Robust Adaptive Scheme for Linear Mixed Models

Gabriel Asare Okyere
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations
Part of the Statistics and Probability Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/445
ROBUST ADAPTIVE SCHEME FOR LINEAR MIXED MODEL

by

Gabriel Asare Okyere

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Statistics
Advisor: Joseph McKean, Ph.D.

Western Michigan University
Kalamazoo, Michigan
June 2011
If the underlying distribution of a statistical model is known then a procedure which maximizes power and efficiency can be selected. For example, if the distribution of errors is known to be normal in a linear model then inference based on least squares maximizes power and efficiency. More generally, if this distribution is known then a ranked based inference based on the appropriate rank score function has maximum efficiency. In practice, though, this distribution is not known. Adaptive schemes are procedures which hopefully select appropriate methods to optimize the analysis.

Hogg (1974) presented an adaptive rank-based scheme for testing in the simple two-sample location model. It consists of a family of distribution free tests, each associated with a rank score function, and a selection procedure which chooses a test from this family. For continuous error distributions, Hogg’s scheme is valid for testing; that is, it retains the level. This, however, is not true for estimation and fitting. Little has been done to extend Hogg's scheme beyond simple location problems. In this research we extend Hogg’s adaptive scheme for testing to mixed models consisting of m clusters of observations.

This is an important practical family of models, including, for example,
repeated measure designs, multi-center clinical designs, and randomized block designs. Under the assumption of exchangeable errors, we establish that our testing scheme is valid. In practice, though, for these models fitting is crucial. Based on the fitting of a model a residual analysis can be performed to check, say, quality of fit and to determine outliers. Further, standard errors of the estimates can then be obtained so that confidence intervals for contrasts of interests can readily be formed. With this in mind, we have also developed several adaptive fitting schemes for these models. These are based on several types of rank-based estimation procedures, including joint ranking and multiple ranking types of fitting.

Our schemes are robust and highly efficient. A large Monte Carlo study over error distributions ranging from heavy-tailed to light-tailed distributions and from symmetric distributions to skewed distributions gives empirical credence to our adaptive procedures. We illustrate our procedures with several real clinical examples.
ACKNOWLEDGMENTS

I am highly indebted to my advisor, Prof. Joseph McKean, whose patience, sacrifice and support motivated me to complete this work. A research of this magnitude could not have been completed without his guidance. It was an honor to work under his shadows. Next I would like to show my gratitude to my committee members for their insightful suggestions during this research. I also wish to thank the Department of Statistics for making supercomputers and relevant facilities available to me for this research.

I am heartily thankful to my parents who have in diverse ways shown tremendous amount of commitment to my upkeep. I wish to thank my friends and anyone who have affected my life in many ways over the years.

Finally I would like to thank Mr. Wilberforce Adjei Sakyi for his support throughout my academic career.

Gabriel Asare Okyere
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .............................................................. ii

LIST OF TABLES ........................................................................ vi

LIST OF FIGURES ...................................................................... vii

CHAPTER

I  INTRODUCTION ...................................................................... 1
   1.1 Model ............................................................................. 2
      1.1.1 Location Model .......................................................... 2
      1.1.2 Mixed Model .............................................................. 3
   1.2 Methodology ..................................................................... 6

II  GENERAL THEORY OF ADAPTATION ................................. 8
   2.1 Hogg’s Adaptive Scheme ................................................ 8
      2.1.1 Two Sample Location Problem and Adaptation ........... 10
      2.1.2 Selector Statistics and Scores ..................................... 13
      2.1.3 Adaptive Test and Test Statistics ............................... 23
      2.1.4 Asymptotic Efficacies and Estimating Equations .......... 29
# Table of Contents-Continued

## CHAPTER

### III MIXED MODEL OVER $M$ CENTERS .............................................. 34

3.1 Parametric Estimation of Mixed Models ............................... 35

3.1.1 Estimation of Fixed Effect and Prediction of Random Effect ..... 35

3.1.2 Maximum Likelihood and Restricted Maximum Likelihood ..... 38

3.2 Adaptation and Linear Mixed Models ................................. 39

3.2.1 Exchangeable Random Variables ................................. 41

3.2.2 Rank-Based Estimates ........................................ 45

3.2.3 Estimation of Variance-Covariance Components ............... 49

3.2.4 Estimation of Scale Parameter ................................. 51

3.2.5 Intercept Estimation ........................................ 54

3.3 Asymptotic Efficacies .............................................. 55

3.3.1 Overall Test Statistic of Adaptation on Sample ............... 60

### IV ADAPTIVE PROCEDURES ................................................... 63

4.1 Procedures ......................................................... 63

4.1.1 Many Rankings ................................................ 63

4.1.2 Hogg-McKean Adaptation on Sample ........................... 65

4.1.3 Hogg-McKean Adaptation on Residuals ........................... 66

4.1.4 Hogg-McKean Adaptation on Sample Ignore Center .......... 68
CHAPTER

V NUMERIC EXAMPLES AND SIMULATION RESULTS 69

5.1 Numeric Example for Weight Loss Data 70

5.2 Simulations for Class of Contaminated Normal and Normal Distributions 72

5.2.1 Simulation Results for Normal 72

5.2.2 Simulation Results for Contaminated Normal CN, \( \epsilon = 0.2, \sigma = 5 \) 75

5.3 Simulations for Contaminated and Truncated Logistic 78

5.3.1 Simulation Results for Contaminated and Truncated Logistic 80

5.4 Results for Empirical Power Based on 10,000 Simulations 82

VI CONCLUSIONS 93

6.1 Summary 93

6.1.1 Future Work 94

REFERENCES 95

APPENDIX 101
LIST OF TABLES

<table>
<thead>
<tr>
<th></th>
<th>Table Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Winsorised Scores</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Type of Distribution, Benchmark along with their Selected Score</td>
<td>29</td>
</tr>
<tr>
<td>3</td>
<td>Trimmed Means</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>Results for the Weight Loss Data</td>
<td>71</td>
</tr>
<tr>
<td>5</td>
<td>Results for the Weight Loss New Data, Ndata</td>
<td>71</td>
</tr>
<tr>
<td>6</td>
<td>Empirical Levels for Normal Data, $\rho = 0, 0.2, 0.4$</td>
<td>73</td>
</tr>
<tr>
<td>7</td>
<td>Empirical Levels for Normal Data, $\rho = 0.6, 0.8$</td>
<td>74</td>
</tr>
<tr>
<td>8</td>
<td>Results for Main Effect Example - Without Outlier</td>
<td>75</td>
</tr>
<tr>
<td>9</td>
<td>Results for Main Effect Example - With Outliers</td>
<td>75</td>
</tr>
<tr>
<td>10</td>
<td>Empirical Levels for CN(2,25) Data, $\rho = 0, 0.2, 0.4$</td>
<td>76</td>
</tr>
<tr>
<td>11</td>
<td>Empirical Levels for CN(2,25) Data, $\rho = 0.6, 0.8$</td>
<td>77</td>
</tr>
<tr>
<td>12</td>
<td>Empirical Levels for CL Data, $\epsilon = 0.2, \rho = 0, 0.2, 0.4$</td>
<td>81</td>
</tr>
<tr>
<td>13</td>
<td>Empirical Levels for CL Data, $\epsilon = 0.2, \rho = 0.6, 0.8$</td>
<td>82</td>
</tr>
<tr>
<td>14</td>
<td>Power Levels of ML Method for Normal Data, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td>90</td>
</tr>
<tr>
<td>15</td>
<td>Power Levels of HMR Method for Normal Data, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td>90</td>
</tr>
<tr>
<td>16</td>
<td>Power Levels of HMS Method for Normal Data, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td>91</td>
</tr>
<tr>
<td>17</td>
<td>Power Levels of HMR Method for CL Data, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td>91</td>
</tr>
<tr>
<td>18</td>
<td>Power Levels of HMS Method CL Data, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td>92</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Plots of Scores on Table 1 for Left Skewed Distributions</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Plots of Scores on Table 1 for Symmetric Distributions</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Plots of Scores on Table 1 for Right Skewed Distributions</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Plot of Benchmarks for n=50</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Power Curves for Normal, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Power Curves for Normal, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Power Curves for Normal, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Power Curves for CL $\epsilon=0.2$, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Power Curves for CL $\epsilon=0.2$, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Power Curves for CL $\epsilon=0.2$, $\rho = 0.2, 0.4, 0.6, 0.8$</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

In this research, we develop an adaptive scheme for a mixed model over $m$ clusters of repeated measures under exchangeable errors. Hájek (1999) and Husková (1995) distinguished between non-restrictive and restrictive adaptive schemes. For the non-restrictive adaptive scheme, the optimal scores for the locally most powerful rank test which depends on an unknown distribution function $F$ and its density $f$ are estimated directly from the data. In the restrictive case, a test is selected from a finite family of distributions with their regression scores, $\varphi_i$.

Al-Shomrani (2003) in his PhD thesis did a comprehensive work on score estimation for some selected distributions for a general linear model. He considered Policello and Hettmansperger’s (1976) robust adaptive procedure which is based on a simple family of rank statistics for one sample location problem. He also considered Kapenga and McKean’s (1989) and Naranjo and McKean’s (1997) robust adaptive procedures which depend on the residuals of initial fits to estimate scores. Kapenga and Mckean’s procedure estimates an optimal score using the ordered residuals of an initial robust fit. Naranjo and McKean’s procedure uses a bandwidth which depends on the sample size to estimate a score function based on its fourier expansion.

In this work we will apply the restrictive case, Hogg (1974, 1975). Our main adap-
tive scheme contains nine Winsorised scores that are classified by tail weight and skewness. Our selector is an extension of Hogg’s (1974) selector statistic. Hogg’s (1974), scheme is for simple location models. In the simple location models, Hogg’s selection rule ensures, under the null hypothesis, that this test is distribution free. Our model of interest is a mixed model. We briefly describe the location and mixed model next.

1.1 Model

1.1.1 Location Model

Consider two independent samples of sizes $n_1$ and $n_2$ from a population with continuous distribution functions $F(x)$ and $F(\frac{x-\Delta}{\vartheta})$ for $\vartheta, \Delta \in R$ respectively. This is a location and scale problem. If $\vartheta=1$, then we have a location problem and the two independent samples differ by $\Delta$. The usual form for this location problem is

$$Y = \mu + \Delta C + e$$

where $Y$ is a vector of the combined samples, $C$ is a design matrix of 0’s (sample 1) and 1’s (sample 2) and $e$ is an independent identically distributed random errors. The most common non parametric and nonadaptive test for $\Delta$, is the Wilcoxon-Mann-Whitney test, see Chapter 2 of Hettmansperger and McKean (2011). If we consider the block random effect, $b$, then we have a mixed model.
1.1.2 Mixed Model

A model of choice in most preclinical and clinical trials is the linear mixed model. Mixed models, often called cluster correlated models are used in analyzing repeated measures, split plot design, and randomized block design. Mixed models have at least one fixed effect and more than one random effects. For example, in a clinical trial, the investigator might be interested in comparing two treatment regimes, say, placebo and a new drug. For such a problem, subjects are randomly assigned to each group, then measurements are taken on each individual at an initial time before the treatment is administered. After administering the treatment, measurements are taken repeatedly over specified periods. Final measurements are taken at the end of the trial. For such experiments, each subject is its own control. Data collected from this type of experiment are called longitudinal data.

Usually such trials are replicated at different centers, blocks, or clusters. So we have a multi-center clinical trial. If we consider centers as random, then we have, in addition to $e$, more than one random term. That is, a model of choice is the mixed model. Mixed model appeals this type of data because within cluster population average curve could be estimated. This is the type of problem we are interested in for this work.

Linear mixed models may be expressed in many equivalent forms. Consider the Laird-Ware’s (1982) matrix representation form,

$$Y = \mu + X\beta + Zb + e,$$

where $\beta = (\Delta, \beta^*)'$ is a vector of fixed effects parameters, $\beta^*$ is a vector of covariate fixed...
effect parameters, \( X = (C', X^*)' \) is the model design, \( b \) is block random effects vector, \( Z \) is an incidence matrix denoting center membership, and \( e \) is a random vector. If there are no covariates, we have

\[
Y = \mu + \Delta C + b + e,
\]

The usual assumptions placed on \( b \) and \( e \) are

\[
b \sim N(0, \Psi)
\]

and \( e \sim N_n(0, \Gamma). \)

where \( \Psi \) is the covariance matrix of block random effect and \( \Gamma \) is the covariance matrix of the errors.

In literature various parametric and non parametric procedures are used in testing and estimating the fixed effect parameter. Under the normality assumptions, the parametric estimates of the fixed effect and the random effect are obtained using the maximum likelihood (ML) or restricted maximum likelihood (REML) estimates. A description of this approach is done in Chapter 3. Once the estimate of the fixed effect and the random effect are obtained, the curves for population average and the marginal curves for each center could be obtained. The problem, however, is that the true model is not known and that the parametric model may be misspecified. Misspecification may come from the distribution assumptions or incorrect covariance structure. Under these assumptions, if the linear mixed model is correctly specified, then the parametric approach is more efficient than the non parametric approach. However, the distributional assumptions or the covariance structure
may be misspecified. To overcome this problem, nonparametric approaches are proposed in literature. Estimation of the fixed and random effect are done using locally weighted mixed model or panelized weighted least squares (PWLS), Bates (2010). Another non-parametric method which was proposed by Rashid et. al. (2011) is many rankings (MRW). This method is a rank-based fitting-type method. It ignores the normality assumption for the random components of the model and estimates the fixed effect parameter by solving a gradient function. The estimates of the fixed effect take into account the dependent error structure within clusters. Usually, Wilcoxon linear scores are used for this process.

The problem with the nonparametric methods is that, under normality, they have higher variance and there could be data irregularities. Rashid et. al’s (2011) method, which ignores the normality assumption and uses Wilcoxon linear score is highly efficient if the error distributions are same from center to center. However, the distribution of random errors , e, from each center in a multi-center clinical trial, may differ from center to center. In particular the underlying structure, tail weight and skewness, of the unknown distributions of the random error term may differ from center to center. Furthermore centers have similar characteristics induced by block random effect. That is within centers or hospitals, treatment protocols are similar. Implementation of these protocols coupled with health care delivery could affect the unknown underlying structure of e. Consequently, test and estimation could be severely impaired and lose efficiency. Therefore there is a need to device an adaptation for linear mixed model under exchangeable errors.

The main objective is to develop a robust adaptive test for
\[ H_0 : \Delta = 0 \]
\[ H_a : \Delta > 0. \]

We further develop estimation procedures for \( \Delta \). We then investigate the power of the adaptive procedures.

### 1.2 Methodology

Six procedures will be considered in this research. Four of which are adaptive procedures, one nonadaptive and nonparametric procedure and one parametric procedure. The adaptive procedures are Hogg-McKean (HMS) adaptation on sample, Hogg-McKean (HMR) adaptation on residuals, Many ranking (MR) and Hogg-Ignoring-Center (HMIC) adaptation on sample. The theory of adaptation will be discussed in Chapter 2.

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 are discussed in Chapters 2 and 4. 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 initial fit is done using lme( linear mixed effects) function in the nlme library in R. Note that in future we will be using a robust fit as our initial fit. The nonparametric procedure, called many
rankings Wilcoxon (MRW) is similar to the many ranking (MR) except that Wilcoxon score is used for each center.

This research divides naturally into four blocks. The first two Chapters provide an introduction, motivation for this research and extensive review of adaptive schemes. Chapters 3 and 4 develop adaptation to linear mixed models and formal algorithm for the adaptive procedures. Testing and estimating procedure for our schemes are also formulated. Further, the asymptotic theory for the estimates and efficacy results are presented. Chapter 5 consists of numeric examples and simulation results. The final chapter presents a discussion of the results, and future work. We begin by reviewing the general theory of adaptation.
CHAPTER II

GENERAL THEORY OF ADAPTATION

2.1 Hogg's Adaptive Scheme

In this section we review the general theory of Hogg-type adaptation. In adaptation, we seek selector statistics. 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, 1976). Suppose we are sampling 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 $\alpha$ for all continuous distributions. Next we state the main theorem behind adaptation.

**Theorem 2.1.1.** (Herbert Büning)

1. Let $\mathcal{F}$ denote the class of distribution functions under consideration. Suppose that each of the $r$ tests based on the statistics $T_1, \cdots, T_r$ is distribution-free over the class $\mathcal{F}$; ie $P_{H_0}(T_i \in C_i|F) = \alpha$ for each $F \in \mathcal{F}$, $i = 1, \cdots, r$, $C_i$ is the critical region of $T_i$.

2. Let $Q$ be some statistic that is independent of $T_1, \cdots, T_r$ under $H_0$ for each $F \in \mathcal{F}$.
Suppose we use $Q$ to decide which test $T_i$ to conduct. Specially, let $S$ denote the set of all values of $Q$ with the following decomposition:

$$S = D_1 \cup D_2 \cup \cdots \cup D_r, \quad D_i \cap D_j = \emptyset \text{ for } i \neq j,$$

so that $Q \in D_i$ corresponds to the decision to use the test $T_i$.

Then a test based on (1) and (2) is distribution-free.

That is,

$$P_{H_0}(\text{reject } H_0 | F) = \sum_{i=1}^{r} P_{H_0}(Q \in D_i, T_i \in C_i | F)$$

$$= \sum_{i=1}^{r} P_{H_0}(Q \in D_i | F) \cdot P_{H_0}(T_i \in C_i | F)$$

$$= \sum_{i=1}^{r} P_{H_0}(Q \in D_i | F) \alpha$$

$$= 1. \alpha$$

$$= \alpha.$$

So the procedure of selecting $T_i$ using an independent statistics $Q$ and then constructing a test of significance level $\alpha$ with test statistic $T_i$ has an overall significance level $\alpha$. Hence the overall testing procedure is defined by, if

$$Q \in D_i \text{ then reject } H_0 \text{ if } T_i \in C_i.$$ 

To further illustrate this theorem, we consider adaptation and two sample location problem next.
2.1.1 Two Sample Location Problem and Adaptation

In this section, our goal is to review the concept of adaptation within in the context of two sample location problem. Let's briefly consider a two independent sample problem.

Let $X_1, \ldots, X_{n_1}$ be independent and identically distributed with distribution $F_X = F(x)$. Let $Y_1, \ldots, Y_{n_2}$ also be independent and identically distributed with $F_Y = F(x - \Delta)$. Assume $X_1, \ldots, X_{n_1} \sim F_X = F(x)$. Let $Y_1, \ldots, Y_{n_2}$ are mutually independent. We want to test,

$H_0 : \Delta = 0$ versus $H_a : \Delta > 0$.

Under $H_0$, the distribution of $X_i$ is same as distribution of $Y_j$. We combine the two samples into one as follows. Define

$$Z_i = \begin{cases} 
X_i, & 1 \leq i \leq n_1 \\
Y_{i-n_1+1}, & n_1 + 1 \leq i \leq n = n_1 + n_2
\end{cases} \quad (2.1)$$

Let $Z_{(i)}$ denote the order statistics for $Z_i$. Under $H_0$, given $Z_{(i)}$, the conditional distribution of $Z_i$ is discrete with probability $\frac{1}{n!}$ on each of the $n!$ permutations of the vector $Z_i$. That is the conditional distribution does not depend on $F(x)$. Hence by definition of sufficiency the order statistics are sufficient for $F$, see Chapter 7 of Hogg, McKean and Craig, (2005). Next we consider the completeness of nonparametric family as in Bhattacharyya, Johnson and Mehrotra (1977).

