9, s_2 = -2, s_3 = 1)$
Right	Medium	$\varphi_{RM} = \varphi_{II}$, with parameters $(s_1 = .7, s_2 = -2, s_3 = 1)$
Right	Heavy	$\varphi_{RH} = \varphi_{I}$, with parameters $(s_1 = .5, s_2 = -2, s_3 = 1)$

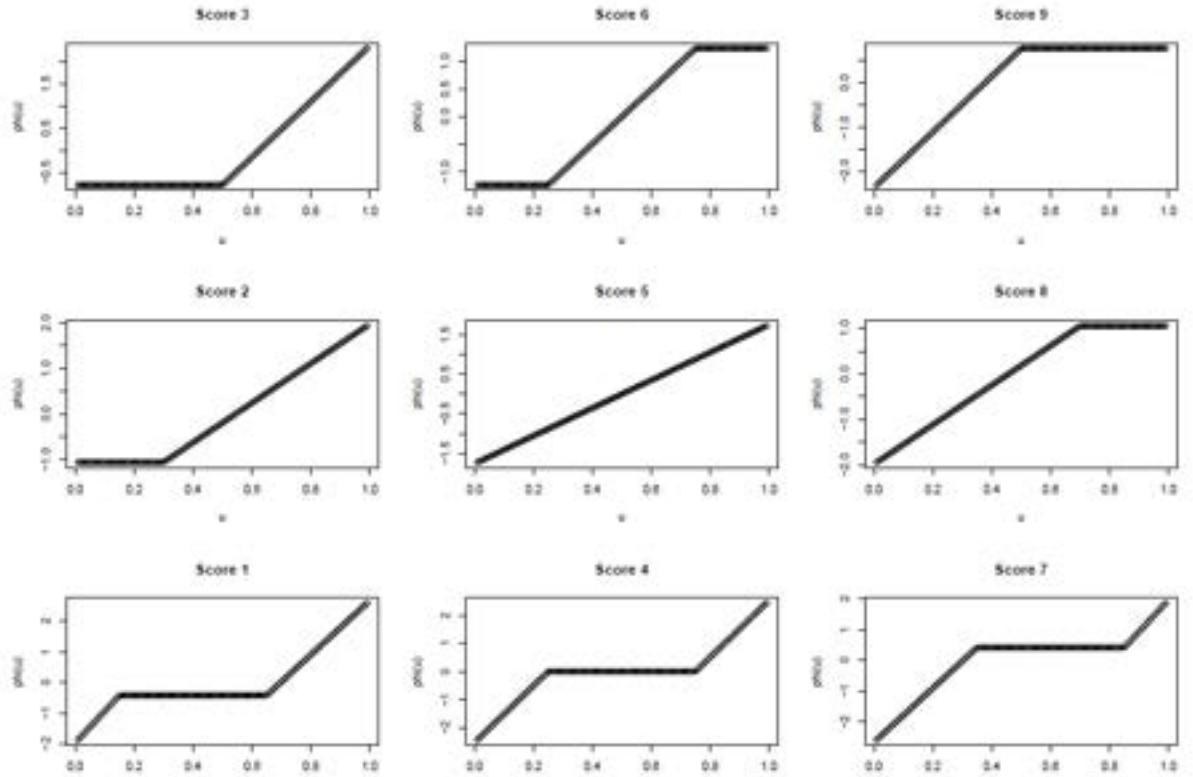

Figure 3.1: Plots of the Nine Winsorised Scores

In the case of adaptation on residuals, initial fit is done. Then the residuals are used for the adaptation. As shown in fig 3.1, $1-3$ represent Left skewed; Left-tailed (LL), Medium-tailed (LM), and Heavy-tailed (LH). $4-6$ are scores for Symmetric with various tailweights. Light-tailed (SL), Medium-tailed (SM) and Heavy-tailed (SH). Finally $7-9$ represent scores for Right skewed with various tail weights; Light-taied (RL), Medium-tailed (RM) and Heavy-tailed (RH).

3.5.4 Adaptive Test and Test Statistics

Let's define the regions based on selector statistic S corresponding to skewness and tailweight of a distribution. The nine regions which depend on the selector statistics $S = (Q_1^*, Q_2^*)$ are defined by:

$D_1 = LL = Q_1^* < \hat{Q}_{1l}^*,\ \ Q_2^* < \hat{Q}_{2l}^*$

$D_2 = LM = Q_1^* < \hat{Q}_{1l}^*,\ \ \hat{Q}_{2l}^* < Q_2^* < \hat{Q}_{2u}^*$

$D_3 = LH = Q_1^* < \hat{Q}_{1l}^*,\ \ Q_2^* > \hat{Q}_{2u}^*$

$D_4 = SL = \hat{Q}_{1l}^* < Q_1^* < \hat{Q}_{1u}^*,\ \ Q_2^* < \hat{Q}_{2l}^*$

$D_5 = SM = \hat{Q}_{1l}^* < Q_1^* < \hat{Q}_{1u}^*,\ \ \hat{Q}_{2l}^* < Q_2^* < \hat{Q}_{2u}^*$

$D_6 = SH = \hat{Q}_{1l}^* < Q_1^* < \hat{Q}_{1u}^*,\ \ Q_2^* > \hat{Q}_{2u}^*$

$D_7 = RL = Q_1^* > \hat{Q}_{1u},\ \ Q_2^* < \hat{Q}_{2l}$

$D_8 = RM = Q_1^* > \hat{Q}_{1u}^*,\ \ \hat{Q}_{2l}^* < Q_2^* < \hat{Q}_{2u}^*$

$D_9 = RH = Q_1^* > \hat{Q}_{1u}^*,\ \ Q_2^* > \hat{Q}_{2u}^*$

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, see Table 3.1 for distributions with their classifications. The regions are shown on figure 3.2.

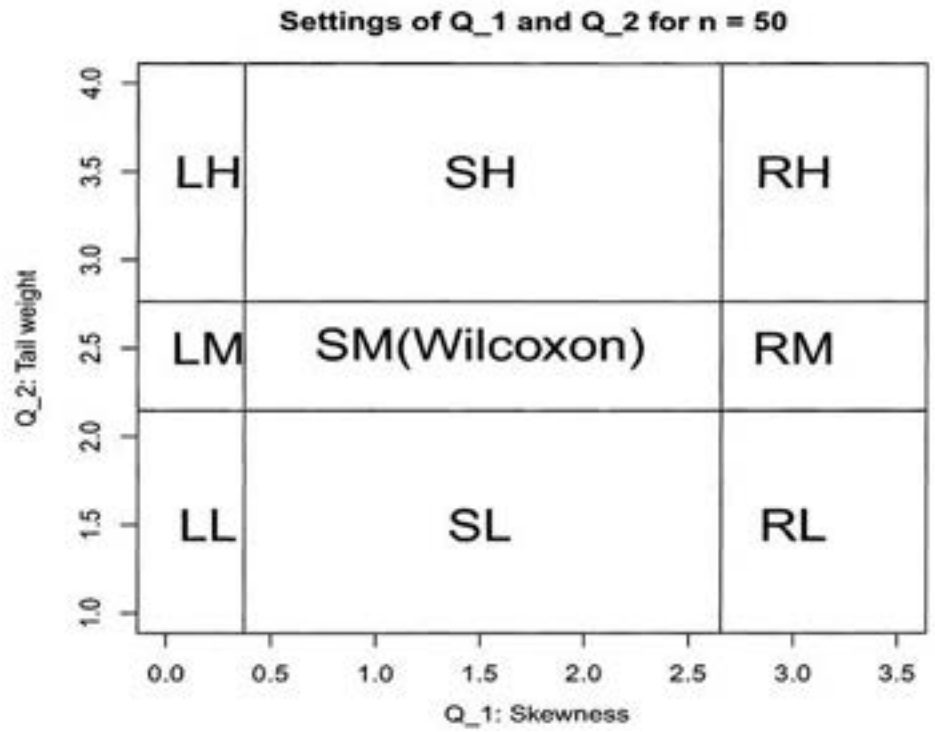

Figure 3.2: Regions of Nine Winsorised scores

Let D_k and φ_k be a region and score selected respectively, with $k = 1, 2, ..., 9$. Then the adaptive test,

$AD(S,\varphi)$, is

$$AD(S,\varphi) = T_{\varphi_k}, \quad S \in D_k \tag{3.35}$$

where

$$T_{\varphi_k}(\Delta) = \sum_{i=1}^{n_2} a_{\varphi_k}[R(y_i - \Delta)] \tag{3.36}$$

is a test statistics based on the ranks and score, φ_k, associated with region D_k and hence distribution-free. Under H_0, the mean of $T_{\varphi_k}(\Delta)$ is zero. Thus,

$$\begin{aligned} E_{H_0}[T_{\varphi k}] &= \sum_{i=1}^{n_2} E_{H_0}[a_{\varphi k}(R(y_i))] \\ &= \sum_{i}^{n_2}\sum_{j=1}^{n} a_{\varphi k}(j)\frac{1}{n} \\ E_{H_0}[T_{\varphi k}] &= 0 \end{aligned}$$

because the ranks of $y_i's$ are uniform on the integers $1, 2, 3, \cdots, n$ and

$$\sum_{j=1}^{n} a_{\varphi k}(j) = 0.$$

Since $E_{H_0}[T_{\varphi k}] = 0$, the variance of $T_{\varphi k}$ is obtained as follows.

$$\begin{aligned} Var_{H_0}[T_{\varphi k}] &= E_{H_0}[T_{\varphi k}^2] \\ &= \sum_{i=1}^{n_2}\sum_{i'=1}^{n_2} E_{H_0}[a_{\varphi k}(R(y_i)a_{\varphi k}(R(y_{i'}))] \\ &= \sum_{i=1}^{n_2} a_{\varphi k}^2(R(y_i)) + \sum\sum E_{H_0}[a_{\varphi k}(R(y_i))a_{\varphi k}(R(y_{i'}))] \\ &= \left[\frac{n_2}{n} - \frac{n_2(n_2-1)}{n(n-1)}\right] s_a^2 \\ Var_{H_0}[T_{\varphi k}] &= \frac{n_1 n_2}{n(n-1)} s_a^2 \end{aligned}$$

where $E_{H_0}[a_{\varphi k}^2(R(y_i))] = \frac{1}{n}s_a^2$, see Okyere (2011) for more details. From literature, $AD(S,\varphi)$ is asymptotically distribution-free. This is because the selector statistic S is based on the order statistics only, $T_{\varphi k}$-statistics is based on the ranks only and asymptotically critical values are used. Thus for the region D_k, the corresponding asymptotic decision rule at level α is reject H_0 if

$$\left|\frac{T_{\varphi k}}{\sqrt{Var_{H_0}(T_{\varphi k})}}\right| \geq z_{\frac{\alpha}{2}}$$

3.5.5 Overall Test Statistic of Adaptation on Residual

Under H_0, it is assumed that the errors in equation (2.14) are exchangeable, thus the order statistics of the combined residual at each time point are sufficient and complete, (Okyere, 2011). In order to obtain the test for the hypothesis after the appropriate scores had been selected, the test statistics for the j time point is developed.

