Rank Regression in Order Restricted Randomized Designs

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of the Ohio State University

By
Jinguo Gao, BS/MS
Graduate Program in Statistics

The Ohio State University
2013

Dissertation Committee:
Dr. Omer Ozturk, Advisor
Dr. Steve MacEachern
Dr. Elizabeth Stasny
ABSTRACT

One of the main principles in designing an experiment is to use blocking factors whenever it is possible. In many studies, blocking information is not precisely defined or may be subjective in nature. Hence, it is usually discarded in the construction of the design and in the analysis of a data set. This research uses a special design, an order restricted randomized design (ORRD), which uses available subjective information in a small set of experimental units, to create a judgment ranked blocking factor. Under this design, we first select a small set of experimental units and rank the units within each set pre-experimentally from smallest to largest based on inherent variation. The ranking process induces a positive correlation structure for the within-set residuals. Under the design, we then use a randomization scheme to assign the treatment levels to the ranked experimental units. Such an assignment with certain restrictions on the randomization scheme, which keeps the design to be balanced, translates the within-set positive dependence structure into a variance reduction technique in the estimation of a contrast parameter.

Chapter 1 provides a literature review on designs that are closely related to the one we proposed. It then introduces two types of designs, Design 1 and Design 2, for ORRDs. Design 1 is constructed for all pairwise contrast parameters, while Design 2 is constructed for all possible contrast parameters. This chapter also provides a discussion on similarities and differences between ORRDs and generalized randomized block designs.
Chapter 2 introduces an additive model to analyze ORRD data. The parameter of this model is estimated with a rank regression estimator. It is shown, under some regularity conditions, that the asymptotic distribution of the rank regression estimator converges to a $p$-dimensional multivariate normal distribution.

Chapter 3 develops statistical inference to test generalized linear hypotheses. We consider three tests: drop, score and Wald tests. Under some regularity conditions, we show that the test statistics of these three tests converge to a chi-squared distribution with appropriate degrees of freedom. For moderate sample sizes, the distributions of these statistics under the null hypothesis can be approximated with an F-distribution with appropriate degrees of freedom.

Chapters 4 and 5 address some computational issues and provide empirical evidence for the performance of the tests. The empirical evidence indicates that the Type I error rates of the tests are reasonably close to the nominal size 0.05 under a wide range of simulation conditions. The empirical power study also indicates that the tests constructed based on ORRDs outperform the tests constructed based on completely randomized designs.

Chapter 6 illustrates the use of the proposed estimators and tests by applying them to a clinical trial data set. Chapter 7 provides concluding remarks and discusses some open problems for future studies.
To my parents, Xin and Jinlan, for their love, support and patience.
ACKNOWLEDGMENTS

I am deeply indebted to my adviser, Dr. Omer Ozturk, for his inspirational guidance, general support and endless patience. Without his help, I could not have finished this work. I also want to thank Dr. Elizabeth Stasny and Dr. Steve MacEachern for serving on my dissertation committee and for their valuable comments and suggestions. My gratitude also goes to Dr. Prem K. Goel for his guidance and help during my two years, working as research assistant with him. I am grateful to all my other professors, fellow students, and staff members for their support and encouragement.
VITA

2003 .............................  B.S. in Biology, Nanjing University

2006 .............................  M.S. in Ecology, Suny Albany

2008 .............................  M.S. in Statistics, The Ohio State University

2006-Present  .....................  Graduate Teaching and Research Associate, The Ohio State University.

PUBLICATIONS


FIELDS OF STUDY

Major Field: Statistics
# TABLE OF CONTENTS

Abstract ................................................................. ii
Dedication ................................................................. iii
Acknowledgments ......................................................... v
Vita ................................................................. vi
List of Tables ............................................................. ix
List of Figures ............................................................ xii

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction .................................................. 1</td>
</tr>
<tr>
<td></td>
<td>1.1 Ranked Set Sampling ....................................... 2</td>
</tr>
<tr>
<td></td>
<td>1.2 Order Restricted Randomized Design (ORRD) ................. 11</td>
</tr>
<tr>
<td></td>
<td>1.3 Generalized Randomized Block Design (GRBD) .................. 24</td>
</tr>
<tr>
<td>2</td>
<td>Linear Models .................................................... 29</td>
</tr>
<tr>
<td></td>
<td>2.1 Model .......................................................... 29</td>
</tr>
<tr>
<td></td>
<td>2.2 Estimation of regression parameter .......................... 31</td>
</tr>
<tr>
<td></td>
<td>2.3 Assumptions ................................................... 34</td>
</tr>
<tr>
<td></td>
<td>2.4 Asymptotic Normality of the Score Function ................. 36</td>
</tr>
<tr>
<td></td>
<td>2.5 Linearity of score function .................................. 54</td>
</tr>
<tr>
<td></td>
<td>2.6 Efficiency Comparison of ORRD to CRD ....................... 66</td>
</tr>
<tr>
<td>3</td>
<td>Test procedures .................................................. 69</td>
</tr>
<tr>
<td></td>
<td>3.1 Drop Test ..................................................... 70</td>
</tr>
<tr>
<td></td>
<td>3.2 Score Test .................................................... 75</td>
</tr>
<tr>
<td></td>
<td>3.3 Wald Test ..................................................... 80</td>
</tr>
<tr>
<td>4</td>
<td>Computational Procedures ......................................... 83</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>5</td>
<td>Simulation Results</td>
</tr>
<tr>
<td></td>
<td>5.1 Simulation Settings</td>
</tr>
<tr>
<td></td>
<td>5.2 Type I Error Rates</td>
</tr>
<tr>
<td></td>
<td>5.3 Empirical Power</td>
</tr>
<tr>
<td>6</td>
<td>Data Study: AIDS Clinical Trials Group Protocol 116B/117 Study</td>
</tr>
<tr>
<td></td>
<td>6.1 Background</td>
</tr>
<tr>
<td></td>
<td>6.2 Illustrative Example</td>
</tr>
<tr>
<td>7</td>
<td>Conclusions and Future Research</td>
</tr>
<tr>
<td></td>
<td>7.1 Conclusions</td>
</tr>
<tr>
<td></td>
<td>7.2 Future Research</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A balanced RSS design with set size $H = 4$ and cycle size $n = 2$.</td>
</tr>
<tr>
<td>1.2</td>
<td>Completely randomized design with two treatment regimes, T and C.</td>
</tr>
<tr>
<td>1.3</td>
<td>Order restricted randomized design with two treatment regimes (T and C) and set size $H = 2$.</td>
</tr>
<tr>
<td>1.4</td>
<td>Brief comparisons of CRD and ORRD designs.</td>
</tr>
<tr>
<td>1.5</td>
<td>Design 1 for all possible pairwise comparisons of treatments $T_1$, $T_2$, and $T_3$.</td>
</tr>
<tr>
<td>1.6</td>
<td>Design 2 for all possible contrasts: A Latin square structure with three treatments is used for randomization.</td>
</tr>
<tr>
<td>1.7</td>
<td>Order restricted randomized designs.</td>
</tr>
<tr>
<td>1.8</td>
<td>The construction of GRBD from Design 1 of ORRD.</td>
</tr>
<tr>
<td>1.9</td>
<td>A list of possible treatment allocations to the EUs in a set for the block design based on Design 1.</td>
</tr>
<tr>
<td>1.10</td>
<td>The construction of GRBD from Design 2 of ORRD.</td>
</tr>
<tr>
<td>1.11</td>
<td>All possible treatment allocations to within-set units in a block design based on Design 2.</td>
</tr>
<tr>
<td>2.1</td>
<td>Asymptotic relative efficiencies (ARE) of rank regression and least square estimators in ORRD and CRD designs, $L$ is the number of treatment levels.</td>
</tr>
<tr>
<td>5.1</td>
<td>Empirical Type I error rates for the score test, Design 1, $H=2$, $L=2$, $K=2$, and simulation size 5000.</td>
</tr>
</tbody>
</table>
5.2 Empirical Type I error rates for the score test, Design 1, H=2, L=3, K=6 and simulation size 5000. .................................................. 98
5.3 Empirical Type I error rates for the score test, Design 1, H=2, L=4, K=12 and simulation size 5000. .................................................. 99
5.4 Empirical Type I error rates for the score test, Design 2, H=3, L=3, K=3 and simulation size 5000. .................................................. 100
5.5 Empirical Type I error rates for the score test, Design 2, when H=4, L=4, K=4 and simulation size 5000. .................................................. 101
5.6 Empirical Type I error rates for the score test, Design 2, H=5, L=5, K=5 and simulation size 5000. .................................................. 102
5.7 Empirical Type I error rates for the Wald test, Design 1, H=2, L=2, K=2 and simulation size 5000. .................................................. 103
5.8 Empirical Type I error rates for the Wald test, Design 1, H=2, L=3, K=6 and simulation size 5000. .................................................. 104
5.9 Empirical Type I error rates for the Wald test, Design 1, H=2, L=4, K=12 and simulation size 5000. .................................................. 105
5.10 Empirical Type I error rates for the Wald test, Design 2, H=3, L=3, K=3, and simulation size 5000. .................................................. 106
5.11 Empirical Type I error rates for the Wald test, Design 2, H=4, L=4, K=4, and simulation size 5000. .................................................. 107
5.12 Empirical Type I error rates for the Wald test, Design 2, H=5, L=5, K=5, and simulation size 5000. .................................................. 108
5.13 Empirical Type I error rates for the drop test, Design 1, H=2, L=2, K=2 and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration. .................. 109
5.14 Empirical Type I error rates for the drop test, Design 1, H=2, L=3, K=6 and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration. .................. 110
5.15 Empirical Type I error rates for the drop test, Design 1, H=2, L=4, K=12 and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration. .................. 111
5.16 Empirical Type I error rates for the drop test, Design 2, H=3, L=3, K=3, and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration. .......................... 112

5.17 Empirical Type I error rates for the drop test, Design 2, H=4, L=4, K=4, and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration. .......................... 113

5.18 Empirical Type I error rates for the drop test, Design 2, H=5, L=5, K=5, and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration. .......................... 114

6.1 ORRD data set for Design 1 with n = 3, H = 2, and K = 6. TRT = treatment, RNK = rank, CDBEG = initial CD4 cell counts, CDEND = 24th week CD4 cell counts, and ID = patient identification number. .......................... 132

6.2 ORRD data set for the Design 2 with n = 4, H = 3, and K = 3. TRT = treatment, RNK = rank, CDBEG = initial CD4 cell counts, CDEND = 24th week CD4 cell counts, and ID = patient identification number. .......................... 133

6.3 Observed statistics and p-values of the score, Wald and drop tests for the ORRD data sets in Tables 6.1 and 6.2. The p-values of Wald and score tests are computed based on an F-approximation. The p-values of the drop tests are computed based on a $\chi^2$ approximation. In the right half, the structure in ORRD is ignored and the data sets are treated as data sets from CRD. .......................... 133

6.4 CRD data set with 3 treatments, each of which has sample size 12. TRT = treatment, CDBEG = initial CD4 cell counts, CDEND = 24th week CD4 cell counts, and ID = patient identification number. .......................... 134

6.5 Observed statistics and p-values of the score, Wald and drop tests for the CRD data set in Table 6.4. The p-values of the Wald and score tests are computed based on an F-approximation. The p-values of the drop test are computed based on a $\chi^2$ approximation. .......................... 134


**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Empirical power curves of the score, Wald and drop tests. The underlying distribution is normal, and total sample size is N=36.</td>
<td>118</td>
</tr>
<tr>
<td>5.2 Empirical power curves of the score, Wald and drop tests. The underlying distribution is normal, and total sample size is N=72.</td>
<td>119</td>
</tr>
<tr>
<td>5.3 Empirical power curves of the score, Wald and drop tests. The underlying distribution is normal, and total sample size is N=108.</td>
<td>120</td>
</tr>
<tr>
<td>5.4 Empirical power curves of the score, Wald and drop tests. The underlying distribution is log-normal, and total sample size is N=36.</td>
<td>121</td>
</tr>
<tr>
<td>5.5 Empirical power curves of the score, Wald and drop tests. The underlying distribution is log-normal, and total sample size is N=72.</td>
<td>122</td>
</tr>
<tr>
<td>5.6 Empirical power curves of the score, Wald and drop tests. The underlying distribution is log-normal, and total sample size is N=108.</td>
<td>123</td>
</tr>
<tr>
<td>5.7 Empirical power curves of the score, Wald and drop tests. The underlying distribution is student’s t with 3 degrees of freedom, and total sample size is N=36.</td>
<td>124</td>
</tr>
<tr>
<td>5.8 Empirical power curves of the score, Wald and drop tests. The underlying distribution is student’s t with 3 degrees of freedom, and total sample size is N=72.</td>
<td>125</td>
</tr>
<tr>
<td>5.9 Empirical power curves of the score, Wald and drop tests. The underlying distribution is student’s t with 3 degrees of freedom, and total sample size is N=108.</td>
<td>126</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Success of a statistical inference depends heavily on the selection of a representative sample to allow inference from the sampling results back to the population. Simple random sampling (SRS) is one of the sampling procedures used for this purpose. SRS has several desirable properties: it does not have a selection bias, it is easy to use, and it produces an unbiased estimator for the population mean. However, on rare occasions SRS can produce unrepresentative samples. In many experimental settings, where making an informal measurement on a unit is far cheaper than making a formal measurement, SRS may also become prohibitively expensive, since every unit needs to be measured in the SRS design. One solution to address these deficiencies in the SRS design is to construct an informative sampling design using all available external information. The external information can be categorized into two different forms. The first form contains information that can be precisely measured or easily categorized. This type of information usually leads to a well-defined covariate variable that can be used in the construction of the design and in the analysis of the data. Many traditional sampling designs, such as stratified and cluster sampling, fall into this modeling category. The second form of external information cannot be precisely measured, could be subjective, imprecise, or even biased. However, if it is used properly, it could be very helpful in constructing an informative sample design
with high efficiency. In this case, the external information of a large group of subjects is used to select a representative smaller group of subjects for full measurement. One of these informative sampling designs, ranked set sampling, provides a collection of techniques with detailed plans to construct such a representative sample.

1.1 Ranked Set Sampling

The ranked set sampling (RSS) procedure was introduced by McIntyre (1952, 2005). It is designed for setting, where the actual measurement of a subject is either expensive or time-consuming, but ranking a few subjects in a set based on subjective information is relatively easy and inexpensive. In RSS, researchers rank subjects in small sets without actual measurement. In RSS, ranking is performed based on either response variable or some auxiliary variable. Then these ranks are used to construct homogenous groups of subjects, and to determine which subjects to select for full measurements. These homogenous sampling groups can be considered as strata. The efficiency improvement of statistical inference, then, can be anticipated from the general theory of the stratified sampling design.

We provide an example to illustrate the construction of the RSS design. Suppose that the variable of interest is the mercury level in a fish population. The exact measurement of mercury level in fish is an expensive and time consuming process requiring chemical and biological analyses. It is reasonable to assume that actual mercury level and the size of a fish are positively correlated, with bigger fish accumulating higher level of mercury. Researchers can then use the size of fish as external information in a small set, and rank the fish based on their size to construct homogenous groups of fish. For a ranked set sample of size \( N \), we first determine a set size \( H \), and select \( NH \) fish at random from the population. These fish are divided randomly into \( N \) sets, each containing \( H \) fish. Each set is ranked from the lowest to
Table 1.1: A balanced RSS design with set size $H = 4$ and cycle size $n = 2$. 

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Set</th>
<th>Rank 1</th>
<th>Rank 2</th>
<th>Rank 3</th>
<th>Rank 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Set 4</td>
<td>$X_{[1]41}$</td>
<td>$X_{[2]41}$</td>
<td>$X_{[3]41}$</td>
<td>$X_{[4]41}$</td>
</tr>
</tbody>
</table>

highest anticipated mercury levels based on the relative sizes of the fish without exact measurement. The $i$-th ($i = 1, \cdots, H$) ranked fish is fully measured for its mercury level in $n_i$ sets, where $\sum_{i=1}^{H} n_i = N$. Note that in each set, only one fish has been fully measured. The others have been used for ranking purposes only. The collection of the fully measured observations, $X_{[ij]}$, $j = 1, \cdots, n_i$, $i = 1, \cdots, H$, is called an unbalanced ranked set sample. If each judgment rank is measured an equal number of times ($n_i \equiv n, i = 1, \cdots, H$), then the sample is called balanced.

An example for a balanced ranked set sample, with $H=4$, and $n=2$, is presented in Table 1.1. In this table, each row represents a set. These sets are ranked from the smallest subject to the largest subject. Only one of the subjects in each set is fully measured. The bold faced entries represent these fully measured subjects. The square brackets are used to denote that the ranking may be in error. If there is no ranking error, the square brackets are replaced with round ones. Then the fully measured observations, $X_{(i)j}$, $j = 1, \cdots, n_i, i = 1, \cdots, H$, become the usual order statistics from a set of size $H$.

To highlight the main difference between an SRS and an RSS, we look at the information content of a single measured subject. Let $X_i$ be the measurement on a subject from an SRS and $X_{[i]}$ be the measurement on a subject from an RSS.
The $X_i$ provides information about the subject on which it was measured, however $X_{[i]}$, in addition to the information that $X_i$ provides, also provides some additional information about the other $(H - 1)$ subjects in the set through the ranking process. Thus, a single measured subject from an RSS design is more informative than a single measured subject from an SRS.

Construction of an RSS relies on a ranking process based on some subjective information. Since the ranking is performed based on rough external information, it could lead to misranking among subjects in a set. The error in the ranking process causes the efficiency to decrease. It has been shown, however, that the efficiency of many statistical procedures under RSS with imperfect ranking is at least as good as the efficiency of the same procedures under SRS (Bohn, 1996; Dell and Clutter, 1972; McIntyre, 1952).