**Theorem 2.1.2.** Let $\mathcal{F}$ be a family of probability distributions for a real or vector random variable $Z$ and let $W = h(Z)$ be a measurable function with $\mathcal{F}_W$ denoting the induced
family of distributions of $W$. If $Z$ is complete with respect to the family $\mathcal{F}$, then $W$ is complete with respect to $\mathcal{F}_W$.

Under $H_0$, $Z_{(1)} < \cdots < Z_{(n)}$, the ordered observations for the combined sample has a common distribution $F \in \mathcal{F}$ where $\mathcal{F}$ is the family of all distributions absolutely continuous with respect to Lebesgue measure, See Bhattacharyya et.al (1977). If we consider the function $h$ such that $h(Z_1 \cdots Z_n) = (Z_{(1)} \cdots Z_{(n)})$ then by Theorem 2.1.2, the combined ordered sample $Z_{(1)} < \cdots < Z_{(n)}$ is complete for $\mathcal{F}$. It is worth mentioning that within blocks the order statistics are complete with respect to the corresponding induced family of distributions.

**Corollary 2.1.1.** Let $X_{(1)} < \cdots < X_{(n)}$ and $Y_{(1)} < \cdots < Y_{(n)}$ be the order statistics for each sample with distributions $F_X$ and $F_Y$ respectively, then $X_{(1)} < \cdots < X_{(n)}$; $Y_{(1)} < \cdots < Y_{(n)}$ are complete with respect to the family distributions induced by $F_X$ and $F_Y$, for $F_X, F_Y \in \mathcal{F}$.

The above corollary allow us to state the next theorem. 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 2.1.3.** Under $H_0 : F_X = F_Y$, the order statistics for the combined sample, $Z_{(1)} < \cdots < Z_{(n)}$ 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 \( T = T[R(X, Y)] \) be such a statistic whose distribution is free of \( F \) under \( H_0 \). Then from Basu's theorem, see appendix, \( T \) is a test statistic and ancillary for \( F \). Furthermore, if the ordered combined sample, \( Z_1, \ldots, Z_n \) is complete and sufficient for \( F \), then it follows that for all measurable functions \( G, G(Z_1, \ldots, Z_n) \) is also complete and sufficient for \( F \). See Appendix.

**Theorem 2.1.4. (Basu's)**

Let \( T = T[R(X, Y)] \) be a statistic whose distribution is free of \( F \), then under \( H_0 \), \( T \) and \( G(Z_1, \ldots, Z_n) \) are independent, for all (measurable) functions \( G \).

Here \( T \)'s are rank tests, thus, \( T = T[R(X, Y)] = \{T_1, \cdots, T_r\} \). This means that the test statistics depend on the joint ranks of \( Z_i \)'s. Applying the Theorem 2.1.1 to the two sample problem, \( \mathcal{F} \) is the class of all continuous distribution functions of \( F \) and \( T_1, \cdots, T_r \) are rank test statistics. So \( T_i \) is distribution-free over \( \mathcal{F} \), for \( i = 1, \cdots, r \). And \( Q \) is a function of the order statistics of the combined sample.

Under \( H_0 \), the order statistics are complete and sufficient for the common, but unknown distribution \( F \), and therefore independent of every statistic whose distribution is free of \( F \), see Basu's theorem. 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.
2.1.2 Selector Statistics and Scores

In practice the distribution $F$ is unknown. So we consider a class of distributions which are classified by the skewness and tail weight. In most proposals four categories are preferred, three for symmetric distributions (short, medium and long tail) and one for distributions skewed to the right see (Herbert Büning, 2009). However we will restrict the categories $D_i$’s to the nine Winsorised. Suppose for some class of distribution, $\mathcal{F}$, the tail weight of $F \in \mathcal{F}$ can vary from light to heavy and may be skewed, then they could well be classified by their skewness and tail weight. So we need a pair of selector statistics $Q_1$ and $Q_2$ to measure the skewness and tail weight respectively. Such a selector statistics must be a function of the order statistics. So that, under $H_0$, the pair selector statistics is independent of each $T_i$’s. A pair of selector statistics which is independent of rank test statistics for the two-sample location problem are proposed in the literature, see Hogg (1974, 1982), Hogg, Fisher and Randles (1975), Ruberg (1986) and Büning (1994, 2009). In this research we use Hogg’s selectors. Hogg (1974) used a pair of selector statistics, $Q_1$ and $Q_2$, which are measures of skewness and tail weight respectively. The measure of skewness $Q_1$ is

$$Q_1 = \frac{U_{.05} - M_{.5}}{M_{.5} - L_{.05}},$$

(2.2)
where $\overline{U}_{.05}$, $\overline{M}_{.5}$, and $\overline{L}_{.05}$ are the averages of the largest 5% of the $Z$s, the middle 50% of the $Z$s, and the smallest 5% of the $Z$s, respectively. The measure of tail weight $Q_2$ is

$$Q_2 = \frac{\overline{U}_{.05} - \overline{L}_{.05}}{\overline{U}_{.5} - \overline{L}_{.5}}. \quad (2.3)$$

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

$$\overline{A}_{\alpha_1} - \overline{B}_{\alpha_2}. \quad (2.4)$$

where $\alpha_1$ and $\alpha_2$ are some fraction to be trimmed from the combined ordered data. Let

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

where $Z_{(i)}$'s are ordered combined sample $t_1 = [n\alpha_1]$, $t_2 = [n\alpha_2]$, $[x]$ denotes the smallest integer greater than $x$, $l = n - t_1 - t_2$ and redefine a measure of skewness $Q_1^*$ and tail weight $Q_2^*$ by

$$Q_1^* = \frac{(m(0.95, 0) - m(0.25, 0.25))}{(m(0.25, 0.25) - m(0, 0.95))} \quad (2.5)$$

$$Q_2^* = \frac{(m(0.95, 0) - m(0, 0.95))}{(m(0.5, 0) - m(0, 0.25))}. \quad (2.6)$$

Suppose we want to adapt on residuals, then we need the combined ordered residuals from an initial fit. The measures of tail weight and skewness of the residuals are obtained by
using $Q_1^*$ and $Q_2^*$ respectively.

In this research, we make use of the benchmarks proposed in the dissertation of Al-Shomrani (2003). In his thesis, the cutoff values for the measures of skewness and tail weight depend on the sample size $n$. This is a modified version of Hogg's (1975). However it converges to Hogg's (1975) as $n \to \infty$. For $Q_1^*$ we have,

\[
\text{lower cutoff} = 0.36 + (0.68/n) \quad (2.7) \\
\text{upper cutoff} = 2.73 - (3.72/n) \quad (2.8)
\]

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

\[
\text{lower cutoff} = 2.17 - (3.01/n) \quad (2.9) \\
\text{upper cutoff} = 2.63 - (3.94/n) \quad (2.10)
\]

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

\[
\text{lower cutoff} = 2.24 - (4.68/n) \quad (2.11) \\
\text{upper cutoff} = 2.63 - (9.37/n). \quad (2.12)
\]

These cutoff points are used to select a rank test which is based on a rank score function corresponding to an unknown distribution. Different scores based on tail weight and/or skewness have been proposed in literature. Most of these scores are selected de-
pending on tail weight and/or skewness. The rank tests that we consider in this thesis are of the form,

$$T_\varphi = \sum_{j=1}^{n} \varphi \left[ \frac{R(Z_j)}{n+1} \right] I(Z_j = Y_j)$$  \hfill (2.13)

where $\varphi$ satisfies the following conditions;

- (S.1) $\varphi$ nondecreasing function and square-integrable on (0,1)
- (S.2) $\varphi$ is differentiable on (0,1)

Since $\varphi$ is square integrable, we assume without loss of generality that,

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

Note that a test statistic is synonymous with score function. We use the two interchangeably. We may also write $a_\varphi(t) = \varphi \left[ \frac{t}{n+1} \right]$ and think of $a_\varphi(1), \cdots, a_\varphi(n)$ as scores. As discussed in chapter 2 of Hettmansperger and McKean (1998), for model

$$Z = \Delta C_t + e_t,$$

where $e_t$ 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 ef-
ficient. For example, Randles and Wolf (1979), Gastwirth (1965), Büning (1994, 1996) proposed rank test based on scores corresponding to some selected distributions. They showed that the scores below with the type of distribution in parenthesis have high power see Büning (2005) over their targeted area of distribution.

Example 2.1.1. (Short Tails)

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

Example 2.1.2. Wilcoxon (Medium Tails)

\[ a(t) = t \]

Example 2.1.3. Hogg Fisher Randles Test (Right Skewed)

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

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

1. \[
\varphi_I(u) = \begin{cases} 
  s_3, & u > s_1 \\
  s_3 + \frac{s_3-s_1}{s_1}(u - s_1), & \text{otherwise}
\end{cases}
\]

2. \[
\varphi_{II}(u) = \begin{cases} 
  -\frac{s_4}{s_1}(u - s_1), & u < s_1 \\
  -\frac{s_4}{s_2-1}(u - 1) + s_4, & u > s_2 \\
  0, & \text{otherwise}
\end{cases}
\]
3.

\[ \varphi_{III}(u) = \begin{cases} 
    s_2, & u < s_1 \\
    s_3 + \frac{s_4-s_3}{s_1-1}(u-1), & \text{otherwise}
\end{cases} \]

4.

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

where \( s_1, s_2, s_3, s_4 \) and \( s_5 \) are parameters and \( a_i(t) = \varphi_1\left(\frac{t}{n+1}\right) \). Table 1 shows distributions and scores with their corresponding parameters.

We may want to adapt on both the samples and residuals (after initial fit). In the case of adaptation on residuals, initial fit is done. Then the residuals are used for the adaptation. The adaptation based on samples under \( H_0 \) has exact level \( \alpha \), while the adaptation based on residuals does not necessarily retain the \( \alpha \) level. Figures (1-3) are the scores we will use for our adaptation.
Figure 1

Plots of Scores on Table 1 for Left Skewed Distributions
Figure 2

Plots of Scores on Table 1 for Symmetric Distributions
Figure 3

Plots of Scores on Table 1 for Right Skewed Distributions
Table 1

Winsorised Scores

<table>
<thead>
<tr>
<th>Tail Weight</th>
<th>Skewness</th>
<th>Score Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light</td>
<td>Left</td>
<td>$\varphi_1=\varphi_{III}$ with parameter($s_1=0.1$, $s_2=-1$ and $s_3=2.0$)</td>
</tr>
<tr>
<td>Medium</td>
<td>Left</td>
<td>$\varphi_2=\varphi_{III}$ with parameter($s_1=0.3$, $s_2=-1$ and $s_3=2.0$)</td>
</tr>
<tr>
<td>Heavy</td>
<td>Left</td>
<td>$\varphi_3=\varphi_{III}$ with parameter($s_1=0.5$, $s_2=-1$ and $s_3=2.0$)</td>
</tr>
<tr>
<td>Light</td>
<td>Symmetric</td>
<td>$\varphi_4=\varphi_{III}$ with parameter($s_1=0.25$, $s_2=0.75$, $s_3=-1$, $s_4=1$ $s_5=0.0$)</td>
</tr>
<tr>
<td>Medium</td>
<td>Symmetric</td>
<td>Wilcoxon Scores $\varphi_5=\sqrt{12\left[u - \frac{1}{2}\right]}$</td>
</tr>
<tr>
<td>Heavy</td>
<td>Symmetric</td>
<td>$\varphi_6=\varphi_{IV}$ with parameter($s_1=0.25$, $s_2=0.75$, $s_3=-1$ and $s_4=1$)</td>
</tr>
<tr>
<td>Light</td>
<td>Right</td>
<td>$\varphi_7=\varphi_{I}$ with parameter($s_1=0.9$, $s_2=-2$ and $s_3=1.0$)</td>
</tr>
<tr>
<td>Medium</td>
<td>Right</td>
<td>$\varphi_8=\varphi_{I}$ with parameter($s_1=0.7$, $s_2=-2$ and $s_3=1$)</td>
</tr>
<tr>
<td>Heavy</td>
<td>Right</td>
<td>$\varphi_9=\varphi_{I}$ with parameter($s_1=0.5$, $s_2=-2$ and $s_3=1.0$)</td>
</tr>
</tbody>
</table>
2.1.3 Adaptive Test and Test Statistics

We present an adaptive test in this subsection. We begin by defining regions based on selector statistic $S$ corresponding to tail weight and skewness of a distribution. Define nine regions $D_k$, for $k = 1, \ldots, 9$, which depends on the selector statistics $S = \{Q_1^*, Q_2^*\}$ by,

\[
D_1 = \{Q_1^* < \hat{Q}_{11}, Q_2^* > \hat{Q}_{2u}\}
\]
\[
D_2 = \{\hat{Q}_{11} < Q_1^* < \hat{Q}_{1u}, Q_2^* > \hat{Q}_{2u}\}
\]
\[
D_3 = \{Q_1^* > \hat{Q}_{1u}, Q_2^* > \hat{Q}_{2u}\}
\]
\[
D_4 = \{Q_1^* < \hat{Q}_{11}^*, \hat{Q}_{21}^* < Q_2^* < \hat{Q}_{2u}\}
\]
\[
D_5 = \{\hat{Q}_{11} < Q_1^* < \hat{Q}_{1u}^*, \hat{Q}_{21}^* < Q_2^* < \hat{Q}_{2u}\}
\]
\[
D_6 = \{Q_1^* > \hat{Q}_{1u}^*, \hat{Q}_{21}^* < Q_2^* < \hat{Q}_{2u}\}
\]
\[
D_7 = \{Q_1^* < \hat{Q}_{11}^*, Q_2^* < \hat{Q}_{2l}\}
\]
\[
D_8 = \{\hat{Q}_{11} < Q_1^* < \hat{Q}_{1u}, Q_2^* < \hat{Q}_{2l}\}
\]
\[
D_9 = \{Q_1^* > \hat{Q}_{1u}^*, Q_2^* < \hat{Q}_{2l}\}
\]

where $\hat{Q}_{11}, \hat{Q}_{1u}, \hat{Q}_{2l},$ and $\hat{Q}_{2u}$ are benchmarks from the ordered samples or residuals (from initial fit), see Al-Shomrani’s PhD thesis (2003). Each region identifies a type of score with their corresponding parameters see Table 1 for distributions with their classifications. The
Settings of Q_1 and Q_2 for n = 50

Figure 4
Plot of Benchmarks for n=50
regions are shown on Figure (4). Let $D_k$ and $\varphi_k$ be a region and score selected respectively.

Then the adaptive test, $AD(S, \varphi)$, is

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

(2.14)

where

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

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

$$E_{H_0}[T_{\varphi_k}] = \sum_{i=1}^{n_2} E_{H_0}[a_{\varphi_k}(R(y_i))]$$

$$= \sum_{i=1}^{n_2} \sum_{j=1}^{n} a_{\varphi_k}(j) \frac{1}{n} = 0,$$

because the ranks of $y_i$'s are uniform on the integers $1, 2, \ldots, n$ 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.
\[ \text{Var}_{H_0}[T_{\varphi_k}] = E_{H_0}[T_{\varphi_k}^2] = \sum_{i=1}^{n_2} \sum_{i' = 1}^{n_2} E_{H_0}[a_{\varphi_k}(R(y_i))a_{\varphi_k}(R(y_{i'}))] \]

\[ = \sum_{i=1}^{n_2} a_{\varphi_k}^2(R(y_i)) + \sum_{i' \neq i} E_{H_0}[a_{\varphi_k}(R(y_i))a_{\varphi_k}(R(y_{i'}))] \]

\[ = \left\{ \frac{n_2}{n} - \frac{n_2(n_2 - 1)}{n(n - 1)} \right\} s_a^2 \]

\[ = \frac{n_1 n_2}{n(n - 1)} s_a^2 \]

where \( E_{H_0}[a_{\varphi_k}^2(R(y_i))] = \frac{1}{n} s_a^2 \), see Hettmansperger and McKean (1998) 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, the \( T_{\varphi_k} \)-statistics is based on the ranks only, and asymptotic critical values are used. Thus for the region \( D_k \), the corresponding asymptotic decision rule at level \( \alpha \) is reject \( H_0 \) if

\[ \left| \frac{T_{\varphi_k}}{\sqrt{\text{Var}_{H_0}(T_{\varphi_k})}} \right| \geq z_{\alpha/2} \]

Note this test is for a two-sided test. As an illustration, we make use of an example in Hogg, McKean and Craig (2005, pages 577-580). Let \( C_{1up}, C_{1lw}, C_{2up} \) and \( C_{2up} \) be some prespecified empirical cutoffs of the measures of skewness and tail weight for a class of distributions. Such that the following rules are needed for score selection:
• $Q_1 \geq C_{1up}$ indicates right-skewed distributions

• $Q_1 < C_{1lw}$ indicates left-skewed distributions

• $C_{1lw} < Q_1 < C_{1up}$ indicates symmetric distributions

• $Q_2 \geq C_{2up}$ indicates heavy tailed distributions

• $Q_2 < C_{2lw}$ indicates light tailed distributions

• $C_{2lw} < Q_2 < C_{2up}$ indicates moderate tailed distributions.