Let φ_{kj} be the score selected at the j^{th} time point and falls in region k, then the test statistic for that time point is

$$T_{\varphi kj} = \sum_{i=1}^{n_h} a_j \left[R_j(Y_i^{(j)}) \right] \tag{3.37}$$

where $T_{\varphi kj}$ has mean

$$E\left[T_{\varphi kj}\right] = 0$$

and variance

$$var\left[T_{\varphi kj}\right] = \frac{n_1 n_h}{n-1} \sum_{l=1}^{n} a_k^2(l)$$

see Okyere (2011) for details.

The test

$$Z = \frac{T_{\varphi kj}}{\sqrt{var\left[T_{\varphi kj}\right]}} \tag{3.38}$$

is asymptotically standard normal and distribution free. Hence we pool the test statistic over time points to obtain the overall test. Thus under H_0, the overall test statistic, T is

$$\begin{aligned} T &= \sum_{j=1}^{t} T_{\varphi hj} \\ T &= \sum_{j=1}^{t} \sum_{i=1}^{n_h} a_j \left[R_j(Y_i^{(j)}) \right] \end{aligned} \tag{3.39}$$

which has asymptotic distribution, $N(0,t)$. Hence for the test

$$H_0 : \Delta = 0$$

$$H_1 : \Delta \neq 0,$$

we have to reject H_0 in favour of H_1 if,

$$T = \frac{\sum_{j=1}^{t} Z_j}{\sqrt{t}} > Z_{\frac{\alpha}{2}}$$

One of the challenges with this test statistic is that the $Z_j's$ may cancel out when summing them over time points so we consider $|Z_j|$. Thus we reject the H_0 if

$$|T| = \left| \frac{\sum_{j=1}^{t} Z_j}{\sqrt{t}} \right| > Z_{\frac{\alpha}{2}} \tag{3.40}$$

CHAPTER FOUR

RESULTS AND DISCUSSION

This chapter focuses on simulation of some distributions and analysis of real data to confirm or otherwise of the efficiency of the performance of the parametric tests and our adaptive scheme.

4.1 Simulation Results

Simulation for distributions with uncorrelated erros were conducted for Normal (Norm), Contaminated Normal (CNorm), Logistic (Logis), Laplace (Lap), Lognormal (LNorm), Exponential (Exp), Cauchy (Cau), Weibull (Wei), Mixture (Mixt) and Pareto (Par) distributions. For correlated errors, simulation was performed for Normal, Laplace and Cauchy distributions. 10,000 simulations were conducted for different sample sizes for each of the distributions to confirm or otherwise of the underlying error distributions.

4.1.1 Classification Table

Table 4.1 displays the classification of (Q_1^*, Q_2^*) of distributions with uncorrelated errors for sample size $n_1 = n_2 = n_3 = 100$ for each of the distributions, where Q_1^* and Q_2^* are respective measures of skewness and tailweight.

Table 4.1: **Skewness and Tailweight Classification of Adaptive Test**

Sel Stats	Norm	CNorm	Logis	Lap	LNorm	Exp	Cau	Wei	Mixt	Par
LH	0	0	0	0	0	0	1702	0	1507	0
LL	0	0	0	0	0	0	0	0	0	0
LM	0	0	0	0	0	0	0	0	0	0
RH	0	0	0	0	9939	2958	1508	2939	1409	10000
RL	0	0	0	0	0	0	0	0	0	0
RM	0	0	0	0	61	7042	0	7058	0	0
SH	5	9	2548	9898	0	0	6790	0	7083	0
SL	1	1	0	0	0	0	0	0	0	0
SM	9994	9990	7452	102	0	0	0	3	1	0

The Normal, Contaminated Normal, Logistic, Laplace, Cauchy and Mixture of distributions were all symmetric but have different tail weights. The Lognormal and Pareto distributions were classified as right skewed and heavy tailed. The Exponential and Weibull distributions were identified as right skewed and medium tailed distributions. It is clear from Table 4.1 that as sample size n increases the skewness and tail weight of the various distributions are identified. Equal sample size of 100 was used for each of the distributions. It is worth noting that the normal and contaminated normal distributions were

correctly classified to be symmetric and medium tailed by 99.94% and 99.90% respectively. The Laplace distribution was correctly classified as symmetric and heavy-tailed by 98.98% and the Lognormal distribution was 98.93% correctly classified as right skewed with heavy tail. The Pareto distribution on the other hand was 100% classified as a right skewed and heavy tailed distribution.

4.2 Simulation Results for Distributions with Uncorrelated Errors

This section presents the results of 10,000 sumulations for the various continuous distributions under uncorrelated errors. A comparison of the parametric F-test and adaptive test was made. The test statistic of F is labelled (value) and the residual standard error (σ) were obtained for the F-test. For the adaptive test, the underlying distribution (score),test statistic (value), scale parameters for the sample (τ_s) and residual (τ_r) were as well obtained.

4.2.1 Normal Distribution

The normal distribution is useful because of the central limit theorem. In its most general form, under some conditions (which include finite variance), it states that averages of samples of observations of random variables independently drawn from independent distributions converge in distribution to the normal, that is, become normally distributed when the number of observations is sufficiently large. The probability density function (pdf) of the normal distribution is

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty$$

where

μ, the mean or expectation of the distribution (and also median and mode) is the location parameter and $-\infty < \mu < \infty$

σ^2 is the variance and

σ, the standard deviation is a scale parameter and $\sigma > 0$.

If $X \sim (\mu, \sigma^2)$, the cumulative distribution function (cdf) of X is

$$F((x|\mu,\sigma^2)) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

The simplest case of a normal distribution is known as the standard normal distribution. This is a special case when $\mu = 0$ and $\sigma = 1$, and it is described by this probability density function:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Using the normal distribution with the location parameter $\mu = 0$ and scale parameter $\sigma = 1$, under H_0, 10,000 simulations were run for the various sample sizes. Table 4.2 displays the simulation results for the

Table 4.2: **Normal Distribution**

Sample Size	F-Test		Adaptive Test			
(n_1, n_2, n_3)	Value	σ	Score	Value	τ_s	τ_r
$(5, 5, 5)$	1.1716	0.9831	SH	2.2919	1.0601	1.0266
$(10, 10, 10)$	1.0788	0.9901	SM	2.1401	1.0297	1.0201
$(15, 15, 15)$	1.0507	0.9935	SM	2.1022	1.0274	1.0231
$(20, 20, 20)$	1.0367	0.9943	SM	2.0485	1.0302	1.0288
$(25, 25, 25)$	1.0237	0.9957	SM	2.0286	1.0307	1.0304
$(30, 30, 30)$	1.0138	0.9974	SM	2.0024	1.0292	1.0294
$(50, 50, 50)$	1.0049	0.9981	SM	2.0023	1.0293	1.0294
$(100, 100, 100)$	1.0201	0.9993	SM	2.0269	1.0271	1.0271

Normal distribution. The selector statistic for the adaptive test identified the normal distribution as a symmetric and medium tailed distribution. The parametric F-test performed better than the adaptive because the residual standard error (σ) reported are less than the estimated scale parameters of the sample (τ_s) and residuals (τ_r) at all levels of the sample sizes considered. However, the two models failed to reject the null hypothesis of no difference in level means at all sample sizes considered at 5% significance level. The asymptotic relative efficiency (ARE) of the parametric F-test over the adaptive test is between 91.7% and 94.7%. Thus, the normal ditribution is more efficient under the parametric F-test than the adaptive test.

4.2.2 Logistic Distribution

The continuous random variable X is said to have a logistic distribution with parameters μ and $\sigma > 0$ if its probability density function (pdf) satisfies

$$f(x|\mu, \sigma) = \frac{e^{-\frac{x-\mu}{\sigma}}}{\sigma\left(1 + e^{-\frac{x-\mu}{\sigma}}\right)^2}, \quad -\infty < x < \infty$$

where μ, the mean, is a location parameter and $\sigma > 0$ is a scale parameter proportional to the standard deviation. The cumulative density function of the logistic distribution is given by

$$F(x;\mu,\sigma) = \frac{1}{1+e^{-\dfrac{x-\mu}{\sigma}}}, \qquad \begin{cases} 0 \leq \mu \leq \infty \\ \sigma > 0 \end{cases}$$

Using the Logistic distribution with a location parameter $\mu = 0$ and scale parameter $\sigma = 1$, 10,000 simulations were carried out. The simulation results are shown in Table 4.3.

Table 4.3: **Logistic Distribution**

Sample Size	F-Test		Adaptive Test			
(n_1, n_2, n_3)	Value	σ	Score	Value	τ_s	τ_r
$(5,5,5)$	1.2139	1.7564	SH	2.3532	1.8070	1.7557
$(10,10,10)$	1.0733	1.7917	SM	2.1579	1.7645	1.7509
$(15,15,15)$	1.0571	1.7977	SM	2.1584	1.7483	1.7414
$(20,20,20)$	1.0367	1.8007	SM	2.1028	1.7446	1.7416
$(25,25,25)$	1.0263	1.8012	SM	2.0828	1.7469	1.7448
$(30,30,30)$	1.0328	1.8053	SM	2.0312	1.7483	1.7470
$(50,50,50)$	1.0067	1.8103	SM	1.9933	0.7459	1.7453
$(100,100,100)$	1.0056	1.8117	SM	2.0098	1.7413	1.7414

The adaptive test classified the logistic distribution as a symmetric and medium tailed distribution. From Table 4.3, the F-test reported greater residual standard errors (σ) than scale parameters (τ_s) and (τ_r) for sample sizes except the sample size 5 where σ reported less value than the τ_s. As a result, the adaptive test performed better than the F-test even though the underlying distribution is symmetric and a medium tailed. Both models however, failed to reject the null hypothesis at significance level of 5%. The asymptotic relative efficiency of the adaptive test over the F-test ranges from 92.4% to 99.9%.