RSS has been widely applied to many nonparametric inferential procedures. Bohn and Wolfe (1992, 1994) developed a rank-sum test for a two-sample problem, and constructed a ranking model to assess the effect of ranking error on the rank-sum test. For brevity of the presentation, we consider the same set size $H$ for the two samples with different cycle sizes, $n$ for the $X$-sample and $m$ for the $Y$-sample. Let $X_{[i]} j, i = 1, \ldots , H, j = 1, \ldots , n$ and $Y_{[t]} s, t = 1, \ldots , H, s = 1, \ldots , n$, be ranked set samples from the $X$- and $Y$-populations, respectively. Assume that the $X$-population has a distribution with a cumulative distribution function (CDF) $F$ and the $Y$-population has a distribution with a CDF $G$. The test statistic for the Bohn-Wolfe rank-sum
test (BW) can be written as

\[ T_{BW} = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{s=1}^{H} \sum_{t=1}^{m} \psi(Y_{(i)j} - X_{(s)t}) \]

\[ = \sum_{i=1}^{H} \sum_{s=1}^{H} \sum_{j=1}^{n} \sum_{t=1}^{m} \psi(Y_{(i)j} - X_{(s)t}) \]

\[ = \sum_{i=1}^{H} \sum_{s=1}^{H} T_{is}, \]

where \( T_{is} = \sum_{j=1}^{n} \sum_{t=1}^{m} \psi(Y_{(i)j} - X_{(s)t}) \), and \( \psi(x) = 1 \) if \( x > 0 \); \( \psi(x) = 0 \), otherwise. The asymptotic distribution of the BW test statistic is normal with mean 0 and variance \( \sigma_F^2 \), where \( \sigma_F^2 \) has a complicated expression given in Bohn and Wolfe (1992). Ozturk (2010) simplifies this expression and provides a compact notation

\[ \sigma_F^2 = \frac{(n + m)^2}{nm} \left[ \frac{1}{3} - \frac{1}{H} \sum_{i=1}^{H} \left( \int F_{[i]}(y) dF(y) \right)^2 \right], \]

where \( F_{[i]} \) stands for the CDF of the \( i \)th ordered statistics. This expression under perfect ranking reduces to

\[ \sigma^2 = \frac{(n + m)^2}{nm} \left[ \frac{1}{3} - \frac{2H + 1}{6(H + 1)} \right]. \]

It is important to notice here that the computation of \( \sigma^2 \) in the BW test relies on the assumption of perfect ranking. When the ranking process is not perfect, the Type I error rate of the BW test rises above the nominal level. To solve this problem, Fligner and MacEachern (2006) introduced a distribution-free rank-sum test (FM) for independent ranked set samples collected from two populations. The proposed
test uses the statistic,

\[ T_{FM} = \sum_{i=1}^{H} T_{ii}, \text{ where } T_{ii} = \sum_{j=1}^{n} \sum_{t=1}^{m} \psi(Y_{(i)j} - X_{(i)t}). \]

This FM test is distribution-free, regardless of the quality of ranking information as long as the ranking mechanism is consistent, which means that the same ranking scheme is used for all sets.

Note that with a consistent ranking mechanism, under the null hypothesis that the X- and Y-populations have the same underlying distribution, we have \( F_{[h]}(x) = G_{[h]}(x) \), for \( h = 1, \ldots, H \), where \( F_{[h]}(x) \) and \( G_{[h]}(x) \) are the \( h \)-th judgment class CDFs of the X- and Y-populations. It is then easy to observe that the null distribution of each \( T_{ii} \) is the same as the distribution of the classical two-sample Mann-Whitney-Wilcoxon (MW) test statistic in SRS, denoted as \( T_{MW} \). Regardless of the judgment ranking mechanism, this null distribution for each \( T_{ii} \) remains the same. Since the \( T_{ii} \) are mutually independent and \( T_{FM} \) is a sum of the individual \( T_{ii} \), the null distribution of \( T_{FM} \) is the convolution of the distributions of \( H \) independent \( T_{MW} \) statistics.

A significant advantage of the proposed statistic \( T_{FM} \) over \( T_{BW} \) is that its null distribution does not depend on the underlying distribution or the quality of ranking information.

Fligner and MacEachern (2006) performed a simulation study to compare the efficiencies of three different rank-sum tests, the MW, BW and FM tests. The simulation study shows that the tests based on RSS designs are preferable to the tests based on an SRS design when the quality of ranking information is moderate to high. The advantage of an RSS design diminishes as the quality of ranking information decreases.

Under perfect ranking, the BW and FM tests are nearly equal. The efficiency of
$T_{FM}$ is sometimes a little better than the efficiency of $T_{BW}$, sometimes nearly equal and sometimes a little worse, depending on the underlying distribution. When the ranking process is imperfect, the actual level of the Type I error rate of the BW test rises above the nominal level. The FM test maintains its level and provides higher power compared to the procedure based on an SRS design. One of the important features of the FM test is that it does not rely on the knowledge of the form of the imperfect ranking model and the underlying distribution of response variable.

Bohn (1998) later introduced a signed rank test for RSS. Ozturk and Wolfe (2000, 2001) constructed an unbalanced RSS design to improve the efficiency of the rank-sum test. In addition to these tests, distribution-free confidence intervals have also drawn attention in the literature. Deshpande, Frey and Ozturk (2006) constructed quantile confidence intervals based on an RSS design in a finite population setting. Chen (2000) and Ozturk and Deshpande (2006) show that the nonparametric confidence intervals based on an RSS design provide a great improvement over the nonparametric confidence intervals based on an SRS design.

Ozturk (2002) used rank regression methodology in the RSS design to estimate regression parameters and to test generalized linear hypothesis. The rank regression estimators have desirable properties, such as, high efficiency at the true model and robustness at a neighborhood model. For example when the underlying distribution is normal, rank regression estimator achieves 95.5% of the asymptotic efficiency of the least squares estimator (McKean, 2004). Unlike least squares estimators, rank regression estimators maintain high efficiency and produce much smaller bias when the model has heavy-tailed error distributions or large outliers. The rank regression model in ranked set sampling can be constructed in different ways. Ozturk (2002) considered an unbalanced ranked set sample of response vector $Y_t$ at each corresponding
covariance vector $\mathbf{x}_t$, $t = 1, \cdots, n$,

$$
Y_t^T = (y_{[1]1}, \cdots, y_{[1]r_1}, \cdots, y_{[H]r_H}),
$$

$$
\epsilon_t^T = (\epsilon_{[1]1}, \cdots, \epsilon_{[1]r_1}, \cdots, \epsilon_{[H]r_H}),
$$

where $\sum_{h=1}^{H} r_h = M$ is the size of RSS at each $\mathbf{x}_t$, $r_h$ is the number of measured observation in the $h$-th judgment class, and $\epsilon_t$ is the residual vector that corresponds to response vector $Y_t$. The regression model for the data structure $(Y_t, \mathbf{x}_t); t = 1, \cdots, n$, can be written as

$$
Y_t = \alpha \mathbf{1}_M + \mathbf{x}_t^T \mathbf{\beta}_M + \epsilon_t, \quad t = 1, \cdots, n,
$$

(1.1.1)

where $\mathbf{1}_M$ is $M \times 1$ dimensional vector whose elements are all one, $\alpha$ is an intercept parameter and $\mathbf{\beta}$ is a $p$-dimensional regression parameter.

The rank regression estimator is defined as the minimizer of the rank dispersion function $D(\mathbf{\beta})$,

$$
D(\mathbf{\beta}) = \frac{1}{N} \sum_{t=1}^{n} \sum_{h=1}^{H} \sum_{k=1}^{r_h} e_{[h]kt} \left\{ R(e_{[h]kt}) - n \sum_{u=1}^{m} \tau_u \tau_{[uh]} - 0.5 \right\}
$$

$$
= \frac{1}{N} \sum_{t=1}^{n} \sum_{h=1}^{H} \sum_{k=1}^{r_h} e_{[h]kt} \left\{ R(e_{[h]kt}) - ER(e_{[h]kt}) \right\},
$$

(1.1.2)

where $\tau_{[uh]} = \int F_{[u]}(y) dF_{[h]}(y)$, $e_{[h]kt} = Y_{[h]kt} - \mathbf{x}_t^T \mathbf{\beta}$, $R(e_{[h]kt})$ is the rank of $e_{[h]kt}$ among all $e_{[h']k't'}$, $h' = 1, \cdots, H$; $k' = 1, \cdots, r_{h'}$; $t' = 1, \cdots, n$, and $N$ is the total sample size.

Let $\hat{\mathbf{\beta}}_{RSS}$ and $\hat{\mathbf{\beta}}_{SRS}$ be the minimizers of the expression (1.1.2) based on the RSS and SRS designs, respectively. Ozturk (2002) showed that the estimator $\hat{\mathbf{\beta}}_{RSS}$ is asymptotically normal and has higher Pitman asymptotic efficiency than the SRS rank regression estimator $\hat{\mathbf{\beta}}_{SRS}$. The asymptotic relative efficiency of $\hat{\mathbf{\beta}}_{RSS}$ with
respect to $\hat{\beta}_{SRS}$ is given by

$$\text{eff}(RSS, SRS) = \frac{[\det\{\sqrt{N}\text{Var}([\hat{\beta}_{SRS}])\}]^{1/p}}{[\det\{\sqrt{N}\text{Var}(\beta_{RSS})\}]^{1/p}} = \frac{\omega^2}{(2\gamma)^2},$$

where $\det(A)$ is the determinant of the matrix $A$, $\rho_i = r_i/M$ for $i = 1, \cdots, m$,

$$\omega = \sum_{u=1}^{H} \sum_{v=1}^{H} \rho_u \rho_v \int f_{[u]}(y) f_{[v]}(y) dy$$

and

$$\gamma = \sum_{j=1}^{m} \sum_{u=1}^{m} \sum_{\nu=1}^{m} \rho_j \rho_u \rho_{\nu} \int [F_{[u]}(y) F_{[\nu]}(y) - \tau_{[u]} \tau_{[\nu]}] dF_{[j]}(y).$$

In general, the asymptotic relative efficiency of an unbalanced RSS estimator with respect to an SRS estimator depends on the underlying distribution of error terms, the allocation procedure, and the judgment ranks. In a balanced RSS, $\rho_h$ equals $1/H$ for $h = 1, \cdots, H$. Under perfect ranking, $\omega$ reduces to $\int f^2(y) dy$. Hence, under a balanced RSS (BRSS) design with perfect ranking, the asymptotic relative efficiency reduces to

$$\text{eff}(BRSS, SRS) = \frac{H + 1}{2}.$$

Ozturk (2002) also introduced testing procedures to test general linear hypotheses under the model (1.1.1). Let $\beta^T = (\beta_1^T, \beta_2^T)$, where $\beta_1^T$ and $\beta_2^T$ are $(p-q)$-dimensional and $q$-dimensional vectors, respectively. With this parameter structure, the linear model (1.1.1) can be written as

$$Y_t = \alpha 1_M + (x_{1t}^T \beta_1 + x_{2t}^T \beta_2) 1_M + \epsilon_t, \quad t = 1, \cdots, n,$$

where $x_{1t}$ and $x_{2t}$ are $(p-q)$- and $q$-dimensional vectors respectively with $x_t^T =$
In this model, the objective is to construct a test for the hypothesis

\[ H_0 : \beta_2 = 0 \text{ versus } H_A : \beta_2 \neq 0, \]

(1.1.3)

where \( \beta_1 \) is unspecified and considered to be a nuisance parameter. Three types of tests, the drop, score and Wald tests, are constructed for the hypotheses (1.1.3). All three test statistics, under the null hypothesis, have asymptotic chi-squared distributions with \( q \)-degrees of freedom as sample size increases. Empirical evidence indicates that the proposed tests provide higher power than corresponding tests based on an SRS design.

Under assumption of perfect ranking, Ozturk (2002) constructed an optimal allocation scheme for an unbalanced RSS design for set sizes less than seven. The optimal allocation scheme chooses middle-ranked units for symmetric unimodal distributions of error terms, and the smallest (largest)-ranked units for right (left) skewed distributions of errors.

In general, RSS is generally better suited for observational studies than for experimental designs. RSS provides improved efficiency over its competitors in the literature. There are, however, some restrictions in its usage in designing an experiment. One major challenge in constructing an experimental design is to recruit enough experimental units (EUs) to participate in the study. RSS require additional \((H-1)\) EUs for each fully-measured unit. This could be a challenge in settings where potential EUs are expensive or limited. Another concern in using the RSS concept in designing an experiment is related to the role of randomization for allocating the EUs to different treatments. Even though in RSS a formal randomization can be performed on the EUs which have the same rank, this may result in some practical difficulties in experiments where EUs come into the study over time.
To address the shortcomings of an RSS design in the construction of an experiment, Ozturk and MacEachern (2004, 2007) introduced the order restricted randomized design (ORRD) that uses all EUs in a set along with a randomization technique to allocate the treatment levels to EUs. In the next section, we introduce ORRD and review the related work in the literature.

1.2 Order Restricted Randomized Design (ORRD)

Order restricted randomized design (ORRD), as in an RSS design, pre-experimentally ranks the within-set EUs using all available subjective information to create judgement blocks or strata. Note that the ranking is performed based on the assessment of EUs pre-experimentally rather than their responses because the experiment has not yet been performed. Unlike an RSS design, however, ORRD uses all within-set EUs to conduct the experiment. The within-set ranking process induces a positive correlation structure among experimental units in the same set. The main objective of ORRD is to capitalize on this positive correlation structure to reduce the variance of a contrast parameter estimator in the experiment. ORRD achieves this objective by performing a carefully planned restricted randomization to assign the treatment levels to the within-set EUs. Restricted randomization makes sure that the treatment groups having opposite signs in the contrast parameter are assigned to positively correlated within-set EUs. This allocation converts the positive correlation among within-set EUs into a negative correlation in the variance of the treatment contrast parameter. Hence, it reduces the variance of the estimator for the contrast parameter.

Before we give a formal description for ORRD, we highlight the difference between a completely randomized design (CRD), and ORRD through a simple example with only 4 subjects. Consider a clinical trial with only two treatment regimes, control
Table 1.2: Completely randomized design with two treatment regimes, T and C.

<table>
<thead>
<tr>
<th>Allocation</th>
<th>C</th>
<th>T</th>
<th>T</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>$X_1$</td>
<td>$Y_1$</td>
<td>$Y_2$</td>
<td>$X_2$</td>
</tr>
</tbody>
</table>

Table 1.3: Order restricted randomized design with two treatment regimes (T and C) and set size $H = 2$.

(C) and treatment (T). Assume that we wish to draw inference about the difference between the control and treatment means, $\Delta = \mu_C - \mu_T$. A CRD would randomize the control (C) and treatment (T) regimes to 4 subjects without considering the form of the contrast. This design is illustrated in Table 1.2, where $X_i$ and $Y_i$ are the response measurements from the $i$-th subject in the control and treatment groups, respectively, where $i = 1, 2$. This design does not consider the structure in the contrast parameter $\Delta$.

The ORRD, on the other hand, focuses on the contrast parameter $\Delta$. It first creates a positive correlation structure between the EUs in a set by ranking them pre-experimentally. The construction of ORRD for this example also requires 4 subjects. These subjects are randomly divided into two sets, each of size 2, and subjects in each set are ranked using all available subjective information. We note that even though the ranking may not be perfect, it still produces a positive correlation among residuals for subjects in each set. To perform the randomization, in one of the sets we randomly assign treatments $C$ and $T$ to the ranked EUs. In the second set, without randomization, we use an opposite allocation to the one performed in the
first set. This design is illustrated in Table 1.3, where control and treatment regimes are assigned to patients having rank 1 and rank 2 in Set 1, and rank 2 and rank 1 in Set 2, respectively. In these sets, the notation, $X_{[i]j}$ and $Y_{[i]j}$, is used to denote the control and treatment responses from the $i$-th ranked subject in the $j$-th set, where $i = 1, 2$ and $j = 1, 2$.

The naive estimators of the mean difference of two treatments for CRD and ORRD can be constructed as follows

$$\hat{\Delta}_{CRD} = \frac{X_1 + X_2}{2} - \frac{Y_1 + Y_2}{2},$$

and

$$\hat{\Delta}_{ORRD} = \frac{X_{[1]1} + X_{[2]2}}{2} - \frac{Y_{[1]2} + Y_{[2]1}}{2}.$$

Both of these estimators are unbiased regardless the quality of ranking information,

$$E(\hat{\Delta}_{CRD}) = E(\hat{\Delta}_{ORRD}) = \Delta.$$

However, the variances of $\hat{\Delta}_{ORRD}$ and $\hat{\Delta}_{CRD}$ are significantly different. We have

$$\text{Var}(\hat{\Delta}_{CRD}) = \sigma^2 = \frac{\sigma_{[1]}^2 + \sigma_{[2]}^2 + 2\sigma_{[1,2]}}{2},$$

Note that $\text{Var}(X_1 + X_2) = \text{Var}(X_{[1]} + X_{[2]})$, and $2\sigma^2 = \sigma_{[1]}^2 + \sigma_{[2]}^2 + 2\sigma_{[1,2]}$. We also have

$$\text{Var}(\hat{\Delta}_{ORRD}) = \frac{\sigma_{[1]}^2 + \sigma_{[2]}^2 - 2\sigma_{[1,2]}}{2},$$

where $\sigma_{[1,2]}$ is the covariance between the response variables of EUs from the ranked subject 1 and the ranked subject 2 in the same set. These variance computations clearly indicate that the estimator $\hat{\Delta}_{ORRD}$ has a significant amount of reduction in
<table>
<thead>
<tr>
<th>Property</th>
<th>CRD</th>
<th>ORRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>Yes</td>
<td>Yes across different sets</td>
</tr>
<tr>
<td>Identically distributed residuals</td>
<td>Yes</td>
<td>Yes for those with the same rank</td>
</tr>
<tr>
<td>Judgment rankings</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Randomization</td>
<td>Yes</td>
<td>Restricted</td>
</tr>
</tbody>
</table>

Table 1.4: Brief comparisons of CRD and ORRD designs.

the variance of the estimator compared to the variance of $\hat{\Delta}_{CRD}$. Table 1.4 provides a brief comparison of CRD and ORRD designs.