We make use of benchmarks proposed by Hogg et al (1975) in this illustration. Let's consider some test statistics which depend on the ranks of the combined sample. Thus let $R(X_1), \ldots, R(X_{n_1}), R(Y_1), \ldots, R(Y_{n_2})$ denote the combined ranks of $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$. Define the elements of a set of test statistics by:

$$T_i = \sum_{j=1}^{n_2} a_i[R(Y_j)], \quad (2.15)$$

where $a_i(t) = \varphi_i(\frac{t}{n+1})$ and $\varphi_i$ is an associated score optimal for a specified distribution. Suppose we think of $F(x)$ can vary from light to heavy one may consider the Wilcoxon test:

$$T_1 = \sum_{j=1}^{n_2} a_1[R(Y_j)], \quad (2.16)$$

where $a_1(t) = \varphi_1(\frac{t}{n+1})$ 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:

\[ T_2 = \sum_{j=1}^{n_2} a_2[R(Y_j)], \]  \hspace{2cm} (2.17)

where \( a_2(t) = \varphi_2(\frac{t}{n+1}) \) and \( \varphi_2(u) = sgn(u-0.5) \) is an associated optimal score for the double exponential distribution. Suppose we are considering right-skewed distributions, we want to down weight the right side.

\[ T_3 = \sum_{j=1}^{n_2} a_3[R(Y_j)], \]  \hspace{2cm} (2.18)

where \( a_3(t) = \varphi_3(\frac{t}{n+1}) \) and \( \varphi_3(u) \) is an associated optimal score for such a distribution. Suppose we are considering light-tailed distributions, we want to emphasize the extreme order statistics.

\[ T_4 = \sum_{j=1}^{n_2} a_4[R(Y_j)], \]  \hspace{2cm} (2.19)

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

1. \( a(t) = \begin{cases} 
1, & t > \frac{n+1}{2} \\
0, & \text{otherwise}
\end{cases} \)
Table 2

Type of Distribution, Benchmark along with their Selected Score

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Distribution</th>
<th>Score Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_2 &gt; 7$</td>
<td>Heavy-tailed symmetric</td>
<td>$\varphi_2(u)$ score (1)</td>
</tr>
<tr>
<td>$Q_1 &gt; 2$ and $Q_2 &lt; 7$</td>
<td>Right-skewed</td>
<td>$\varphi_3(u)$ score (2)</td>
</tr>
<tr>
<td>$Q_1 \leq 2$ and $Q_2 \leq 2$</td>
<td>Light-tailed symmetric</td>
<td>$\varphi_4(u)$ score (3)</td>
</tr>
<tr>
<td>elsewhere</td>
<td>Moderate heavy-heavy-tailed</td>
<td>$\varphi_1(u)$ score (4)</td>
</tr>
</tbody>
</table>

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

3. $a(t) = \begin{cases} 
  t - \left[ \frac{n+1}{2} \right] - \frac{1}{2}, & t \leq \frac{n+1}{4} \\
  t - n + \left[ \frac{n+1}{4} \right] - \frac{1}{2}, & t \geq n - \left[ \frac{n+1}{4} \right] + 1 \\
  0, & \text{otherwise} 
\end{cases}$

4. $a(t) = \begin{cases} 
  t, & 1 \leq t \leq n \\
  0, & \text{otherwise} 
\end{cases}$

2.1.4 Asymptotic Efficacies and Estimating Equations

In this subsection we examine the asymptotic efficacies which are vital for asymptotic theory of our estimates. Suppose after the adaptation, $\varphi_k$ is a score selected for region $k$. The asymptotic Pitman efficacies of $T_{\varphi_k}$ under the alternative $\Delta_n = \frac{\delta}{\sqrt{n}}$ is as follows.
We have

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

so

\[ \mu_{\varphi_k}(\Delta) = E_\Delta[T_{\varphi_k}(0)] = E_0[T_{\varphi_k}(-\Delta)] = \frac{1}{n} E_0 \sum_{i=1}^{n_2} a_{\varphi_k}(R(y_i + \Delta)) \]

Under \( H_0 \),

\[ \mu_{\varphi_k}(0) = E_0[T_{\varphi_k}(0)] = 0. \]

Suppose we have \( \hat{F}_{n_1} \) and \( \hat{F}_{n_2} \) as the empirical CDF of the two independent samples respectively, then

\[ R(y_i + \Delta) = n_1 \hat{F}_{n_1}(y_i + \Delta) + n_2 \hat{F}_{n_2}(y_i) \]

this implies

\[ \mu_{\varphi_k}(\Delta) = \frac{1}{n} \sum_{i=1}^{n_2} E_0 \left\{ \frac{n_1}{n + 1} \hat{F}_{n_1}(y_i + \Delta) + \frac{n_2}{n + 1} \hat{F}_{n_2}(y_i) \right\}. \]
For efficacy we need the derivative of $\mu$, thus we have,

$$
\mu'_{\varphi_k}(\Delta) = \lambda_2 \int_{-\infty}^{\infty} \varphi_k' \{\lambda_2 F(y) + \lambda_1 F(y + \Delta)\} f(y) dy
$$

$$
= \lambda_1 \lambda_2 \tau_{\varphi_k}^{-1}
$$

see Hogg, Mckean and Craig (2005) for a detailed proof. Thus the efficacy $c_{\varphi_k}$ of $T_{\varphi_k}$ is

$$
c_{\varphi_k} = \lim_{n \to \infty} \frac{\mu'_{\varphi_k}(0)}{\sqrt{n\sigma_{\varphi_k}}}
$$

where $\sigma_{\varphi_k}^2 = \text{Var}_0(T_{\varphi_k}(0))$. For a sequence of alternatives, the $\alpha$ level asymptotic test based on $T_{\varphi_k}(0)$ has a power function define by,

$$
\lim_{n \to \infty} \beta_{\varphi_k}(\Delta_n) = 1 - \Phi\left(z_{\alpha/2} - c_{\varphi_k} \delta\right)
$$

where $\Phi$ is a standard normal. The corresponding estimator of $\Delta$ solves the estimating equations

$$
T_{\varphi_k}(\hat{\Delta}) = 0
$$

The asymptotic distribution of the estimate $\hat{\Delta}$ is approximate normal with mean $\Delta$ and vari-
As a real life example, we consider a data set from Rhoads see Laird (2004). The data, from 100 children, are made blood lead levels from treatment of lead exposed children trial. The children were randomly assigned to a placebo group and treatment (chelation) group. Measurements were taken at the baseline (week 0), week 1, week 4, and week 6. We chose our response to be the difference in measurements at baseline and week 6. Suppose we are interested in testing an effect \( \Delta \); that is

\[
H_0: \Delta = 0 \\
\text{versus} \\
H_0: \Delta \neq 0,
\]

where \( \Delta \) is the difference in treatment effect between the placebo and chelation.

**Results for the Blood Lead Levels Data**

distribution (score)

"SH(6)"

selector

0.7882984 3.3411101

cutoffs

0.3668 2.6928 2.1932 2.8563

Trimmed means
Table 3

Trimmed Means

<table>
<thead>
<tr>
<th>$\hat{U}_{0.05}$</th>
<th>$\hat{L}_{0.05}$</th>
<th>$\hat{M}_{0.5}$</th>
<th>$\hat{U}_{0.5}$</th>
<th>$\hat{L}_{0.05}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.95</td>
<td>-11.72</td>
<td>4.312</td>
<td>8.451</td>
<td>-0.13</td>
</tr>
</tbody>
</table>

16.95000 -11.72000 4.31200 8.45098 -0.13000

test statistic 16.77052
z test statistics is 3.345706
P_value is 0.0008207347
estimated effect is 3.3
a 95% confidence interval is (1.4, 5.25)
estimated tau is 5.030221

So the estimated selector statistics is $S = \{0.7882984, 3.3411101\}$. This means that score 6 is selected. From our results, the adaptive test statistic is 16.77052 with an approximate $p$-value 0.00082. The estimated effect is 3.3 with a 95% confidence interval (1.4, 5.25).

In the next section we consider the general theory of mixed models and adaptation under exchangeable errors.
CHAPTER III

MIXED MODEL OVER \( M \) CENTERS

Clinical trials to compare treatment groups are often conducted at different clusters (sites or centers). For each site patients (clients) are randomly assigned to treatment groups. Treatments are replicated several times at each center. The centers are often drawn from different areas. They may even be nationwide or partly international. This allows for investigations over a good representative sample and conditions (different medical practices). We presume that sites are randomly selected, so that the design is viewed as a mixed model with treatment groups as one factor (fixed factor) and the random centers (stratified factor). For example one might be interested in the difference in effect of a placebo and new drug, Succimer, when given orally to treat lead-exposed children over different centers. That is measures of blood lead levels are taken repeatedly overtime at different centers. For each site patients (clients) are randomly assigned to the treatment groups.

Chakravarty and Grizzle (1975), Boos and Brownie (1992), Khatri and Patel (1992) and Rashid (2003), and Kloke, McKean and Rashid (2009) discussed how multi-center clinical trials could be analyzed using ranks for mixed models. In this research, we present adaptive schemes for estimating and testing fixed effects for mixed models. We begin by showing that adaptation works for mixed models. To the best of our knowledge little work has been done in this area. We start by reviewing some parametric estimation of mixed
model.

3.1 Parametric Estimation of Mixed Models

The usual linear model has only 1 random term, the random error. In multi-center clinical trials, if we assume centers are random, then we have two random effects- center and error. The generic term linear model is consistent with linear models because the fixed and random components are additive and each term is in linear form.

3.1.1 Estimation of Fixed Effect and Prediction of Random Effect

Consider the matrix representation of the $i$th center or block model

$$Y_i = \mu 1_n + X_i \beta + b_i + \varepsilon_i$$  \hspace{1cm} (3.1)

where $Y$ is combined response vector at center $i$, $X$ is the model matrix for the fixed effects, $\beta$ is a vector of fixed effect unknown parameters, $b$ is the $i$th block random effect and $\varepsilon$ is also random experimental error.

Let $\mathcal{Y}$ be stack combined samples over all centers. Thus for $i = 1, \ldots, m$, let $n_i$ be the number of patients at the $i$th center and we let $Y_i$ denote the responses, $n_i \times 1$ vector, of combined samples. Then $\mathcal{Y} = (Y_1', \ldots, Y_m')'$ is the $n \times 1$ vector of all responses, where $n = n_1 + n_2 + \cdots + n_m$ is the total sample size. Consequently, $\mathcal{Y}$ follows a mixed model (overall),
\( \mathcal{Y} = \mu 1_n + \mathcal{X} \beta + \mathcal{Z} b + e, \) 

(3.2)

where,

\[
\begin{align*}
\mathcal{Y} &= \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}, \text{ stacked combined responses,} \\
\mathcal{X} &= \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}, \text{ stacked model matrices,} \\
b &= \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \text{ stacked block random effects,} \\
e &= \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}, \text{ stacked random errors and}
\end{align*}
\]
a block matrix where the diagonal matrices are the block incidence matrices. The usual
distributional assumption about the random terms is that, both the random effect and the
random error are normally distributed. That is

\[
b \sim N(0, \Psi)
\]

and \(e \sim N_n(0, \Gamma)\), and \(\text{cov}(b_i, e_i) = 0\). Under these assumptions, if the linear mixed model is correct, then the
following statements are true about the marginal and conditional distribution about the
mean,

\[
E(\mathcal{Y}|b) = X\beta + Zb
\]

\[
E(\mathcal{Y}) = X\beta
\]

\[
\text{Var}(\mathcal{Y}|b) = \Lambda
\]

\[
\text{Var}(\mathcal{Y}) = \Lambda + Z\Theta Z' = \Lambda^*
\]

It is worth noting that the conditional distribution of \(Y|b\) is normal with mean \(X\beta + Zb\) and
variance-covariance matrix \(\Lambda\). Given the assumptions about the mixed model, the estimator
of $\beta$ and the best linear unbiased predictor $b$ are

$$\hat{\beta} = (X'\Lambda^{-1}X)^{-1}X'\Lambda^{-1}Y$$

and

$$\hat{b} = \Theta Z'\Lambda^{-1}(Y - X\hat{\beta})$$

respectively. For each block, the predicted response is

$$\hat{Y}_i = X_i\hat{\beta} + \hat{b}_i, \quad i = 1, \ldots, m.$$

If the inverse does not exist then we use the generalized inverse. In practice, $\Lambda$ and $\Theta$, and hence $\Lambda^*$ are not known. They could be estimated through restricted maximum likelihood.

3.1.2 Maximum Likelihood and Restricted Maximum Likelihood

Variance components of clustered data are estimated by restricted maximum likelihood (REML) and maximum likelihood method. Within blocks or clusters, some form of correlation structure is introduced into responses by the block random effect. This structure occurs even if all of the random effects are independent from one another. For example, in
a multi-center trials, two observations from the same center (hospital) are marginally correlated because they share the same center random error. The form they take are not limited to, first-order autoregressive AR(1), compound symmetry, saturated and unsaturated.

Let $\beta^{**}$ be an unknown vector of all the variance components, then the log-likelihood function for the fixed effect $\beta$ and $\beta^{**}$ is

$$
\ell(\beta, \beta^{**}|y) = -\frac{1}{2} \log |\Lambda^*(\beta^{**})| - \frac{1}{2} (y - \mathcal{X}\beta)' \Lambda^*(\beta^{**-1}) (y - \mathcal{X}\beta) - \frac{1}{2} n \log(2\pi).
$$

The estimates of $\beta^{**}$, is obtained by maximizing $\ell (\hat{\beta}, \beta^{**}|y)$ with respect to $\beta^{**}$. This is done through Newton-Raphson algorithm. Suppose $\hat{\beta}^{**}$ is the estimate of $\beta^{**}$, then the ML estimate of $\beta$ is

$$
\hat{\beta} = (\mathcal{X}' \Lambda^{**-1}(\hat{\beta}^{**}) \mathcal{X})^{-1} \mathcal{X}' \Lambda^{**-1}(\hat{\beta}^{**}) y
$$

### 3.2 Adaptation and Linear Mixed Models

In this section we develop an adaptive procedure for mixed model under exchangeable errors. We consider the model,

$$
\mathcal{Y} = \mu 1_n + \xi \Delta + \mathcal{Z}b + e,
$$

(3.3)

where $\Delta$ is the fixed effect parameter. The usual normality restriction on $e$ and $b$ are relaxed however exchangeability is assumed for $e$. Further we assume that, there are no covariates...
so we have a design matrix $C$. Note that the design matrix is made up of 0’s and 1’s. Under $H_0$, consider a vector of distribution functions,

$$
\mathcal{F} = \begin{bmatrix}
    F^{(1)} \\
    F^{(2)} \\
    \vdots \\
    F^{(m)}
\end{bmatrix}
$$

stacked unknown distribution for $m$ centers.

Basically, we are considering multi-center designs where the distributions unknown. Even if they are known, may vary from center to center. We are developing adaptive procedures that adapt within centers. The 'unadaptive' many rankings procedure (MRW) would seem less efficient than our adaptive procedures. It is worth noting that we can only exchange errors within centers. That is the sub-vectors, $e_i$'s, of the stacked random errors,

$$
e = \begin{bmatrix}
    e_1 \\
    e_2 \\
    \vdots \\
    e_m
\end{bmatrix}
$$

are not exchangeable. On the contrary, within a center, the elements of $e_i$'s are exchangeable. Consequently, the assumption of exchangability is applicable within centers. Because of this we consider the $i$th block for the development of our schemes. Consider the $i$th block mixed model,

$$
Y_{ij} = \mu + c_i^T \Delta + b_i + \epsilon_{ij}, \quad j = 1, \ldots, n, \ i = 1, \ldots, m,
$$

(3.4)
where $Y_{ij}$ combined response samples, $e_{ij}$ are elements of a design matrix $C$, which are 0's and 1's and $\Delta$ fixed effects parameters, $\varepsilon_{ij}$ are independent and identically distributed with distribution $F$ and $b_i$ is a random variable (block random effect) independent of $\varepsilon_{ij}$ and with distribution $G$. The block random errors, $\varepsilon_{ij}$, may be correlated hence the form of covariance structure of $\varepsilon_{ij}$ needs to be considered for any valid analysis. Next, we look at exchangeable random variables and it's relationship with the cluster correlated model.

3.2.1 Exchangeable Random Variables

**Definition 3.2.1.** The random variables $\xi_1^{(i)}, \ldots, \xi_n^{(i)}$ are exchangeable if the $n!$ permutations $\xi_{k_1}^{(i)}, \ldots, \xi_{k_n}^{(i)}$ have the same $n$-dimensional probability distribution. The variables of an infinite sequence $\xi_{n}^{(i)}$ are exchangeable if $\xi_1^{(i)}, \ldots, \xi_n^{(i)}$ are exchangeable for each $n$. As an example we consider (Polya's Urn). An urn containing $r$ red balls and $w$ white balls. A ball drawn at random is replaced together with $a$ balls of the same color. This process is repeated infinitely such that $\xi_i^{(i)} = 1$ if a red ball is picked at the $i$th trial otherwise $\xi_i^{(i)} = 0$. Let $\xi_1^{(i)} = 1, \xi_2^{(i)} = 0, \xi_3^{(i)} = 1, \xi_4^{(i)} = 1, \xi_5^{(i)} = 0, \xi_6^{(i)} = 1$ for the first six picks then we have

$$P(1, 0, 1, 1, 0, 1) = P(1, 1, 0, 0, 1, 1) = P(1, 1, 1, 0, 1)$$

$$P(1, 0, 1, 1, 0, 1) = \frac{r}{r + w} \frac{w}{r + w} \frac{r + a}{r + w} \frac{r + 2a}{r + w} \frac{w + a}{r + w} \frac{r + 3a}{r + w}$$

$$P(1, 1, 0, 0, 1, 1) = \frac{r}{r + w} \frac{r + a}{r + w} \frac{w}{r + w} \frac{a}{r + w} \frac{r + 2a}{r + w} \frac{w + a}{r + w}$$

$$P(1, 1, 1, 0, 0) = \frac{r}{r + w} \frac{r + a}{r + w} \frac{r + 2a}{r + w} \frac{r + 3a}{r + w} \frac{w + a}{r + w} \frac{w + a}{r + w}$$
Thus for all $6!$ permutations, we have the same distribution. For these binary sequences, the absolute of pairwise differences are also exchangeable.

**Lemma 3.2.1.** If $\xi_1^{(i)}, \ldots, \xi_{n_i}^{(i)}, \ldots$ are independent and identically distributed, then they are exchangeable, but not conversely.

In the example of Polya's Urn, the $\xi_1^{(i)}, \ldots, \xi_6^{(i)} \ldots$ are not independent but are exchangeable.

**Theorem 3.2.1.** *(de Finetti’s)* To every infinite sequence of exchangeable variables $\xi_{n_i}^{(i)}$ there corresponds a probability distribution $F$ concentrated on $[0,1]$ such that

$$P\{\xi_1 = 1, \ldots, \xi_k = 1, \xi_{k+1} = 0, \ldots, \xi_{n_i} = 0\} = \int_0^1 \theta^k (1 - \theta)^{n-k} F\{d\theta\} \quad (3.5)$$

Within our context, exchangeability implies that two samples of equal size have the same response distribution regardless of the covariates see Peter McCullagh (2005). Define the treatment effect by a functional $H$, thus

$$\Delta = H\left( F_{X_2}^{(i)} \right) - H\left( F_{X_1}^{(i)} \right) \quad (3.6)$$

such that

$$H\left( F_{X_2}^{(i)} \right) = H\left( F_{X_1}^{(i)} \right) \text{ implies } F_{X_2}^{(i)} = F_{X_1}^{(i)} \forall, F_{X}^{(i)} \in \mathcal{F}.$$
Then $\Delta$ could be difference in median, mean or any parameter of interest. The functional, $H$ is also selected such that under the block model, $\Delta$ is constant over all covariates. If $\Delta = 0$, then there is no treatment effect hence equality in distribution. If $F_{X_2}^{(i)} = F_{X_1}^{(i)}$, it implies the model does not depend on treatment levels and the treatment effect is identical to zero. Thus for $\Delta = 0$, the errors of the combined sample are exchangeable. Next we establish that Hogg’s adaptive procedure works for a mixed model.

It will be more convenient to rewrite the $i$th block model as

$$Y_j^{(i)} = \mu + \xi_j^{(i)} + b_i + \epsilon_j^{(i)}, \quad j = 1, \ldots, n_i,$$

where $\xi_j^{(i)}$ has a distribution $F$. Let $\xi_j^{(i)}$ denote the order statistics for $\epsilon_j^{(i)}$. One may want to test fixed effects of the mixed model. Hence we test, $H_0: \Delta = 0$.