4.2.3 Laplace Distribution

The Laplace distribution is also known as double exponential distribution. This distribution is characterized by location μ (any real number) and scale λ (has to be greater than 0) parameters. The probability density function of Laplace (μ, λ) is

$$f((x|\mu,\lambda)) = \frac{1}{2\lambda} e^{\left(-\dfrac{|\,\mu - x\,|}{\lambda}\right)}.$$

Alternatively, the pdf of Laplace (μ, λ) is given as

$$f(x|\mu,\lambda) = \frac{1}{2\lambda} \begin{cases} e^{-\dfrac{\mu - x}{\lambda}}, & if x < \mu \\ e^{-\dfrac{x-\mu}{\lambda}}, & x \geq \mu \end{cases}$$

The μ is a location parameter and $\lambda > 0$, which is sometimes referred to as the diversity, is a scale parameter. If $\mu = 0$ and $\lambda = 1$, the positive half-line is exactly an exponential distribution scaled by $\frac{1}{2}$. Thus

$$f(x|\mu,\lambda) = \frac{1}{2}\begin{cases} e^{x} \\ \\ e^{-x} \end{cases} .$$

The probability density function of the Laplace distribution is also reminiscent of the normal distribution; however, whereas the normal distribution is expressed in terms of the squared difference from the mean μ, the Laplace density is expressed in terms of the absolute difference from the mean. Consequently, the Laplace distribution has fatter tails than the normal distribution.

The cumulative density function looks even more impressive, yet rather easy to integrate because of the absolute value in the formula

$$F(x|\mu,\lambda) = \begin{cases} \frac{1}{2}e^{-\frac{x-\mu}{\lambda}}, & if x \leqslant \mu \\ 1 - \frac{1}{2}e^{-\frac{\mu-x}{\lambda}}, & if x \geqslant \mu \end{cases} .$$

The expected value of a Laplace distribution

$$E(X) = \mu$$

As in the case of other symmetric distributions such as the Normal and the Logistic distributions. Laplace's location is the same as its mean, median and mode. The variance is

$$Var(X) = 2\lambda^2$$

Simulation results for 10,000 simulations carried out for Laplace distribution with the location parameter $\mu = 0$ and a scale parameter $\lambda = 1$ are displayed in Table 4.4.

Table 4.4: **Laplace Distribution**

Sample Size	F-Test		Adaptive Test			
(n_1, n_2, n_3)	Value	σ	Score	Value	τ_s	τ_r
$(5, 5, 5)$	1.1641	1.3535	SH	2.3008	1.2634	1.2377
$(10, 10, 10)$	1.0680	1.3856	SH	2.1494	1.1955	1.1915
$(15, 15, 15)$	1.0593	1.3958	SH	2.1568	1.1501	1.1496
$(20, 20, 20)$	1.0445	1.3988	SH	2.1604	1.1276	1.1270
$(25, 25, 25)$	1.0177	1.4048	SH	2.0619	1.1308	1.1305
$(30, 30, 30)$	1.0383	1.4050	SH	2.0784	1.1300	1.1301
$(50, 50, 50)$	1.0250	1.4063	SH	2.0568	1.1121	1.1120
$(100, 100, 100)$	1.0112	1.4112	SH	2.0220	1.0956	1.0955

The selector statistic for the adaptive test classified the Laplace distribution as a symmetric and heavy tailed distribution. The adaptive test outperformed the F-test as the scale parameter for the adaptive test (τ_r) is smaller than the F-test (σ). The asypmtotic relative efficiency of the adaptive test over the

F-test is between 60.3% and 83.6%. At 5% or 1% significance level, the two tests failed to reject the null hypothesis at all sample sizes under consideration.

4.2.4 Lognormal Distribution

A positive random variable X is log-normally distributed if the logarithm of X is normally distributed,

$$ln(X) \sim N(\mu, \sigma^2).$$

A random variable X is said to have the lognormal distribution with parameters $\mu \in R$ and $\sigma > 0$ if $ln(X)$ has the normal distribution with mean μ and standard deviation σ. Equivalently, $X = e^Y$ where Y normally distributed with mean μ and standard deviation σ. The probability density function of the lognormal distribution with parameters μ and σ is given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{[ln(x)-\mu]^2}{2\sigma^2}}, \qquad x > 0$$

The cumulative distribution function (cdf) of the lognormal distribution is

$$F_X(x) = \Phi\left(\frac{lnx - \mu}{\sigma}\right)$$

where Φ is the cumulative distribution function of the standard normal distribution ($i.e. N(0,1)$). The expected value of the lognormal distribution is given by

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2}$$

and the variance

$$Var(X) = (e^{\sigma^2} - 1)(e^{2\mu+\sigma^2})$$

Using the lognormal distribution with location parameter $\mu = 0$ and a scale parameter $\sigma = 1$, under H_0, 10,000 simulations were conducted. Results for the simulation are presented in Table 4.5.

Table 4.5: **Lognormal Distribution**

Sample Size	F-Test		Adaptive Test			
(n_1, n_2, n_3)	Value	σ	Score	Value	τ_s	τ_r
$(5,5,5)$	1.1463	1.8080	RH	1.9577	0.9671	0.9663
$(10,10,10)$	1.0568	1.9121	RM	1.7346	0.8513	0.8599
$(15,15,15)$	1.0415	1.9783	RH	2.0403	0.7311	0.7353
$(20,20,20)$	1.0214	1.9874	RH	2.0129	0.6781	0.6755
$(25,25,25)$	1.0288	2.0203	RH	2.0420	0.6748	0.6712
$(30,30,30)$	1.0125	2.0398	RM	2.0243	0.6984	0.6951
$(50,50,50)$	1.0079	2.0753	RH	1.9625	0.6556	0.6528
$(100,100,100)$	0.9914	2.1125	RH	1.9459	0.6027	0.6026

The selector statistic for the adaptive test identified the lognormal distribution as a right skewed and heavy tailed distribution. The adaptive test performed better than the F-test per the values of the scale parameter (τ_r) and the residual standard error (σ). The asymptotic relative efficiency of the two tests are between 8.1% and 28.6%. Both tests however, failed to reject the null hypothesis of no difference in level means at all sample sizes considered at 5% significance level.

4.2.5 Exponential Distribution

The continuous random variable X is said to have the exponential distribution with positive parameter λ if its probability density function (pdf) is given by

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The cumulative density function (cdf) of the exponential distribution is given by

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The expected value of the random variable X which has exponential distribution with positive parameter λ is given by

$$E(X) = \frac{1}{\lambda}$$

and the variance is

$$Var(X) = \frac{1}{\lambda^2}$$

Under H_0, 10,000 simulations were run for the exponential distribution with a scale parameter $\lambda = 1$. The simulation results are shown in Table 4.6. The selector statistic for the adaptive test showed that

Table 4.6: **Exponential Distribution**

Sample Size	F-Test		Adaptive Test			
(n_1, n_2, n_3)	Value	σ	Score	Value	τ_s	τ_r
$(5, 5, 5)$	1.1735	0.9422	RH	2.1081	0.7123	0.7057
$(10, 10, 10)$	1.0874	0.9731	RM	1.7320	0.6302	0.6491
$(15, 15, 15)$	1.0415	0.9772	RM	2.0308	0.5913	0.6090
$(20, 20, 20)$	1.0347	0.9851	RM	2.0645	0.5660	0.5682
$(25, 25, 25)$	1.0410	0.9904	RM	2.0337	0.5672	0.5726
$(30, 30, 30)$	1.0132	0.9884	RM	1.9999	0.5586	0.5682
$(50, 50, 50)$	1.0129	0.9941	RM	1.9841	0.5534	0.5544
$(100, 100, 100)$	1.0112	0.9976	RM	1.9744	0.5250	0.5220

exponential distribution is a right skewed and medium tailed distribution. The estimated scale parameters (τ_r) and τ_s indicate that the adaptive test performed better than the F-test because the estimated scale

parameters reported less values than the residual standard error (σ) even though, both tests failed to reject the null hypothesis at 5% or 1% significance level at all the sample sizes. The asymptotic relative efficiency of the adaptive test over the F-test is between 27.4% to 56.1%.

4.2.6 Weibull Distribution

The Weibull distribution is a continuous probability distribution. The probability density function (pdf) of a Weibull random variable X is given by

$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter of the distribution. The form of the density function of the Weibull distribution changes drastically with the value of k. For $0 < k < 1$, the density function tends to ∞ as x approaches zero from above and is strictly decreasing. For $k = 1$, the density function tends to $\frac{1}{\lambda}$ as x approaches zero from above and is strictly decreasing. For $k > 1$, the density function tends to zero as x approaches zero from above, increases until its mode and decreases after it. The density function has infinite negative slope at $x = 0$ if $0 < k < 1$, infinite positive slope at $x = 0$ if $1 < k < 2$ and null slope at $x = 0$ if $k > 2$. For $k = 2$ the density has a finite positive slope at $x = 0$. As k goes to infinity, the Weibull distribution converges to a Dirac delta distribution centered at $x = \lambda$. Moreover, the skewness and coefficient of variation depend only on the shape parameter.

The cumulative distribution function (cdf) for the Weibull distribution is

$$F(x;\lambda,k) = \begin{cases} 1 - e^{-\left(\frac{x}{\lambda}\right)^{k}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

If $x = \lambda$, then $F(x;\lambda,k) = 1 - e^{-1} = 0.632$.

The expected value of the Weibull distribution is given by

$$E(X) = \lambda\Gamma(1 + \frac{1}{k})$$

and the variance as

$$Var(X) = \lambda^2 \left[\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2\right]$$

Simulation results for Weibull distribution with the shape parameter $k > 0$ and a scale parameter $\lambda = 1$ are shown in Table 4.7. The underlying error distribution of the Weibull distribution was identified by the

Table 4.7: **Weibull Distribution**

Sample Size	F-Test		Adaptive Test			
(n_1, n_2, n_3)	Value	σ	Score	Value	τ_s	τ_r
$(5, 5, 5)$	1.1757	0.9423	RH	2.0644	0.7075	0.7002
$(10, 10, 10)$	1.0605	0.9705	RM	1.7301	0.6291	0.6486
$(15, 5, 15)$	1.0598	0.9788	RM	2.0527	0.5945	0.6121
$(20, 20, 20)$	1.0154	0.9834	RM	2.0755	0.5643	0.5665
$(25, 25, 25)$	1.0306	0.9840	RM	2.0731	0.5641	0.5689
$(30, 30, 30)$	1.0252	0.9870	RM	1.9691	0.5590	0.5692
$(50, 50, 50)$	1.0183	0.9926	RM	2.0214	0.5530	0.5542
$(100, 100, 100)$	0.9953	0.9961	RM	1.9813	0.5252	0.5225

selector statistic as a right skewed and medium tailed. The adaptive test outperformed the F-test because the estimated scale parameters (τ_s) and (τ_r) both reported less values than the residual standard error (σ). It is worth noting that as the sample size increases the residual standard error (σ) increases whereas the estimated scale parameters (τ_s) and (τ_r) decrease. Notwithstanding, both tests failed to reject the null hypothesis at 5% or 1% significance level. The asymptotic relative efficiency of the adaptive test (τ_r) over the F-test (σ) is between 27.5% and 56.2%.