Ozturk and MacEachern (2013) introduced two different ORRDs. One of the designs is constructed to draw inference for all possible pairwise contrast parameters, while the other design is constructed for all possible contrast parameters. We provide a detailed description of these two designs.

**Design 1 (All possible pairwise contrast design):** The main objective of this design is to use the within-set positive correlation structure to reduce the variance of the pairwise contrast parameter estimates. Let $\mu_1, \ldots, \mu_L$ be the treatment means for $L$ non-overlapping populations. In general, $\binom{L}{2}$ different pairwise contrast combinations are possible. For each pair of contrasts, $\Delta_{ij} = \mu_i - \mu_j$, $i < j$, the ORRD requires two sets of EUs, each of size 2. Thus, each replicate needs $2\binom{L}{2}$ different sets of EUs. EUs in each set are ranked pre-experimentally based on available external information, as ranked unit 1 ($R_1$) and ranked unit 2 ($R_2$). For a pair of treatment regimes ($T_i, T_j$), ORRD specifies that we select two sets. In one of the sets we randomly assign treatment regimes $T_i$ and $T_j$ to the ranked experimental units. Namely, a particular randomization may assign $T_i$ to the ranked unit 1 ($R_1$) and $T_j$ to the ranked unit 2 ($R_2$). In the second set, without a randomization, we select the
opposite assignment so that the treatment $T_i$ is assigned to the ranked unit 2 ($R_2$) and the treatment $T_j$ is assigned to the ranked unit 1 ($R_1$).

To clarify the construction of this design, we consider an ORRD with three treatments, $T_1$, $T_2$ and $T_3$. In this case, three pairwise comparisons are possible, $T_1T_2$, $T_1T_3$ and $T_2T_3$. For the treatment pair $T_1T_2$, we select two sets of EUs, each of size 2. EUs within sets are ranked as smaller ($R_1$) and larger ($R_2$). In the first set, we perform a randomization to allocate treatments $T_1$ and $T_2$ to the ranked units. Assume that the random allocation assigns $T_1$ to the ranked unit $R_2$ and $T_2$ to the ranked unit $R_1$. In the second set, we perform an opposite allocation without a randomization, i.e, we assign $T_1$ to the ranked unit $R_1$ and $T_2$ to the ranked unit $R_2$. We do a similar procedure for the other pairs, $T_1T_3$ and $T_2T_3$. The construction of Design 1 for one replication is illustrated in Table 1.5 for the treatment levels $T_1$, $T_2$, $T_3$. Usually this basic design is repeated $n$ times to increase the sample size. In Table 1.5 $Y_{l[h]kt}$ is used to denote the measured response, for treatment $l$, ranked unit $h$, set $k$ and replication $t$, where $l = 1, \ldots, L; h = 1, \ldots, H; k = 1, \ldots, K; \text{and } t = 1, \ldots, n$.

**Design 2 (All contrasts design):** The main motivation behind this design is to draw inference for all possible contrasts in a designed experiment having $L$ different treatments. Let $\Delta = \sum_{i=1}^{L} c_i \mu_i$ be the contrast of interest with $\sum c_i = 0$, where for robustness purposes, we require that $c_i$ is either 1 or -1. This design requires that all treatment levels in the contrast are assigned to EUs in the same set. For a contrast of this type, $K$ ($K = L$) different sets, each of size $H$ ($H = K = L$), are needed for the construction of an ORRD. As in Design 1, all $H$ EUs in each set are pre-experimentally judgement ranked. For random allocation of treatment levels to experimental units, we use an available randomization technique which preserves the structure of order restriction. A Latin-square design is used for this purpose by treating sets as rows and ranking orders as columns. To perform the randomization
for Design 2, we select a Latin-square design at random from all available Latin-square designs for a given number of treatments \((L)\). We assign the treatment levels based on this selected Latin-square structure. An ORRD, however, is different from a Latin-square design because of the within-set correlation structure. The Latin-square structure is used only to perform randomized allocation of the treatment levels to the ranked EUs in the ORRD.

One particular ORRD for set size \(H = 3\) and number of treatments \(L = 3\) is illustrated for one replication in Table 1.6. In the first set, we have three ranked EUs. We assign the treatment level \(T_1\) to the \(R_1\) unit, the treatment level \(T_2\) to the \(R_2\) unit and the treatment level \(T_3\) to the \(R_3\) unit. In the second set, we assign the treatment levels \(T_3, T_1, T_2\) to the ranked units 1, 2, 3, respectively. Finally in Set 3, the treatment levels \(T_2, T_3, T_1\) are assigned to the ranked units 1, 2, 3, respectively. In this table, \(Y_{l[h]ji}\) is the measured response for treatment \(l\), set \(j\), replication \(i\) and ranked unit \(h\). Because of the Latin-square structure, each treatment \((T_i)\) appears only once in any column and row. Again this basic design is usually repeated \(n\) times to increase the sample size by selecting a design at random from all possible Latin square designs.

In both Design 1 and Design 2, the EUs need to be grouped into sets of size \(H\). Sets can be constructed either as a group of sequentially admitted EUs or as a random selection from the pool of all available units for the study. In both designs, the subjective information is pre-experimentally used to rank the EUs within a set. Under fairly general conditions, this ranking process induces a positive correlation structure among within-set errors. The ORRDs manipulate this positive correlation structure through a randomization scheme under two restrictions. The first restriction is that the treatment levels having opposite signs in the contrast are equally distributed to the EUs within a set. The second restriction is that each treatment level is applied
Table 1.5: Design 1 for all possible pairwise comparisons of treatments \( T_1, T_2, \) and \( T_3. \)

Table 1.6: Design 2 for all possible contrasts: A Latin square structure with three treatments is used for randomization.
to all ranks. The first restriction turns the positive within-set correlation into a negative correlation in the estimation of the contrast parameter and, hence, reduces the variance of the contrast estimate. The second restriction makes sure that the estimators are unbiased for the contrast parameters.

Design 1 and Design 2 are equivalent when $L = 2$, but they differ when $L > 2$. While Design 1 always requires two units in each set ($H = 2$), Design 2 requires $H$ units in each set ($H = L = K$). When $L > 2$, Design 1 requires more sets and experimental units than Design 2 for one repetition of the basic design.

Ozturk and MacEachern (2004) developed an ORRD within the context of a control versus treatment multiple comparison procedure. The ORRD is introduced in an ANOVA setting to test if the control group is different from any of the other $L - 1$ treatment groups, where $L \geq 2$. Let $F(x - \theta_i)$, $i = 1, \cdots, L$, be the CDF of the $i$-th treatment population with the unique median $\theta_i$. Let the subscripts C and 1 indicate the control group, and subscripts 2, $\cdots$, $L$ denote the treatment groups. The control versus the treatments comparison testing procedure has the following null hypothesis

$$H_0 : \theta_1 = \theta_i; \quad i = 2, \cdots, L,$$

against the alternative hypothesis,

$$H_A : \theta_1 \neq \theta_i \text{ for at least one } i, i \geq 2.$$

The testing procedure first constructs $(1 - \alpha_i)100\%$ median confidence intervals $I_i = [L_i, U_i], i = 1, \cdots, L$, for the treatment medians, where $L_i$ and $U_i$ are the lower and upper limits of the median confidence interval for $\theta_i$. The test then rejects the null hypothesis $H_0$ in favor of the alternative hypothesis $H_A$ if, for some $i \geq 2$, $I_i$ and $I_1$ are disjoint. The family-wise error rate is controlled by the appropriate choices of the
confidence coefficients, $1 - \alpha_i$, $i = 1, \cdots, L$. Ozturk and MacEachern (2004) showed that their procedure is superior to corresponding procedures based on CRD or RSS designs. The superiority appears both in the asymptotic relative efficiency and in the empirical power for finite sample sizes. For the two-sample median test, this new procedure provides asymptotic relative efficiencies of 2.0 and 1.5 with respect to the same procedure applied to SRS and RSS designs, respectively. Similar improvement is observed in empirical power comparisons. This test does not rely on a perfect ranking assumption.

Ozturk and MacEachern (2007) further developed inference for a two-sample problem based on an ORRD with set size $H \geq 2$. In this setting, the parameter of interest is the difference between the control and treatment means, $\Delta = \mu_C - \mu_T$. In this case, the construction of the ORRD requires two sets and is not unique because the ranked EUs in each set can be partitioned into two disjoint sets in more than one way. To account for all possible designs, Ozturk and MacEachern (2007) partitioned the ranks $R_1, \cdots, R_H$ in each set into two disjoint subsets $a = (a_1, a_2, \cdots, a_r)$ and $b = (b_1, b_2, \cdots, b_{H-r})$, such that $(R_1, \cdots, R_H) = (a, b)$. Each partition defines a design. If $H=2$, the design is unique. If $H > 2$, there are $2^{H-1} - 1$ possible ORRDs.

<table>
<thead>
<tr>
<th>Sets</th>
<th>$H = 2$</th>
<th>$H &gt; 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Treatment</td>
</tr>
<tr>
<td>I Ranks</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Resp $Y_1[1]$</td>
<td>$Y_2[2]$</td>
<td>$Y_1[a_1]$</td>
</tr>
<tr>
<td>II Ranks</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Resp $Y_1[2]$</td>
<td>$Y_2[1]$</td>
<td>$Y_1[b_1]$</td>
</tr>
</tbody>
</table>

Table 1.7: Order restricted randomized designs.
These treatment allocations are illustrated in Table 1.7 which was given by Ozturk and MacEachern (2007).

Treatment allocation in each replication is still subject to restricted randomization on the subsets \( a \) and \( b \). In the first set of each replication, treatment and control regimes are randomly assigned to the EUs in subsets \( a \) and \( b \). In the second set, an opposite allocation is performed without randomization. For example, if the treatment and control regimes are assigned to subsets \( a \) and \( b \) in the first set, they are assigned to subsets \( b \) and \( a \) in the second set, respectively. To increase the sample size, the basic design is repeated \( n/2 \) times.

A naive estimator for the contrast parameter, \( \Delta \), can be obtained by taking the difference between the average response measurements on the two treatment regimes. Let

\[
d_{1j} = \sum_{h \in a} Y_{1[h]j} - \sum_{h \in b} Y_{2[h]j}, \quad j = 1, \ldots, n/2,
\]

and

\[
d_{2j} = \sum_{h \in b} Y_{1[h]j} - \sum_{h \in a} Y_{2[h]j}, \quad j = 1, \ldots, n/2,
\]

be the differences between control and treatment responses in set 1 and set 2 in each repetition \( j = 1, \ldots, n/2 \), respectively. Note that \( d_{1j} \) and \( d_{2j} \) are independent of each other, since they represent difference measurements from two different sets. Then the estimator \( \hat{\Delta}_{ORR} \) is given by

\[
\hat{\Delta}_{ORR} = \frac{2}{Hn} \sum_{j} [d_{1j} + d_{2j}].
\]

Ozturk and MacEachern (2007) also looked at the optimal allocation of the ranks in a set. The optimal design distributes the elements of \( a \) and \( b \) as evenly as possible in a set. For example, the optimal design for \( H = 6 \) must have \( a = (1, 3, 5) \) and
This leads to a design with the property that EUs with adjacent judgment ranks are assigned to different treatments in the contrast parameter. Hence, this design provides the maximum reduction in the variance of the contrast parameter estimator.

Under fairly general conditions, the distribution of \( \hat{\Delta}_{ORR} \) asymptotically converges to a normal distribution. A natural statistic to test the null hypothesis \( H_0 : \Delta = \Delta_0 \) against the alternative hypothesis \( H_A : \Delta \neq \Delta_0 \) is given by

\[
Z_n = \frac{\hat{\Delta}_{ORR} - \Delta_0}{\sqrt{\hat{V}(\hat{\Delta}_{ORR})}},
\]

where \( Z_n \) converges to a standard normal distribution as \( n \to \infty \).

To reduce the effect of ranking error, Ozturk and MacEachern (2007) constructed a consistent estimator, \( \hat{V}(\hat{\Delta}_{ORR}) \), for the variance of \( \hat{\Delta}_{ORR} \). Applying Slutsky’s theorem, they show that

\[
T_n^{**} = \frac{\hat{\Delta}_{ORR} - \Delta_0}{\sqrt{\hat{V}(\hat{\Delta}_{ORR})}}
\]

converges in distribution to the standard normal distribution when the null hypothesis is true. For finite sample sizes, they showed that a Student’s t-distribution provides a better approximation for the null distribution of the test statistic. Their simulation study shows that the proposed design has some advantages compared to its competitors in the literature for both high and low quality of ranking information. They also point out that the efficiency of the new design increases with the quality of ranking information.

Ozturk and Sun (2009) used the ORRD in Ozturk and MacEachern (2007) to develop two-sample inference based on a rank-sum test statistic for the difference between two population medians. The 2009 paper develops statistical inference for
the location shift between treatment and control population medians based on an arbitrary, but consistent, ranking scheme.

Ozturk and Sun show that under the null hypothesis $H_0 : \Delta = 0$, the rank-sum test is asymptotically distribution free, but the size of the test rises substantially if the design has judgment ranking error. To reduce the impact of ranking error on the Type I error rate, the test statistic is modified by replacing the null variance of the test statistic with its estimate. The estimated variance calibrates the test and the calibrated test performs quite well even under imperfect ranking. Empirical evidence shows that the rank-sum test in an ORRD has substantially higher power than the power of the Mann-Whitney-Wilcoxon rank-sum test in a CRD with the same size, as long as there is some useful information to rank the EUs. If the ranking information is poor, the rank-sum test based on an ORRD is at least as good as the corresponding rank-sum test in CRD.

Markiewicz (2008) used a linear model to explain the data structure in an ORRD. The linear model for ORRD is given by

\[
Y_{l[h]kt} = X_{l[h]kt} \beta + \epsilon_{l[h]kt}, \tag{1.2.1}
\]

where $Y_{l[h]kt}$ and $X_{l[h]kt}$ are the response measurement and the covariate vector that correspond to the $l$-th treatment, $h$-th judgment class, $k$-th set type, and the $t$-th replication. The vector $\beta$ is a $p$-dimensional vector of unknown parameters. The error term, $\epsilon_{l[h]kt}$, has CDF $F_{[h]}$ with an unique median zero. Note that even though $\epsilon_{l[h]kt}$ is a judgment order statistic, it is centered so that its median is zero. The parameter vector $\beta$ contains treatment effects as well as the shift parameters for each judgment class distribution to account for the departure from the centered median zero. We use $N$, where $N$ equals $nHK$, to denote the total sample size in the
experiment. The model (1.2.1) can be written in a compact notation as,

\[ Y = X \beta + \epsilon. \]  

(1.2.2)

Markiewicz (2008) used the \( L_1 \)-norm to estimate the regression parameter \( \beta \), defining

\[ D^*(\beta) = \| Y - X \beta \|_{L_1} = \sum_{t=1}^{n} \sum_{h=1}^{H} \sum_{k=1}^{K} |Y_{t[h]kt} - X_{t[h]kt} \beta|. \]

Note that \( D^*(\beta) \) is a convex and continuous function with respect to \( \beta \). Hence, it has a unique minimizer. The estimate of \( \beta \) is obtained by minimizing \( D^*(\beta) \) over the parameter vector \( \hat{\beta} = \arg \min D^*(\beta) \), or equivalently by solving the estimating equation,

\[ S^*(\hat{\beta}) = -\frac{\partial D^*(\beta)}{\partial \beta}\bigg|_{\beta=\hat{\beta}} = X^\top S^*(\hat{\beta}) = 0, \]

where \( S^*(\hat{\beta}) \) is given by

\[
S^*(\hat{\beta}) = \begin{pmatrix}
\text{sign}(Y_{1[1]1} - X_{1[1]1} \hat{\beta}) \\
\vdots \\
\text{sign}(Y_{1[H]Kn} - X_{1[H]Kn} \hat{\beta})
\end{pmatrix}.
\]

Under some regularity conditions, Markiewicz (2008) showed that \( \sqrt{N}(\hat{\beta} - \beta_0) \) converges to a \( p \)-dimensional multivariate normal distribution.

Markiewicz also developed testing procedures for the general linear hypotheses in the form (1.1.3). The tests include the score, drop and Wald tests. Under mild regularity conditions, null distributions of these tests can be approximated by an appropriate chi-squared distribution. A simulation study showed that the empirical sizes of the tests are reasonably close to nominal size, and the powers of the tests are higher than the powers of the corresponding tests in a CRD.
1.3 Generalized Randomized Block Design (GRBD)

The ORRDs introduced in the previous section look similar to a generalized randomized block design (GRBD). The most apparent connection can be made by matching the ranking groups in the ORRD to the blocks in the GRBD, so that there would be $H$ blocks in the GRBD. However, the randomization is different in the block and ORRD designs. In a block design, the treatment regimes are randomly assigned to all experimental units within each block. In the ORRD, each treatment regime is assigned to one of the units in each ranking group (block in block design), imposing a restriction on the randomization. The GRBD lacks this restriction. There is another significant difference between these two designs. In the block design, between-block responses are independent, but in the ORRD between-block (judgment class) responses are positively correlated.