By Theorem (2.1.3), $\xi_j^{(i)}$ are sufficient and complete for $F$. Let $b_i + \epsilon_j^{(i)} = \zeta_j^{(i)}$ then from the block model, $\zeta_1^{(i)}, \ldots, \zeta_{n_i}^{(i)}$ are exchangeable.

Therefore under $H_0$, the block model becomes,

$$Y_j^{(i)} = \mu + b_i + \zeta_j^{(i)}, \quad j = 1, \ldots, n_i. \quad (3.7)$$

Let $P$ be the distribution of $\zeta_j^{(i)}$, then $\mu + b_i + \zeta_j^{(i)}$ has the same characteristics in distribution within block. If $\zeta_j^{(i)}$ is exchangeable then the distribution of $Y_1^{(i)}, \ldots, Y_{n_i}^{(i)}$ is exchangeable and, hence, rank statistics using ranks of $Y_1^{(i)}, \ldots, Y_{n_i}^{(i)}$ are distribution free.

Let $T = T[R(Y_1^{(i)}, \ldots, Y_{n_i}^{(i)})]$ be such a rank statistic. So $T$ is distribution free under $H_0$. 

But note that

$$R\left(Y_j^{(i)}\right) = R\left(\mu + b_i + \varepsilon_j^{(i)}\right) = R\left(\varepsilon_j^{(i)}\right)$$ (3.8)

Thus

$$T\left[R(Y_1^{(i)}, \ldots, Y_n^{(i)})\right] = T\left[R(\varepsilon_1^{(i)}, \ldots, \varepsilon_n^{(i)})\right].$$ (3.9)

Therefore, under $H_0$, $T$ is independent of the $\varepsilon_j^{(i)}$s.

Now consider the selector statistics based on $Y_j^{(i)}$, as noted above they are functions of the form:

$$\bar{A}_{1\alpha_1} - \bar{A}_{2\alpha_2} = \frac{1}{\alpha_{\alpha_1}} \sum_{j=1}^{\alpha_1} Y_j^{(i)} - \frac{1}{\alpha_{\alpha_2}} \sum_{j=2}^{\alpha_2} Y_j^{(i)}$$

$$= \frac{1}{\alpha_{\alpha_1}} \sum_{j=1}^{\alpha_1} [\mu + b_i + \varepsilon_j^{(i)}] - \frac{1}{\alpha_{\alpha_2}} \sum_{j=2}^{\alpha_2} [\mu + b_i + \varepsilon_j^{(i)}]$$

$$= \frac{1}{\alpha_{\alpha_1}} \sum_{j=1}^{\alpha_1} [\varepsilon_j^{(i)}] - \frac{1}{\alpha_{\alpha_2}} \sum_{j=2}^{\alpha_2} [\varepsilon_j^{(i)}].$$

Therefore, under $H_0$, the selector statistics are independent of $T$ (from Chapter 2). So Hogg’s adaptive scheme works under exchangeable errors.

In our development of an adaptive scheme, we apply rank scores to the residuals which provide consistent and highly efficient tests for fixed effects of the mixed model. These scores are selected based on the adaptation. Since the distribution of $F$ and $G$ for $b_i$ and $\varepsilon_j^{(i)}$ respectively are unknown, and that $\xi_j^{(i)}$ could be either asymmetric or symmetric the general rank scores are used, see Hettmansperger and Mckean (1998). If sign-rank scores are used then the underlying distribution of $\xi_j^{(i)}$ will have to be symmetric, see
3.2.2 Rank-Based Estimates

The general rank-based procedure is robust. Its applicability to our set up follows from the fact that the overall dispersion function, $D(\Delta)$, is convex. Note that adaptation could only be done within centers since the errors are exchangeable within centers. We want to develop an estimator for the fixed effect using the pseudo-norm. Like a classical ranked based estimates some assumptions have to be made about the design matrix of the block model.

Consider the general overall linear mixed model (3.3) and assume that for each center, the design matrix has a full column rank. Without loss of generality we will let $x_j^{(s)}'$ instead of $c_j^{(s)}'$ be the design matrix. Let the centered matrix for each center be denoted by,

$$X_c^{(s)'} = [I_{n_s} - \frac{1}{n_s}J_{n_s}]x_j^{(s)'}$$

then for all the $m$ centers, we stack these centered design matrices as follows,

$$X_c = (X_c^{(1)'}, \ldots, X_c^{(m)'})'$$

Define the matrix $W$ by,

$$W = \frac{1}{n}X_c'X_c$$

**Theorem 3.2.2.** (Regularity Conditions) Let $H_n$ be a sequence of a hat matrix define by $H = x_j^{(s)}x_j^{(s)'}$ and denote $h_{jjn}$ as the $j$th diagonal entry of $H_n$. Then,
\[(D.2) \lim_{x \to \infty} \max_j h_{jjn} = 0\]
\[(D.3) \lim_{x \to \infty} n^{-1} x_j^{(i)} x_j^{(i)\prime} = \Sigma\]

where \(\Sigma\) is a \(p \times p\) positive definite design matrix.

Condition (D.2) is called the Huber's condition which implies the that; Under the above regularity conditions, the least estimate of the fixed effect for the \(i\)th block is

\[\hat{\Delta}_{LS} \sim N_p(\Delta, \sigma(x_j^{(i)} x_j^{(i)\prime})^{-1}).\]

\(x_j^{(i)}\) is assumed to be centered. Based on our assumptions \(f\) is unknown so adaptation is done within centers to select a score. We assume that the selected score is optimal.

**Theorem 3.2.3.** Suppose that \(a_i(1) \leq \ldots \leq a_i(n_i), \sum a(t) = 0\) and \(a_i(t) = -a_i(n_i + 1 - t)\). Then the function \(\|v\|_p\) is a pseudo-norm.

McKean and Schrader (1980), established that R estimate could be obtained by replacing the euclidean norm \(\|V\|_2^2\) by the pseudo-norm

\[\|v\|_\varphi = \sum_{t=1}^{n} a[R(v_t)]v_t, \quad v \in R^n\]

where \(R(v_t)\) is the rank of \(v_t\) among \(v_1, \ldots, v_{n_i}\) and \(a_i(t)\) is a set of rank scores for the \(i\)th center. Note that \(a_i(t) = \varphi_i[\frac{t}{n_i+t}]\), \(t = 1, \ldots, n_i\) where \(\varphi_i(u)\) is a specified nondecreas-
ing, square-integrable function defined on the interval $(0,1)$ and standardized such that
\[ \int_0^1 \varphi_i(u) \, du = 0 \quad \text{and} \quad \int_0^1 \varphi_i(u)^2 \, du = 1 \]
be a set of rank scores for the $i$th center. For example, the Wilcoxon pseudo-norm is generated by the linear score function $\varphi(u) = \sqrt{12}(u - \frac{1}{2})$ and the sign score is generated by $\varphi(u) = \text{sgn}(u - \frac{1}{2})$. Using the generalized score, the robust (R)-estimate if there no centers is given by,

\[ \hat{\Delta}_R = \text{Argmin} \| Y - X \Delta \|_{\varphi}. \]

In the independent error case, Rashid et al. (2011) showed that under regularity conditions $\hat{\Delta}_{\varphi_i}$ has an asymptotic normal distribution

\[ \hat{\Delta}_{\varphi_i} \sim N_p(\Delta, \tau_i(X_c'X_c)^{-1}) \]

where

\[ \tau_i^{-1} = \int_0^1 \varphi_i(u) \left\{ -\frac{f(F^{-1}(u))}{f[F^{-1}(u)]} \right\} \, du. \]

If centers are not ignored, we have an $MR$ process with an estimator

\[ \hat{\Delta}_R = \text{Argmin} \| Y - X \Delta \|_R \]

where

\[ \| v \|_R = \sum_{i=1}^m \| v \|_{\varphi_i}. \]
For the \( m \) centers, the fixed effect estimate, \( \hat{\Delta}_R \), for the overall model (3.2.3) has an asymptotic normal distribution

\[
\hat{\Delta}_R \sim N_p(\Delta, \tau(\mathcal{X}_c'\mathcal{X}_c)^{-})
\]

where

\[
\tau = \frac{1}{m} \sum_{i=1}^{m} \tau_i,
\]

and the generalized inverse of \( \mathcal{X}_c'\mathcal{X}_c \) is \( \mathcal{X}_c'\mathcal{X}_c^{-} \). So our dispersion function for center \( i \) becomes,

\[
D_i(\Delta) = \left\| Y^{(i)} - \mathcal{X}_c^{(i)} \Delta \right\|_{\phi_i}
\]

(3.10)

where \( \phi_i \) is the score function selected at the \( i \)th center and the pseudo-norm is given by

\[
\|v\|_{\phi_i} = \sum_{j=1}^{n_i} a_i[R_i(v_j)]v_j, \quad v \in \mathbb{R}^{n_i}
\]

and \( R_i(v_j) \) denotes the ranks of \( v_j \) among \( v_1, \ldots, v_{n_i} \). The ranks are not invariant to the centers however the overall dispersion function based on the norm is invariant to the centers.

**Theorem 3.2.4.** Let \( h_i \) be convex functions on \([a,b]\) for \( i = 1, \ldots, m \), then \( \sum_{i=1}^{m} h_i \) is convex.

**Theorem 3.2.5.** Let \( D_i \) be the dispersion function for \( i = 1, \ldots, m \), then \( \sum_{i=1}^{m} D_i \) is convex.
From the Theorem (3.2.5), the fixed effect of the mixed model could be obtained by minimizing

$$D(\Delta) = \sum_{i=1}^{m} D_i(\Delta).$$

(3.11)

This estimator also satisfies the regression equivariances. Thus if \( T = uY + X\nu \) then \( \hat{\Delta}_R(T) = u\hat{\Delta}_R(Y) + \nu \) where \( \hat{\Delta}_R \) is R-estimate of \( \Delta \). To minimize the overall dispersion function we need to find \( \hat{\Delta} \) that solves the gradient function,

$$S(\hat{\Delta}) = \sum_{i=1}^{m} S_i(\hat{\Delta}) = 0$$

(3.12)

where

$$S_i(\Delta) = \sum_{i=1}^{m} x_{ij} a_i [R_i(y_{ij} - x_{ij}\Delta)],$$

\( x_{ij} \) is the \( i \)th row of \( X \), and \( R_i \) denotes intra-center rankings see (Rashid et al 2011).

3.2.3 Estimation of Variance-Covariance Components

In this subsection we develop an asymptotic variance-covariance matrix of \( \hat{\Delta}_R \). This is similar to the one developed by Rashid et al (2011) except that scores at centers may differ. The asymptotic variance-covariance matrix of \( \hat{\Delta}_R \) is

$$\mathcal{V}_R = \tau^2 (\mathcal{X}_c'\mathcal{X}_c)^{-1} \left( \sum_{i=1}^{m} X_{ei}'\Sigma_{-\hat{\phi}_i}X_{ei} \right) (\mathcal{X}_c'\mathcal{X}_c)^{-1}$$
Let

\[ K = \sum_{i=1}^{m} \frac{n_i!}{(n_i - 2)!} - p \]

where \( p \) is the dimension of \( \Delta_R \). Then the estimator of \( p \) is

\[ \hat{p} = \frac{1}{K} \sum_{i=1}^{m} \sum_{r>s} \Phi(r,s) \]

where

\[ \Phi(r,s) = a_i[R(\hat{c}_{ir})]a_i[R(\hat{c}_{is})]. \]

Which implies the robust estimate of the asymptotic covariance matrix is obtained.

Next we consider the variance component estimator of \( \sigma_b^2 \) and \( \sigma_\varepsilon^2 \), see Rashid et. al (2011). From the block model let \( \xi^{(i)} = b_i + \varepsilon^{(i)} \) and define \( \text{Var}(\varepsilon^{(i)}) \) and \( \text{Var}(b_i) \) as follows

\[
\begin{align*}
\text{Var}(\varepsilon^{(i)}) &= \sigma_\varepsilon^2 \\
\text{Var}(b_i) &= \sigma_b^2 \\
\sigma_T^2 &= \sigma_\varepsilon^2 + \sigma_b^2
\end{align*}
\]

where \( \sigma_T^2 \) is the total variance. We assume that the covariance matrix of \( \xi^{(i)} \) exist. Then \( \text{Var}(\xi^{(i)}) = \sigma_T^2 (1 - \rho) I_n + \rho J_n \), where \( \rho = \frac{\sigma_b^2}{\sigma_T^2} \). For the estimator of the variance com-
ponents, $\sigma_b^2$ and $\sigma_T$ we consider the block model, hence $\hat{\xi}^{(i)} = Y^{(i)} - \hat{\mu} - X_c^{(i)}\hat{\beta}$. The predictor of $b_i$ is $\hat{b}_i = \text{med,} \left\{ \xi_j^{(i)} \right\}$, see Rashid and Nandram (1998). Hence the estimate of $\sigma_b^2$ is

$$\hat{\sigma}_b^2 = \frac{\hat{\rho}}{1 - \hat{\rho}} \hat{\sigma}_e^2$$

where $\hat{\xi}_j^{(i)} = \hat{\xi}_j - \hat{b}_i$ and $\text{MAD}_i(v_i) = 1.483\text{med} \left\{ v_i - \text{med,} \left\{ v_j \right\} \right\}$. Note that for each center $b_i$ is a $1 \times 1$ random variable. Thus estimate of $\sigma_b^2$ is

$$\hat{\sigma}_b^2 = \frac{\hat{\rho}}{1 - \hat{\rho}} \hat{\sigma}_e^2$$

3.2.4 Estimation of Scale Parameter

In this subsection we develop a consistent estimator of $\tau$, see Hettmansperger and McKean (1998). A consistent estimator $\hat{\tau}_i$ of $\tau_i$ for the $i$th center is obtained from the differences in residuals at that center. This approach was proposed by Koul et al. (1987). Let $\hat{\Delta}_R$ be the overall robust estimate of $\Delta$ for the linear mixed model. Then the vector of residuals of robust fit at the $i$th center is given by,

$$\hat{\varphi}^{(i)} = Y^{(i)} - X_c^{(i)}\hat{\Delta}_R.$$ 

Aubuchon and Hettmansperger (1984,1989) obtained a density type estimator for $\tau_{\varphi_i}$, $\varphi_i$ using wilcoxon score , which is a function differences of residuals. They established that
the density type estimator was consistent for symmetric and asymmetric error distributions. Let \( \varphi_i \) be the score at the \( i \)th center, satisfying S.1 and S.2. Since \( \varphi_i \) is bounded we consider the standardization,

\[
\varphi_i^*(u) = \frac{\varphi_i(u) - \varphi_i(0)}{\varphi_i(1) - \varphi_i(0)}
\]

(3.13)

which is a linear function of \( \varphi_i \) and hence a distribution function on (0,1). Then the inference made using either score is the same. We have established that linear combination of exchangeable random variables is exchangeable and since we seek a density-type estimator which is a function of difference in residuals we need the distribution of \( |e_k^{(i)} - e_l^{(i)}|, k \neq l \) for the \( i \)th center. We use the results obtained in Hettmanspeger and McKean (1998). Under \( H_0 \), the marginal distribution are the same. Let \( \varepsilon_1^{(i)} \) and \( \varepsilon_2^{(i)} \) be independent random variable with distribution \( F_i(x) \) and \( \varphi_i(F_i(x)) \) respectively. Then the distribution of \( |e_k^{(i)} - e_l^{(i)}| \), see Hettmanspeger and McKean (1998) is

\[
H(y) = \begin{cases} 
  P(|\varepsilon_1^{(i)} - \varepsilon_2^{(i)}| \leq y) = \int_{-\infty}^{\infty} [F_i(\varepsilon_2^{(i)} + y) - F_i(\varepsilon_2^{(i)} - y)]d\varphi_i^*(F_i(\varepsilon_2^{(i)})), & y > 0 \\
  0, & \text{otherwise} 
\end{cases}
\]

(3.14)

We know that

\[
\tau_i^{-1} = \int_0^1 \varphi_i(u) \varphi_i(u) du
\]
where \( \varphi_{i,f} = \frac{-f'(F^{-1}(u))}{f(F^{-1}(u))} \), \( f \) and \( F \) are the error density and distribution at the \( i \)th center respectively. Let

\[
c = \int_0^1 \varphi_i^*(u)\varphi_{i,f}(u)du
\]

\[
= (\varphi_i(1) - \varphi_i(0))^{-1} \int_0^1 \varphi_i^*(u)\varphi_{i,f}(u)du
\]

After making change of variable and simplifying, (see Hettmanspeger and McKean (1998)) we obtain

\[
c = \int_{-\infty}^{\infty} f(x)d\varphi_i^*(F(x))
\]

To get the density type estimator, we need to differentiate under the integral sign of the \( H(y) \). After simplification we have

\[
h(0) = 2c \quad (3.15)
\]

Note that for some empirical distribution of \( H(y) \) (see Hettmanspeger and McKean (1998)) for the \( i \)th center, consider the sequence , \( t_{n,\ell} = \frac{\hat{t}_{n,\ell}}{\sqrt{n_i}} \) where \( \hat{t}_{n,\ell} \) is the \( \ell \)th quantile of \( \hat{H}_{n,i} \).

Choose \( t_{n,\ell} \) close to zero then the estimate of \( h(0) \) and hence \( c \) is of the form \( \hat{h}_{n,i}(t_{n,\ell})/(2t_{n,\ell}) \).

. This implies the consistent estimator of \( c \) is

\[
\hat{c}_{n,\ell} = \frac{(\varphi_i(1) - \varphi_i(0))^{-1}\hat{H}(t_{n,\ell}/\sqrt{n_i})}{2t_{n,\ell}/\sqrt{n_i}} \quad (3.16)
\]

It follows that the consistent estimator of \( \tau_i \) is
As indicated in Hettmanspeger and McKean (1998), when \( \frac{n_a}{p} \) exceeds 5 the value \( \ell = 0.8 \) gives valid estimate and \( \ell = 0.9 \) yielded valid estimates when \( \frac{n_a}{p} \) less than 5.

Since the fixed effect estimates are invariant to random effect (thus centers) the differences of residuals are also invariant. So we have the consistent estimator of \( \tau \) as

\[
\hat{\tau} = \frac{1}{\hat{\ell}_{m,\ell}}
\]  

(3.17)

3.2.5 Intercept Estimation

The intercept cannot be estimated from the dispersion function. Once \( \Delta \) is estimated, \( \hat{\mu} \) the estimate of the intercept \( \mu \) could be found by the median of the residuals of the overall fit. Our main goal is to obtain an asymptotic distribution of \( (\mu, \hat{\Delta}_R) \), where \( \hat{\Delta}_R \) is the overall estimate of the fixed effect parameter. Then the nondecreasing step function of \( \Delta_R \) is,

\[
S^*(Y - \mu 1 - \chi \hat{\Delta}_R) = \sum_{j=1}^{n_t} \sum_{i=1}^{m} \text{sign}(y_j^{(i)} - \mu - x_j^{(i)' \hat{\Delta}_R})
\]  

(3.19)

This implies the estimate, \( \hat{\mu} \) of \( \mu \) satisfies,

\[
S^*(Y - \mu 1 - \chi \hat{\Delta}_R) = 0 \text{ which is the median of the residuals , thus }
\]

\[ \hat{\mu} = \text{med}_{ij} \hat{r}_{ij} \]

where \( \hat{r}_{ij} \) are the elements of \( Y - \chi c \hat{\Delta}_\varphi \).
3.3 Asymptotic Efficacies

For Pitman efficacies, we consider the gradient function of our overall set up. As shown in Chapter 2 of Hemmansperger and Mckean (1998), the gradient function associated with each center is Pitman regular. We assume that the error density at each center has finite Fisher information. We consider the mixed model without covariates to develop the efficacy results. For the overall gradient function $S$, we have

$$ S(\Delta) = \sum_{i=1}^{m} S_i(\Delta) $$

where $S_i$ is the gradient function at the $i$th center. We know that $S_i$ is pitman regular for each center. We state a proposition about $S$

**Proposition 3.3.1.** Let $S_i$ be Pitman regular for each $i$ then

$$ S(\Delta) = \sum_{i=1}^{m} S_i(\Delta) $$

is Pitman regular.