4.2.7 Cauchy Distribution

The Cauchy distribution is a continuous probability distribution. The Cauchy distribution $f(x; x_0, \gamma)$ is the distribution of the x-intercept of a ray issuing from (x_0, γ) with a uniformly distributed angle. It is also the distribution of the ratio of two independent normally distributed random variables if the denominator distribution has mean zero. The Cauchy distribution does not have finite moments of order greater than or equal to one; only fractional absolute moments exist. The Cauchy distribution has no moment generating function. The Cauchy distribution has the probability density function (pdf)

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]}$$

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma}\left[\frac{\gamma^2}{(x - x_0)^2 + \gamma^2}\right]$$

where x_0 is a location parameter, specifying the location of the peak of the distribution, and γ is the scale parameter which specifies the half-width at half-maximum (HWHM), alternatively γ^2 is full-width at half-maximum. The special case when $x_0 = 0$ and $\gamma = 1$ is called the standard Cauchy distribution with the probability density function

$$f(x; x_0, \gamma) = \frac{1}{\pi(1 + x^2)}.$$

The cumulative density function of Cauchy distribution is given by

$$F(x; x_0, \gamma) = \frac{1}{\pi} tan^{-1}\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}.$$

The Cauchy distribution is an example of a distribution which has no mean, variance or higher moments defined. Its mode and median are well defined and are both equal to x_0. The standard Cauchy distribution was simulated 10,000 under H_0. The results of the simulation are shown in Table 4.8.

Table 4.8: **Cauchy Distribution**

Sample Size	F-Test		Adaptive Test			
(n_1, n_2, n_3)	Value	σ	Score	Value	τ_s	τ_r
$(5, 5, 5)$	1.0899	26.2994	LH	2.0829	2.7886	2.7844
$(10, 10, 10)$	1.0320	39.1821	LH	2.2975	2.3935	2.4009
$(15, 15, 15)$	1.0158	44.0567	SH	2.1192	1.9799	1.9824
$(20, 20, 20)$	1.0171	64.7010	SH	2.0545	1.9060	1.9073
$(25, 25, 25)$	1.0109	61.5071	SH	2.0691	1.8858	1.8865
$(30, 30, 30)$	1.0104	76.9902	SH	2.0246	1.9384	1.9386
$(50, 50, 50)$	1.0076	64.0498	SH	2.0381	1.8584	1.8588
$(100, 100, 100)$	1.0076	110.054	SH	2.0110	1.7335	1.7339

The Cauchy distribution is identified by the selector statistic as a symmetric and heavy tailed distribution. The adaptive test outperformed the traditional F-test as the estimated scale parameters (τ_r) and (τ_s) reported by the adaptive test are smaller than the residual standard error (σ) reported by the F-test for all sample sizes. Both tests however, failed to reject the null hypothesis of no difference in level means at all the sample sizes at 5% or 1% significance level.

4.2.8 Contaminated Normal Distribution

Let Z be random samples drawn from normal distributions, $I_{1-\epsilon}$ be a discrete random variable defined by

$$I_{1-\epsilon} = \begin{cases} 1, & with\ \ probability \quad 1-\epsilon \\ 0, & with\ \ probability \quad \epsilon \end{cases}$$

and assume that Z and $I_{1-\epsilon}$ are independent, see Hogg et al. (2013).

Let $Q = ZI_{1-\epsilon} + \sigma_c Z(1 - I_{1-\epsilon})$, then by the independence of Z and $I_{1-\epsilon}$ the cdf of Q is given by

$$F_Q(q) = Pr(Q \leq q)$$

$$F_Q(q) = Pr[Q \leq q, I_{1-\epsilon} = 1] + Pr[Q \leq q, I_{1-\epsilon} = 0]$$

$$F_Q(q) = Pr[Q \leq \frac{q}{I_{1-\epsilon}} = 1]Pr[I_{1-\epsilon} = 1] + Pr[Q \leq \frac{q}{I_{1-\epsilon}} = 0]Pr[I_{1-\epsilon} = 0]$$

$$F_Q(q) = Pr[Q \leq q](1 - \epsilon) + Pr[Q \leq \frac{q}{\sigma_c}]\varepsilon$$

$$F_Q(q) = \Phi(q)(1 - \epsilon) + \Phi(\frac{q}{\sigma_c})\epsilon$$

where σ_c is the standard deviation of contamination, $I_{1-\epsilon}$ is the characteristics function, ϵ is the percentage of contamination. Thus, the pdf of the contaminated distribution is given by

$$\frac{dF_Q(q)}{dq} = (1-\epsilon)f_Q(q) + \frac{\epsilon}{\sigma_c} f_Q(\frac{q}{\sigma_c})$$

Simulation results for normal distribution with the location parameter $\mu = 0$ and scale parameter $\sigma = 1$, contaminated at 5%, 10%, 15% and 20% are displayed in Table 4.9. The underlying error distribution

Table 4.9: **Contaminated Normal Distribution**

Sample Size	Level	F-Test		Adaptive Test			
(n_1, n_2, n_3)	%	Value	σ	Score	Value	τ_s	τ_r
(5, 5, 5)	5%	1.1915	0.9795	SH	2.3221	1.0557	1.0211
(5, 5, 5)	10%	1.1755	0.9821	SH	2.2739	1.0558	1.0207
(5, 5, 5)	15%	1.2079	0.9777	SH	2.3325	1.0579	1.0227
(5, 5, 5)	20%	1.2075	0.9803	SH	2.3639	1.0583	1.0239
(10, 10, 10)	5%	1.0626	0.9894	SM	2.0979	1.0299	1.0191
(10, 10, 10)	10%	1.1050	0.9915	SM	2.1870	1.0320	1.0204
(10, 10, 10)	15%	1.1039	0.9896	SM	2.1712	1.0292	1.0204
(10, 10, 10)	20%	1.0744	0.9912	SM	2.1312	1.0294	1.0209
(15, 15, 15)	5%	1.0609	0.9926	SM	2.1151	1.0270	1.0229
(15, 15, 15)	10%	1.0496	0.9936	SM	2.1003	1.0271	1.0240
(15, 15, 15)	15%	1.0577	0.9939	SM	2.1007	1.0281	1.0245
(15, 15, 15)	20%	1.0589	0.9936	SM	2.1004	1.0271	1.0233
(20, 20, 20)	5%	1.0410	0.9945	SM	2.0631	1.0296	1.0281
(20, 20, 20)	10%	1.0306	0.9966	SM	2.0549	1.0325	1.0313
(20, 20, 20)	15%	1.0357	0.9955	SM	2.0460	1.0320	1.0307
(20, 20, 20)	20%	1.0307	0.9950	SM	2.0533	1.0294	1.0269
(25, 25, 25)	5%	1.0220	0.9972	SM	2.0182	1.0341	1.0335
(25, 25, 25)	10%	1.0210	0.9971	SM	2.0271	1.0348	1.0335
(25, 25, 25)	15%	1.0270	0.9980	SM	2.0416	1.0334	1.0331
(25, 25, 25)	20%	1.0256	0.9970	SM	2.0170	1.0345	1.0339
(30, 30, 30)	5%	1.0320	0.9962	SM	2.0514	1.0276	1.0275
(30, 30, 30)	10%	1.0232	0.9962	SM	2.0219	1.0283	1.0283
(30, 30, 30)	15%	1.0227	0.9964	SM	2.0371	1.0288	1.0286
(30, 30, 30)	20%	1.0203	0.9967	SM	2.0231	1.0282	1.0280
(50, 50, 50)	5%	0.9994	0.9975	SM	1.9847	1.0287	1.0286
(50, 50, 50)	10%	1.0024	0.9990	SM	2.0011	1.0295	1.0297
(50, 50, 50)	15%	1.0235	0.9986	SM	2.0327	1.0294	1.0295
(50, 50, 50)	20%	1.0204	0.9982	SM	2.0236	1.0297	1.0297
(100, 100, 100)	5%	1.0070	0.9989	SM	2.0026	1.0271	1.0272
(100, 100, 100)	10%	0.9908	0.9997	SM	1.9789	1.0279	1.0279
(100, 100, 100)	15%	1.0160	0.9990	SM	2.0254	1.0269	1.0269
(100, 100, 100)	20%	1.0003	0.9986	SM	1.9861	1.0267	1.0267

of Contaminated Normal distribution was classified by the selector statistic as a symmetric and heavy tailed distribution when the sample is 5. The underlying error distribution for the remaining samples was identified as symmetric and medium tailed. The F-test performed better than the adaptive because the residual standard error (σ) reported for the F-test is smaller than the estimated scale parameters (τ_r) and (τ_s) reported by the adaptive test. Notwithstanding, the two tests failed to reject the null hypothesis at a significance level of 5% or 1%. The asymptotic relative efficiency of the F-test over the adaptive for the sample size is between 91.6 % to 94.3%.

4.2.9 Mixture of Distributions

In this subsection, 10,000 simulations of mixture of distributions for three samples were generated from the Normal, Laplace and Cauchy distributions respectively and is shown in Table 4.10. The Normal distribution was generated with the parameter $\mu = 0$ and $\sigma = 1$. The parameters for the Laplace distribution were $\mu = 0$ and $\lambda = 1$ and the Cauchy with location parameter $x_0 = 0$ and scale parameter $\gamma = 1$

Table 4.10: **Mixture of Distributions**

Sample Size	F-Test		Adaptive Test			
(n_1, n_2, n_3)	Value	σ	Score	Value	τ_s	τ_r
$(5, 5, 5)$	1.1740	10.6971	SH	2.6740	1.6893	1.7001
$(10, 10, 10)$	1.0934	10.0139	LH	2.6780	1.6295	1.6261
$(15, 15, 15)$	1.0369	88.1010	SH	2.6322	1.5095	1.5088
$(20, 20, 20)$	1.0345	14.2196	SH	2.6094	1.4928	1.4922
$(25, 25, 25)$	1.0229	14.5802	SH	2.6327	1.5079	1.5082
$(30, 30, 30)$	1.0280	16.3163	SH	2.5979	1.5414	1.5415
$(50, 50, 50)$	1.0151	32.4284	SH	2.7024	1.5213	1.5211
$(100, 100, 100)$	0.9995	39.4103	SH	3.0333	1.4866	1.4865

The mixture of distributions was classified by the selector statistic as a symmetric and heavy tailed distribution. The estimated scale parameters (τ_s) and (τ_r) suggest that the adaptive test outperformed the F-test because both (τ_s) and (τ_r) reported less values than the residual standard error (σ) at all levels of the sample sizes. However, with the exception of sample size $(100, 100, 100)$ for the adaptive test, the two tests failed to reject the null hypothesis of no difference in level means at 5% significance level.