To highlight these differences, Ozturk and MacEachern (2013) provided alternative constructions for the GRBD in the framework of Design 1 and Design 2. Design 1, with $L = 3$ treatments, $H = 2$ units per set, $K = 6$ different sets and $n = 1$, has two ranking classes, $R_1$ and $R_2$. These ranking classes can be considered as block 1 and block 2 in a GRBD. To preserve the correlation structure of Design 1, we first identify two randomized treatment sequences with size $n \times K$ ($1 \times 6 = 6$), one for each block. These treatment sequences are independent of one another. Each sequence is balanced, with every treatment appearing $nK/L$ ($6/3 = 2$) times. In the $k$-th set of Design 1, $k = 1, \cdots, 6$, the EU having rank $R_h$, $h = 1, 2$, is assigned to the treatment in the $k$-th position of the randomized sequence in the block $h$. Construction of the GRBD for the two particular randomized sequences is illustrated in Table 1.8. In this table, the experimental units having ranks $R_1$ and $R_2$ in set $k$, say $k=3$, are assigned to treatments $T_1$ and $T_3$ in the randomized sequences 1 and 2, respectively. The others 5 sets can be interpreted in a similar fashion.
This design preserves the correlation structure of the within-set units of the Design 1. The main difference between ORRD Design 1 and the traditional GRBD design is the correlation structure of units within sets. The within-set response measurements from block 1 and block 2 are positively correlated in Design 1, where they are independent in the traditional GRBD. This positive correlation structure in the ORRD design acts differently on the within-set treatment allocation. To understand the impact of their correlation structure, Ozturk and MacEachern listed all possible within-set treatment allocations. In the block design with one replication, there are nine different possible treatment allocations. These treatment allocations are presented in Table 1.9. In this table, there are two groups of treatment allocations. Group I contains the treatment allocations such that both units in a set receive the same treatment. Group II contains the treatment allocations such that units in a set receive different treatments.

The randomization scheme in a GRBD selects six columns out of the nine possible allocations in Table 1.9 at random. Columns are not selected independently, however; the marginal probability of selecting each column for a particular set is 1/9. The randomization scheme in the ORRD Design 1 always selects the last six


Table 1.9: A list of possible treatment allocations to the EUs in a set for the block design based on Design 1.

<table>
<thead>
<tr>
<th>Group I</th>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td>block 1 ( (R_1) )</td>
<td>( T_1 ) ( T_2 ) ( T_3 )</td>
</tr>
<tr>
<td>block 2 ( (R_2) )</td>
<td>( T_1 ) ( T_2 ) ( T_3 )</td>
</tr>
</tbody>
</table>

columns (Group II allocations), because of the restricted randomization. We note that if we are interested in pairwise differences of treatment means, it is better not to select a treatment allocation from the first three columns in Table 1.9 because they provide less information about the contrast parameter. In this respect, the restricted randomization forces the ORRD to select treatment allocations only from Group II. Hence, it reduces the variance of the estimator of the contrast parameter \( (\mu_i - \mu_j) \). This restriction on the randomization makes the ORRD outperform the GRBD in the estimation of the contrast parameter.

For the construction of a GRBD based on Design 2, Ozturk and MacEachern (2013) used \( L = 3 \), \( H = 3 \), \( K = 3 \), and \( n = 1 \). In this case, since the set size is \( H = 3 \), the design requires three blocks: block 1, block 2 and block 3 corresponding to ranks \( R_1 \), \( R_2 \), and \( R_3 \). Each block contains three experimental units. For Design 2 construction, the block design needs three randomized treatment sequences of size 3, one for each block. Each sequence contains each treatment only once. As in Design 1, in the \( k \)-th set, \( k = 1, 2, 3 \), the experimental unit that has rank \( R_h \), \( h = 1, 2, 3 \), is assigned to the treatment in the \( k \)-th position in block \( h \). Construction of this design for three particular randomized sequences is illustrated in Table 1.10. In this table, experimental units having ranks \( R_1 \), \( R_2 \) and \( R_3 \) in set \( k \), say \( k = 2 \), are assigned treatment levels \( T_2 \), \( T_3 \) and \( T_1 \) from randomized treatment sequences 1, 2 and 3, respectively.
The randomization scheme in an all contrast GRBD design creates 27 different possible treatment allocations to within-set units. These 27 allocations can be grouped into three different groups, Group I, Group II and Group III, which contain the treatment allocations such that all three, two out of three and none of the three units in the set receive the same treatment regime, respectively. These treatment allocations are presented in Table 1.11.

In the randomization scheme for this block design, we select 3 of these 27 treatment allocations at random. Again columns are not selected independently, but the marginal probability of selecting each column for a particular set is 1/27. The randomization scheme in the ORRD, on the other hand, selects three treatment allocations without replacement from the last group (Group III allocations) and applies
it to within-set EUs. Again, the ORRD excludes Group I and II allocations through restricted randomization. Hence it increases the variance of the contrast estimators.

The construction of a GRBD from the perspective of ORRD highlights the difference between the full randomization in GRBD and restricted randomization in ORRD. It is clear from Tables 1.9 and 1.11 that restricted randomization in ORRD is used to emphasize the inference on contrast parameters. In the following chapters, we develop statistical inference for the contrast parameters in a linear model based on a rank dispersion function.
2.1 Model

In this chapter, we fit an additive model to analyze the data obtained from ORRDs which were introduced in the previous chapter. In this additive model, each response measurement contains two components: one deterministic and one random. The deterministic part accounts for the treatment effects, while the random component provides the covariance structure of the ORRD. In a general form, the additive model is given by

\[
Y_{u[w]ij} = \mu_u + \epsilon_{u[w]ij} = \alpha + X_{u[w]ij}^\top \beta + \epsilon_{u[w]ij},
\]

where \(Y_{u[w]ij}\) and \(X_{u[w]ij}^\top\) are, respectively, the response measurement and the covariate vector for the treatment \(u\), ranked unit \(w\), set \(i\) and replication \(j\). The scaler \(\alpha\) is an intercept parameter. The vector \(\beta\) is a \(p\)-dimensional vector of unknown parameters. It is clear that the terms \(\mu_u\) and \(\epsilon_{u[w]ij}\) represent the deterministic and random components of the additive model, respectively. The error term \(\epsilon_{u[w]ij}\) is assumed to come from a continuous distribution with a CDF \(F\) having a finite Fisher information based on an ORRD. We assume that \(\epsilon_{u[w]ij}\) has a CDF \(F_{[w]}\), where \(F_{[w]}\) is the CDF of the \(w\)-th judgment order statistic in a set of size \(H\), from a distribution \(F\) with mean
0 and scale $\sigma$. The ORRD induces a positive correlation structure among within-set error terms. Without loss of generality, we assume that $\epsilon_{u[w]ij}$, $\epsilon_{w'[w']ij}$ are positively correlated.

Let $1_H$ be an $H$-dimensional vector of ones. The additive model (2.1.1) then can be written in a compact notation to highlight the correlation structure of the within-set error terms

$$Y_{ij} = \alpha 1_H + X_{ij} \beta + \epsilon_{ij},$$

(2.1.2)

where $Y_{ij}$, $X_{ij}$ and $\epsilon_{ij}$ represent the $H \times 1$ dimensional response vector, $H \times p$ dimensional design matrix and $H \times 1$ dimensional within-set error vector, for the set $i$ and replication $j$, respectively. In this model, although $\epsilon_{ij}$, $i = 1, \ldots, K$; $j = 1, \ldots, n$, are all independent identically distributed random vectors, the components of $\epsilon_{ij}$ are positively correlated. The design matrix $X_{ij}$ is assumed to be a known constant matrix. The model (2.1.2) can also be written in matrix notation as follows:

$$Y = \alpha 1_N + X \beta + \epsilon = \mu + \epsilon,$$

(2.1.3)

where $X$ is an $N \times p$ dimensional design matrix with $N = nKH$. This model looks like an ordinary linear model, but it is different because of the error structure in $\epsilon$. In an ordinary regression model, the components of $\epsilon$ are all independent. In the ORRD model, however, while errors between different sets are independent, errors within the same set are usually positively correlated. Throughout this dissertation, we also assume that the covariates $X_{ij}$ are centered so that $\sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top = 0$. This model for Design 1, the design for all pairwise comparisons, in Table 1.5 reduces to

$$Y_{u[w]ij} = \theta + \beta_u + \epsilon_{u[w]ij}, \quad u = 1, 2, 3; \quad w = 1, 2; \quad i = 1, \ldots, 6; \quad j = 1,$$
where $\theta$ is the overall mean and $\beta_u$ is the $u$-th treatment effect. This model uses the usual constraint $\sum_{u=1}^{3} \beta_u = 0$ for the treatment effects.

### 2.2 Estimation of regression parameter

We consider drawing inference for the regression parameters $\alpha$ and $\beta$ in the additive model (2.1.2). To develop a theory that covers both estimation and testing, we define a subspace $V$, and use the projection theory to construct the estimators and tests.

Let $Y = \mu + \epsilon$, $\mu \in V$, where $V$ is a $p$-dimensional subspace spanned by the columns of the design matrix $X$. This model can be considered as the coordinate-free version of the model (2.1.2). We first consider the estimation of the parameter vector $\mu$. The estimation of $\mu$ can be achieved by projecting the response vector $Y$ onto subspace $V$ with respect to a metric that measures the distance between the subspace $V$ and response vector $Y$. We define this distance in terms of a norm. For a general norm $|| \cdot ||$, we have the estimator $\hat{\mu}$ as

$$\|Y - \hat{\mu}\| = \min_{\mu \in V} \|Y - \mu\|,$$

where $\hat{\mu}$ is considered as the projection of $Y$ with respect to the norm $|| \cdot ||$ on to the subspace $V$. After we estimate $\hat{\mu}$, we can estimate $\beta$ by solving $X\hat{\beta} = \hat{\mu}$. The choice of the norm $|| \cdot ||$ determines the properties of the estimator.

One of the commonly used metrics in linear models is the $L_2$-norm which leads to the least square analysis in linear models. Least square analysis has some desirable properties. For example it has high efficiency when the random component of the additive model has a normal distribution. On the other hand, it performs poorly when the random part deviates from the normal distribution or when the data has some
outliers. The ORRD, because of judgement ranking error in the random component of the additive model (2.1.1), may be sensitive to this model mis-specification. Thus, it is desirable to use a metric that is less sensitive to the presence of within-set judgement ranking error in the ORRD.

In this dissertation, we consider a rank-based dispersion function,

\[ D_n(\beta) = \frac{1}{N} \sum_{u=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} \epsilon_{u[w]ij} \{ R(\epsilon_{u[w]ij}) - \tau_w \}, \]

where \( R(\epsilon_{u[w]ij}) \) is the rank of \( \epsilon_{u[w]ij} \) among all errors \( \epsilon_{u[w]ij}, u = 1, \cdots, L, \)
\( w = 1, \cdots, H, \quad i = 1, \cdots, K, \quad j = 1, \cdots, n, \) and \( \tau_w \) is the expected value of \( R(\epsilon_{u[w]ij}) \). The rank of \( \epsilon_{u[w]ij} \) can be written in an explicit form

\[ R(\epsilon_{u[w]ij}) = \sum_{q=1}^{H} \sum_{d=1}^{K} \sum_{t=1}^{n} I(\epsilon_{p[q]dt} \leq \epsilon_{u[w]ij}), \]

where \( I(\cdot) \) is an indicator function.

The quantity \( D_n(\beta) \) is an extension of the Jaeckel’s dispersion function to ORRD settings with Wilcoxon scores (Jaeckel, 1972). This dispersion function is similar to the \( L_2 \)-norm. The main difference is that while the \( L_2 \)-norm considers the sum of squared residuals, \( D_n(\beta) \) considers the sum of the weighted errors. These weights are proportional to the differences between the ranks of the errors and their expected values.

**Lemma 2.2.1.** The dispersion function \( D_n(\beta) \) is a convex function with respect to \( \beta \) and almost everywhere differentiable.

**Proof.** See the book by Hettmansperger and McKean (2011, page 168). \( \square \)

If the design matrix \( X \) has full rank, the dispersion function \( D_n(\beta) \) attains its
minimum and the set of minimizers is bounded. We then define our estimator \( \hat{\beta} \) as any value that minimizes \( D_n(\beta) \) in this set

\[
\hat{\beta} = \text{argmin} D_n(\beta).
\]

The estimator can also be defined equivalently by solving the gradient function of \( D_n(\beta) \)

\[
S_n(\hat{\beta}) = -\frac{\partial D_n(\beta)}{\partial \beta}|_{\beta = \hat{\beta}} = \frac{1}{N} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} X_{u[w]ij} \{ R(\hat{\epsilon}_{u[w]ij}) - \tau_w \} = 0,
\]

where \( \hat{\epsilon}_{u[w]ij} = Y_{u[w]ij} - \alpha - X_{u[w]ij}^T \hat{\beta} \).

This gradient function \( S_n(\hat{\beta}) \) will be called the score function throughout this dissertation. We call \( S_n(\hat{\beta}) = 0 \) the estimating equation. Since the estimating equation is invariant with respect to a location shift, the intercept parameter \( \alpha \) can not be estimated from the rank-dispersion function \( D_n(\beta) \). So without loss of generality, we assume that \( \alpha = 0 \).

The score function can be written in compact notation as a function of the error vector \( \epsilon_{ij} \), where \( \epsilon_{ij}^T = (\epsilon_{u_1[1]ij}, \epsilon_{u_2[2]ij}, \cdots, \epsilon_{u_H[H]ij}) \) is error vector in the \( i \)-th set and \( j \)-th replication. Let

\[
h_R^T(\epsilon_{ij}) = (R(\epsilon_{u_1[1]ij}) - \tau_1, R(\epsilon_{u_2[2]ij}) - \tau_2, \cdots, R(\epsilon_{u_H[H]ij}) - \tau_H)
\]

be the vector of centered ranks in the \( i \)-th set and \( j \)-th replication, and

\[
A_R^T(\epsilon) = (h_R^T(\epsilon_{11}), h_R^T(\epsilon_{21}), \cdots, h_R^T(\epsilon_{K1}), \cdots, h_R^T(\epsilon_{ij}), \cdots, h_R^T(\epsilon_{Kn}))
\]

be the \( N \)-dimensional centered ranks of all error terms. With this compact notation,
the score function reduces to

\[ S_n(\beta_0) = \frac{1}{N} X^\top A_R(\epsilon) = \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top h_R(\epsilon_{ij}), \]

where \( \beta_0 \) is the true regression parameter.

### 2.3 Assumptions

We need the following assumptions throughout this dissertation to develop the asymptotic theory. These assumptions state the required conditions on the ranking process, error distribution and design matrix.

**J1:** The judgment ranking process is consistent. That is to say, the ranking process assigns a rank to each unit in a set, and the same ranking process is used in all sets.

**A1:** We have \( \omega_h = \int f(y)f[y_h(y)]dy < \infty, h = 1, \cdots, H, \) which is satisfied if the underlying distribution has an absolutely continuous density \( f \) with finite Fisher information.

**A2:** \( \lim_{N \to \infty} \frac{1}{N} X^\top \Gamma X = \Sigma, \) where \( \Gamma = \text{Cov}(A_R(\epsilon)) \), and \( \Sigma \) is a positive definite matrix.

**A3:** \( \lim_{N \to \infty} \frac{m(X_{ij})}{\sqrt{N}} = 0, \) where \( m(A) = \max_{p,q}(|a_{pq}|), \) and \( A \) is a matrix whose entry in column \( q \) and row \( p \) is \( a_{pq}. \)

**A4:** \( \lim_{N \to \infty} \frac{1}{N} X^\top X = \Xi, \) and \( \Xi \) is a positive definite matrix.

**A5:** \( \lim_{N \to \infty} \frac{1}{N} X^\top \Omega X = \Psi > 0, \) where \( \Omega \) is a diagonal matrix of \( (\omega_1, \cdots, \omega_{nK}), \omega_i = (\omega_1, \omega_2, \cdots, \omega_H), i = 1, \cdots, nK. \)

Under assumption **J1**, Dell and Clutter (1972) provide a fundamental equality

\[ F(t) = \frac{1}{N} \sum_{h=1}^{H} F[y_h(t)], \quad (2.3.1) \]
which establishes the connection between the judgment class and the population CDFs. Using equation (2.3.1), we can give an explicit expression for \( \tau_w \). We first note that

\[
\tau_{qw} = E(I(\epsilon_{p[q]}dt \leq \epsilon_{u[w]ij})) = \begin{cases} 
\tau_{qw}^* & \text{if } d=i \text{ and } t=j, \\
\tau_{qw}^* & \text{otherwise,}
\end{cases}
\]

where \( \tau_{qw}^* = E(I(\epsilon_{p[q]}ij \leq \epsilon_{u[w]ij}) \) and \( \tau_{qw}^* = \int F_{[q]}(y)F_{[w]}(y)dy. \)

In the expression for \( \tau_{qw} \), it is clear that while \( \tau_{qw}^* \) denotes the expectation of the indicator function of the errors from the same set, \( \tau_{qw}^* \) denotes the expectation of the indicator function of the residuals from different sets. Using this notation, under assumption \( J1 \), the expression \( \tau_w \) can be written as

\[
\tau_w = \sum_{q=1}^{H} \sum_{d=1}^{K} \sum_{t=1}^{n} (E(I(\epsilon_{p[q]}dt \leq \epsilon_{u[w]ij}))) = (nK - 1) \sum_{q=1}^{H} \tau_{qw}^* + \sum_{q=1}^{H} \tau_{qw}^*.
\]

From the fundamental equality (2.3.1), it is not difficult to show that

\[
\sum_{q=1}^{H} \tau_{qw}^* = \sum_{q=1}^{H} \int F_{[q]}(y)dF_{[w]}(y) = H \int F(y)dF_{[w]}(y).
\]

Let \( \tau_w = \int F(y)dF_{[w]}(y) \). Then the quantity \( \tau_w \) reduces to

\[
\tau_w = (nK - 1)H \tau_w + \sum_{q=1}^{H} \tau_{qw}^*.
\]

We note that the expression above holds for any consistent judgement ranking scheme.
2.4 Asymptotic Normality of the Score Function

The estimator is implicitly defined by the solution of the estimating equation \( S_n(\hat{\beta}) = 0 \). The asymptotic behavior of \( \hat{\beta} \) strongly depends on the asymptotic behavior of the score function \( S_n(\hat{\beta}) \). For a known value of the regression parameter \( \beta_0 \), we first find the expected value and limiting variance of \( S_n(\beta_0) \).