To conceptually obtain the efficacy results, we will need the null mean and variance of the overall gradient function. In the process we will establish that the overall gradient function is Pitman regular. This argument can be made rigorously using the asymptotic
linearity result in Ranshid et. al. (2011). From the overall gradient function we have

\[
S(\Delta) = \sum_{i=1}^{m} S_i(\Delta)
\]

\[
= \sum_{i=1}^{m} \sum_{j=1}^{n_{2i}} a_i \left( R_i(Y_j^{(i)} - \Delta) \right)
\]

\[
= \sum_{i=1}^{m} \sum_{j=1}^{n_{2i}} \varphi_i \left[ \frac{n_{1i}}{n_i + 1} F^{(i)}_{n_{1i}}(Y_j^{(i)}) + \frac{n_{2i}}{n_i + 1} F^{(i)}_{n_{2i}}(Y_j^{(i)} - \Delta) \right]
\]

where \( F^{(i)}_{n_{1i}} \) and \( F^{(i)}_{n_{2i}} \) are the empirical distributions of each of the samples. For the null mean we have

\[
\mu(\Delta) = E_0 \left[ \frac{1}{n} S(\Delta) \right] = \frac{1}{n} E_0 [S(\Delta)]
\]

\[
= \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_{2i}} E_0 \left[ \varphi_i \left[ \frac{n_{1i}}{n_i + 1} F^{(i)}_{n_{1i}}(Y_j^{(i)}) + \frac{n_{2i}}{n_i + 1} F^{(i)}_{n_{2i}}(Y_j^{(i)} - \Delta) \right] \right]
\]

Note that under \( H_0 \) we have 2m limits of \( \hat{F}^{(i)}_{n_k} \rightarrow F^{(i)} \) for \( i = 1 \ldots m \) and \( k = 1, 2 \). Again under \( H_0 \) \( \mu(0) = 0 \) since for each center the null mean is zero. Assume that the sample sizes are equal for all centers. Then \( n = mn^* \) where \( n^* \) is the sample size at each center. In practice, this condition is often satisfied. Consequently for a different distribution at each center, the null mean becomes,
\[ \mu(\Delta) = \frac{1}{m n} \sum_{i=1}^{m} \sum_{j=1}^{n_{2i}} E_0 \left[ \varphi_i \left[ \frac{n_{1i}}{n + 1} F^{(i)} (Y_j^{(i)}) + \frac{n_{2i}}{n + 1} F^{(i)} (Y_j^{(i)} - \Delta) \right] \right] \]

hence

\[ \mu(\Delta) \rightarrow \frac{1}{m n} \sum_{i=1}^{m} \sum_{j=1}^{n_{2i}} E_0 \left[ \varphi_i \left[ \frac{n_{1i}}{n + 1} F^{(i)} (Y_j^{(i)}) + \frac{n_{2i}}{n + 1} F^{(i)} (Y_j^{(i)} - \Delta) \right] \right] \]

\[ \rightarrow \frac{1}{m} \sum_{i=1}^{m} \lambda_2^{(i)} E_0 \left[ \varphi_i \left[ \lambda_1^{(i)} F^{(i)} (Y_j^{(i)}) + \lambda_2^{(i)} F^{(i)} (Y_j^{(i)} - \Delta) \right] \right] \]

where \( \lambda_1^{(i)} \rightarrow \frac{n_{1i}}{n} \) and \( \lambda_2^{(i)} \rightarrow \frac{n_{2i}}{n} \). Since \( \varphi_i \) is a nondecreasing function, continuous function and \( \varphi'_i \) on \((0,1)\), we can obtained the efficacy results similar to Hettmanspeger and McKean (1998) as follows,

\[ \frac{d \mu(\Delta)}{d \Delta} \bigg|_{\Delta=0} = \frac{1}{m} \sum_{i=1}^{m} \lambda_2^{(i)} \lambda_1^{(i)} \int_{-\infty}^{\infty} \varphi_i' (F^{(i)}(t)) f^{(i)}(t) dt \]

\[ = \frac{1}{m} \sum_{i=1}^{m} \lambda_2^{(i)} \lambda_1^{(i)} \int_{-\infty}^{\infty} \varphi_i (F^{(i)}(t)) \left( \frac{-f^{(i)}(t)}{f^{(i)}(t)} \right) f^{(i)}(t) dt \]

\[ = \frac{1}{m} \sum_{i=1}^{m} \lambda_2^{(i)} \lambda_1^{(i)} \int_{0}^{1} \varphi_i (u) \varphi_i (u) du \]

\[ = \frac{1}{m} \sum_{i=1}^{m} \lambda_2^{(i)} \lambda_1^{(i)} \varphi_i^{-1} du. \]

Since for each \( i \lambda_2^{(i)} \lambda_1^{(i)} \varphi_i^{-1} > 0 \) it implies that \( \mu'(0) > 0 \). Hence, the second of Pitman regular is satisfied. It is worth stating that if the proportion of treatment group is kept same for each center having same distribution, thus \( \lambda_1^{(i)} = \lambda_1^{(h)} \), \( \lambda_2^{(i)} = \lambda_2^{(h)} \) and \( \varphi_i = \varphi_h \) for all
$i \neq h$, then $\mu'(0) = \lambda_2 \lambda_1 \tau^{-1}$. For asymptotic linearity results we need the null variance.

Our null mean is zero so for the null variance we have,

$$\text{Var}_0(S) = E_0(S^2)$$

$$= E_0 \left[ \sum_{i=1}^{m} S_i^2 + 2 \sum_{i \neq j} S_i S_j \right]$$

$$= E_0 \left[ \sum_{i=1}^{m} S_i^2 \right] + E_0 \left[ 2 \sum_{i \neq j} S_i S_j \right]$$

$$= \sum_{i=1}^{m} E_0(S_i^2),$$

because for $i \neq j$, $S_i$ and $S_j$ are independent. Let $E_0[a_i(R(y_j^{(i)}))] = \frac{1}{n_i} s_{a_i}^2$ and $\text{Var}_0(S_i) = \frac{n_1n_2}{n_i(n_i - 1)} s_{a_i}^2$ where $s_{a_i}^2 = \sum_{j=1}^{n_i} a_i^2(j)$ then we have,

$$\text{Var}_0(S) = \sum_{i=1}^{m} \text{Var}_0(S_i^2)$$

$$= \sum_{i=1}^{m} \frac{n_1n_2}{n_i(n_i - 1)} s_{a_i}^2$$

$$= \sum_{i=1}^{m} \frac{n_1n_2}{n_i(n_i - 1)} s_{a_i}^2$$

Thus for equal sample sizes at each center, we have $n = n * m$. The null variance $\sigma^2(0) =$
\[ \lim n\text{Var}_0(\tilde{S}(0)) \text{ becomes} \]

\[ \sigma(0) = \frac{mn^*}{m^2n^*} s^* \]

so

\[ \sqrt{n}\sigma(0) \to \frac{1}{\sqrt{m}} \sqrt{\sum_{i=1}^{m} \lambda_1^{(i)} \lambda_2^{(i)}} \]

This means that the efficacy \( C \), of a test based on \( S(\Delta) \) is,

\[ C = \frac{1}{m} \sum_{i=1}^{m} \lambda_1^{(i)} \lambda_2^{(i)} \tau_\phi^{-1} \sqrt{\frac{1}{m} \sum_{i=1}^{m} \lambda_1^{(i)} \lambda_2^{(i)}} \]

Thus asymptotic linear results is obtained hence \( S \) is pitman regular. If proportion of those enrolled at each center is same and the error distribution at each center is the same, then the efficacy reduces to,

\[ C = \frac{\lambda_1 \lambda_2 \tau^{-1}}{\sqrt{\lambda_1 \lambda_2}} \]

which is same to the results obtained in Hettmaansparger and McKean(1998). Based on this result, assuming we have an optimal score for each center then the meta analysis may be efficient.

In general we assume for a fixed number of centers and for \( n_i \), we have \( \frac{n_i}{n} \to \lambda_i \) and \( \sum_{i=1}^{m} \lambda_i = 1 \). Because of invariant, without loss of generality we assume that the true parameters of the model are zero.
As discussed by Rashid et. al. (2011), under these assumptions coupled with Theorem 3.1 of Brunner and Dunker (1994)

$$\frac{1}{\sqrt{n}} S_{X_i}(0) \rightarrow N_p(0, \Sigma_{\varphi_i})$$

where

$$\Sigma_{\varphi_i} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{m} X_i' \Sigma_{\varphi_i} X_i.$$  

This means that the linearity and quadraticity results could be applied to our model, see Hettemansperger and McKean (1998 chapters 3,4,5). That is we have for the overall model,

$$S_{X_c}(\Delta) = S_{X_c}(0) - \tau^{-1} n^{-1} X_c' X_c \Delta + o_p(\sqrt{n})$$

uniform for $\sqrt{n} \parallel \Delta \parallel_2 \leq l$ for $l > 0$, where $\tau$ the scale parameter. Subsequently we obtain the asymptotic representation of the fixed effect estimates as,

$$\hat{\Delta}_R = \Delta + \tau (X_c' X_c)^{-1} \sum_{i=1}^{m} X_c' \varphi_i [F^{(i)}(\xi^{(i)})] + o_p(1/\sqrt{n}), \quad (3.20)$$

see Rashid et. al. (2011) for a rigorous proof.

### 3.3.1 Overall Test Statistic of Adaptation on Sample

In this subsection we develop an overall test statistic on sample from the $m$ centers. The error distribution for the $i$th center, $F^{(i)}$, may be known, and/or may differ. If it is
unknown, our classification by tail weight and skewness may select different scores. Under $H_0$, we can be certain that the errors of the block model are exchangeable. Between centers, the errors are not exchangeable and distributions may be different, the order statistic will not be sufficient and complete. In view of this limitation, we develop the overall adaptive test as follows.

Let $\varphi_{ki}$ be the score selected at the $i$th center and falls in region $k$, then the test statistic of that center is

$$T_{\varphi_{ki}} = \sum_{j=1}^{n_{2i}} a_t \left( R_t(Y_j^{(i)}) \right).$$

Note however that $T_{\varphi_{ki}}$ is asymptotic standard normal and distribution free. Hence we pool the test statistic over centers to obtain the overall test. Thus under $H_0$, the overall test statistic, $T$ is

$$T = \sum_{i=1}^{m} T_{\varphi_{ik}} = \sum_{i=1}^{m} \sum_{j=1}^{n_{2i}} a_t \left( R_t(Y_j^{(i)}) \right)$$

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

$$H_0 : \Delta = 0 \text{ versus } H_a : \Delta > 0,$$
we have

\[
\text{reject } H_0 \text{ in favor of } H_a \text{ if, } T = \frac{\sum_{i=1}^{m} Z_i}{\sqrt{m}} > Z_{\alpha}.
\]

So we have a meta-test. Note that for each center, the asymptotic test statistic is distribution free hence the overall test statistic is also distribution free. One of the problems with this test statistic is that the $Z_i$'s may cancel out when summing them over centers so we have to consider $|Z_i|$. 
CHAPTER IV

ADAPTIVE PROCEDURES

4.1 Procedures

In this chapter we give a formal algorithm for testing and estimating the fixed effect parameter for each adaptive scheme for a linear mixed model over \( m \) clusters. Since adaption could only be within centers, the block model is used, then overall test formulated for,

\[
H_0 : \Delta = 0
\]

versus

\[
H_a : \Delta > 0.
\]

Based on the scores selected for each center an overall estimate is obtained. References will be made to Chapters 1 and 3.

4.1.1 Many Rankings

We will call this process an MR. In the MR process, rankings are done at each center separately. The \( i \)th cluster subset has \( n_i \) elements. Chapters 2 and 3 gives a theoretical framework for this process. An algorithm for the MR process is as follows; The residuals
from the initial fit of the mixed model using lme in R package is obtained. Initial estimates \( \hat{\tau}_i^{(0)} \) of \( \tau_i^{(0)} \) and subsequently \( \hat{\tau}^{(0)} \) of \( \tau^{(0)} \) are computed from equation (3.17) and equation (3.18) respectively. The initial estimate of the dispersion \( \hat{\mathcal{D}}^{(0)} \), is also computed by using equation (3.11). For each center the combined residuals are ordered, the estimate of the selector statistic \( \text{Sel} = \{ Q_1^*, Q_2^* \} \) is obtained. The unknown distribution of the residuals from each center is classified using both the selector statistic \( \text{Sel} = \{ \hat{Q}_1^*, \hat{Q}_2^* \} \) and the cutoffs. Based on the selector statistic and the cutoffs, corresponding score to the unknown distribution is selected at each center. Using the scores selected for the initial fit, refitting is done iteratively through a Newton-type algorithm. The estimate of \( \hat{\Delta} \) for \( k \)-th Newton-type step is

\[
\hat{\Delta}^{(k)} = \hat{\Delta}^{(k-1)} + \frac{\hat{\tau}^{(k-1)}}{n} \mathbf{W}^{-} \mathbf{S} \left( \hat{\Delta}^{(k-1)} \right),
\]

where \( \mathbf{W}^{-} \) is the generalized inverse and

\[
\mathbf{W} = \frac{1}{n} \mathcal{X}_e' \mathcal{X}_e
\]

The shrinkage factor,

\[
\frac{\hat{\tau}^{(k-1)}}{n} \mathbf{W}^{-} \mathbf{S} \left( \hat{\Delta}^{(k-1)} \right)
\]

is used to update \( \hat{\Delta}^{(k-1)} \) until \( \hat{\Delta}^{(k-1)} \) and \( \hat{\Delta}^{(k)} \) is within an acceptable tolerance. This
tolerance is measured between $D \left[ \hat{\Delta}^{(k-1)} \right]$ and $D \left[ \hat{\Delta}^{(k)} \right]$. Scores selected at each center after the initial fit are used throughout the iterative process. Once $\hat{\Delta}$ is obtained, residuals for the refit at each center are calculated. Estimates of $\tau$, $\hat{\tau}$, and $D \left( \hat{\Delta} \right)$ are recorded. Goodness of fit based on the residuals may be performed. The test statistic, $T$ for this process is

$$T = \frac{\hat{\Delta}}{\tau \sqrt{[C' C]}}.$$  

(4.1)

We consider adaptation on sample next.

4.1.2 Hogg-McKean Adaptation on Sample

We will call this adaptation an HMS process. For this process, an adaptation is done on the responses at each center. Chapter 2 gives the theoretical framework for this process. The HMS process is as follows; For each center, the response samples are ordered and classification is carried out using both the estimated selector $\text{Sel} = \{ \hat{Q}_1^*, \hat{Q}_2^* \}$ from responses and cutoff points based on sample sizes. A corresponding score, for each site is selected. An R estimate of $\Delta$ is obtained at each center. Estimates at each center may differ and to resolve this a weighted estimate is used. Suppose we have $n_i$ subjects at each center and $n_{ij}$ number of patients randomly assigned to the treatment group at each center. Let $\hat{\Delta}_i$ be the fixed effect estimate at the $i$th center, then the overall estimate obtained by,

$$\hat{\Delta} = \sum_{i=1}^{m} \frac{n_{ij}}{n_i} \hat{\Delta}_i$$
Under the null hypothesis, the distribution of the gradient function $S_x(\Delta)$, for each center is asymptotically normal. Furthermore the test statistic is distribution free. Hence for overall test statistic, we use meta-test on the standardized test at each center. So if for center $i$, we have $T_{\varphi_i}$ as the test statistic then, asymptotically the test becomes,

$$Z_i = \frac{T_{\varphi_i}}{\sqrt{\text{Var}_0 (T_{\varphi_i})}}.$$ 

Consequently, the overall test statistic is

$$Z = \frac{\sum_{i=1}^{m} Z_i}{\sqrt{m}}$$

Note that under $H_0$, $Z$ is asymptotically normal with mean 0 and variance 1.

4.1.3 Hogg-McKean Adaptation on Residuals

Chapter 2 gives the theoretical framework for this process. Initial residuals are obtained after an R-fit of each center. The Wilcoxon score is used for the R-fit. An adaptation is carried out within centers. Selector statistic from the ordered combined residuals of the R-fit and cutoff points based on the sample size are computed at each center. A corresponding score is selected for the unknown distribution of R-fit residuals at each center. To obtain the estimate $\hat{\Delta}$ of $\Delta$, we solve
\( T_{\psi_1} \left( \hat{\Delta} \right) \overset{\Delta}{=} 0, \)

by numerical iterative search methods like bisection or false position for each center. Suppose we have \( n_i \) subjects at each center and \( n_{ij} \) number of patients randomly assigned to the treatment group at each center. Let \( \hat{\Delta}_i \) be the fixed effect estimate at the \( i \)th center, then the overall estimate obtained by,

\[
\hat{\Delta} = \sum_{i=1}^{m} \frac{n_{ij}}{n_i} \hat{\Delta}_i
\]

An overall \( \hat{\Delta} \) is also computed for using equation (3.18). A meta-test based on the standardized test of \( \hat{\Delta}_i \) is used. That is,

\[
\text{reject } H_0 \text{ in favor of } H_a \text{ if } \left| \sum_{i=1}^{m} \frac{Q_i}{\sqrt{\text{Var}_{H_0} Q_i}} \right| > Z_{\alpha},
\]

where \( Q_i = \sum_{j=1}^{n} \varphi_i \left[ \frac{R(Z_j)}{n+1} \right] I(Z_j = Y_j) \) is the test statistic at the \( i \)th center, \( Z_{\alpha} = \Phi^{-1} (1 - \alpha) \) and \( \Phi \) is the standard normal distribution.
4.1.4 Hogg-McKean Adaptation on Sample Ignore Center

We will call this process HMIC. For this process, centers are ignored and samples combined. The combined data from all the centers are ordered. The selector statistic is calculated from the ordered samples and cutoff benchmarks obtained. Classification is done by the selector statistics and a score selected. Under this process one score, \( \varphi \), is selected. The test statistic for this process is

\[
Z = \frac{T_{\varphi}}{\sqrt{\text{Var}_0 (T_{\varphi})}}.
\]

Given a prespecified convergence tolerance level, \( \hat{\Delta} \) is obtained by solving

\[
T_{\varphi} \left( \hat{\Delta} \right) = 0,
\]

using a bisection method or false position method.