4.2.10 Pareto Distribution

The results of 10,000 simulations for a Pareto distribution with the shape parameter $\alpha > 0$ and a scale parameter $x_m > 0$ are displayed in Table 4.11.

Table 4.11: **Pareto (Type 1) Distribution**

Sample Size	F-Test		Adaptive Test			
(n_1, n_2, n_3)	Value	σ	Score	Value	τ_s	τ_r
$(5, 5, 5)$	1.0151	1.9317^{51}	RH	0.6143	3.7535^{35}	1.6493^{51}
$(10, 10, 10)$	1.0006	2.8310^{56}	RL	1.6782	2.8655^{40}	2.0832^{40}
$(15, 15, 15)$	1.0025	2.1706^{59}	RH	17.255	1.9854^{43}	3.2230^{42}
$(20, 20, 20)$	1.0003	5.4607^{59}	RH	61.172	1.1046^{43}	1.1046^{43}
$(25, 25, 25)$	1.0030	4.3094^{56}	RH	73.991	1.4160^{40}	6.0965^{39}
$(30, 30, 30)$	1.0016	9.0949^{51}	RL	66.600	2.2733^{35}	2.2479^{35}
$(50, 50, 50)$	1.0027	2.5032^{58}	RL	347.03	1.7899^{41}	3.8992^{41}
$(100, 100, 100)$	1.0010	7.7567^{66}	RH	2741.1	5.1532^{49}	5.1532^{49}

The selector statistic classified the underlying error distribution of a Pareto distribution as a right skewed and heavy tailed distribution. The estimated scale parameters (τ_s) and (τ_r) suggest that the adaptive

test is a better option than the traditional parametric F test because the estimated scale parameters (τ_s) and (τ_r) reported less values than the residual standard error (σ). The adaptive test failed to reject the null hypothesis of no difference in level means at sample sizes $(5, 5, 5)$ and $(10, 10, 10)$ but rejected the null hypothesis at the remaining sample sizes at 5% or 1% significant level. However, the F-test failed to reject the null hypothesis at all the sample sizes at 5% or 1% level of significance.

4.3 Simulation Results for Distributions with Correlated Errors

In this section, simulation studies were conducted for the adaptive test and the repeated measures ANOVA test. These two tests were compared under Normal, Laplace and Cauchy distributions. In the simulation of the repeated measures data, consideration of three time points with three treatment groups were made. Samples of equal sizes were generated for each treatment group at each of the time points. In each case, intra-class correlation coefficients of $\rho = 0.0, 0.3, 0.5, 0.75$ were considered. At each time point, data was generated under H_0 for equation (2.13).

4.3.1 Normal Distribution

Using the normal distribution, under H_0, 10,000 simulations were conducted for sample sizes $5, 10, 15$ and 20 subjects each with correlation coefficient ρ being $0.0, 0.3, 0.5, 0.75$. The results are displayed in Table 4.12. The selector statistics for the adaptive test identified the underlying error distribution of a normal

Table 4.12: **Normal Distribution**

Sample Size	Corr	F-Test		Adaptive Test		
(n_1, n_2, n_3)	ρ	Value	σ	Score	Value	τ
$(5, 5, 5)$	0.00	57.378	0.977	SH	12.009	1.024
	0.30	58.228	0.975	SH	11.977	1.022
	0.50	60.094	0.975	SH	12.006	1.030
	0.75	67.583	0.959	SM	11.672	0.995
$(10, 10, 10)$	0.00	102.038	0.990	SM	179.439	1.021
	0.30	102.853	0.989	SM	181.592	1.011
	0.50	105.369	0.986	SM	187.641	1.015
	0.75	109.944	0.979	SM	229.992	0.991
$(15, 15, 15)$	0.00	148.685	0.993	SM	292.075	1.024
	0.30	149.086	0.992	SM	292.339	1.026
	0.50	151.293	0.991	SM	296.332	1.022
	0.75	154.688	0.988	SM	314.040	1.002
$(20, 20, 20)$	0.00	194.333	0.996	SM	372.332	1.032
	0.30	195.218	0.995	SM	374.670	1.030
	0.50	197.545	0.993	SM	378.247	1.029
	0.75	201.011	0.993	SM	397.022	1.014

distribution as a symmetric and medium tailed distribution as the sample size increases. From Table

4.12, the parametric F-test performed better than the adaptive test at all the levels of the sample sizes considered because the residual standard error (σ) reported less value than the estimated scale parameter (τ) . Thus, it is confirmed that the parametric F-test is more efficient under normal distribution than our adaptive test.

4.3.2 Laplace Distribution

Simulation results for Laplace distribution with the location vector μ and definite-postive $k \times k$ covariance matrix ($\sum$) are displayed in Table 4.13.

Table 4.13: **Laplace Distribution**

Sample Size	Corr	F-Test		Adaptive Test		
(n_1, n_2, n_3)	ρ	Value	σ	Score	Value	τ
$(5, 5, 5)$	0.00	76.842	0.944	SH	206.132	0.887
	0.30	78.560	0.940	SH	209.047	0.884
	0.50	84.738	0.929	SH	224.142	0.878
	0.75	94.936	0.924	SH	237.974	0.878
$(10, 10, 10)$	0.00	119.907	0.964	SH	342..521	0.849
	0.30	120.031	0.972	SH	343.324	0.848
	0.50	123.409	0.969	SM	354.430	0.844
	0.75	126.324	0.968	SM	358.524	0.833
$(15, 15, 15)$	0.00	163.381	0.984	SH	507.221	0.822
	0.30	165.608	0.983	SH	514.363	0.816
	0.50	171.576	0.973	SH	525.284	0.813
	0.75	178.682	0.907	SH	546.037	0.810
$(20, 20, 20)$	0.00	210.538	0.985	SH	677.762	0.801
	0.30	213.086	0.983	SH	685.978	0.780
	0.50	217.272	0.980	SH	688.290	0.800
	0.75	222.108	0.980	SH	704.636	0.799

From Table 4.13, the selector statistics classified the underlying error distribution of the Laplace distribution as a symmetric and heavy- tailed. The Laplace distribution performed better under the adaptive test than the F test because the estimated scale parameter (τ) reported less value than the residual standard error (σ).

4.3.3 Cauchy Distribution

Simulation results for Cauchy distribution with the location vector μ and a positive-definite of $k \times k$ scale matrix sigma ($\sum$) are shown in Table 4.14.

The selector statistic identified the underlying error distribution of Cauchy distribution as symmetric and heavy-tailed distribution. However, at sample size 10 and intraclass correlation coefficient $0.30, 0.50, 0.75$, the selector statistic indicated left skewed and heavy tailed. From Table 4.14, it is clear that adaptive test is a better option than the F test for a data known to have been generated from a Cauchy distribution.

Table 4.14: **Cauchy Distribution**

Sample Size	Corr	F-Test		Adaptive Test		
(n_1, n_2, n_3)	ρ	Value	σ	Score	Value	τ
$(5, 5, 5)$	0.00	13.782	21.512	SH	41.133	10.409
	0.30	13.655	17.519	SH	42.421	5.821
	0.50	15.125	16.961	SH	46.607	7.333
	0.75	16.502	12.916	SH	48.872	5.503
$(10, 10, 10)$	0.00	12.956	20.028	SH	66.789	2.971
	0.30	12.713	20.178	LH	68.602	2.795
	0.50	13.382	19.474	LH	71.184	2.400
	0.75	14.203	19.598	LH	76.273	2.364
$(15, 15, 15)$	0.00	12.439	35.722	SH	102.049	2.069
	0.30	12.421	34.535	SH	104.162	2.048
	0.50	12.785	33.594	SH	104.610	2.025
	0.75	13.431	62.974	SH	110.002	2.012
$(20, 20, 20)$	0.00	12.500	26.679	SH	114.631	1.846
	0.30	12.500	29.653	SH	142.210	1.870
	0.50	12.673	75.662	SH	142.602	1.878
	0.75	13.226	24.884	SH	143.318	1.896

In all the cases, the adaptive test reported less estimated scale parameter (τ) value than the residual standard error (σ) so the adaptive test is more efficient than the parametric F-test.

4.4 Application of Adaptive Scheme

In this section, the adaptive scheme is applied to real datasets for one-way ANOVA and repeated measures ANOVA to acsertain the efficiency of the two tests.

4.4.1 One-way ANOVA

In this subsection, analyses of real datasets on One-way ANOVA for equal and unequal sample sizes were considered. The datasets for the equal sample size and the unequal sample sizes are shown in Appendices A and B respectively.

Figure 4.1 clearly indicates that the assumption of normality of the pain refief data is not appropriate. An outlier in Drug B is obvious.

The results of data in Appendix A is displayed in Table 4.15

Table 4.15: **Time of Relief for Migraine Headache Sufferers**

Sample Size	F-Test			Adaptive Test			
(n_1, n_2, n_3)	Value	p-value	σ	Score	Value	p-value	τ
$(9, 9, 9)$	11.91	0.0003	1.089	SH	13.6881	0.0001	0.8788

The selector statistic classified the underlying error distribution of the data as a symmetric and heavy tailed distribution. The adaptive test reported the less scale parameter (τ) value compared with residual standard error (σ). Hence, the adaptive test is more efficient than the F-test. However, both models

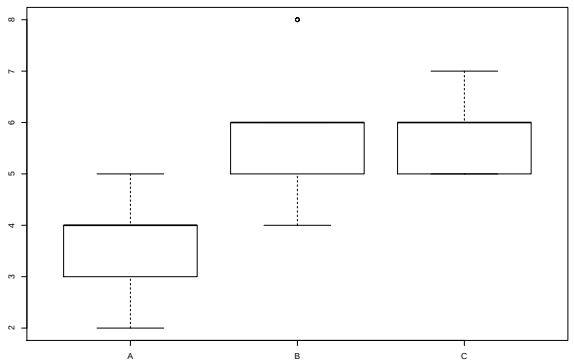

Figure 4.1: Time of Relief for Migraine Headache Sufferers

rejected the null hypothesis, H_0. Thus, there are differences in mean time for the pain relief. The asymptotic relative efficiency (ARE) of the adaptive test over the F-test is about 65.1%

The result of the effect of four catalysts on concentration of a liquid mixture is presented in Table 4.16,

Table 4.16: **Effect of Four Catalysts on Liquid Mixture**

Sample Size	F-Test			Adaptive Test			
(n_1, n_2, n_3, n_4)	Value	p-value	σ	Score	Value	p-value	τ
$(5, 4, 3, 4)$	9.9157	0.0014	1.6971	SM	13.6749	0.0002	2.4666

The selector statistic classified the underlying error distribution of the data in Appendix B as a symmetric and medium tailed distribution. The F-test performed better than the adaptive test because the residual standard error ($\sigma = 1.6971$) is less than the estimated scale parameter ($\tau = 2.4666$). Both tests however, rejected the null hypothesis at 5% significance level. As a consequence, the four catalysts do not have the same mean effect on the concentration. The asymptotic relative efficiency of the F-test over the adaptive test is about 47.3%.