**Theorem 2.4.1.** Under assumptions \( J1, A1 \) and \( A2 \),

(i) \( E(S_n(\beta_0)) = 0 \)

(ii) \( \lim_{N \to \infty} \text{Var}\left( \frac{1}{\sqrt{N}} S_n(\beta_0) \right) = \lim_{N \to \infty} \frac{1}{N} X^\top \Gamma X = \Sigma. \)

**Proof.** (i) The proof follows from the fact that the score function \( S_n(\beta_0) \) is centered correctly.

(ii) We first partition the variance of \( \frac{1}{\sqrt{N}} S_n(\beta_0) \) into within- and between-set covariances as follows:

\[
\text{Var}(\frac{1}{\sqrt{N}} S_n(\beta_0)) = \frac{1}{N^3} \text{Var}\left( \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top h_R(\epsilon_{ij}) \right) = \frac{1}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} \text{Cov}(X_{ij}^\top h_R(\epsilon_{ij}), X_{dt}^\top h_R(\epsilon_{dt})) = \frac{1}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} \text{Cov}(X_{ij}^\top h_R(\epsilon_{ij}), X_{ij}^\top h_R(\epsilon_{ij})) + \frac{1}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} \sum_{\{i=d,j=t\}^c} \text{Cov}(X_{ij}^\top h_R(\epsilon_{ij}), X_{dt}^\top h_R(\epsilon_{dt}))
\]

\[= B_1 + B_2, \quad (2.4.1)\]

where the notation \( \{i = d, j = t\}^c \) is used to indicate the index set that either \( i \neq d \) or \( j \neq t \). In the expression \( B_2 \), the random vectors \( \epsilon_{ij} \) and \( \epsilon_{dt} \) are mutually independent since they are from different sets. We now evaluate the expressions \( B_1 \) and \( B_2 \).
We first consider the expression $B_1$:

$$
B_1 = \frac{1}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} \text{Cov}(X_{ij}^T h_R(\epsilon_{ij}), X_{ij}^T h_R(\epsilon_{ij}))
$$

$$
= \frac{1}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^T \text{Cov}(h_R(\epsilon_{ij}), h_R(\epsilon_{ij})) X_{ij}.
$$

We note that $\text{Cov}(h_R(\epsilon_{ij}), h_R(\epsilon_{ij}))$ is an $H \times H$ matrix. Let

$$
\gamma^{**} = \text{Cov}(h_R(\epsilon_{ij}), h_R(\epsilon_{ij})) = \begin{bmatrix}
\gamma^{**}_{[1],[1],n} & \cdots & \gamma^{**}_{[1],[H],n} \\
\vdots & \ddots & \vdots \\
\gamma^{**}_{[W],[q],n} & \cdots & \gamma^{**}_{[H],[H],n}
\end{bmatrix},
$$

where $\gamma^{**}_{[w],[q],n} = E((R(\epsilon_{u[w]ij}) - \tau_w)(R(\epsilon_{p[q]ij}) - \tau_q))$. In the covariance matrix $\gamma^{**}$, it is sufficient to look at an arbitrary entry $\gamma^{**}_{[w],[q],n}$

$$
\gamma^{**}_{[w],[q],n} = \sum_{\eta=1}^{H} \sum_{g=1}^{K} \sum_{m=1}^{n} \sum_{b=1}^{H} \sum_{c=1}^{K} \sum_{f=1}^{n} E(T_{\zeta[\eta]gm}(u[w]ij) T_{a[\eta]cf}(p[q]ij)),
$$

where $T_{\zeta[\eta]gm}(u[w]ij) = I(\epsilon_{\zeta[\eta]gm} \leq \epsilon_{u[w]ij}) - \tau_{\eta w}^*$. We now partition $\gamma^{**}_{[w],[q],n}$ into several components to isolate the independent components from the dependent ones

$$
\gamma^{**}_{[w],[q],n} = \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{g=1}^{K} \sum_{m=1}^{n} E(T_{\zeta[\eta]gm}(u[w]ij) T_{a[\eta]gm}(p[q]ij))
$$

$$
+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{g=1}^{K} \sum_{m=1}^{n} \sum_{\{g=e,m=f\}^c} E(T_{\zeta[\eta]gm}(u[w]ij) T_{a[\eta]cf}(p[q]ij)).
$$
We further partition $\gamma^{**}_{[u],[q],n}$ to yield

\[
\gamma^{**}_{[u],[q],n} = \sum_{\eta=1}^{H} \sum_{b=1}^{H} E(T_{\zeta[j]ij}(u[w]ij)T_{a[b]ij}(p[q]ij)) \\
+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{\{g=i, m=j\}^c} E(T_{\zeta[j]gm}(u[w]ij)T_{a[b]gm}(p[q]ij)) \\
+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{\{c=i, f=j\}^c} E(T_{\zeta[j]if}(u[w]ij)T_{a[b]if}(p[q]ij)) \\
+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{\{g=i, m=j\}^c} \sum_{\{c=i, f=j\}^c} E(T_{\zeta[j]gm}(u[w]ij)T_{a[b]cf}(p[q]ij)),
\]

where $I_{gmcf}^c(ij)$ is the index set that

\[
I_{gmcf}^c(ij) = \{(g, m, c, f) : \{g = c, m = f\}^c \cap \{g = i, m = j\}^c \cap \{c = i, f = j\}^c\}.
\]

In the equation $\gamma^{**}_{[w],[q],n}$, the first four terms are of order $O(n)$, which becomes negligible in the computation of $B_1$. The covariance $\gamma^{**}_{[w],[q],n}$ then simplifies to

\[
\gamma^{**}_{[w],[q],n} = \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{\{g=i, m=j\}^c} \sum_{\{c=i, f=j\}^c} E(T_{\zeta[j]gm}(u[w]ij)T_{a[b]cf}(p[q]ij)) + O(n).
\]

Let $C_{n,K} = (nK - 1)(nK - 2)$. The summation over the index set $I_{gmcf}^c(ij)$ yields

\[
\gamma^{**}_{[w],[q],n} = C_{n,K} \sum_{\eta=1}^{H} \sum_{b=1}^{H} E(T_{\zeta[j]gm}(u[w]ij)T_{a[b]cf}(p[q]ij)) + O(n) \\
= C_{n,K} \sum_{\eta=1}^{H} \sum_{b=1}^{H} E((I(\epsilon_{\zeta[j]gm} \leq \epsilon_{u[w]ij}) - \tau^*_{ju})(I(\epsilon_{a[b]cf} \leq \epsilon_{p[q]ij}) - \tau^*_{bq})) + O(n).
\]

38
Conditioning on the random vector \( \epsilon_{ij} \), the expression for \( \gamma_{[w],[q],n}^{**} \) reduces to

\[
\begin{align*}
\gamma_{[w],[q],n}^{**} &= C_{n,K} \sum_{\eta=1}^{H} \sum_{b=1}^{H} E\{E\left( T_{\zeta[n]gm}(u[w]ij)T_{a[b]f}(p[q]ij)|\epsilon_{ij}\right)\} + O(n) \\
&= C_{n,K} \sum_{\eta=1}^{H} \sum_{b=1}^{H} E\{(F_{[\eta]}(\epsilon_{u[w]ij}) - \tau_{\eta w}^*)(F_{[b]}(\epsilon_{p[q]ij}) - \tau_{\eta q}^*)\} + O(n) \\
&= C_{n,K} \sum_{\eta=1}^{H} \sum_{b=1}^{H} \int (F_{[\eta]}(x) - \tau_{\eta w}^*)(F_{[b]}(y) - \tau_{\eta q}^*)dF_{[w,q]}(x,y) + O(n) \\
&= C_{n,K} \int \sum_{\eta=1}^{H} (F_{[\eta]}(x) - \tau_{\eta w}^*) \sum_{b=1}^{H} (F_{[b]}(y) - \tau_{\eta q}^*)dF_{[w,q]}(x,y) + O(n).
\end{align*}
\]

The quantity \( \gamma_{[w],[q],n}^{**} \) can be simplified further using the fundamental equality (2.3.1)

\[
\begin{align*}
\gamma_{[w],[q],n}^{**} &= C_{n,K} \int (HF(x) - H\tau_{.w})(HF(y) - H\tau_{.q})dF_{[w,q]}(x,y) + O(n) \\
&= H^2 C_{n,K} \int (F(x) - \tau_{.w})(F(y) - \tau_{.q})dF_{[w,q]}(x,y) + O(n) \\
&= H^2 C_{n,K} \gamma_{[w],[q]}^{**} + O(n),
\end{align*}
\]

where, \( \gamma_{[w],[q]}^{**} = \int (F(x) - \tau_{.w})(F(y) - \tau_{.q})dF_{[w,q]}(x,y) \). Let

\[
\gamma = \begin{bmatrix}
\gamma_{[1],[1]}^{**} & \gamma_{[1],[2]}^{**} & \cdots & \gamma_{[1],[H]}^{**} \\
\gamma_{[2],[1]}^{**} & \gamma_{[2],[2]}^{**} & \cdots & \gamma_{[2],[H]}^{**} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{[H],[1]}^{**} & \gamma_{[H],[2]}^{**} & \cdots & \gamma_{[H],[H]}^{**}
\end{bmatrix}.
\]
With this notation, expression $B_1$ in equation (2.4.1) reduces to

$$B_1 = \frac{1}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} \text{Cov}(X_{ij}^\top h_R(\epsilon_{ij}), X_{ij}^\top h_R(\epsilon_{ij}))$$

$$= \frac{1}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top \text{Cov}(h_R(\epsilon_{ij}), h_R(\epsilon_{ij})) X_{ij}$$

$$= \frac{H^2 C_{n,K}}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top \gamma X_{ij} + o(1).$$

We then conclude that

$$\lim_{N \to \infty} B_1 = \lim_{N \to \infty} \frac{H^2 C_{n,K}}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top \gamma X_{ij} = \lim_{N \to \infty} \frac{1}{N} X^\top \Gamma X = \Sigma,$$

where $\Gamma$ is an $N \times N$ block diagonal matrix of $\gamma$.

We now consider the expression $B_2$ in equation (2.4.1):

$$B_2 = \frac{1}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} \text{Cov}(X_{ij}^\top h_R(\epsilon_{ij}), X_{dt}^\top h_R(\epsilon_{dt}))$$

$$= \frac{1}{N^3} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} X_{ij}^\top \text{Cov}(h_R(\epsilon_{ij}), h_R(\epsilon_{dt})) X_{dt}. $$

Our interest in this expression is the covariance matrix $\text{Cov}(h_R(\epsilon_{ij}), h_R(\epsilon_{dt}))$. It is again sufficient to look at the $(w, q)$-th entry of this matrix. For given values of $i, j, d$ and $t$, we write

$$\gamma^*_{[w],[q],n} = \sum_{\eta=1}^{H} \sum_{g=1}^{K} \sum_{m=1}^{n} \sum_{b=1}^{H} \sum_{c=1}^{K} \sum_{f=1}^{n} E(T_{\zeta [v] \gamma m}(u[w] ij) T_{a[b] \epsilon f}(p[q] dt))$$

$$= \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{g=1}^{K} \sum_{m=1}^{n} E(T_{\zeta [v] \gamma m}(u[w] ij) T_{a[b] \gamma m}(p[q] dt))$$

$$+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{g=e}^{K} \sum_{m=f}^{n} E(T_{\zeta [v] \gamma m}(u[w] ij) T_{a[b] \epsilon f}(p[q] dt)).$$

40
We further partition these two sums and write

$$
\gamma^*_{[w],[q],n} = \sum_{\eta=1}^{H} \sum_{b=1}^{H} E(T_{\zeta[n]ij}(u[w]ij)T_{a[b]ij}(p[q]dt))
+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} E(T_{\zeta[n]dt}(u[w]ij)T_{a[b]dt}(p[q]dt))
+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} K \sum_{n=1}^{n} \sum_{\{g=d, m=t\}^c} E(T_{\zeta[n]gm}(u[w]ij)T_{a[b]gm}(p[q]dt))
+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} K \sum_{n=1}^{n} \sum_{\{g=c, m=f\}^c} E(T_{\zeta[n]gm}(u[w]ij)T_{a[b]cf}(p[q]dt)).
$$

It is clear that the first two expressions in the above equation do not involve $n$. Therefore, they become asymptotically negligible and can be put in $O(1)$. With this $O(1)$ notation, expression $\gamma^*_{[w],[q],n}$ reduces to

$$
\gamma^*_{[w],[q],n} = \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{\{g=i, m=j\}^c \cap \{g=d, m=t\}^c} E(T_{\zeta[n]gm}(u[w]ij)T_{a[b]gm}(p[q]dt))
+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} K \sum_{n=1}^{n} \sum_{\{g=c, m=f\}^c} E(T_{\zeta[n]gm}(u[w]ij)T_{a[b]cf}(p[q]dt)) + O(1)
= b_{21} + b_{22} + O(1).
$$

Since the expression in $b_{21}$ is constant over the set

$$
\{\{g = i, m = j\}^c \cap \{g = d, m = t\}^c\},
$$

the sum simplifies to

$$
b_{21} = (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} E(T_{\zeta[n]12}(u[w]ij)T_{a[b]12}(p[q]dt)),
$$

where expected value is defined over the set $\{i = 1, j = 2\}^c \cap \{d = 1, t = 2\}^c$. We
We further simplify

\[ b_{22} = \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{K=} \sum_{n=1}^{n} E(T_{\zeta^{[\eta]}_{ij}}(u[w]i,j)T_{a[b]c|f}(p[q]dt)) \]

\[ + \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{K=} \sum_{n=1}^{n} \sum_{K=} \sum_{n=1}^{n} E(T_{\zeta^{[\eta]}_{gm}}(u[w]i,j)T_{a[b]c|f}(p[q]dt)). \]

### Now consider \( b_{22} \)

\[ b_{22} = \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{K=} \sum_{n=1}^{n} E(T_{\zeta^{[\eta]}_{dt}}(u[w]i,j)T_{a[b]ij}(p[q]dt)) \]

\[ + \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{K=} \sum_{n=1}^{n} \sum_{K=} \sum_{n=1}^{n} E(T_{\zeta^{[\eta]}_{gm}}(u[w]i,j)T_{a[b]cf}(p[q]dt)). \]

\[ + \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{K=} \sum_{n=1}^{n} \sum_{K=} \sum_{n=1}^{n} E(T_{\zeta^{[\eta]}_{gm}}(u[w]i,j)T_{a[b]cf}(p[q]dt)). \]

### Partitioning of the first term in the above expression, we have

\[ b_{22} = \sum_{\eta=1}^{H} \sum_{b=1}^{H} E(T_{\zeta^{[\eta]}_{dt}}(u[w]i,j)T_{a[b]ij}(p[q]dt)) \]

\[ + \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{K=} \sum_{n=1}^{n} E(T_{\zeta^{[\eta]}_{dt}}(u[w]i,j)T_{a[b]ij}(p[q]dt)) \]

\[ + \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{K=} \sum_{n=1}^{n} \sum_{K=} \sum_{n=1}^{n} E(T_{\zeta^{[\eta]}_{gm}}(u[w]i,j)T_{a[b]cf}(p[q]dt)). \]

Again the first expression in the above equation does not contain \( n \), and it can be put in \( O(1) \). The expression of \( b_{22} \) then reduces to

\[ b_{22} = \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{K=} \sum_{n=1}^{n} E(T_{\zeta^{[\eta]}_{dt}}(u[w]i,j)T_{a[b]cf}(p[q]dt)) \]

\[ + \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{K=} \sum_{n=1}^{n} \sum_{K=} \sum_{n=1}^{n} E(T_{\zeta^{[\eta]}_{gm}}(u[w]i,j)T_{a[b]cf}(p[q]dt)). \]

\[ + O(1). \]

We further simplify \( b_{22} \) by observing that expectation in the first sum is constant over
We insert this expression in equation (2.4.2) to obtain two expressions are zero. The quantity
\[ b_{22} \]
where the summation in the first term runs over the index set \( S_{gm,c,f} \), which is defined as

\[ S_{gm,c,f} = \{(g, m, c, f) : (g, m, c, f) \in \{g = c, m = f\}^c \cap \{g = i, m = j\}^c \cap \{g = d, m = t\}^c \cap \{c = i, f = j\}^c\}. \]

In the equation \( b_{23} \), since the error terms from different sets are independent, the first two expressions are zero. The quantity \( b_{23} \) thus simplifies to

\[ b_{23} = (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} \sum_{K}^{n} \sum_{n}^{K} E(T_{\zeta[n]}(u[w]ij)T_{a[b]}(p[q]dt)) + O(1). \]
We finally combine $b_{21}$ and $b_{22}$ to compute $\gamma^*_w, \gamma^*_u, \gamma^*_n = b_{21} + b_{22}$, as
\[
\gamma^*_w, \gamma^*_u, \gamma^*_n = \begin{align*}
&= (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} E(T_{\zeta[\eta]gm}(u[w]ij)T_{a[b]gm}(p[q]dt)) \\
&+ (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} E(T_{\zeta[\eta]dt}(u[w]ij)T_{a[b]cf}(p[q]dt)) \\
&+ (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} E(T_{\zeta[\eta]gm}(u[w]ij)T_{a[b]ij}(p[q]dt)) + O(1).
\end{align*}
\]