The danger though for using meta-analysis is the tendency to combined results of the multi-center trials while ignoring potential differences from center to center, see Meinert (1989).
CHAPTER V

NUMERIC EXAMPLES AND SIMULATION RESULTS

In this section we present a numerical example and simulation study of the adaptive procedures. We will compare these adaptive procedures to a non parametric 'unadaptive' procedure called the multiple ranking Wilcoxon (MRW) and the maximum likelihood method (ML). The MRW was proposed by Rashid et. al. (2011). This is similar to the MR, except that adaption is not done and Wilcoxon score is used for all centers. We consider simulation results for normal, contaminated normal, truncated normal, truncated logistic and contaminated logistic distributions. The error distributions range from heavy-tailed to light-tailed distributions and from symmetric distributions to skewed distributions. We assume that there are three centers. Data are generated from these distributions at each of the three centers. For each of the three centers, we have two treatment groups. Equal samples are generated at each center. The ratio of subjects in treatment group to sample size at each center is maintain at the same value for all centers. We also considered the set of interclass correlation coefficient \( \{0, 0.2, 0.4, 0.6, 0.8\} \). Under \( H_0 \), data were generated from this model,

\[
Y_j^{(i)} = \mu + \beta_j^{(i)} \Delta + b_i + e_j^{(i)}, \quad j = 1, \ldots, n_t.
\]

All calculations were done using the software package, R. We start with a real life example.
5.1 Numeric Example for Weight Loss Data

In this example, we consider a placebo controlled appetite suppressant study taken from O’Gorman (2004). The data consist of weights of clients who were randomly assigned to placebo and appetite suppressant drug. The investigator wanted to know whether the suppressant will reduce appetite and subsequently weight. Twenty clients were randomly assigned to each of the six centers. At each center, randomization to each treatment group was done. Measurements were taken at baseline and at the end of twelve week period of the experiment. The response is the difference between the baseline and the final measurement of weight in kilograms. Negative values indicate weight gain, see the appendix for data. To investigate the robustness of our schemes, we obtained a second data set called (Ndata) by replacing the maximum weight loss and gain by 3 at center 3 and 5 at center 5 respectively in the original data.

Results for our adaptive schemes are displayed of on Tables (4) and (5). From our adaptive scheme, the underlying structure of the unknown distribution at centers 1, 2, 3 and 5 is symmetric and has heavy tail. The distributions for centers 4 and 6 are left skewed with heavy and moderate tails respectively. A formal test at centers 3 and 5 for both the new data and original data indicates no statistical significant difference in weight loss for those taking placebo and the suppressant. This implies, the overall test statistics and estimates for both the new and old data should not be affected that much. However the maximum likelihood method changes drastically as compared to the nonparametric procedures. While the overall test for the old data shows significance difference between placebo and suppressant
Table 4
Results for the Weight Loss Data

<table>
<thead>
<tr>
<th>Method</th>
<th>P-value</th>
<th>Test</th>
<th>Est</th>
<th>SE</th>
<th>τ or σ</th>
<th>Distribution(Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMS</td>
<td>0.00</td>
<td>3.37</td>
<td>2.92</td>
<td>0.86</td>
<td>1.45</td>
<td>SH(6) SH(6) SH(6) LH(3) SH(6) LM(2)</td>
</tr>
<tr>
<td>HMR</td>
<td>0.00</td>
<td>3.09</td>
<td>2.87</td>
<td>0.93</td>
<td>1.54</td>
<td>SH(6) SH(6) SH(6) LH(3) SH(6) LM(2)</td>
</tr>
<tr>
<td>HMIC</td>
<td>0.00</td>
<td>3.25</td>
<td>1.10</td>
<td>0.34</td>
<td>2.30</td>
<td>SH(6)</td>
</tr>
<tr>
<td>MR</td>
<td>0.00</td>
<td>3.18</td>
<td>1.00</td>
<td>0.31</td>
<td>1.78</td>
<td>SH(6) SH(6) SH(6) LH(3) SH(6) LM(2)</td>
</tr>
<tr>
<td>MRW</td>
<td>0.00</td>
<td>3.18</td>
<td>0.80</td>
<td>0.25</td>
<td>1.36</td>
<td>Wilcoxon</td>
</tr>
<tr>
<td>ML</td>
<td>0.07</td>
<td>1.83</td>
<td>0.81</td>
<td>0.44</td>
<td>2.38</td>
<td>Normal (Not Applicable)</td>
</tr>
</tbody>
</table>

Table 5
Results for the Weight Loss New Data, Ndata

<table>
<thead>
<tr>
<th>Method</th>
<th>P-value</th>
<th>Test</th>
<th>Estimate</th>
<th>SE</th>
<th>τ or σ</th>
<th>Distribution(Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMS</td>
<td>0.00</td>
<td>3.55</td>
<td>2.62</td>
<td>0.74</td>
<td>1.58</td>
<td>SH(6) SH(6) SM(5) LH(3) SM(5) LM(2)</td>
</tr>
<tr>
<td>HMR</td>
<td>0.00</td>
<td>3.19</td>
<td>2.85</td>
<td>0.81</td>
<td>1.38</td>
<td>SH(6) SH(6) SM(5) LH(3) SM(5) LM(2)</td>
</tr>
<tr>
<td>HMIC</td>
<td>0.00</td>
<td>3.17</td>
<td>1.00</td>
<td>0.32</td>
<td>1.95</td>
<td>SH(6)</td>
</tr>
<tr>
<td>MR</td>
<td>0.00</td>
<td>2.95</td>
<td>0.86</td>
<td>0.29</td>
<td>1.58</td>
<td>SH(6) SH(6) SM(5) LH(3) SM(5) LM(2)</td>
</tr>
<tr>
<td>MRW</td>
<td>0.00</td>
<td>3.01</td>
<td>0.70</td>
<td>0.23</td>
<td>1.26</td>
<td>Wilcoxon (5)</td>
</tr>
<tr>
<td>ML</td>
<td>0.19</td>
<td>1.33</td>
<td>0.49</td>
<td>0.37</td>
<td>1.99</td>
<td>Normal (Not Applicable)</td>
</tr>
</tbody>
</table>

at 1% level of significance for all methods, the least squares method for the new data shows otherwise. It is worth-noting that even though at 5% level of significance, the least squares (ML) results shows no effect, it is not robust. Tables (4) and (5) show the results of the methods discussed in this research.
5.2 Simulations for Class of Contaminated Normal and Normal Distributions

5.2.1 Simulation Results for Normal

In this example, we generated data from a normal distribution under $H_0$. That assuming center is normal and $\Delta = 0$, normal random variables generated for $\varepsilon_j^{(i)}$ in the model,

$$Y_j^{(i)} = \mu + a_j^{(i)} \Delta + b_i + \varepsilon_j^{(i)}, \quad j = 1, \ldots, n_i,$$

at each center. This was done for each $\rho = \{0, 0.2, 0.4, 0.6, 0.8\}$. We considered sample sizes 15, 20, 25, 30, 50, and 100 for each center for each simulation. The tables (6) and (7) shows the results of the empirical levels at 5% nominal level. For a sample size of 10,000, the confidence interval of 5% nominal level is (0.0456, 0.0543). At normal distribution, if there is no interclass correlation, the empirical levels for the Hogg-McKean’s adaptation on sample, HMS, Hogg-McKean’s Ignore Center adaptation HMIC and the maximum likelihood, ML, perform well for all sample sizes. This is because the empirical levels are within the 5% nominal level interval, (0.0456, 0.0543). Hogg-McKean adaptation on Residuals, HMR and Many Ranking MR adaptation performs well at 5% nominal level for sample size 50 or more. The MR is conservative for sample size below 50. Many ranking Wilcoxon MRW is conservative for all sample sizes. Similar results are obtained when there is an interclass correlation, except that there is a sharp increase in empirical levels of the HMIC adaptive procedure.
Table 6

Empirical Levels for Normal Data, $\rho = 0, 0.2, 0.4$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$n_1$</th>
<th>$n_{tt}$</th>
<th>HMS</th>
<th>HMR</th>
<th>HMIC</th>
<th>MR</th>
<th>MRW</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>15</td>
<td>8</td>
<td>0.0498</td>
<td>0.0757</td>
<td>0.0481</td>
<td>0.0642</td>
<td>0.0863</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td>0.0489</td>
<td>0.0678</td>
<td>0.049</td>
<td>0.0665</td>
<td>0.0811</td>
<td>0.0514</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>12</td>
<td>0.0479</td>
<td>0.0626</td>
<td>0.0481</td>
<td>0.0602</td>
<td>0.0762</td>
<td>0.0478</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15</td>
<td>0.0503</td>
<td>0.0644</td>
<td>0.0487</td>
<td>0.0617</td>
<td>0.0801</td>
<td>0.0507</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.0471</td>
<td>0.053</td>
<td>0.0463</td>
<td>0.0519</td>
<td>0.0723</td>
<td>0.0465</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.0487</td>
<td>0.0508</td>
<td>0.049</td>
<td>0.0503</td>
<td>0.0682</td>
<td>0.0485</td>
</tr>
<tr>
<td>0.2</td>
<td>15</td>
<td>8</td>
<td>0.0511</td>
<td>0.0768</td>
<td>0.071</td>
<td>0.0602</td>
<td>0.0832</td>
<td>0.0489</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td>0.0491</td>
<td>0.0676</td>
<td>0.096</td>
<td>0.0688</td>
<td>0.0817</td>
<td>0.0494</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>12</td>
<td>0.0536</td>
<td>0.0666</td>
<td>0.096</td>
<td>0.0669</td>
<td>0.0848</td>
<td>0.0526</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15</td>
<td>0.0517</td>
<td>0.0632</td>
<td>0.1138</td>
<td>0.0633</td>
<td>0.0801</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.0442</td>
<td>0.0511</td>
<td>0.0934</td>
<td>0.0525</td>
<td>0.0711</td>
<td>0.0454</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.0466</td>
<td>0.0514</td>
<td>0.103</td>
<td>0.0495</td>
<td>0.0682</td>
<td>0.0435</td>
</tr>
<tr>
<td>0.4</td>
<td>15</td>
<td>8</td>
<td>0.0482</td>
<td>0.0744</td>
<td>0.1023</td>
<td>0.062</td>
<td>0.0822</td>
<td>0.0446</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td>0.0473</td>
<td>0.0667</td>
<td>0.0939</td>
<td>0.0694</td>
<td>0.0804</td>
<td>0.0449</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>12</td>
<td>0.0502</td>
<td>0.0613</td>
<td>0.094</td>
<td>0.064</td>
<td>0.0792</td>
<td>0.0461</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15</td>
<td>0.0503</td>
<td>0.0628</td>
<td>0.1059</td>
<td>0.0595</td>
<td>0.0775</td>
<td>0.0466</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.0483</td>
<td>0.0531</td>
<td>0.1242</td>
<td>0.0538</td>
<td>0.0739</td>
<td>0.0482</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.042</td>
<td>0.0493</td>
<td>0.1122</td>
<td>0.0513</td>
<td>0.0615</td>
<td>0.0441</td>
</tr>
</tbody>
</table>
Table 7
Empirical Levels for Normal Data, $\rho = 0.6, 0.8$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$n_i$</th>
<th>$n_{it}$</th>
<th>HMS</th>
<th>HMR</th>
<th>HMIC</th>
<th>MR</th>
<th>MRW</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>15</td>
<td>8</td>
<td>0.0494</td>
<td>0.0754</td>
<td>0.0853</td>
<td>0.0623</td>
<td>0.0813</td>
<td>0.0486</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td>0.0487</td>
<td>0.0668</td>
<td>0.0917</td>
<td>0.0681</td>
<td>0.0808</td>
<td>0.0473</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>12</td>
<td>0.0504</td>
<td>0.0662</td>
<td>0.1014</td>
<td>0.0666</td>
<td>0.084</td>
<td>0.0484</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15</td>
<td>0.0494</td>
<td>0.0629</td>
<td>0.143</td>
<td>0.06</td>
<td>0.0809</td>
<td>0.0505</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.0479</td>
<td>0.0539</td>
<td>0.164</td>
<td>0.0536</td>
<td>0.0731</td>
<td>0.0474</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.0411</td>
<td>0.0523</td>
<td>0.12</td>
<td>0.0512</td>
<td>0.0613</td>
<td>0.0466</td>
</tr>
<tr>
<td>0.8</td>
<td>15</td>
<td>8</td>
<td>0.052</td>
<td>0.0751</td>
<td>0.104</td>
<td>0.0603</td>
<td>0.0802</td>
<td>0.0502</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td>0.0541</td>
<td>0.071</td>
<td>0.193</td>
<td>0.0695</td>
<td>0.0847</td>
<td>0.0506</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>12</td>
<td>0.0507</td>
<td>0.0624</td>
<td>0.11</td>
<td>0.0659</td>
<td>0.0833</td>
<td>0.0485</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15</td>
<td>0.0499</td>
<td>0.0624</td>
<td>0.1084</td>
<td>0.0607</td>
<td>0.0787</td>
<td>0.0494</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.0518</td>
<td>0.0575</td>
<td>0.1106</td>
<td>0.0577</td>
<td>0.0764</td>
<td>0.0494</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.047</td>
<td>0.0517</td>
<td>0.1</td>
<td>0.0524</td>
<td>0.0601</td>
<td>0.0474</td>
</tr>
</tbody>
</table>
Table 8

Results for Main Effect Example - Without Outlier

<table>
<thead>
<tr>
<th>Method</th>
<th>Test</th>
<th>P-values</th>
<th>$\tau$ or $\sigma$</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMS</td>
<td>4.20</td>
<td>0.00</td>
<td>1.522</td>
<td>4.1802</td>
</tr>
<tr>
<td>HMR</td>
<td>4.421</td>
<td>0.00</td>
<td>1.522</td>
<td>4.1802</td>
</tr>
<tr>
<td>HMIC</td>
<td>3.97</td>
<td>0.00</td>
<td>1.33</td>
<td>2.5426</td>
</tr>
<tr>
<td>MR</td>
<td>4.258</td>
<td>0.00</td>
<td>1.425</td>
<td>2.4868</td>
</tr>
<tr>
<td>MRW</td>
<td>4.307</td>
<td>0.00</td>
<td>2.223</td>
<td>2.4721</td>
</tr>
<tr>
<td>ML</td>
<td>2.166</td>
<td>0.00</td>
<td>8.718</td>
<td>3.9601</td>
</tr>
</tbody>
</table>

Table 9

Results for Main Effect Example - With Outliers

<table>
<thead>
<tr>
<th>Method</th>
<th>Test</th>
<th>P-values</th>
<th>$\tau$ or $\sigma$</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMS</td>
<td>4.10</td>
<td>0.00</td>
<td>2.35</td>
<td>3.76</td>
</tr>
<tr>
<td>HMR</td>
<td>4.10</td>
<td>0.00</td>
<td>2.40</td>
<td>3.75</td>
</tr>
<tr>
<td>HMIC</td>
<td>3.61</td>
<td>0.00</td>
<td>2.65</td>
<td>2.37</td>
</tr>
<tr>
<td>MR</td>
<td>4.12</td>
<td>0.00</td>
<td>2.38</td>
<td>2.53</td>
</tr>
<tr>
<td>MRW</td>
<td>4.34</td>
<td>0.00</td>
<td>2.28</td>
<td>2.55</td>
</tr>
<tr>
<td>ML</td>
<td>1.57</td>
<td>0.12</td>
<td>8.03</td>
<td>3.32</td>
</tr>
</tbody>
</table>

5.2.2 Simulation Results for Contaminated Normal CN , $\epsilon = 0.2$, $\sigma = 5$

In this example we generated a right skewed heavy tailed for two centers, and left skewed heavy tailed for one center for 20 subjects at each center and 10 in each treatment group. Block random effect is assumed to be normally distributed and $\Delta = 2.5$. The results on tables (8) and (9) shows that the ML method was not robust when an outlier was introduced. Again the non parametric methods gave a correct analysis.
A simulation study was conducted for heavy tailed to light distributions and skewed to symmetric distributions for contaminated and truncated normal. Under, $\Delta=0$, 10,000 simulations were done for 15, 20, 25, 30, 50 and 100 subjects for each center. These were done for each interclass correlation $\rho\in\{0, 0.2, 0.4, 0.6, 0.8\}$. The empirical levels at 5% nominal for the proposed methods when $\rho\in\{0, 0.2, 0.4, 0.6, 0.8\}$ are shown on the tables (10) and (11). The simulations result show that all the adaptive schemes perform very well but the ML method. The HIMC procedure seem to fluctuate.
Table 11

Empirical Levels for CN(.2,25) Data, $\rho = 0.6, 0.8$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$n_1$</th>
<th>$n_{1t}$</th>
<th>HMS</th>
<th>HMR</th>
<th>HMIC</th>
<th>MR</th>
<th>MRW</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>15</td>
<td>8</td>
<td>0.051</td>
<td>0.0605</td>
<td>0.07</td>
<td>0.0583</td>
<td>0.053</td>
<td>0.1392</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td>0.0448</td>
<td>0.0528</td>
<td>0.071</td>
<td>0.0503</td>
<td>0.053</td>
<td>0.1043</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>12</td>
<td>0.048</td>
<td>0.0555</td>
<td>0.075</td>
<td>0.0535</td>
<td>0.053</td>
<td>0.0925</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15</td>
<td>0.041</td>
<td>0.0505</td>
<td>0.0775</td>
<td>0.0545</td>
<td>0.0625</td>
<td>0.0725</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.0525</td>
<td>0.0605</td>
<td>0.0815</td>
<td>0.0613</td>
<td>0.065</td>
<td>0.1025</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.0475</td>
<td>0.049</td>
<td>0.0825</td>
<td>0.0627</td>
<td>0.0607</td>
<td>0.0945</td>
</tr>
<tr>
<td>0.8</td>
<td>15</td>
<td>8</td>
<td>0.045</td>
<td>0.0575</td>
<td>0.0855</td>
<td>0.0513</td>
<td>0.0615</td>
<td>0.1075</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td>0.052</td>
<td>0.0605</td>
<td>0.086</td>
<td>0.0575</td>
<td>0.0575</td>
<td>0.1125</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>12</td>
<td>0.0468</td>
<td>0.0513</td>
<td>0.085</td>
<td>0.0519</td>
<td>0.0517</td>
<td>0.0895</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15</td>
<td>0.05</td>
<td>0.0565</td>
<td>0.0825</td>
<td>0.0515</td>
<td>0.0615</td>
<td>0.0825</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.0475</td>
<td>0.0535</td>
<td>0.0725</td>
<td>0.0502</td>
<td>0.055</td>
<td>0.0955</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.051</td>
<td>0.0512</td>
<td>0.085</td>
<td>0.0525</td>
<td>0.0525</td>
<td>0.0975</td>
</tr>
</tbody>
</table>
5.3 Simulations for Contaminated and Truncated Logistic

In this section, we examine simulated data from contaminated and truncated logistics distributions (CL). Consider the block mixed model,

\[ Y_{jw}^{(t)} = \mu^* + x_{jw}^{(t)} \Delta + b_i + \varepsilon_{jw}^{(t)}, \quad j = 1, \ldots, n_a, \]

and suppose \( \varepsilon_{jw}^{(t)} \), under \( H_0 \) and for different centers, data are drawn from contaminated and truncated logistics distributions. Then our generated data at each of the three centers, may be left skewed or/and right skewed distribution. The tail weight may be heavy or light. Let us briefly review contaminated distributions and truncated distributions.