4.4.2 Repeated Measures ANOVA for One Sample

Repeated measures ANOVA for one sample in respect of small and large samples were considered in this subsection. In each case, the ANOVA F-test and adaptive test were performed for with and without outliers. The datasets are displayed in Appendices C and D respectively. Tables 4.17 and 4.18 show the results of the ANOVA F-test and adaptive test of the pain tolerance measured under four different drugs administered on four subjects.

From Table 4.17, the total error sum of squares of 83.5 if the dependency is ignored is split into two components. The part which is due to individual differences (70.25) is removed from the error sum of

Table 4.17: **Repeated Measures ANOVA for Pain Tolerance under Four Different Drugs**

Source	Df	SS	MS	F	p-value
Within-subjects Effect					
Drug	3	50.25	16.75	11.38	0.002
Residuals	9	13.25	1.472		
Between-subject Effects					
Residuals	3	70.25	23.42		

squares for the drug effect. The residual 13.25 reflects the differences between the four drugs. At 5% or 1% significance level, there is a significant change in pain tolerance under the four different drugs.

The adaptive test for the pain tolerance of the four drugs administered on the four subjects is presented in Table 4.18.

Table 4.18: **Adaptive Test for Pain Tolerance under Four Different Drugs**

Method	Test Statistics	σ or τ	p-value	Distribution (scores)
Adaptive Test	0.0672	1.3876	0.9981	SL(4),SM(5),SL(4),SL(4)
F-test	11.38	1.2133	0.002	Normal (Not Applicable)

The adaptive scheme displays the structure of the underlying error distributions of the data as Symmetric and light-tailed for subjects 1, 3 and 4. The structure of the underlying error distribution for subject 2 is a symmetric with medium-tailed distribution, see Table 4.18.

The extract from Twisk (2013)as shown in (Appendix C) was contaminated as follows: Drug 1 for subject 3 was recorded as 1.1. For drug 2 subject 2, it was recorded as 1.2 and drugs 3 and 4 were recorded as 50 and 41 for subjects 4 and 3 respectively. The analyses of this data containing the outliers are shown in Tables 4.19 and 4.20.

Table 4.19: **Repeated Measures ANOVA for Pain Tolerance under Four Different Drugs with Outliers**

Source	Df	SS	MS	F	p-value
Within-subjects Effect					
Drug	3	612.4	204.1	1.025	0.427
Residuals	9	1792.8	199.2		
Between-subject Effects					
Residuals	3	376.8	125.6		

From Tables 4.18 and 4.20, the adaptive tests for both the original data and the one with outliers are not statistically significant under the null hypothesis H_0. However, the p-values changed slightly. On the other hand, the ANOVA F-test is statistically significant under null hypothesis H_0 for the original data (Table 4.17) but H_0 was not rejected in the contaminated data (Table 4.19). This is an indication that our adaptive scheme is insensitive to outliers. Hence, the adaptive test is robust for size when outliers are found in the data.

Table 4.20: **Adaptive Test for Pain Tolerance under Four Different Drugs with Outliers**

Method	Test Statistics	σ or τ	p-value	Distribution (scores)
Adaptive Test	0.0710	5.2949	0.9978	SL(4),SM(5),SL(4),SL(4)
F-test	1.025	14.11	0.427	Normal (Not Applicable)

The data shown in Appendix D is an example of large sample size for repeated measures ANOVA for One Sample. The result is displayed in Table 4.21.

Table 4.21: **Repeated Measures ANOVA for Fatigue Time**

Source	Df	SS	MS	F	p-value
Within-subjects Effects					
Time	4	5900	1474.9	16.48	9.23^{-08}
Residual	36	3221	89.5		
Between-subjects Effects					
Residual	9	4359	484.3		

The ANOVA F-test shown in Table 4.21 indicates the total error sum of squares of 7580 if the dependency is ignored is split into two components. The part which is due to individual differences (4359) is removed from the error sum of squares for the time effect. The residual 3221 reflects the differences between the five-time points.

The result of the adaptive test for the data in Appendix D is displayed in Table 4.22

Table 4.22: **Adaptive Test for Fatigue Time**

Method	Test Statistics	τ or σ	p-value	Distribution (scores)
Adaptive test	2.5121	12.955	0.0146	SM(5),SL(4),SL(4),SL(4), RL(7)
F-test	16.48	9.460	9.23^{-08}	Normal (Not Applicable)

As shown in Table 4.22, the adaptive scheme indicates that the structure of the underlying error distributions for responses on Minutes 3 is symmetric and medium-tailed (SM), Minutes 6, 9 and 12 are symmetric and light-tailed (SL). For Minutes 15, it is right-skewed and light-tailed (RL).
From Tables 4.21 and 4.22, and at 5% or 1% significance level, the result is statistically significant. Thus, H_0 is rejected and we conclude that there is significant change in fatigue over time. The asymptotic relative efficiency of the adaptive test over the F test is about 68.1%.

The data on balance error of fatique time measured over the 15 minutes period (Appendix D) was contaminated as follows: For Min 3, subjects 6 and 9 were entered as 31 and 50, Min 6 and subject 4 was recorded as 1.8. For Min 9, subjects 1 and 2 were entered as 2.3 and 62. Min 12 had its entries for subjects 3 and 5 as 3.1 and 2.8. Finally, for Min 15, subjects 6 and 10 were recorded as 6.5 and 5.7 respectively.
The analyses of the contaminated data are displayed in Tables 4.23 and 4.24.
From Tables 4.22 and 4.24, the adaptive tests for both the original data ($p - value = 0.0146$) and

Table 4.23: **Repeated Measures ANOVA for Fatigue Time with Outliers**

Source	Df	SS	MS	F	p-value
Within-subjects Effects					
Time	4	1794	448.6	1.794	0.151
Residual	36	9000	250.0		
Between-subjects Effects					
Residual	9	2945	327.2		

Table 4.24: **Adaptive Test for Fatigue Time with Outliers**

Method	Test Statistics	τ or σ	p-value	Distribution (scores)
Adaptive test	3.1112	8.8959	0.0035	RL(7),RL(7),SL(4),SL(4), RL(7)
F-test	1.794	15.0	0.151	Normal (Not Applicable)

the one with outliers ($p - value = 0.0035$), the null hypothesis H_0 was rejected. However, there is a slight change in their p-values. On the other hand, the null hypothesis H_0, was rejected for the original data ($p - value = 9.23^{-08}$) under the ANOVA F-test (see (Table 4.21) but for the contaminated data ($p - value = 0.151$), the null hypothsis H_0 was not rejected as displayed in Table 4.23. This indicates that ANOVA F-test is sensitive to outliers. Hence, the adaptive test is robust for size when outliers are found in the data.

4.4.3 Repeated Measures ANOVA for Multiple Samples

Repeated measures ANOVA for multiple samples were performed on small and large samples. In each of the examples, the ANOVA F-test and adaptive test were conducted for with and without outliers in this subsection. Data on small and large samples are shown in Appendices E and F respectively.
Tables 4.25 and 4.26 display the results of the mean depression level for the repeated measures ANOVA and the adaptive test respectively.

Table 4.25: **Repeated-Measures ANOVA for Mean Depression Level**

Source	Df	SS	MS	F	P-value
Within-Subjects					
Time	2	1.0	0.50	0.079	0.925
Group:Time	2	1736	868.2	137.079	5.44^{-09}
Residuals	12	76.0	6.3		
Betwwen-subject					
Group	1	2542	2542	629.0	2.65^{-07}
Residuals	6	24.3	4.05		

From Table 4.25, the between groups test indicates that the variable treatment group is significant. Thus, there is significant difference in mean depression level between groups. The within-subject test as well indicates that there is a significant interaction effect between treatment group and time, that is, the treatment groups are changing over time but they are changing in different directions. This can be inferred from figure 2 that one group is increasing in depression level over time while the other group is decreasing in depression level over time.

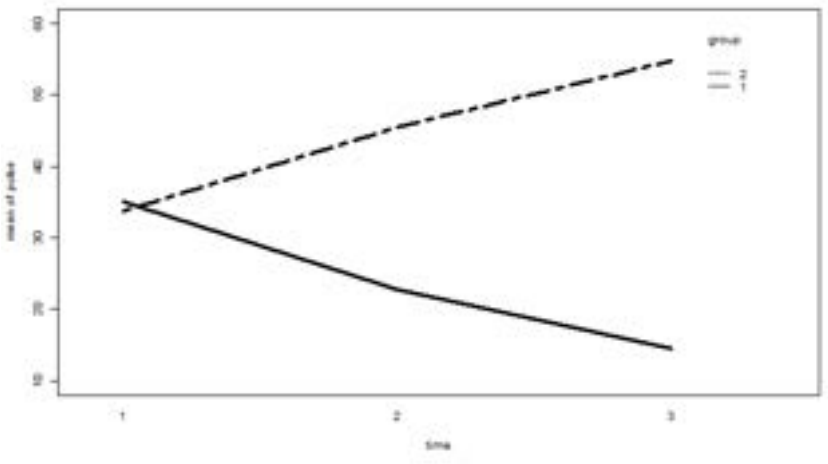

Figure 4.2: Mean Depression Level

From Table 4.26, the adaptive scheme shows that the structure of the underlying error distribution for

Table 4.26: **Adaptive Test for Mean Depression Level**

Method	Test Statistics	σ or τ	P-value	Distribution (scores)
Adaptive Test	9.6009	8.1349	0.0004	SM(5),SL(4),SL(4)
F-test	137.079	2.51	5.44^{-09}	Normal (Not Applicable)

the three-time points is symmetric and medium tailed for Week 1. The structure for the underlyng error distribution for Weeks 2 and 3 are symmetric and light tailed.