Note that $T_{\zeta[\eta]gm}(u[w]ij) = I(\epsilon_{\zeta[\eta]gm} \leq \epsilon_{u[w]ij}) - \tau_{uw}$. By inserting this into $\gamma^*_w, \gamma^*_u, \gamma^*_n$, we write
\[
\gamma^*_w, \gamma^*_u, \gamma^*_n = \begin{align*}
&= (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} E((I(\epsilon_{\zeta[\eta]gm} \leq \epsilon_{u[w]ij}) - \tau_{uw})(I(\epsilon_{a[b]gm} \leq \epsilon_{p[q]dt}) - \tau_{bq})) \\
&+ (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} E((I(\epsilon_{\zeta[\eta]dt} \leq \epsilon_{u[w]ij}) - \tau_{uw})(I(\epsilon_{a[b]cf} \leq \epsilon_{p[q]dt}) - \tau_{bq})) \\
&+ (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} E((I(\epsilon_{\zeta[\eta]gm} \leq \epsilon_{u[w]ij}) - \tau_{uw})(I(\epsilon_{a[b]ij} \leq \epsilon_{p[q]dt}) - \tau_{bq})) \\
&+ O(1).
\end{align*}
\]

Using the conditional argument as in $B_1$, the expression $\gamma^*_w, \gamma^*_u, \gamma^*_n$ reduces to
\[
\gamma^*_w, \gamma^*_u, \gamma^*_n = \begin{align*}
&= (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} \int (1 - F_{[w]}(x) - \tau_{uw})(1 - F_{[q]}(y) - \tau_{bq})dF_{[\eta,b]}(x, y) \\
&+ (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} \int (1 - F_{[w]}(x) - \tau_{uw})(F_{[b]}(y) - \tau_{bq})dF_{[\eta,q]}(x, y) \\
&+ (nK - 2) \sum_{\eta=1}^{H} \sum_{b=1}^{H} \int (F_{[\eta]}(x) - \tau_{uw})(1 - F_{[q]}(y) - \tau_{bq})dF_{[w,b]}(x, y) + O(1).
\end{align*}
\]
Let

$$\gamma^*_{[w],[q]} = \sum_{\eta=1}^{H} \sum_{b=1}^{H} \int (1 - F_{[w]}(x) - \tau_{\eta w})(1 - F_{[q]}(y) - \tau_{bq})dF_{[\eta,b]}(x,y)$$

$$+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} \int (1 - F_{[w]}(x) - \tau_{\eta w})(F_{[b]}(y) - \tau_{bq})dF_{[\eta,b]}(x,y)$$

$$+ \sum_{\eta=1}^{H} \sum_{b=1}^{H} \int (F_{[\eta]}(x) - \tau_{\eta w})(1 - F_{[q]}(y) - \tau_{bq})dF_{[w,b]}(x,y),$$

and

$$\gamma^* = \begin{bmatrix} \gamma^*_{[1],[1]} & \cdots & \gamma^*_{[1],[H]} \\ \vdots & \ddots & \vdots \\ \gamma^*_{[H],[1]} & \cdots & \gamma^*_{[H],[H]} \end{bmatrix}.$$
to a compact form:

\[
B_2 = \frac{(nK - 2)}{N^3} \left\{ \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} X_{ij}^\top \gamma^* X_{dt} + \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top \gamma^* X_{ij} \right. \\
\left. - \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top \gamma^* X_{ij} \right\}
\]

\[
= \frac{(nK - 2)}{N^3} \left\{ \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} X_{ij}^\top \gamma^* X_{dt} - \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top \gamma^* X_{ij} \right\}
\]

\[
= \frac{(nK - 2)}{N^3} \left\{ \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top \sum_{d=1}^{K} \sum_{t=1}^{n} \gamma^* X_{dt} - \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top \gamma^* X_{ij} \right\}
\]

\[
= \frac{(nK - 2)}{N^3} \left\{ - \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top \gamma^* X_{ij} \right\}.
\]

The last equality follows from the fact that the \(X_{ij}\) are centered, so that \(\sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top = 0\). It is now easy to observe that \(B_2\) converges to \(0\), as \(n\) goes to infinity. By combining the \(B_1\) and \(B_2\) terms, we conclude that \(\text{Var}(\frac{1}{\sqrt{N}}S_n(\beta)) = B_1 + B_2\) converges to \(\Sigma\) as \(n\) goes to infinity, which completes the proof of the theorem. \(\square\)

It is obvious that the score function \(S_n(\beta_0)\) consists of a weighted sum of the centered error terms, where the weights are the known covariates in the model (2.1.1). Since the ranks are not independent, the components in the linear combination of \(S_n(\beta_0)\) are correlated. In this case, the regular central limit theorem does not apply. For the asymptotic distribution of the score function, we first project \(S_n(\beta_0)\) onto a sum of independent observations, and show that the projected statistic is asymptotically equivalent to \(S_n(\beta_0)\). The limiting distribution of \(S_n(\beta_0)\) then follows from the sum of independent observations.

**Theorem 2.4.2.** Under assumptions \(J1, A1, A2\) and \(A3\), as \(N \to \infty\), \(\frac{1}{\sqrt{N}}S_n(\beta_0)\)
is equivalent in distribution to $V_p$, where $V_p$ is the projection of the estimating equation \( \frac{1}{\sqrt{N}}S_n(\beta_0) \) onto the space of the sum of independent random variables, 

$$
V_p = \frac{1}{\sqrt{N}} \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^\top h_F(\epsilon_{ij}) \quad \text{and} \quad h_F(\epsilon_{ij}) = \begin{pmatrix}
F(\epsilon_{u[1]ij}) - \tau_1 \\
\vdots \\
F(\epsilon_{u[w]ij}) - \tau_w \\
\vdots \\
F(\epsilon_{u[H]|ij}) - \tau_H
\end{pmatrix}.
$$

A key feature of the projected statistic $V_p$ is that it allows us to replace the centered rank error vector $h_R(\epsilon_{ij})$ in $S_n(\beta_0)$ with $h_F(\epsilon_{ij})$. The random vectors $h_F(\epsilon_{ij})$, $i = 1, \ldots, K$, $j = 1, \ldots, n$, unlike $h_R(\epsilon_{ij})$, are mutually independent. The asymptotic distribution of $V_p$ then follows from the independent sum in $V_p$.

**Proof of Theorem 2.4.2.** Note that

$$
\frac{1}{\sqrt{N}}S_n(\beta_0) = \frac{1}{N^{\frac{3}{2}}} \sum_{g=1}^{K} \sum_{m=1}^{n} X_{gm}^\top h_R(\epsilon_{gm}).
$$

The projection of \( \frac{1}{\sqrt{N}}S_n(\beta_0) \) is given by the sum of the conditional expected values of \( \frac{1}{\sqrt{N}}S_n(\beta_0) \) given $\epsilon_{ij}$, $i = 1, \ldots, K$, $j = 1, \ldots, n$, as $V_p = \frac{1}{\sqrt{N}} \sum_{i} \sum_{j} \mathbb{E}(S_n(\beta_0)|\epsilon_{ij})$. We first need to evaluate the following conditional expectation

$$
\frac{1}{\sqrt{N}}E(S_n(\beta_0)|\epsilon_{ij}) = \frac{1}{N^{\frac{3}{2}}} \left( X_{ij}^\top \mathbb{E}(h_R(\epsilon_{ij})|\epsilon_{ij}) + \sum_{\{g=i,m=j\}^c} X_{gm}^\top \mathbb{E}(h_R(\epsilon_{gm})|\epsilon_{ij}) \right)
$$

$$
\equiv \frac{1}{N^{\frac{3}{2}}} \left( G_{ij} + \sum_{\{g=i,m=j\}^c} D_{gm} \right).
$$
In the above equation, to simplify the notation we write

\[ G_{ij} = X_{ij}^T \mathbb{E}(h_R(e_{ij})|e_{ij}) = X_{ij}^T g_{ij} \quad \text{and} \quad D_{gm} = X_{gm}^T \mathbb{E}(h_R(e_{gm})|e_{ij}) = X_{gm}^T d_{ij}, \]

where \( g_{ij} \) and \( d_{ij} \) represent the \( H \)-dimensional vectors of the conditional expectations of \( h_R(e_{ij}) \) and \( h_R(e_{gm}) \) given the within-set residual vector \( e_{ij} \). In these conditional expectations, we first evaluate an arbitrary component \( g_{ij}(w) \),

\[ w = 1, \cdots, H, \] in \( g_{ij} \),

\[ g_{ij}(w) = E(R(\epsilon_{u[w]ij}))|e_{ij} - \tau_w \]

\[ = \sum_{q=1}^{H} \sum_{d=1}^{K} \left( \sum_{t=1}^{n} E(I(\epsilon_{p[q]}dt \leq \epsilon_{u[w]ij})|e_{ij}) - E(I(\epsilon_{p[q]}dt \leq \epsilon_{u[w]ij})) \right) \]

\[ = \sum_{q=1}^{H} \sum_{d=1}^{K} E\left( \left( I(\epsilon_{p[q]}dt \leq \epsilon_{u[w]ij}) - \tau_{qw}^* \right) | e_{ij} \right) \]

\[ + \sum_{q=1}^{H} \sum_{d=1}^{K} E\left( \left( I(\epsilon_{p[q]}d_j \leq \epsilon_{u[w]ij}) - \tau_{qw}^* \right) | e_{ij} \right) - E\left( I(\epsilon_{p[q]}dt \leq \epsilon_{u[w]ij}) \right) \].

We further partition the second expression above to separate the components containing the within-set error terms in \( e_{ij} \),

\[ g_{ij}(w) = \sum_{q=1}^{H} \sum_{d=1}^{K} \sum_{t \neq j}^{n} \left( \sum_{t=1}^{n} E[I(\epsilon_{p[q]}dt \leq \epsilon_{u[w]ij}) - \tau_{qw}^*]|e_{ij}] \right) \]

\[ + \sum_{q=1}^{H} \sum_{d \neq i}^{K} E[I(\epsilon_{p[q]}d_j \leq \epsilon_{u[w]ij}) - \tau_{qw}^*]|e_{ij}] \]

\[ + \sum_{q=1}^{H} \left( E[I(\epsilon_{p[q]}ij \leq \epsilon_{u[w]ij})]|e_{ij}] - \tau_{qw}^{**} \right). \]

The conditional expectation of the indicator function simplifies the above expression
\[ g_{ij}(w) = (n - 1) \sum_{q=1}^{H} \sum_{d=1}^{K} (F_{[q]}(\epsilon_{u[w]ij}) - \tau_{qw}^{*}) \]
\[ + (K - 1) \sum_{q=1}^{H} (F_{[q]}(\epsilon_{u[w]ij}) - \tau_{qw}) + \sum_{q=1}^{H} (I(\epsilon_{p[q]}ij \leq \epsilon_{u[w]ij}) - \tau_{qw}^{**}). \]

Finally, using the fundamental equation (2.3.1), we write

\[ g_{ij}(w) = (n - 1)K(\sum_{q=1}^{H} F_{[q]}(\epsilon_{u[w]ij}) - H\tau_{w}) + (K - 1)(\sum_{q=1}^{H} F_{[q]}(\epsilon_{u[w]ij}) - H\tau_{w}) \]
\[ + \sum_{q=1}^{H} (I(\epsilon_{p[q]}ij \leq \epsilon_{u[w]ij}) - \tau_{qw}^{**}). \]

Combining the first two terms, we obtain

\[ g_{ij}(w) = (nK - 1)H(\sum_{q=1}^{H} F_{[q]}(\epsilon_{u[w]ij}) - \tau_{w}) + \sum_{q=1}^{H} (I(\epsilon_{p[q]}ij \leq \epsilon_{u[w]ij}) - \tau_{qw}^{**}). \]

We now consider an arbitrary element \( d_{ij}(\eta), \eta = 1, \cdots, H, \) in \( d_{ij}, \)

\[ d_{ij}(\eta) \]
\[ = E[R(\epsilon_{\zeta[ij]gm}|\epsilon_{ij})] - \tau_{\eta} \]
\[ = \sum_{q=1}^{H} \sum_{d=1}^{K} \sum_{t=1}^{n} (E[I(\epsilon_{p[q]}dt \leq \epsilon_{\zeta[ij]gm}|\epsilon_{ij}] - E(I(\epsilon_{p[q]}dt \leq \epsilon_{\zeta[ij]gm}))) \]
\[ = \sum_{q=1}^{H} \sum_{d=g,t=m}^{K} \sum_{t=1}^{n} (E[I(\epsilon_{p[q]}dt \leq \epsilon_{\zeta[ij]gm}|\epsilon_{ij}] - E(I(\epsilon_{p[q]}dt \leq \epsilon_{\zeta[ij]gm}))) \]
\[ + \sum_{q=1}^{H} (E[I(\epsilon_{p[q]}gm \leq \epsilon_{\zeta[ij]gm}|\epsilon_{ij}] - E(I(\epsilon_{p[q]}gm \leq \epsilon_{\zeta[ij]gm}))) \]
\[ + \sum_{q=1}^{H} (E[I(\epsilon_{p[q]}ij \leq \epsilon_{\zeta[ij]gm}|\epsilon_{ij}] - E(I(\epsilon_{p[q]}ij \leq \epsilon_{\zeta[ij]gm}))) \].

In the above equation, the first two terms are zero since error vectors \( \epsilon_{ij} \) and \( \epsilon_{gm} \) are
from two different sets. We then write that

\[
    d_{ij}(\eta) = 0 + 0 + \sum_{q=1}^{H} \left( E[I(\epsilon_{p[q]ij} \leq \epsilon_{\zeta_0[\eta, gm]}_i)] - E(I(\epsilon_{p[q]ij} \leq \epsilon_{\zeta_0[\eta, gm]})) \right)
\]

\[
    = \sum_{q=1}^{H} \left( 1 - F_{[\eta]}(\epsilon_{p[q]ij}) - \tau_{\eta}^\ast \right), \text{ where } \eta = 1, \cdots, H.
\]

We now combine \( G_{ij} \) and \( D_{ij} \) to write

\[
    \frac{1}{\sqrt{N}} E(S_n(\beta_0) | \epsilon_{ij}) = \frac{1}{N^2} \sum_{g=1}^{K} \sum_{m=1}^{n} X_{gm}^\top E(h_R(\epsilon_{gm}) | \epsilon_{ij})
\]

\[
    = \frac{1}{N^2} (G_{ij} + \sum_{\{g=i,m=j\}^c} D_{gm})
\]

\[
    = \frac{1}{N^2} (G_{ij} - D_{ij} + \sum_{g=1}^{K} \sum_{m=1}^{n} D_{gm})
\]

\[
    = \frac{1}{N^2} X_{ij}^\top g_{ij} - \frac{1}{N^2} X_{ij}^\top d_{ij} + \frac{1}{N^2} \sum_{g=1}^{K} \sum_{m=1}^{n} X_{gm}^\top d_{ij}
\]

\[
    = \frac{1}{N^2} X_{ij}^\top (g_{ij} - d_{ij}) + \frac{1}{N^2} d_{ij} \sum_{g=1}^{K} \sum_{m=1}^{n} X_{gm}^\top.
\]

It is important to note that the design matrix is centered so that we can write

\[
    \sum_{g=1}^{K} \sum_{m=1}^{n} X_{gm}^\top = 0.
\]

Using this property, the conditional expectation of \( \frac{1}{\sqrt{N}} S_n(\beta_0) \) given \( \epsilon_{ij} \) can be written as

\[
    \frac{1}{\sqrt{N}} E(S_n(\beta_0) | \epsilon_{ij}) = \frac{1}{N^2} X_{ij}^\top (g_{ij} - d_{ij}).
\]
Putting the lower order terms in $o\left(\frac{1}{\sqrt{N}}\right)$, we write that

$$\frac{1}{\sqrt{N}} E(S_n(\beta_0) | \epsilon_{ij}) = \frac{1}{\sqrt{N}} X_{ij}^T h_F(\epsilon_{ij}) + o\left(\frac{1}{\sqrt{N}}\right),$$

where $h_F(\epsilon_{ij})$ is given in page 47. The projection of $\frac{1}{\sqrt{N}} S_n(\beta_0)$ then reduces to

$$V_p(\beta_0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{K} \sum_{i=1}^{n} E(S(\beta_0) | \epsilon_{ij}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{K} \sum_{i=1}^{n} X_{ij}^T h_F(\epsilon_{ij})$$

$$= \frac{1}{\sqrt{N}} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} (F(\epsilon_{u[w][ij]} - \tau_w) X_{u[w][ij]}.$$

The expression for $V_p(\beta_0)$ can be written in compact notation,

$$V_p(\beta_0) = \frac{1}{\sqrt{N}} X^T A_F,$$

where $A_F = (h_F(\epsilon_{11}), \ldots, h_F(\epsilon_{Kn})).$

To finish this proof, we need to show that the difference between the variances of $\frac{1}{\sqrt{N}} S_n(\beta_0)$ and $V_p(\beta_0)$ approaches 0 as $N$ goes to infinity.

Note that $\text{Var}(h_F(\epsilon_{ij})) = \gamma$ for all $i, j$. It is then clear that $\text{Var}(A_F) = \Gamma$, where $\Gamma$ is an $N \times N$ block diagonal matrix. Each block corresponds to the covariance matrix of $h_F(\epsilon_{ij})$. The covariance matrix of the projected statistic $V_p$ is then given by

$$\text{Var}(V_p) = \text{Var}\left(\frac{1}{\sqrt{N}} X^T A_F\right) = \frac{1}{N} X^T \Gamma X = \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n} X_{ij}^T \gamma X_{ij}.$$

It is now easy to observe from Theorem 2.4.1 that

$$\lim_{N \to \infty} \text{Var}\left(\frac{1}{\sqrt{N}} S_n(\beta_0)\right) - \lim_{N \to \infty} \text{Var}(V_p(\beta_0)) = 0$$

which completes the proof. \qed

Theorem 2.4.2 establishes that $\frac{1}{\sqrt{N}} S_n(\beta_0)$ is asymptotically equivalent to $V_p$. Since $V_p$ is a sum of independent random variables, the central limit theorem for
independent (not necessarily identically distributed) random variables can be used to
derive the limiting distribution of \( V_p \).