Let \( Z_1 \) and \( Z_2 \) be random sample from a logistic distribution, define \( X \) by

\[ X = (1 - I_\varepsilon)Z_1 + I_\varepsilon(\sigma_c Z_2 + \mu) \quad (5.1) \]

where \( \sigma_c \) is standard deviation of contamination, \( I_\varepsilon \) a characteristic function, \( \varepsilon \) is the percentage of contamination and \( \mu \) a location parameter. The distribution of contaminated logistic is obtained as follows,
\[ P[X \leq x] = P[X = (1 - I_\epsilon)Z_1 + I_\epsilon(\sigma_c Z_2 + \mu) \leq x] \]
\[ = E[P[X = (1 - I_\epsilon)Z_1 + I_\epsilon(\sigma_c Z_2 + \mu) \leq x]|I_\epsilon] \]
\[ = P[Z_1 \leq x|I_\epsilon = 0] = F_Z(x) \]
\[ = P[\sigma_c Z_2 + \mu \leq x|I_\epsilon = 1] = F_Z\left(\frac{x - \mu}{\sigma_c}\right) \]
\[ P[X \leq x] = (1 - \epsilon)F_Z(x) + \epsilon F_Z\left(\frac{x - \mu}{\sigma_c}\right). \]

Hence the density \( f_X \) is

\[
\frac{dP[X \leq x]}{dx} = (1 - \epsilon)f_Z(x) + \frac{\epsilon}{\sigma_c} f_Z\left(\frac{x - \mu}{\sigma_c}\right)
\]

when \( \mu > 0 \) we have right skewed and \( \mu < 0 \) we have left skewed.

Consider the logistic distribution, \( F(x) = \frac{1}{1 + e^{-x}} \), let \( F(x) = u \) then \( x = \log\left(\frac{u}{1-u}\right) \).

Thus \( F^{-1}(u) = \log\left(\frac{1}{1-u}\right) \), for some given \( \alpha_1 \) and \( \alpha_2 \) the quantiles are

\[ q_{\alpha_1} = F^{-1}(\alpha_1) = \log\left(\frac{\alpha_1}{1 - \alpha_1}\right) \text{ and} \]
\[ q_{\alpha_2} = F^{-1}(\alpha_2) = \log\left(\frac{1 - \alpha_2}{\alpha_2}\right) \]

respectively. The distribution of a random variable, \( X \) is
\[
\begin{align*}
P(X \leq x) &= P[X \leq x | q_{\alpha_1} \leq x \leq q_{\alpha_2}] \\
&= \begin{cases} 
0 & \text{if } x < q_{\alpha_1} \\
\gamma & \text{if } q_{\alpha_1} \leq x \leq q_{\alpha_2} \\
1 & \text{otherwise}
\end{cases}
\end{align*}
\]

where

\[
\begin{align*}
\gamma &= \frac{P[q_{\alpha_1} \leq X \leq x]}{P[q_{\alpha_1} \leq X \leq q_{\alpha_2}]} \\
&= \frac{\frac{1}{1+e^{-x}} - \frac{1}{1+e^{-q_{\alpha_1}}}}{F(q_{\alpha_1}) - F(q_{\alpha_2})} \\
&= \frac{\frac{1}{1+e^{-x}} - \alpha_1}{1 - \alpha_2 - \alpha_1}
\end{align*}
\]

5.3.1 Simulation Results for Contaminated and Truncated Logistic

Simulation study was conducted for symmetric light, right heavy and left heavy distributions. Three centers were considered for the simulations. For this class, we make use of \( \varepsilon=0.2 \) and \( \mu = 4 \). For the null model, \( \Delta = 0 \), data are generated for each center. Empirical levels and power levels at a nominal 5% for each method were recorded. We performed 10,000 simulations for each of the interclass correlation \( \rho=\{0.2, 0.4, 0.6, 0.8\} \). Tables (12-13) and (15-18) show the results of empirical levels and power at 5% nominal for the 10,000 simulations respectively.
Table 12

Empirical Levels for CL Data, $\epsilon = 0.2$, $\rho = 0, 0.2, 0.4$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$n_i$</th>
<th>$n_{it}$</th>
<th>HMS</th>
<th>HMR</th>
<th>HMIC</th>
<th>MR</th>
<th>MRW</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>15</td>
<td>8</td>
<td>0.0525</td>
<td>0.0603</td>
<td>0.075</td>
<td>0.0617</td>
<td>0.0620</td>
<td>0.0919</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td>0.0515</td>
<td>0.0509</td>
<td>0.0844</td>
<td>0.0605</td>
<td>0.0732</td>
<td>0.0636</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>12</td>
<td>0.0466</td>
<td>0.0501</td>
<td>0.0761</td>
<td>0.0503</td>
<td>0.075</td>
<td>0.1003</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15</td>
<td>0.0505</td>
<td>0.0523</td>
<td>0.0819</td>
<td>0.0592</td>
<td>0.0678</td>
<td>0.0997</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.05</td>
<td>0.0528</td>
<td>0.0706</td>
<td>0.0531</td>
<td>0.0764</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.0499</td>
<td>0.0498</td>
<td>0.0832</td>
<td>0.0623</td>
<td>0.0715</td>
<td>0.0916</td>
</tr>
<tr>
<td>0.2</td>
<td>15</td>
<td>8</td>
<td>0.0469</td>
<td>0.0618</td>
<td>0.1022</td>
<td>0.0638</td>
<td>0.0623</td>
<td>0.0921</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td>0.0522</td>
<td>0.0556</td>
<td>0.0712</td>
<td>0.0541</td>
<td>0.0729</td>
<td>0.0861</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>12</td>
<td>0.0530</td>
<td>0.0521</td>
<td>0.0881</td>
<td>0.0519</td>
<td>0.0692</td>
<td>0.0813</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15</td>
<td>0.0517</td>
<td>0.0508</td>
<td>0.083</td>
<td>0.0472</td>
<td>0.0743</td>
<td>0.0822</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.0526</td>
<td>0.0511</td>
<td>0.0734</td>
<td>0.0591</td>
<td>0.0618</td>
<td>0.1001</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.0499</td>
<td>0.0527</td>
<td>0.071</td>
<td>0.0544</td>
<td>0.0635</td>
<td>0.0925</td>
</tr>
<tr>
<td>0.4</td>
<td>15</td>
<td>8</td>
<td>0.0501</td>
<td>0.0515</td>
<td>0.0703</td>
<td>0.0601</td>
<td>0.0671</td>
<td>0.0912</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10</td>
<td>0.0503</td>
<td>0.0525</td>
<td>0.0824</td>
<td>0.0621</td>
<td>0.0688</td>
<td>0.1116</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>12</td>
<td>0.0495</td>
<td>0.0495</td>
<td>0.0835</td>
<td>0.0603</td>
<td>0.0722</td>
<td>0.0815</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15</td>
<td>0.0516</td>
<td>0.0469</td>
<td>0.0995</td>
<td>0.0633</td>
<td>0.0647</td>
<td>0.0933</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>25</td>
<td>0.0525</td>
<td>0.0469</td>
<td>0.0835</td>
<td>0.0578</td>
<td>0.0615</td>
<td>0.0939</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>0.0523</td>
<td>0.0477</td>
<td>0.0712</td>
<td>0.0595</td>
<td>0.0662</td>
<td>0.0886</td>
</tr>
</tbody>
</table>
5.4 Results for Empirical Power Based on 10,000 Simulations

In this section, we present the empirical power of our adaptive schemes for 10,000 simulations. A nominal level of 5% is used. Data were generated for each center when $\sigma^2 = 1$ and $\Delta = \{0.1, 0.2, 0.3, 0.4, 0.5\}$. Samples of sizes 20, 30 and 50 at each center were considered for these simulation studies. The resulting power curves are in Figures (5-10). We considered HMS, HMR and ML methods for the data generated from normal and contaminated logistic because from our empirical levels, HMS and HMR are the two competing adaptive procedures. The red color represents HMS, the green color HMR and the blue ML. For a given sample size, power increases as $\rho$ increases (read rowwise). Overall power increases with increasing sample size and increasing $\rho$. At normal, the ML outperform the HMS and HMR methods. For contaminated logistics, HMR procedure is has higher power.
than ML but has less power than HMS. The results of the power calculations are displayed on Tables (15-18).
Power Curves for Normal, $\rho = 0.2, 0.4, 0.6, 0.8$
Figure 6

Power Curves for Normal, $\rho = 0.2, 0.4, 0.6, 0.8$
Figure 7

Power Curves for Normal, $\rho = 0.2, 0.4, 0.6, 0.8$
Figure 8

Power Curves for CL $\varepsilon = 0.2, \rho = 0.2, 0.4, 0.6, 0.8$
Figure 9

Power Curves for CL $\epsilon = 0.2, \rho = 0.2, 0.4, 0.6, 0.8$
Figure 10

Power Curves for CL $\epsilon = 0.2$, $\rho = 0.2, 0.4, 0.6, 0.8$
### Table 14

**Power Levels of ML Method for Normal Data, \( \rho = 0.2, 0.4, 0.6, 0.8 \)**

<table>
<thead>
<tr>
<th>((\rho, n_i))</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.2, 20)</td>
<td>0.3880</td>
<td>0.5837</td>
<td>0.7553</td>
<td>0.7999</td>
<td>0.8011</td>
</tr>
<tr>
<td>(0.4, 20)</td>
<td>0.4843</td>
<td>0.6045</td>
<td>0.7608</td>
<td>0.8001</td>
<td>0.8492</td>
</tr>
<tr>
<td>(0.6, 20)</td>
<td>0.4940</td>
<td>0.6215</td>
<td>0.7955</td>
<td>0.8137</td>
<td>0.8587</td>
</tr>
<tr>
<td>(0.8, 20)</td>
<td>0.5094</td>
<td>0.6584</td>
<td>0.8012</td>
<td>0.8603</td>
<td>0.8984</td>
</tr>
<tr>
<td>(0.2, 30)</td>
<td>0.4406</td>
<td>0.6147</td>
<td>0.7741</td>
<td>0.8217</td>
<td>0.9139</td>
</tr>
<tr>
<td>(0.4, 30)</td>
<td>0.4965</td>
<td>0.6896</td>
<td>0.7830</td>
<td>0.8473</td>
<td>0.9220</td>
</tr>
<tr>
<td>(0.6, 30)</td>
<td>0.5488</td>
<td>0.6950</td>
<td>0.7970</td>
<td>0.8742</td>
<td>0.9510</td>
</tr>
<tr>
<td>(0.8, 30)</td>
<td>0.6079</td>
<td>0.7453</td>
<td>0.8185</td>
<td>0.8998</td>
<td>0.9560</td>
</tr>
<tr>
<td>(0.2, 50)</td>
<td>0.5221</td>
<td>0.6314</td>
<td>0.7616</td>
<td>0.8331</td>
<td>0.9213</td>
</tr>
<tr>
<td>(0.4, 50)</td>
<td>0.5302</td>
<td>0.6899</td>
<td>0.7892</td>
<td>0.8567</td>
<td>0.9345</td>
</tr>
<tr>
<td>(0.6, 50)</td>
<td>0.5612</td>
<td>0.7123</td>
<td>0.7996</td>
<td>0.8812</td>
<td>0.9612</td>
</tr>
<tr>
<td>(0.8, 50)</td>
<td>0.7121</td>
<td>0.7523</td>
<td>0.8353</td>
<td>0.9013</td>
<td>0.9812</td>
</tr>
</tbody>
</table>

### Table 15

**Power Levels of HMR Method for Normal Data, \( \rho = 0.2, 0.4, 0.6, 0.8 \)**

<table>
<thead>
<tr>
<th>((\rho, n_i))</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.2, 20)</td>
<td>0.3338</td>
<td>0.4484</td>
<td>0.5334</td>
<td>0.6336</td>
<td>0.7496</td>
</tr>
<tr>
<td>(0.4, 20)</td>
<td>0.3808</td>
<td>0.4837</td>
<td>0.5456</td>
<td>0.7107</td>
<td>0.8454</td>
</tr>
<tr>
<td>(0.6, 20)</td>
<td>0.4196</td>
<td>0.5297</td>
<td>0.6224</td>
<td>0.7363</td>
<td>0.8606</td>
</tr>
<tr>
<td>(0.8, 20)</td>
<td>0.5366</td>
<td>0.6106</td>
<td>0.6418</td>
<td>0.7607</td>
<td>0.8625</td>
</tr>
<tr>
<td>(0.2, 30)</td>
<td>0.4314</td>
<td>0.4944</td>
<td>0.5559</td>
<td>0.6859</td>
<td>0.8014</td>
</tr>
<tr>
<td>(0.4, 30)</td>
<td>0.4429</td>
<td>0.5066</td>
<td>0.6122</td>
<td>0.7285</td>
<td>0.8478</td>
</tr>
<tr>
<td>(0.6, 30)</td>
<td>0.4570</td>
<td>0.5488</td>
<td>0.6506</td>
<td>0.7591</td>
<td>0.8726</td>
</tr>
<tr>
<td>(0.8, 30)</td>
<td>0.5606</td>
<td>0.6422</td>
<td>0.7155</td>
<td>0.7740</td>
<td>0.8960</td>
</tr>
<tr>
<td>(0.2, 50)</td>
<td>0.4614</td>
<td>0.5135</td>
<td>0.5619</td>
<td>0.6991</td>
<td>0.8614</td>
</tr>
<tr>
<td>(0.4, 50)</td>
<td>0.4713</td>
<td>0.5342</td>
<td>0.6612</td>
<td>0.7430</td>
<td>0.8898</td>
</tr>
<tr>
<td>(0.6, 50)</td>
<td>0.4810</td>
<td>0.5782</td>
<td>0.6744</td>
<td>0.7711</td>
<td>0.8927</td>
</tr>
<tr>
<td>(0.8, 50)</td>
<td>0.5816</td>
<td>0.6722</td>
<td>0.8255</td>
<td>0.8921</td>
<td>0.9960</td>
</tr>
</tbody>
</table>
Table 16

Power Levels of HMS Method for Normal Data, $\rho = 0.2, 0.4, 0.6, 0.8$

<table>
<thead>
<tr>
<th>$(\rho, n_t)$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.2,20)</td>
<td>0.3285</td>
<td>0.4504</td>
<td>0.5314</td>
<td>0.6822</td>
<td>0.7714</td>
</tr>
<tr>
<td>(0.4,20)</td>
<td>0.3502</td>
<td>0.4795</td>
<td>0.5835</td>
<td>0.7171</td>
<td>0.8307</td>
</tr>
<tr>
<td>(0.6,20)</td>
<td>0.3714</td>
<td>0.5260</td>
<td>0.6114</td>
<td>0.7660</td>
<td>0.8568</td>
</tr>
<tr>
<td>(0.8,20)</td>
<td>0.4442</td>
<td>0.5756</td>
<td>0.7490</td>
<td>0.7852</td>
<td>0.8575</td>
</tr>
<tr>
<td>(0.2,30)</td>
<td>0.4146</td>
<td>0.4956</td>
<td>0.5403</td>
<td>0.7011</td>
<td>0.8186</td>
</tr>
<tr>
<td>(0.4,30)</td>
<td>0.4744</td>
<td>0.5053</td>
<td>0.5846</td>
<td>0.7485</td>
<td>0.8513</td>
</tr>
<tr>
<td>(0.6,30)</td>
<td>0.4887</td>
<td>0.5258</td>
<td>0.6397</td>
<td>0.7609</td>
<td>0.8802</td>
</tr>
<tr>
<td>(0.8,30)</td>
<td>0.5228</td>
<td>0.5890</td>
<td>0.7512</td>
<td>0.8126</td>
<td>0.9332</td>
</tr>
<tr>
<td>(0.2,50)</td>
<td>0.4731</td>
<td>0.5051</td>
<td>0.5761</td>
<td>0.8001</td>
<td>0.8443</td>
</tr>
<tr>
<td>(0.4,50)</td>
<td>0.5144</td>
<td>0.5654</td>
<td>0.6001</td>
<td>0.7717</td>
<td>0.8801</td>
</tr>
<tr>
<td>(0.6,50)</td>
<td>0.5189</td>
<td>0.5774</td>
<td>0.6812</td>
<td>0.7919</td>
<td>0.899</td>
</tr>
<tr>
<td>(0.8,50)</td>
<td>0.5765</td>
<td>0.6001</td>
<td>0.7892</td>
<td>0.8886</td>
<td>0.9901</td>
</tr>
</tbody>
</table>

Table 17

Power Levels of HMR Method for CL Data, $\rho = 0.2, 0.4, 0.6, 0.8$

<table>
<thead>
<tr>
<th>$(\rho, n_t)$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.2,20)</td>
<td>0.3752</td>
<td>0.7628</td>
<td>0.7927</td>
<td>0.8114</td>
<td>0.8324</td>
</tr>
<tr>
<td>(0.4,20)</td>
<td>0.4042</td>
<td>0.7876</td>
<td>0.8048</td>
<td>0.8534</td>
<td>0.8613</td>
</tr>
<tr>
<td>(0.6,20)</td>
<td>0.4236</td>
<td>0.7965</td>
<td>0.8556</td>
<td>0.8660</td>
<td>0.8678</td>
</tr>
<tr>
<td>(0.8,20)</td>
<td>0.5776</td>
<td>0.8019</td>
<td>0.8678</td>
<td>0.8791</td>
<td>0.8828</td>
</tr>
<tr>
<td>(0.2,30)</td>
<td>0.4281</td>
<td>0.7840</td>
<td>0.8041</td>
<td>0.8310</td>
<td>0.8343</td>
</tr>
<tr>
<td>(0.4,30)</td>
<td>0.4310</td>
<td>0.8032</td>
<td>0.8634</td>
<td>0.8716</td>
<td>0.8753</td>
</tr>
<tr>
<td>(0.6,30)</td>
<td>0.5840</td>
<td>0.8021</td>
<td>0.8766</td>
<td>0.8774</td>
<td>0.8751</td>
</tr>
<tr>
<td>(0.8,30)</td>
<td>0.6209</td>
<td>0.8135</td>
<td>0.8780</td>
<td>0.8837</td>
<td>0.9885</td>
</tr>
<tr>
<td>(0.2,50)</td>
<td>0.4412</td>
<td>0.7990</td>
<td>0.8130</td>
<td>0.8442</td>
<td>0.9010</td>
</tr>
<tr>
<td>(0.4,50)</td>
<td>0.4921</td>
<td>0.8222</td>
<td>0.8916</td>
<td>0.9020</td>
<td>0.9133</td>
</tr>
<tr>
<td>(0.6,50)</td>
<td>0.5910</td>
<td>0.8347</td>
<td>0.8972</td>
<td>0.9142</td>
<td>0.9200</td>
</tr>
<tr>
<td>(0.8,50)</td>
<td>0.6409</td>
<td>0.8725</td>
<td>0.9001</td>
<td>0.9407</td>
<td>0.9901</td>
</tr>
</tbody>
</table>
Table 18