The data in Appendix E was contaminated as follows: For Id 1 Group 1 Week 1, the entry was made 53. For Id 3 Group 1 Week 3 it was recorded as 81. Again, for Id 4, Group 1 Week 3, it was recorded as 51. Similarly, Id 5 Group 2 Week 2, 34 was recorded. Furthermore, Id 6 Group 2 Week 2, it was entered as 64 and Id 7 Group 2 Week 3 was recorded as 5.1. Finally, Id 8 Group 2 Week 1 was recorded as 23. The analyses of this data containing outliers are shown in Tables 4.27 and 4.28

Table 4.27: **Repeated-Measures ANOVA for Mean Depression Level with Outliers**

Source	Df	SS	MS	F	P-value
Within-Subjects					
Time	2	185	92.6	0.247	0.785
Gropu:Time	2	1116	557.9	1.49	0.264
Residuals	12	4493	374.4		
Betwwen-subject					
Group	1	294.7	294.7	0.101	0.335
Residuals	6	1606.5	267.8		

From Table 4.27, in the presence of outliers, none of the tests was significant. This shows that ANOVA F-test is sensitive to outliers because in the original data, interaction effect and group effect were significant. As a result ANOVA F-test is not robust for size if a data contains outliers. On the contrary, the adaptive test (Tables 4.26 and 4.28) maintained non significance for both the original data and the

Table 4.28: **Adaptive Test for Mean Depression Level with Outliers**

Method	Test Statistics	σ or τ	P-value	Distribution (scores)
Adaptive Test	0.3849	13.7386	0.6852	LL(1),SL(4),SL(4)
F-test	1.490	19.3494	0.2640	Normal (Not Applicable)

contaminted one. Consequently, the adaptive test is robust for size.

The results of repeated measures ANOVA and adaptive test of the data on the pulse rate of participants during types of exercise (Appendix F) are displayed in Tables 4.29 and 4.30 respectively.

Table 4.29: **Repeated Measures ANOVA for Pulse Rate of Participants**

Source	Df	Sum of Squares	Mean Squares	F	P-value
Within Subjects					
Time	2	2067	1033.3	23.54	4.45^{-08}
Exercise Type:Time	4	2723	680.8	15.51	1.65^{-08}
Residuls	54	2370	43.9		
Between Subjects					
Exercise Type	2	8326	4163	27.0	3.62^{-07}
Residuals	27	4163	154		

From Table 4.29, the between subject test indicates that the variable exercise type is significant. The within subject test indicates that there is significant time effect and interaction of time and exercise type is significant.

Table 4.30: **Adaptive Test for Pulse Rate of Participants**

Method	Test Statistics	σ or τ	P-value	Distribution (Scores)
Adaptive Test	0.0601	11.5127	0.9417	SM(5),SM(5),RM(8)
F-Test	15.51	6.6257	1.65^{-08}	Normal (Not Applicable)

The adaptive test, Table 4.30, indicates that the structure of the underlying error distribution for responses time 1 and 2 are symmetric and medium-tailed whereas Week 3 is a right-skewed and medium-tailed.

The data in Appendix F was contaminated as follows: For the exercise type "rest", ID 2 and Time 1 was recorded as 9.0; ID 3 and Time 3, 49 was recorded and ID 7 and Time 2 was entered as 8. For the exercise type "Walking leisurely" ID 17 and Time 1 the data was recorded as 130; ID 16 and Time 3 was entered as 98 and ID 14 and Time 2 was recocrded as 69. For the exercise type "Running", ID 21 and Time 3, ID 26 and Time 2 and ID 27 and Time 1 were recorded as 11, 162 and 10 respectively.

The analyses of the contaminated data with outliers are shown in Tables 4.31 and 4.32 respectively. From Table 4.31, the interaction of time and exercise type is not significant. This goes to confirm that ANOVA F-test is not robust in the presence of outliers. From the original data (Table 4.29), all the tests were significant. However, the adaptive test remained robust for size with outliers in the data.

Table 4.31: **Repeated Measures ANOVA for Pulse Rate of Participants with Outliers**

Source	Df	Sum of Squares	Mean Squares	F	P-value
Within Subjects					
Time	2	2525	1262.5	2.50	0.0916
Exercise Type:Time	4	4288	1071.9	2.122	0.0906
Residuls	54	27275	505.1		
Between Subjects					
Exercise Type	2	8644	4322	7.668	0.0023
Residuals	27	15219	564		

Table 4.32: **Adaptive Test for Pulse Rate of Participants with Outliers**

Method	Test Statistics	σ or τ	P-value	Distribution (Scores)
Adaptive Test	0.0172	10.2645	0.9830	LH(3),SH(6),SH(6)
F-Test	2.122	22.4744	0.0906	Normal (Not Applicable)

4.5 Covariance Structure

The variance-covariance matrices were computed for the mean depression level (W) and pulse rate of participants (V), for data in Appendices E and F respectively. Thus,

$$W = \begin{bmatrix} 3.9821 & -5.9107 & -15.2679 \\ -5.9107 & 150.9821 & 258.3393 \\ -15.2679 & 258.3393 & 4705536 \end{bmatrix}$$

and

$$V = \begin{bmatrix} 37.8437 & 48.7885 & 60.2851 \\ 48.7885 & 212.1195 & 233.7609 \\ 60.2851 & 233.7609 & 356.3230 \end{bmatrix}$$

Four different covariance structures namely; compound symmetry (CS), unstructured (UN), first order autoregressive (AR(1)) and autoregressive with heterogeneous variance (ARH(1)) are used to determine the most suitable covariance structure for the data in Appendices E and F.

Four fit statistics; Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), loglikelihood and number of covariance parameters were calculated for each of the covariance structures.

Table 4.33: **Covariance Structure for Mean Derpression Level**

Covariance Structure	AIC	BIC	Loglikelihood	No. of Covariance Parameters
CS	105.9294	113.0524	−44.9647	8
Un	103.8572	114.5417	−39.9286	12
AR(1)	106.1264	113.2494	−45.0632	8
ARH(1)	108.7845	117.6882	−44.3923	10

From Tables 4.33, the most suitable covariance structure for the mean depression level is the unstructured

(Un) because it has the minimum AIC value of 103.8572.

Table 4.34: **Covariance Structure for Pulse Rate of Participants**

Covariance structure	AIC	BIC	loglikelihood	No. of Covariance Parameters
CS	612.8316	639.1706	−295.4158	11
Un	607.7365	643.6532	−288.8682	15
AR(1)	612.1163	638.4553	−295.0582	11
ARH(1)	605.7693	636.8971	−289.8846	13

The best covariance structure for the pulse rate of participants (Appendix F) is an Autoregressive with heterogeneous variance (ARH(1)) because it has the minimum AIC value of 605.7693.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This chapter presents the summary of the findings of the study, draws conclusions on the study and finally make some recommendations to address the challenges of the usage of the adaptive scheme by researchers and data analysts. The focus of this atudy is to find a robust adaptive scheme for Gauss Markov Model. The adaptive scheme is a non-parametric whereas the Gauss Markov Model is a parametric statistics.

5.1 Summary

The researcher has been able to extend Hogg's adaptive scheme to the Gauss Markov model. The Guass Markov model belongs to the general classes of linear models. Several models commonly used in statistics, according to Monahan (2008), are examples of the general linear model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ These include, but not limited to linear regression models and analysis of variance (ANOVA) models. Regression models generally refer to those for which X is full rank, while ANOVA models refer to those for which X consists of zeros and ones. Another area of application of Gauss Markov model is the analysis of covariance (ANCOVA). The asymptotic properties of statistical estimates and tests solely rely on the Central Limit Theorem, however, sample sizes are often not large as in clinical trials. The ANOVA-F test employs the assumptions of normality, homogeneity of variance and large sample size of the data.

To overcome the problems of the assumptions of the ANOVA F-test, the adaptive scheme is adopted. The two dimensional selector statistic $S = (Q_1^*, Q_2^*)$ where Q_1^* and Q_2^* which are respective measures of skewness and tailweight of the unknown distribtion function was employed for this work. The nine winsorised scores proposed by Hettmansperger (1984) as the most appropriate set of rank scores for testing hypothesis and estimating statistics were used. The usage of the nine winsorised scores accommodated a wide range of distributions which are either symmetric or asymmetric with varying tailweights as shown in figure 3.1 and in Table 3.1

The procedures for the adaptive test are as follows

1 The combined ordered residuals are obtained from independent random samples from continuous distribution function $f(t)$ with some amount of variations denoted by Δ among the samples, that is $f(t - \Delta)$.

2 Adaptation is based on the residuals after an initial fit of the winsorised Wilcoxon scores on the observed samples has been done.

3 The nine winsorised scores are incorporated and extended to the HFR test in the context of Gauss Markov model.

4 Residuals were obtained from an initial R-fit using Wilcoxon scores.

5 The residuals were ordered and their distribution classified by using both Q_1^* and Q_2^*.

6 Once the distribution of these residuals is classified, a corresponding score function was selected based on the scores presented in Table 3.1.

7 After the selection of the score function, the model was refit using this selected score function and an inference such as estimates of parameters were obtained.

To specify the cut off points for skewness and tailweight the benchmarks for cut-off values proposed by Al-Shomrani (2003) which are dependent on the sample size for skewness as shown by equation (3.32), and tail-weight by equations (3.33) and (3.34) were used. The R-package, Rfit by Kloke and Mckean (2016) was used for the study.

5.2 Conclusions

From the 10,000 simulations for both uncorrelated and correlated error distributions, the structure of the underlying error distributions were confirmed. Thus, the normal, contaaminated normal and logistic distributions indicated symmetric and medium-tailed distributions. The other distributions considered in this study showed either symmetric or asymmetric with varying tailweights.

Again, the simulations from other distributions which have symmetric or asymmetric with varying tailsweights such as Laplace, Lognormal, Exponential, Weibull, Pareto distributions and mixture of distribution confirmed the superiority of the Adaptive test over the F-test because the estimated scale parameters (τ_s) and (τ_r) on each occasion reported less value than the residual standard error (σ).

For the real data analyses, the Adaptive test exhibited its robustness in cases where datasets were contaminated with outliers because the level of significance still remained close to one each other. However, the F-test displayed some weaknesses in the performance of such data. This results confirms the study by Hill et al. (1988) when they used lung cancer data to demonstrate the dominance of their adaptive procedures over the parametric and rank based procedures when the size of each sample was at least 20.

The study further revealed that the adaptive test is suitable for any type of data irrespective of the sample size and whether is it sampled from normal or non-normal population. A sample size as low as 15 is suitable for the adaptive test. This claim is evident in the simulation and the application of the

adaptive scheme.

Finally, the relative efficiency of the two tests indicates that the adaptive test has an urge over a broad class of distributions as compared to the Gauss Markove model which is a parametric test.