**Theorem 2.4.3.** Under assumptions \( J1, A1, A2 \) and \( A3 \), \( V_p \) converges to a \( p \)-
dimensional multivariate normal distribution with mean 0 and covariance matrix \( \Sigma \).

**Proof.** Let

\[
W_n = t^T V_p = \frac{1}{\sqrt{N}} \sum_{i=1}^{K} \sum_{j=1}^{n} t^T X_{ij}^T h_F(\epsilon_{ij}) = \sum_{i=1}^{K} \sum_{j=1}^{n} w_{ij,n},
\]

where \( t \) is an arbitrary \( p \)-dimensional vector, with \( ||t|| = 1 \) and

\[
w_{ij,n} = \frac{1}{\sqrt{N}} t^T X_{ij}^T h_F(\epsilon_{ij}).
\]

The proof follows from the Cramer-Wold device, which is equivalent to showing that
\( W_n \) converges to a univariate normal distribution with mean 0 and variance \( t^T \Sigma t > 0 \).

We first note that \( E(W_n) = 0 \). To complete the proof, it is sufficient to show that
the conditions of the Lindeberg-Feller theorem hold:

1. \( \text{Var}(W_n) = \frac{1}{N} t^T X^T \Gamma X t \to t^T \Sigma t > 0, \) as \( N \to \infty \),

2. \( \lim_{n \to \infty} \sum_{i=1}^{K} \sum_{j=1}^{n} E\{w_{ij,n}^2 I_{\epsilon}(w_{ij,n})\} = 0, \) for any \( \epsilon > 0 \), where

\[
I_{\epsilon}(w_{ij,n}) = \begin{cases} 
1 & \text{if } |w_{ij,n}| > \epsilon \\
0 & \text{if } |w_{ij,n}| \leq \epsilon.
\end{cases}
\]

The proof of (1) follows immediately from assumption \( A2 \). For the proof of (2),
we let \( C_N = E\{\sum_{i=1}^{K} \sum_{j=1}^{n} w_{ij,n}^2 I_{\epsilon}(w_{ij,n})\} \). Then \( w_{ij,n} \) can be rewritten as

\[
w_{ij,n} = \frac{1}{\sqrt{N}} t^T X_{ij}^T h_F(\epsilon_{ij}) = \frac{1}{\sqrt{N}} t^T (\gamma^\frac{1}{2} X_{ij})^T \gamma^{-\frac{1}{2}} h_F(\epsilon_{ij}).
\]

Using the Cauchy-Schwarz inequality, we can find an upper bound for \( |w_{ij,n}| \) (and
\( w_{ij,n}^2 \) as well):

\[
\left( \frac{1}{\sqrt{N}} t^T (\gamma^\frac{1}{2} X_{ij})^T \gamma^{-\frac{1}{2}} h_F(\epsilon_{ij}) \right)^2 \leq \frac{1}{N} t^T X_{ij}^T \gamma X_{ij} th_F(\epsilon_{ij}) \gamma^{-1} h_F(\epsilon_{ij}). \quad (2.4.3)
\]
It is also easy to observe that

\[ I_{\epsilon}(w_{ij,n}) = I_{\epsilon}^2(w_{ij,n}) = \begin{cases} 1 & \text{if } w_{ij,n}^2 > \epsilon^2 \\ 0 & \text{if } w_{ij,n}^2 \leq \epsilon^2 \end{cases} \]

and

\[
I_{\epsilon}^2(w_{ij,n}) \leq I_{\epsilon}^2\left( \frac{1}{N} t^T X_{ij}^\gamma X_{ij} t h_{F}(\epsilon_{ij})\gamma^{-1} h_{F}(\epsilon_{ij}) \right) \\
\leq \frac{I_{\epsilon}^2}{m*} (h_{F}(\epsilon_{ij})\gamma^{-1} h_{F}(\epsilon_{ij})), \tag{2.4.4}
\]

where \( m^* = \max_{ij} \frac{1}{N} t^T X_{ij}^\gamma X_{ij} t \). Inequalities (2.4.3) and (2.4.4) above provide an upper-bound for \( C_N \):

\[
C_N = \mathbb{E} \sum_{i=1}^{K} \sum_{j=1}^{n} w_{ij,n}^2 I_{\epsilon}(w_{ij,n}) \\
\leq \mathbb{E} \sum_{i=1}^{K} \sum_{j=1}^{n} \frac{1}{N} t^T X_{ij}^\gamma X_{ij} t h_{F}(\epsilon_{ij})\gamma^{-1} h_{F}(\epsilon_{ij}) I_{\epsilon}^2 (h_{F}(\epsilon_{ij})\gamma^{-1} h_{F}(\epsilon_{ij})) \\
= \sum_{i=1}^{K} \sum_{j=1}^{n} \frac{1}{N} t^T X_{ij}^\gamma X_{ij} t \mathbb{E} h_{F}(\epsilon_{ij})\gamma^{-1} h_{F}(\epsilon_{ij}) I_{\epsilon}^2 (h_{F}(\epsilon_{ij})\gamma^{-1} h_{F}(\epsilon_{ij})) \\
= \sum_{i=1}^{K} \sum_{j=1}^{n} \frac{1}{N} t^T X_{ij}^\gamma X_{ij} t \mathbb{E} h_{F}(\epsilon_{11})\gamma^{-1} h_{F}(\epsilon_{11}) I_{\epsilon}^2 (h_{F}(\epsilon_{11})\gamma^{-1} h_{F}(\epsilon_{11})) \\
= \frac{1}{N} t X^T \Gamma X t \mathbb{E} h_{F}(\epsilon_{11})\gamma^{-1} h_{F}(\epsilon_{11}) I_{\epsilon}^2 (h_{F}(\epsilon_{11})\gamma^{-1} h_{F}(\epsilon_{11})).
\]

From assumption \( A2 \), we have \( \lim_{N \to \infty} \frac{1}{N} t^T X^T \Gamma X t = t^T \Sigma t > 0 \). We now observe that

\[
h_{F}(\epsilon_{11})\gamma^{-1} h_{F}(\epsilon_{11}) I_{\epsilon}^2 (h_{F}(\epsilon_{11})\gamma^{-1} h_{F}(\epsilon_{11})) \leq h_{F}(\epsilon_{11})\gamma^{-1} h_{F}(\epsilon_{11})
\]

and

\[ \mathbb{E} h_{F}^\top(\epsilon_{11})\gamma^{-1} h_{F}(\epsilon_{11}) = 1. \]
Therefore, by the dominated convergence theorem, we can interchange the expectation and limit in the following inequality

\[
\lim_{n \to \infty} C_N \leq \lim_{n \to \infty} \frac{1}{N} t^\top X^\top \Gamma X t \lim_{n \to \infty} E[h_F^\top(\epsilon_{11})\gamma^{-1}h_F(\epsilon_{11})I_{\mathcal{F}}(\gamma^{-1}h_F(\epsilon_{11}))].
\]

Since by assumption \(A_3, m^* \to 0\) as \(N \to \infty\), we then conclude that \(\lim_{N \to \infty} C_N = 0\), which completes the proof of (2) and the theorem.

From the previous two theorems, it is easy to conclude that the asymptotic distribution of the score function converges to a \(p\)-dimensional normal distribution.

**Theorem 2.4.4.** Under assumptions \(J1, A1, A2\) and \(A3\), \(\frac{1}{\sqrt{N}} S_n(\beta_0)\) converges to a \(p\)-dimensional multivariate normal distribution with mean \(0\) and covariance matrix \(\Sigma\).

### 2.5 Linearity of score function

To investigate the limiting behavior of the estimator of the parameter vector \(\beta\) and the test statistics, we need to focus on the asymptotic behavior of \(S_n(\beta)\). It is clear that \(S_n(\beta)\) is not a linear function of the parameter vector \(\beta\) and has a finite number of discontinuities with respect to the parameter \(\beta\). Without loss of generality, we assume that the true regression parameter vector is zero, \(\beta_0 = 0\). With these in mind, we wish to approximate the score function \(S_n(\beta)\) with a linear function of \(\beta\). To achieve this goal, we reparametrize our linear model as follows:

\[
Y = X\beta + \epsilon = X(X^\top X)^{-\frac{1}{2}}(X^\top X)^{\frac{1}{2}} \beta = C\Delta + \epsilon,
\]

(2.5.1)

where

\[
C = X(X^\top X)^{-\frac{1}{2}} \quad \text{and} \quad \Delta = (X^\top X)^{\frac{1}{2}} \beta.
\]
This reparameterization yields some useful properties that will be used in the construction of a linear approximation. These properties can be listed as follows:

1) \( C^T C = I_p \)

2) \( \Delta = O(1) \), assuming that \( \sqrt{N}\beta = O(1) \)

3) \( H_X = X(X^T X)^{-1}X^T = C(C^T C)^{-1}C^T = H_C \),

where \( H_C \) is the projection matrix onto the column space of \( C \). Let \( h_{u[w]ij} \) be the entries of \( H_X \). From the model (2.5.1), we can also define

\[
q_{u[w]ij} = C_{u[w]ij}^T \Delta = C_{u[w]ij,1} \Delta_1 + \cdots + C_{u[w]ij,p} \Delta_p,
\]

and

\[
q_{ij}^T = (q_{u[1]ij}, \cdots, q_{u[H]ij}),
\]

where \( C_{u[w]ij} = (C_{u[w]ij,1}, \cdots, C_{u[w]ij,p}) \) is the row vector in matrix \( C \) that corresponds to the \( u \)-th treatment, \( w \)-th judgment class, \( i \)-th set and \( j \)-th replication and \( \Delta_i \) is the \( l \)-th component of \( \Delta \). Using assumption \( A3 \) and the fact that the design matrix is centered, we can easily write that

\[
\hat{q} = \frac{1}{N} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} q_{u[w]ij} = 0,
\]

\[
\|C_{u[w]ij}\|^2 = h_{u[w]ij}^2,
\]

\[
\sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} q_{u[w]ij}^2 \leq \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} \|C_{u[w]ij}\|^2 \|\Delta\|^2 = p \|\Delta\|^2,
\]

and

\[
\max q_{u[w]ij} \leq \|\Delta\|^2 \max \|C_{u[w]ij}\|^2 = \|\Delta\|^2 \max \ h_{u[w]ij}^2 \to 0,
\]

as \( n \to \infty \).
Under the reparametrized model (2.5.1), the score function can be written as

\[
S_n^* (\Delta) = \frac{1}{N} \sum_{u=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[w]ij} \left[ R(Y_{u[w]ij} - q_{u[w]ij}) - \tau_w \right].
\]

We note that \( S_n^* (\Delta) \) is a \( p \)-dimensional vector. For \( \nu = 1, \cdots, p \), we define,

\[
S_{n,\nu}^* (\Delta) = \frac{1}{N} \sum_{u=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[w]ij,\nu} \left[ R(Y_{u[w]ij} - q_{u[w]ij}) - \tau_w \right],
\]

where \( C_{u[w]ij,\nu} \) is the \( \nu \)-th component of the vector \( C_{u[w]ij} \). Let

\[
U_{n,\nu} (\Delta) = S_{n,\nu}^* (\Delta) - S_{n,\nu}^* (0).
\]

In the following theorem, we will study the asymptotic behavior of \( U_{n,\nu} \).

**Theorem 2.5.1.** Under assumptions \( J1, A1-A5 \),

\[
U_{n,\nu} (\Delta) = -C_{\nu}^T \Omega C_{\nu} \Delta_{\nu} + o_p(1), \quad \nu = 1, \cdots, p,
\]

where \( \Delta_{\nu} \) is the \( \nu \)-th row of \( \Delta \), and \( C_{\nu} \) is the \( \nu \)-th column of \( C \).

**Proof.** The proof can be divided into two steps. In step I, we show that

\[
\lim_{n \to \infty} E[U_{n,\nu} (\Delta)] = -C_{\nu}^T \Omega C_{\nu} \Delta_{\nu},
\]

and in step II, we show that the variance of \( U_{n,\nu} (\Delta) \) converges to zero as \( n \) goes to infinity.
Proof of step I: We first look at the expected value of \( U_{n,\nu}(\Delta) \)

\[
E[U_{n,\nu}(\Delta)] = \frac{1}{N} \sum_{u=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[ij],\nu} E \left[ R(Y_{u[ij]} - q_{u[ij]}) - R(Y_{u[ij]}) \right]
\]

\[
= \sum_{u=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[ij],\nu} \int [F(y - q_{u[ij]}) - F(y)] dF_{w}(y).
\]

With a simple application of the mean value theorem, we have

\[
E[U_{n,\nu}(\Delta)] = - \sum_{u=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[ij],\nu} q_{u[ij]} \int f(y - q_{u[ij]}^{*}) dF_{w}(y)
\]

\[
= - \sum_{u=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[ij],\nu} q_{u[ij]} \int [f(y - q_{u[ij]}^{*}) - f(y) + f(y)] dF_{w}(y)
\]

\[
= - \sum_{u=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[ij],\nu} q_{u[ij]} \int f(y) dF_{w}(y)
\]

\[
- \sum_{u=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[ij],\nu} q_{u[ij]} \int [f(y - q_{u[ij]}^{*}) - f(y)] dF_{w}(y),
\]

where \( |q_{u[ij]}^{*}| \) is between 0 and \( |q_{u[ij]}| \).

Note that the maximum of \( q_{u[ij]} \) goes to zero as \( n \to \infty \). We then conclude that the second term in the equation above goes to zero as \( n \) gets large. The expected
value of $U_{n,\nu}(\Delta)$ is then asymptotically equal to

$$E[U_{n,\nu}(\Delta)]$$

$$= - \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[w]ij,\nu} q_{u[w]ij} \int f(y) dF_{[w]}(y)$$

$$= - \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[w]ij,\nu} [C_{u[w]ij,1}\Delta_{1} + \cdots + C_{u[w]ij,\nu}\Delta_{\nu}] \int f(y) dF_{[w]}(y)$$

$$= - \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[w]ij,\nu} \Delta_{\nu} \int f(y) dF_{[w]}(y)$$

$$- \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{s=1, s\neq \nu} C_{u[w]ij,\nu} \int f(y) dF_{[w]}(y) C_{u[w]ij,s} \Delta_{s}.$$ 

The matrix $C$ is orthogonal, and each treatment is applied to all ranks in each replication, so that $\sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[w]ij,\nu} C_{u[w]ij,s} = 0$ for $\nu \neq s$. The limiting value of $E[U_{n,\nu}(\Delta)]$ then reduces to

$$\lim_{n \to \infty} E[U_{n,\nu}(\Delta)] = - \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[w]ij,\nu}^{2} \Delta_{\nu} \int f(y) dF_{[w]}(y)$$

$$= - C_{\nu}^{T} \Omega C_{\nu} \Delta_{\nu},$$

which completes the proof of step I.
Proof of step II: The variance of $U_{n,\nu}(\Delta)$, for $\nu = 1, \cdots, p$, can be written as

$$\text{Var}(U_{n,\nu}(\Delta)) = \text{Var}(S_{n,\nu}^*(\Delta) - S_{n,\nu}^*(0))$$

$$= \text{Var}(\sum_{u=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} C_{u[ij,\nu]} \left[ R(Y_{u[ij]} - q_{u[ij]}) - R(Y_{u[ij]}) \right])$$

$$= \text{Var}(\sum_{i=1}^{K} \sum_{j=1}^{n} C_{ij,\nu}^T h_R^*(\epsilon_{ij}))$$

$$= \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} \text{Cov}(C_{ij,\nu}^T h_R^*(\epsilon_{ij}), C_{d,t,\nu}^T h_R^*(\epsilon_{dt}))$$

$$= \sum_{i=1}^{K} \sum_{j=1}^{n} \text{Cov}(C_{ij,\nu}^T h_R^*(\epsilon_{ij}), C_{ij,\nu}^T h_R^*(\epsilon_{ij}))$$

$$+ \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} \sum_{\{i=d, j=t\}} \text{Cov}(C_{ij,\nu}^T h_R^*(\epsilon_{ij}), C_{d,t,\nu}^T h_R^*(\epsilon_{dt}))$$

$$\equiv A_{\nu}^* + B_{\nu}^*,$$

where

$$h_R^*(\epsilon_{ij}) = \begin{pmatrix}
\frac{R(Y_{1[ij]} - q_{1[ij]}) - R(Y_{1[ij]})}{N} \\
\vdots \\
\frac{R(Y_{H[ij]} - q_{H[ij]}) - R(Y_{H[ij]})}{N}
\end{pmatrix}$$

and $C_{ij,\nu}^T = (C_{1[ij,\nu]}, \cdots, C_{H[ij,\nu]})$.

It can be shown, with partitions similar to the ones presented in Theorem 2.4.1, that $A_{\nu}^*$ and $B_{\nu}^*$ are asymptotically equal to

$$A_{\nu}^* = \sum_{i=1}^{K} \sum_{j=1}^{n} C_{ij,\nu}^T \gamma_u(q_{ij}) C_{ij,\nu} + o(1), \quad \nu = 1, \cdots, p,$$

and $B_{\nu}^* = o(1)$, where $\gamma_u(q_{ij})$ is an $H \times H$ matrix whose entry at cell $(r, r')$ is given
by
\[
\gamma_{u,r,r'}(q_{w,[r]ij}, q_{w,[r']ij}) = \int [F(x - q_{w,[r]ij}) - F(x)] \\
[F(y - q_{w,[r']ij}) - F(y)] dF_{r,r'}(x, y).
\]

Note that the maximum of \(|q_{w,[r]ij}|\) converges to zero as \(n \to \infty\). We then conclude that matrix \(\gamma_u(q_{ij})\) converges to zero. Since \(C_{ij,\nu}^T C_{ij,\nu} = 1\), we conclude that \(A^*_\nu\) converges to zero as \(n \to \infty\), which completes the proof of step II and the theorem. \(\square\)

Let \(U_n = (U_{n,1}, \ldots, U_{n,p})\). It is clear that component-wise convergence in probability implies the convergence of a vector. Using this fact along with Theorem 2.5.1, we state the following result.