Power Levels of HMS Method CL Data, $\rho = 0.2, 0.4, 0.6, 0.8$

<table>
<thead>
<tr>
<th>$(\rho, n_i)$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.2, 20)</td>
<td>0.5168</td>
<td>0.6193</td>
<td>0.7361</td>
<td>0.8150</td>
<td>0.8283</td>
</tr>
<tr>
<td>(0.4, 20)</td>
<td>0.5880</td>
<td>0.6562</td>
<td>0.8072</td>
<td>0.8110</td>
<td>0.8449</td>
</tr>
<tr>
<td>(0.6, 20)</td>
<td>0.6340</td>
<td>0.6610</td>
<td>0.8112</td>
<td>0.8289</td>
<td>0.8608</td>
</tr>
<tr>
<td>(0.8, 20)</td>
<td>0.6683</td>
<td>0.7144</td>
<td>0.8216</td>
<td>0.8442</td>
<td>0.8840</td>
</tr>
<tr>
<td>(0.2, 30)</td>
<td>0.5920</td>
<td>0.7503</td>
<td>0.8120</td>
<td>0.8580</td>
<td>0.8685</td>
</tr>
<tr>
<td>(0.4, 30)</td>
<td>0.5976</td>
<td>0.7960</td>
<td>0.8524</td>
<td>0.8738</td>
<td>0.8804</td>
</tr>
<tr>
<td>(0.6, 30)</td>
<td>0.6013</td>
<td>0.8460</td>
<td>0.8680</td>
<td>0.8833</td>
<td>0.8887</td>
</tr>
<tr>
<td>(0.8, 30)</td>
<td>0.6759</td>
<td>0.8511</td>
<td>0.8770</td>
<td>0.8897</td>
<td>0.8979</td>
</tr>
<tr>
<td>(0.2, 50)</td>
<td>0.6001</td>
<td>0.7842</td>
<td>0.8441</td>
<td>0.8621</td>
<td>0.8822</td>
</tr>
<tr>
<td>(0.4, 50)</td>
<td>0.6172</td>
<td>0.7999</td>
<td>0.8650</td>
<td>0.8892</td>
<td>0.8976</td>
</tr>
<tr>
<td>(0.6, 50)</td>
<td>0.6512</td>
<td>0.8434</td>
<td>0.8810</td>
<td>0.8941</td>
<td>0.9017</td>
</tr>
<tr>
<td>(0.8, 50)</td>
<td>0.7011</td>
<td>0.8602</td>
<td>0.8887</td>
<td>0.8997</td>
<td>0.9391</td>
</tr>
</tbody>
</table>
CHAPTER VI

CONCLUSIONS

6.1 Summary

We have developed a highly efficient adaptive scheme for mixed models under exchangeable errors. We have demonstrated that an adaptive procedure is applicable to mixed models under exchangeable errors. Besides the exchangeability assumption and the classification of skewness and tail weight no other condition is required for these schemes. Notwithstanding this limitation, our adaptive schemes are well behaved over a broad class of distributions ranging from heavy-tailed to light-tailed distributions and from symmetric distributions to skewed distributions. The schemes are robust. At normal distribution, based on the empirical levels, the HMS is as efficient as the ML method. There is a good agreement of our adaptive test (meta test) for all our schemes, and the ML test at normal. When the distribution of random errors are not normal, our 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 we have developed pave way for diagnostics checking of the fitted model.
We have developed an asymptotic theory and linearity results for the linear mixed model under exchangeable error. With asymptotic efficiency between two estimates in mind, efficacy results has also been developed for a balanced designed.

6.1.1 Future Work

The MR method showed promise and we plan to investigate further using residuals from the initial R-fit. From our efficacy results, we intend to investigate the asymptotic relative efficiency of the estimates from our schemes. Other areas we hope our adaptive procedures will work are profile analysis for repeated measures, interrupted time series and vector autoregressive process.
REFERENCES

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


McCullagh P. (2004), Exchangeability and regression models *Manuscript*.


APPENDIX

The discussion and proofs in this are in Pitman Regularity Consider the estimating equation,

\[ S(\Delta) = \sum_{i=1}^{m} S_i(\Delta) \]

then we define Pitman regular as follows see Hettmansperger and McKean (2011) page 23:

**Definition 0.0.1.** \( S(\Delta) \) is said to be Pitman regular if the following four conditions hold;

1. \( S(\Delta) \) is a nonincreasing function in \( \Delta \)

2. \( \bar{S}(\Delta) = S(\Delta)/n^\gamma \), for some \( \gamma > 0 \) there exist a function \( \mu(\Delta) \), such that \( \mu(0) = 0 \), \( \mu'(\Delta) \) is continuous at 0, \( \mu'(0) > 0 \) and either \( \bar{S}(0) \overset{P}{\rightarrow} \mu(\Delta) \) or \( E_\Delta(\bar{S}(0)) = \mu(\Delta) \)

3. The asymptotic linearity of \( S(\Delta) \) \( \sup_{|b| \leq B} \left| \sqrt{n} \bar{S}(\frac{b}{\sqrt{n}}) - \sqrt{n} \bar{S}(0) + \mu'(0)b \right| \overset{P}{\rightarrow} 0 \), for any \( B > 0 \)

4. there is a constant \( \sigma(0) \) such that \( \frac{\sqrt{n} \bar{S}(0)}{\sigma(0)} \overset{D}{\rightarrow} N(0,1) \).

The efficacy of \( S(\Delta) \) is given by,

\[ c = \frac{\mu'(0)}{\sigma(0)} \]

where \( \sigma^2(0) = \lim[n Var_0(\bar{S}(0))] \) and \( \mu'(0) = \frac{d}{d\Delta} E_\Delta[\bar{S}(0)]|_{\Delta=0} \)
Theorem 0.0.1. Suppose

\[ S(\Delta) = \sum_{i=1}^{m} S_i(\Delta) \]

is a Pitman regular. Then

\[
\frac{\sqrt{n} \bar{S}(b/\sqrt{n})}{\sigma(0)} \overset{D}{\to} N(0, 1) - cb
\]

\[
\frac{\sqrt{n} S(0)}{\sigma(0)} \overset{D}{\to} N(0, 1) - cb
\]

Functionals

To be able to investigate the robustness of our adaptive process under exchangeable errors over \( m \) clusters, we need to understand the concept of functionals. This is done within the context of location model see Hettmansperger and McKean (2011).

Definition 0.0.2. Let \( T(H) \) be a function defined on a set of distribution functions. We say \( T(H) \) is a location functional if

1. If \( G \) is stochastically larger than \( F \), i.e. \( G(x) \leq F(x) \) for all \( x \), then \( T(G) \geq T(F) \);

2. \( T(H_{aX+b}) = aT(H_X) + b, a > 0 \);

3. \( T(H_{-X}) = -T(H_X) \)

Suppose \( X \) has a location functional \( \theta = T(H) \). Then a statistical location model of
$X_i$ is given by,

$$X_i = \theta + e_i, \; i = 1, \ldots, n,$$

where $e_1, \ldots, n$ are independent and identically distributed random variables with distribution and density functions $F(x)$ and $f(x)$ respectively. $T(F)$ is the location of $e$. Now consider the block model,

**Basu Theorem**

The Indian Journal of Statistics Basu's theorem is not just useful in the context of adaptation but also in a wide range of applications like distribution theory, theory of estimation and hypothesis testing. We give some basic definitions needed for Basu's theorem. This will be followed by some examples. Applications and some converses of Basu's theorem will be discussed see Ghosh (2002).

*Definition 0.0.3. (sufficiency)*

Consider a random variable $X \in \mathbb{R}^n$ on some measurable space $\Omega$. Let $F$ denote a family of distribution of $X$ such that

$$F = \{F(X, \theta), \theta \in \Theta\}.$$

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

*Definition 0.0.4. (Minimal Sufficient)*

A statistic $T \equiv T(X)$ is minimal sufficient if
• $T$ is sufficient

• no other function of $T$ is sufficient unless it is a one-to-one function of $T$

**Definition 0.0.5. (Completeness)**

A statistic $T(X)$ is complete for $\theta \in \Theta$ if and only if, for any real-valued function (Borel) $h$,

- $E[h(T)] = 0$, $\forall \theta \in \Theta$, implies $P[h(T) = 0] = 1$

**Definition 0.0.6. (Ancillary)**

A statistic $T(x)$ is ancillary if the distribution of $T$ does not depend on $\theta$ for every $F(X; \theta)$, $\theta \in \Theta$.

It is worth remarking that sufficient statistic $T$ appears to be most successful in reducing data if no non constant function of $T$ is ancillary. Furthermore, a complete statistic boundedly complete. If $T$ is complete the $h(T)$ is also complete for any measurable function $h$. A complete sufficient statistic should be minimal sufficient however the converse is not necessarily true. For example, $T = (X(1), X(n)) = (\min(X_i), \max(X_i))$ is minimal sufficient but not complete.

**Lemma 0.0.1.** If $\theta$ is in exponential family of full rank with density

$$f_\theta(X) = \exp\{\theta^T T(X) - g(\theta)\} h(X)$$

then $T(X)$ is complete and sufficient for $\theta \in \Theta$.

For example suppose $X_i$, $i = 1, \ldots, n$ are random variables from $N(\mu, \sigma^2)$then the joint p.d.f is
\[
\frac{1}{(2\pi)^{n/2}} \exp \{ \theta_1 T_1 + \theta_2 T_2 - ng(\theta) \}
\]

where \( T_1 = \sum_{i=1}^{n} X_i, T_2 = -\sum_{i=1}^{n} X_i^2 \) and \( \theta = (\theta_1, \theta_2), \theta_1 = \frac{\mu}{\sigma^2} \) and \( \theta_1 = \frac{1}{2\sigma^2} \). Hence the joint distribution is a natural exponential family of full rank. This implies \( T = (T_1, T_2) \) is complete and sufficient for \( \theta \), since there is a one-to-one correspondence between \( (\mu, \sigma^2) \) and \( \theta \). Thus \( (\bar{X}, S^2) \) is complete and sufficient for \( (\mu, \sigma^2) \)

**Theorem 0.0.2.** If \( G \) is boundedly complete sufficient for \( F \), and \( T \) is ancillary, then \( G \) and \( T \) are independently distributed (conditionally) on every \( \theta \in \Theta \)

**Proof.** We need to show that

\[
P[T \leq t, G \leq g | \theta] = P[T \leq t | \theta] P[G \leq g | \theta]
\]

Since \( T \) is ancillary and \( G \) is sufficient, we write \( v \) as,

\[
v = P[T \leq t] = P[T \leq t | \theta] \quad \forall \quad \theta \in \Theta
\]

and

\[
h(g) = P[T \leq t | G = g] = E[I_{[T \leq t]} | G = g]
\]

where \( I \) is indicator function define on the range of \( T \). By the iterative formula of conditional expectation,

\[
E[h(G) | \theta] = E[E[I_{[T \leq t]} | G] | \theta]
\]

\[
= P[T \leq t | \theta]
\]

\[
= v \quad \forall \quad \theta \in \Theta.
\]
and by the bounded completeness of $G$

$$P[h(G) = v|\theta] = 1 \text{ } \forall \theta \in \Theta.$$ 

$$P[T \leq t, G \leq g|\theta] = E[h(G)I_{G \leq g}|\theta]$$

$$= vP[G \leq g]$$

$$= P[T \leq t|\theta]P[G \leq g|\theta]$$

\[{}^\square\]

By Basu's theorem, no ancillary statistic other than constant can be computed from a statistics which is sufficient and complete.

Below are the Data from O’Gorman (2004)
Weight Loss Data, WL= weight loss, S= site, T= treatment

<table>
<thead>
<tr>
<th></th>
<th>WL</th>
<th>S</th>
<th>T</th>
<th></th>
<th>WL</th>
<th>S</th>
<th>T</th>
<th></th>
<th>WL</th>
<th>S</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.40</td>
<td>1</td>
<td>0</td>
<td></td>
<td>4.40</td>
<td>2</td>
<td>0</td>
<td></td>
<td>2.10</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2.70</td>
<td>1</td>
<td>0</td>
<td></td>
<td>4.30</td>
<td>2</td>
<td>0</td>
<td></td>
<td>1.20</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>-1.50</td>
<td>1</td>
<td>0</td>
<td></td>
<td>4.90</td>
<td>2</td>
<td>0</td>
<td></td>
<td>2.90</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3.70</td>
<td>1</td>
<td>0</td>
<td></td>
<td>4.00</td>
<td>2</td>
<td>0</td>
<td></td>
<td>4.60</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2.70</td>
<td>1</td>
<td>0</td>
<td></td>
<td>13.30</td>
<td>2</td>
<td>0</td>
<td></td>
<td>1.60</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2.30</td>
<td>1</td>
<td>0</td>
<td></td>
<td>2.40</td>
<td>2</td>
<td>0</td>
<td></td>
<td>10.50</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>3.40</td>
<td>1</td>
<td>0</td>
<td></td>
<td>1.80</td>
<td>2</td>
<td>0</td>
<td></td>
<td>1.00</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>3.60</td>
<td>1</td>
<td>0</td>
<td></td>
<td>4.00</td>
<td>2</td>
<td>0</td>
<td></td>
<td>2.60</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>4.90</td>
<td>1</td>
<td>0</td>
<td></td>
<td>5.70</td>
<td>2</td>
<td>1</td>
<td></td>
<td>-6.20</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>4.70</td>
<td>1</td>
<td>0</td>
<td></td>
<td>5.70</td>
<td>2</td>
<td>1</td>
<td></td>
<td>2.70</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>2.20</td>
<td>1</td>
<td>1</td>
<td></td>
<td>-0.70</td>
<td>2</td>
<td>1</td>
<td></td>
<td>4.10</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>4.70</td>
<td>1</td>
<td>1</td>
<td></td>
<td>4.20</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3.80</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>2.90</td>
<td>1</td>
<td>1</td>
<td></td>
<td>4.10</td>
<td>2</td>
<td>1</td>
<td></td>
<td>5.50</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>2.80</td>
<td>1</td>
<td>1</td>
<td></td>
<td>5.10</td>
<td>2</td>
<td>1</td>
<td></td>
<td>5.60</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>5.30</td>
<td>1</td>
<td>1</td>
<td></td>
<td>2.70</td>
<td>2</td>
<td>1</td>
<td></td>
<td>4.40</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>4.10</td>
<td>1</td>
<td>1</td>
<td></td>
<td>4.90</td>
<td>2</td>
<td>1</td>
<td></td>
<td>2.60</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>4.50</td>
<td>1</td>
<td>1</td>
<td></td>
<td>4.70</td>
<td>2</td>
<td>1</td>
<td></td>
<td>5.60</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>2.50</td>
<td>1</td>
<td>1</td>
<td></td>
<td>5.50</td>
<td>2</td>
<td>1</td>
<td></td>
<td>5.00</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>2.00</td>
<td>1</td>
<td>1</td>
<td></td>
<td>3.50</td>
<td>2</td>
<td>1</td>
<td></td>
<td>4.70</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>3.90</td>
<td>2</td>
<td>0</td>
<td></td>
<td>4.50</td>
<td>2</td>
<td>1</td>
<td></td>
<td>2.90</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>
### Weight Loss Data, WL= weight loss, S= site, T= treatment

<table>
<thead>
<tr>
<th></th>
<th>WL</th>
<th>S</th>
<th>T</th>
<th></th>
<th>WL</th>
<th>S</th>
<th>T</th>
<th></th>
<th>WL</th>
<th>S</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>1.30</td>
<td>4</td>
<td>0</td>
<td></td>
<td>3.10</td>
<td>5</td>
<td>0</td>
<td></td>
<td>4.40</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>62</td>
<td>4.90</td>
<td>4</td>
<td>0</td>
<td></td>
<td>4.60</td>
<td>5</td>
<td>0</td>
<td></td>
<td>2.20</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>63</td>
<td>2.50</td>
<td>4</td>
<td>0</td>
<td></td>
<td>4.50</td>
<td>5</td>
<td>0</td>
<td></td>
<td>1.10</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>64</td>
<td>2.40</td>
<td>4</td>
<td>0</td>
<td></td>
<td>4.90</td>
<td>5</td>
<td>0</td>
<td></td>
<td>4.00</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>65</td>
<td>3.70</td>
<td>4</td>
<td>0</td>
<td></td>
<td>4.50</td>
<td>5</td>
<td>0</td>
<td></td>
<td>-1.30</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>66</td>
<td>1.20</td>
<td>4</td>
<td>0</td>
<td></td>
<td>1.80</td>
<td>5</td>
<td>0</td>
<td></td>
<td>3.70</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>67</td>
<td>2.80</td>
<td>4</td>
<td>0</td>
<td></td>
<td>1.30</td>
<td>5</td>
<td>0</td>
<td></td>
<td>4.80</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>68</td>
<td>1.20</td>
<td>4</td>
<td>0</td>
<td></td>
<td>3.60</td>
<td>5</td>
<td>0</td>
<td></td>
<td>4.20</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>69</td>
<td>2.60</td>
<td>4</td>
<td>0</td>
<td></td>
<td>3.80</td>
<td>5</td>
<td>1</td>
<td></td>
<td>4.70</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>70</td>
<td>1.90</td>
<td>4</td>
<td>1</td>
<td></td>
<td>2.60</td>
<td>5</td>
<td>1</td>
<td></td>
<td>2.00</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>71</td>
<td>5.00</td>
<td>4</td>
<td>1</td>
<td></td>
<td>4.20</td>
<td>5</td>
<td>1</td>
<td></td>
<td>4.20</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>72</td>
<td>2.40</td>
<td>4</td>
<td>1</td>
<td></td>
<td>5.30</td>
<td>5</td>
<td>1</td>
<td></td>
<td>-0.60</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>73</td>
<td>1.90</td>
<td>4</td>
<td>1</td>
<td></td>
<td>5.00</td>
<td>5</td>
<td>1</td>
<td></td>
<td>5.00</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>74</td>
<td>-5.60</td>
<td>4</td>
<td>1</td>
<td></td>
<td>14.70</td>
<td>5</td>
<td>1</td>
<td></td>
<td>4.60</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>75</td>
<td>2.00</td>
<td>4</td>
<td>1</td>
<td></td>
<td>5.20</td>
<td>5</td>
<td>1</td>
<td></td>
<td>5.00</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>76</td>
<td>4.70</td>
<td>4</td>
<td>1</td>
<td></td>
<td>5.70</td>
<td>5</td>
<td>1</td>
<td></td>
<td>4.30</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>77</td>
<td>4.20</td>
<td>4</td>
<td>1</td>
<td></td>
<td>2.70</td>
<td>5</td>
<td>1</td>
<td></td>
<td>2.90</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>78</td>
<td>2.90</td>
<td>4</td>
<td>1</td>
<td></td>
<td>5.60</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>4.60</td>
<td>4</td>
<td>1</td>
<td></td>
<td>1.80</td>
<td>6</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>4.00</td>
<td>5</td>
<td>0</td>
<td></td>
<td>2.30</td>
<td>6</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>