5.3 Recommendations

Based on the findings of the study and conclusions drawn therefrom, the following reccommendations are made:

1. That adaptive test may be performed alongside the parametric tests and comparative efficiency will inform a better choice of test.
2. That statisticians, researchers and data analysts are to be encouraged to use adaptive schemes as an alternative to parametric tests because they are applicable to every data.
3. That adaptive test be extended to cover other research areas such as Aitken Model, time series, factor analysis, growth curve analysis among others.
4. That the adaptive procedures be included in the common statistical software packages such as Statistical Package for Social Sciences (SPSS), MINITAB among others.

REFERENCES

Afrifa-Yamoah, E., Okyere, G. A., Asare-Bediako, M., and Bashiru, I. I. S. (2016). A robust procedure for the fit of oneway anova model under adaptation on the observed samples. *Journal of Mathematics and Statistical Science*, pages 534 – 553.

Al-Shomrani, A. A. (2003). *A Comparison of Different Schemes for Selecting and Estimating Score Functions Based on Residuals*. PhD thesis, Western Michigan University.

Bandyopadhyay, U. and Dutta, D. (2007). Adaptive nonparametric tests for a single sample location problem. *Statistical Methodology, .*, 4(4):423 – 433.

Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*. Wiley, N. Y.

Bhattacharyya, G. K., Johnson, R. A., and Mehrotra, K. (1977). On the completeness of minimal sufficient statistics with censored observations. *Ann. Statist.*, 5, Number 3 (1977):547 – 553.

Bolton, S. and Bon, C. (2009). *Pharmaceutical Statistics: Practical and Clinical Application. 5th Edition*. Marcel Dekker, Inc.,.

Büning, H. (1994). Robust and adaptive tests for the two-sample location problem. *OR Spektrum*, 16:33 – 39.

Büning, 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.

Büning, H. (2009). Adaptive test for the c-sample location problem. *In Statistical Inference, Econometric Analysis and Matrix Algebra,*, 1:3 – 17.

Christensen, R. (2001). *Plane Answers to Complex Questions: The Theory of Linear Models*. Albuquerque, New Mexico.

Crowder, M. and Hand, D. (1990). *Analysi of Repeated Measures*. Chapman and Hall, London.

Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer-Verlag, New York, Inc.

Gastwirth, J. L. (1965). Percentile modification of two sample rank tests. *Journal of the American Statistical Association*, 60:1127 – 1140.

Glen, S. (2018). Gauss markov theorem and assumptions. Retrieved March 30 2022.

Good, P. I. (2002). Extensions of the concept of exchangeability and their applications. *Journal of Modern Applied Statistical Methods*, 1(2):243 – 247.

Hájek, J. and Sidák, Z. S. (1967). *Theory of Rank Tests.* New York, Academic Press.

Hájek, J., Sidák, Z. S., and Sen, P. K. (1999). *Theory of Rank Tests.* Academic, New York.

Hao, L. and Houser, D. (2012). Adaptive procedures for nonparametric test: Seven decades of advances. *Communications in Statistics-Theory and Methods.*

Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks.* John Wiley & Sons, Inc., New York,.

Hettmansperger, T. P. and McKean, J. W. (1998). *Robust Nonparametric Statistical Methods.* Arnold, Londom Great Britain.

Hettmansperger, T. P. and McKean, J. W. (2011). *Robust Nonparametric Statistical Methods. 2nd Edition.* Chapman-Hill, New York.

Hill, N. J., Padmanabhan, A., and Puri, M. L. (1988). Adaptive nonparametric procedures and applications. *Jounal of the Royal Statistical Society*, 37(2):205 – 218.

Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, 69:909 – 923.

Hogg, R. V., Craig, A. T., and W., M. J. (2013). *Introduction to Mathematical Statistics. Seventh Edition.* Pearson Prentice Hall.

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.

Huber, J. P. (1981). *Robust statistics.* Wiley, New York.

Huber, J. P. and Ronchetti, E. M. (2009). *Robust Statistics.* John Wiley & Sons Inc., Hoboken, New Jersey.

Husková, M. (1985). Partial review of adaptive procedures. *In: Sequential Methods in Statistics, 16, Banach Center Publicatons, Warschau.*

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. (2016). R packages for rank based estimates. *R Journal.*

Kössler, W. (2010). Max-type rank tests, u-tests and adaptive tests for the two-sample location problem - an assymptotic power study. *Jounal of Computational Statistics and Data Analysis*, 54(9):2053 – 2065.

Kössler, W. and Kumar, N. (2008). An adaptive test for the two-sample location problem based on u-statistics. *Communication in Statistics - Simulation and Computation,*, 37:1329 – 1346.

Mahmoud, H. (2008). *Pólya Urn Models.* Chapman & Hall/ CRC Press Texts in Statistical Science.

Miller, Jr, . R. G. (1997). *Beyond ANOVA. Basics of Applied Statistics.* Chapman & Hill, London.

Monahan, J. F. (2008). *A Primer on Linear Models.* Chapman & Hall/ CRC, Taylor & Francis Group, Boca Raton.

Montgomery, D. C. (2013). *Design and Analysis of Experiments. 8th Edition.* John Wiley & Sons, New York,.

Muhammed Di, N, F., Syed Yahya, S. S., and Abdullah, S. (2014). Comparing groups using robust h statistic with adaptive trimmed mean. *Sains Malaysiana*, 43(4):643 – 648.

Oehlert, G. (2010). *A First Course in Design and Analysis of Experiments.* Library of Congress Cataloging-in-Publication Data. University of Minnesota.

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. (2004). *Applied Adaptive Statistical Methods:Tests of Significance and Confidence Interval.* Society for Industrial and Applied Mathematics, Philadelphia.

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

Okyere, G. A. (2011). *Robust Adaptive Scheme for Linear MIxed Models.* PhD thesis, Western Michigan University.

Okyere, G. A., Bruku, S., Tawiah, R., and Biney, G. (2018). Adaptive robust profile analysis of a longitudinal data. *Journal of Advances in Mathematics and Computer Science*, 28(1):1 – 18.

Pitman, E. J. G. (1949). *Notes on Non-Parametric Statistical Inference.* North Carolina University. Dapartment of Statistics. Institute of Statistics Mimeo Series 27.

Randels, R. H. and Wolfe, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics.* Wiley, New York.

Rashid, M. M., McKean, J. W., and Kloke, J. D. (2012). R estimates and associated inferences for mixed models with covariates in a multi-center clinical trial. *Statistics in Biopharmaceutical Research,*, 4(1):37 – 49.

Saleem, S. and Sherani, R. A. K. (2020). Selecting and estimating rank score functions based on residuals for linear mixed models. *Punjab University Journal of Mathematics*, 52(9):47 – 63.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. in olkin, i, et al. (eds). *Contributions to Probability and Statistics, Essays in Honour of Harold Hotelling. Stanford: Stanford University Press.*, pages 448 – 485.

Twisk, J. W. R. (2013). *Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide. 2nd Edition.* Cambridge University.

Wilcox, R. R. and Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods*, 8:254 – 274.

Winsor, C. P., Hastings, Jr., C., Mosteller, F., and Tukey, J. W. (1947). "low momemts for small samples: A comparative studyy of order statistics". *Annals of Mathematical Statistics*, 18(3):413 – 426.

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

Xie, J. and Priebe, C. E. (2002). A weighted generalization of the mann-whitney-wilcoxon statistics. *Journal of Statistical Planning and Inference.*, 102:441 – 466.

Zar, J. H. (1996). *ANOVA Assumptions.Biostatistical Analysis.* New Jersey: Prentice Hall Inc.

APPENDIX A

Three different pain relief drugs were administered on 27 patients suffering from migraine headache and time of relief in minutes were recorded.

Table 1: **Time of Relief for Migraine Headache Sufferers**

Drug A	Drug B	Drug C
4	6	6
5	8	7
4	4	6
3	5	6
2	4	7
4	6	5
3	5	6
4	8	5
4	6	5

Source: **Oehlert (2010)**

APPENDIX B

Four catalyts that may affect the concentration of one component in a three-component liquid mixture are being investigated.

Table 2: **Effect of Four Catalysts on Liquid Mixture**

1	2	3	4
58.2	56.3	50.1	52.9
57.2	54.5	54.2	49.9
58.4	57.0	55.4	50.0
55.8	55.3		51.7
54.9			

Source: **Montgomery (2013)**

APPENDIX C

Four different drugs were administered on four subjects to measure their pain tolerance

Table 3: **Pain Tolerance**

	Treatment			
Subject	Drug 1	Drug 2	Drug 3	Drug 4
1	5	9	6	11
2	7	12	8	9
3	11	12	10	14
4	3	8	5	8

Source: **Twisk (2013)**

APPENDIX D

Measurements of balance errors were taken of 10 bicycle riders at five times levels of fatigue. Fatigue is a within subjects with five levels. Subjects rode for 15 minutes divided into five 3-minute periods for the purpose of collecting data.

Table 4: **Fatigue Time**

	Time				
Subject	Min 3	Min 6	Min 9	Min 12	Min 15
1	7	7	23	36	20
2	12	22	26	26	20
3	11	6	9	31	25
4	10	18	16	28	37
5	6	12	9	55	65
6	13	21	30	10	11
7	5	0	2	37	42
8	15	18	22	16	11
9	0	2	0	32	57
10	6	8	27	70	40

Source: **Davis (2002)**

APPENDIX E

The data are measurements of depression level assessed under two treatment groups over 3 time periods.

Table 5: **Mean Drepression Level**

		Time		
Id	Group	Week 1	Week 2	Week 3
1	1	35	25	12
2	1	34	22	13
3	1	36	21	18
4	1	35	23	15
5	2	31	43	57
6	2	35	46	58
7	2	37	48	51
8	2	32	45	53

Source: **Crowder and Hand (1990)**

APPENDIX F

The dataset consists of people who were randomly assigned to three different types of exercise:,at rest, walking leisurely and running. Their pulse rate was measured at three different time points during their assigned exercise at 1 minute, 15 minutess and 30 minutes.

Table 6: **Pulse Rate of Participants**

		Time		
Exercise Type	ID	1	2	3
At Rest	1	85	85	88
	2	90	92	94
	3	97	97	14
	4	80	82	83
	5	91	92	91
	6	83	83	83
	7	87	88	90
	8	92	94	95
	9	97	99	96
At Rest	10	100	97	100
Walking Leisurely	11	86	86	84
	12	93	103	104
	13	90	92	93
	14	95	95	100
	15	89	96	95
	16	84	86	89
	17	103	109	90
	18	92	96	101
	19	97	98	100
Walking Leisurely	20	102	104	103
Running	21	93	98	100
	22	98	104	112
	23	96	105	99
	24	87	132	120
	25	94	110	116
	26	95	126	143
	27	100	126	140
	28	103	124	140
	29	94	135	130
Running	30	99	111	150

Source:**Davis (2002)**