**Theorem 2.5.2.** Under assumptions \(J1, A1-A5\), for \(\epsilon > 0\) and for an arbitrary \(\Delta\)

\[
\lim_{n \to \infty} P(||S_n^*(\Delta) - (S_n^*(0) - C^T \Omega C \Delta)|| \geq \epsilon) = 0.
\] (2.5.2)

The result of Theorem 2.5.2 holds for an arbitrary, but fixed \(\Delta\). To develop the asymptotic theory of rank-based inference, we need to strengthen this result by providing a uniform convergence result on a compact set

\[
\lim_{n \to \infty} P(\sup_{||\Delta|| \leq c} ||S_n^*(\Delta) - (S_n^*(0) - C^T \Omega C \Delta)|| \geq \epsilon) = 0,
\] (2.5.3)

where \(c > 0\). We note that the dispersion function for the reparametrized model (2.5.1) is given by

\[
D_n(\Delta) = \frac{1}{N} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} (Y_{u[w]ij} - C_{u[w]ijj}^T \Delta) \left[ R(Y_{u[w]ij} - C_{u[w]ijj}^T \Delta) - \tau_w \right].
\]

We now wish to establish a quadratic approximation to the dispersion function \(D_n(\Delta)\).
Let
\[ Q_n(\Delta) = D_n(0) - \Delta^\top S_n^*(0) + \frac{1}{2}\Delta^\top C^\top \Omega C \Delta \]
be a quadratic function in $\Delta$. This quadratic approximation can be anticipated from the integration of the asymptotic expression of $S_n^*(\Delta)$ with respect to $\Delta$ in equation (2.5.3).

We then need to establish a uniform convergence between $D_n(\Delta)$ and $Q_n(\Delta)$ in a compact set,

\[ \lim_{n \to \infty} P\left( \sup_{||\Delta|| \leq c} |D_n(\Delta) - Q_n(\Delta)| \geq \epsilon \right) = 0. \] (2.5.4)

**Theorem 2.5.3.** Under assumptions $J1$, $A1$ to $A5$, the equations (2.5.2), (2.5.3) and (2.5.4) are equivalent.

**Proof.** Proof of the theorem relies only on the convexity of $D_n(\Delta)$ and $Q_n(\Delta)$, and is given in theorem A.3.7 of Hettmansperger and McKean (2011, page 464) in the context of a simple random sampling design. \qed

The quadratic approximation $Q_n(\Delta)$ can be written in a general form when
\[ \Delta_0 = (X^\top X)^{1/2} \beta_0 \neq 0, \]
\[ Q_n(\Delta) = D_n(\Delta_0) - (\Delta - \Delta_0)^\top S_n^*(\Delta_0) + \frac{1}{2}(\Delta - \Delta_0)^\top C^\top \Omega C(\Delta - \Delta_0). \]

We now state $Q_n(\Delta)$, expressions (2.5.3) and (2.5.4) in terms of the original model.

**Corollary 2.5.4.** Under assumptions $J1$, and $A1$- $A5$, for $\epsilon > 0$ and for any $c > 0$,

\[ \lim_{n \to \infty} P\left( \sup_{||\beta - \beta_0|| \leq \frac{\epsilon}{\sqrt{n}}} \frac{1}{\sqrt{n}} \|S_n(\beta) - (S_n(\beta_0) - X^\top \Omega X(\beta - \beta_0))\| \geq \epsilon \right) = 0, \] (2.5.5)
\[
\lim_{n \to \infty} P\left( \sup_{\|\beta - \beta_0\| \leq \epsilon \sqrt{n}} |D_n(\beta) - Q_n(\beta)| \geq \epsilon \right) = 0, 
\] (2.5.6)

and

\[
Q_n(\beta) = D_n(\beta_0) - (\beta - \beta_0)^\top S_n(\beta_0) + \frac{1}{2}(\beta - \beta_0)^\top X^\top \Omega X (\beta - \beta_0). 
\] (2.5.7)

The expression in equation (2.5.7) provides a quadratic approximation for \(D_n(\beta)\). The convexity of \(Q_n(\beta)\) ensures that there exists a minimizer for \(Q_n(\beta)\). Let \(\tilde{\beta}\) be a value that minimizes this quadratic function:

\[
\tilde{\beta} = \text{argmin } Q_n(\beta).
\]

An equivalent expression for \(\tilde{\beta}\) can also be constructed from the estimating equation:

\[
\frac{\partial}{\partial \beta} Q_n(\beta)|_{\beta = \tilde{\beta}} = -S_n(\beta_0)\frac{1}{\sqrt{N}} + \frac{1}{N} X^\top \Omega X (\tilde{\beta} - \beta_0).
\]

The solution of the above equation for \(\tilde{\beta}\) yields

\[
\tilde{\beta} = \left( \frac{1}{N} X^\top \Omega X \right)^{-1} \frac{1}{\sqrt{N}} S_n(\beta_0) + \beta_0. 
\] (2.5.8)

The representation in equation (2.5.8) provides an iterative algorithm to compute \(\tilde{\beta}\). Using initial values for \(\Omega\) and \(\beta_0\), the estimator \(\tilde{\beta}\) can be iteratively updated through equation (2.5.8). This process is explained in detail in Chapter 4. Note that \(\sqrt{N}(\tilde{\beta} - \beta_0)\) is a linear function of \(\frac{1}{\sqrt{N}} S_n(\beta_0)\). Thus, the asymptotic distribution of \(\sqrt{N}(\tilde{\beta} - \beta_0)\) immediately follows from Theorem 2.4.4.

**Theorem 2.5.5.** Under model (2.1.1) and assumptions J1, A1-A5,
\[ \sqrt{N}(\beta - \beta_0) \] converges to a \( p \)-dimensional multivariate normal distribution with mean \( 0 \) and covariance matrix \( \Psi^{-1} \Sigma \Psi^{-1} \), i.e \( \sqrt{N}(\beta - \beta_0) \xrightarrow{D} N_p(0, \Psi^{-1} \Sigma \Psi^{-1}) \), where \( \Psi = \lim_{N \to \infty} \frac{1}{N} X^\top \Omega X \) and \( \Sigma = \lim_{N \to \infty} \frac{1}{N} X^\top \Gamma X \).

Our estimator, \( \hat{\beta} \), minimizes \( D_n(\beta) \). To establish its asymptotic distribution, we use the asymptotic properties of \( \tilde{\beta} \). The quadratic function \( Q_n(\beta) \) provides a local approximation to \( D_n(\beta) \) uniformly on a compact set; see the expression (2.5.6). The next theorem shows that the minimizers of \( D_n(\beta) \) and \( Q_n(\beta) \) are also asymptotically equivalent.

**Theorem 2.5.6.** Under model (2.1.1) and assumptions \( J1, A1-A5 \),

\[ \sqrt{N}(\tilde{\beta} - \hat{\beta}) \xrightarrow{P} 0. \]

**Proof.** See Hettmansperger and McKean (2011, page 183).

Combining this result with Theorem 2.5.5, we establish the asymptotic distribution of \( \hat{\beta} \).

**Corollary 2.5.7.** Under model (2.1.1) and assumptions \( J1, A1-A5 \),

\[ \sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{D} N_p(0, \Psi^{-1} \Sigma \Psi^{-1}). \]

It is obvious that the limiting distribution of \( \hat{\beta} \) depends on the quality of ranking information as well as on the underlying distribution. Under assumption \( J1 \), the quantities \( \Psi \) and \( \Sigma \) are given by

\[ \Psi = \lim_{N \to \infty} \frac{1}{N} X^\top \Omega X \quad \text{and} \quad \Sigma = \lim_{N \to \infty} \frac{1}{N} X^\top \Gamma X, \]

where \( \Gamma \) and \( \Omega \) are defined in assumptions \( A2 \) and \( A5 \), respectively. Note that \( \Gamma \) is
an $N \times N$ block diagonal matrix. Each block $\gamma$ corresponds to the covariance matrix of $h_F(\epsilon_{ij})$. Under assumption $J1$, the entry in the $w$-th row and $q$-th column of $\gamma$ is given by

$$
\gamma_{wq} = \int (F(x) - \tau_w)(F(y) - \tau_q)dF_{[w,q]}(x, y) \\
= \int F(x)F(y)dF_{[w,q]}(x, y) - \tau_w\tau_q \\
= C - \tau_w\tau_q,
$$

where $\tau_w = \int F(y)dF_{[w]}(y)$. Under a perfect ranking assumption, the expression $\tau_w$ reduces to $\tau_w = \frac{w}{H+1}$. To find an explicit expression for $C$, without loss of generality we assume that $q \geq w > 0$. Using the joint distribution of the $w$-th and $q$-th order statistics in a set of size $H$, the expression for $C$ can be reduced to a simple form

$$
C = \iiint_{x<y} F(x)F(y)dF_{(w,q)}(x, y) \\
= \iiint_{x<y} F(x)F(y) \frac{H!}{(w-1)!(q-w-1)!(H-q)!} f(x)f(y) \\
\times F^{w-1}(x)(F(y) - F(x))^{q-w-1}(1 - F(y))^{H-q}dxdy \\
= \frac{H!}{(w-1)!(q-w-1)!(H-q)!} \iiint_{x<y} F^{w}(x)(F(y) - F(x))^{q-w-1} \\
\times F(y)(1 - F(y))^{H-q}dF(x)dF(y).
$$
Let $F(x) = u$, $F(y) = v$. We can write $C$ as

$$C = \frac{H!}{(w-1)!(q-w-1)!(H-q)!} \int_0^1 \int_0^v u^w(v-u)^{q-w-1}v(1-v)^{H-q}dudv$$

$$= \frac{H!}{(w-1)!(q-w-1)!(H-q)!} \int_0^1 \int_0^v (1-v)^{H-q} \{ \int_0^v u^w(v-u)^{q-w-1}du \}dv$$

$$= \frac{H!}{(w-1)!(q-w-1)!(H-q)!} \int_0^1 (1-v)^{H-q} \{ \int_0^v u^w(v-u)^{q-w-1}du \}dv$$

$$= \frac{H!}{(w-1)!(q-w-1)!(H-q)!} \int_0^1 (1-v)^{H-q} \ D \ dv,$$

where

$$D = \int_0^v u^w(v-u)^{q-w-1}du = v^q \int_0^1 z^w(1-z)^{q-w-1}dz = v^q \frac{\Gamma(w+1)\Gamma(q-w)}{\Gamma(q+1)}.$$

We insert $D$ in the above equation to get

$$C = \frac{(q+1)w}{(H+2)(H+1)}.$$ By inserting this in equation (2.5.9), we obtain $\gamma_{wq} = \frac{w(H-q+1)}{(H+1)^2(H+2)}$. In general, we can write

$$\gamma_{wq} = \begin{cases} \frac{w(H-q+1)}{(H+1)^2(H+2)} & \text{if } q \geq w > 0 \\ \frac{q(H-w+1)}{(H+1)^2(H+2)} & \text{if } w > q > 0. \end{cases}$$

Thus, under perfect ranking, $\Phi$ (and $\Sigma$) has a simple form and only depends on the set size $H$.

To compare the asymptotic variance of $\hat{\beta}$, we also need to evaluate the diagonal matrix $\Omega$. The component $\omega_h$ in $\Omega$ is not distribution free. It depends on the underlying distribution $F$ and is given by $\omega_h = \int f(y)f_{(h)}(y)dy$. Even under perfect ranking, there is no closed-form expression for $\omega_h$ for many distributions. On the
Table 2.1: Asymptotic relative efficiencies (ARE) of rank regression and least square estimators in ORRD and CRD designs, $L$ is the number of treatment levels.

<table>
<thead>
<tr>
<th>$L$</th>
<th>$(D_1, CRD)$</th>
<th>$(D_1, LS)$</th>
<th>$(D_2, CRD)$</th>
<th>$(D_2, D_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.000</td>
<td>2.865</td>
<td>3.000</td>
<td>1.000</td>
</tr>
<tr>
<td>3</td>
<td>2.000</td>
<td>1.910</td>
<td>4.000</td>
<td>2.000</td>
</tr>
<tr>
<td>4</td>
<td>1.800</td>
<td>1.719</td>
<td>5.250</td>
<td>2.917</td>
</tr>
<tr>
<td>5</td>
<td>1.714</td>
<td>1.637</td>
<td>6.708</td>
<td>3.914</td>
</tr>
</tbody>
</table>

other hand, a numerical value for $\omega_h$ can be obtained using a numerical integration procedure.

### 2.6 Efficiency Comparison of ORRD to CRD

In this section we look at the efficiency of our proposed estimator. Let $\hat{\beta}_{D_1}$, $\hat{\beta}_{D_2}$, $\hat{\beta}_{CRD}$, and $\hat{\beta}_{LS}$ be the rank regression estimators based on Design 1, Design 2, CRD and the least squares regression estimator of CRD, respectively.

The asymptotic variances of ORRD estimators (Design 1 and Design 2) are given in Corollary 2.5.7,

$$\text{Var}(\sqrt{N}(\hat{\beta}_{ORR} - \beta_0)) = \Psi^{-1}\Sigma\Psi^{-1}. $$

The asymptotic variances of $\sqrt{N}\hat{\beta}_{CRD}$ and $\sqrt{N}\hat{\beta}_{LS}$ are given in Hettmansperger and McKean (2011, page 183)

$$\text{Var}(\sqrt{N}\hat{\beta}_{CRD}) = \frac{1}{12(f(y)d(dy))^2}\left(\frac{1}{N}X^TX\right)^{-1} \text{ and Var}(\sqrt{N}\hat{\beta}_{LS}) = \sigma^2\left(\frac{1}{N}X^TX\right)^{-1}. $$

Since $\beta$ is a $p$-dimensional vector, we use the generalized variance to explore
the asymptotic relative efficiency of the proposed estimators. Asymptotic relative
efficiency between any two estimators, say $\hat{\beta}_{D_1}$ and $\hat{\beta}_{D_2}$ is given by

$$\text{ARE}(D_1, D_2) = \frac{[\det\{\text{Var}(\sqrt{N} \hat{\beta}_{D_2})]\}]^{\frac{1}{p}}}{[\det\{\text{Var}(\sqrt{N} \hat{\beta}_{D_1})]\}]^{\frac{1}{p}},$$

where $\det(A)$ is the determinant of the matrix $A$.

Table 2.1 presents the asymptotic relative efficiencies of ORRD rank regression
estimators compared to CRD rank regression and least square (LS) regression estimators. The efficiency values are computed for a normal distribution under the perfect
ranking assumption. It is clear from Table 2.1 that estimators based on ORRDs
yield higher efficiency than the efficiency of the estimators based on CRD. Within
the ORRDs, the Design 2 is more efficient than the Design 1 when $L > 2$.

Chapter 1 provides construction methods for GRBDs in the framework of Design
1 and Design 2. Since this construction preserves the same within-set correlation
structure as in ORRD, the results in Chapter 2, with a modified design matrix,
apply to the GRBD directly. Let $X_B$ be the design matrix in a GRBD design. The
asymptotic variance of the rank regression estimator of $\hat{\beta}_{GRBD}$ follows from Theorem
2.5.5:

$$\text{Var}(\sqrt{N}(\hat{\beta}_{GRBD} - \beta_0)) = \Psi_B^{-1} \Sigma_B \Psi_B^{-1}$$

(2.6.1)

$$= \left(\frac{1}{N}X_B^T \Omega X_B\right)^{-1}\left(\frac{1}{N}X_B^T \Gamma X_B\right)\left(\frac{1}{N}X_B^T \Omega X_B\right)^{-1}.$$  

The asymptotic variance of the estimator of the rank regression parameters in
ORRDs and GRBDs can easily be obtained from Theorem 2.5.5 and equation (2.6.1).
For Design 1, with set size \( H = 2 \) and the number of treatments \( L = 3 \), the asymptotic variances of \( \hat{\beta}_{D1} \) and \( \hat{\beta}_{GRBD1} \) are

\[
\text{Var}(\sqrt{N}(\hat{\beta}_{D1} - \beta_0)) = \begin{pmatrix}
1.0471976 & -0.5235988 \\
-0.5235988 & 1.0471976
\end{pmatrix},
\]

and

\[
\text{Var}(\sqrt{N}(\hat{\beta}_{GRBD1} - \beta_0)) = \begin{pmatrix}
1.3962634 & -0.6981317 \\
-0.6981317 & 1.3962634
\end{pmatrix},
\]

where \( \hat{\beta}_{GRBD1} \) is the regression estimator of the GRBD design in the framework of Design 1. The comparison of these two covariance matrices indicates that Design 1 in ORRD is 1.33 times more efficient than the corresponding GRBD design.

The asymptotic variances for Design 2, with \( H = 3, \ L = 3 \) are given by

\[
\text{Var}(\sqrt{N}(\hat{\beta}_{D2} - \beta_0)) = \begin{pmatrix}
0.5235988 & -0.2617994 \\
-0.2617994 & 0.5235988
\end{pmatrix},
\]

and

\[
\text{Var}(\sqrt{N}(\hat{\beta}_{GRBD2} - \beta_0)) = \begin{pmatrix}
1.0471976 & -0.5235988 \\
-0.5235988 & 1.0471976
\end{pmatrix},
\]

where \( \hat{\beta}_{GRBD2} \) is the rank regression estimator of \( \beta \) based on the GRBD design in the framework of Design 2. Again the efficiency comparison indicates that the estimator for ORRD Design 2 is two times more efficient than the corresponding GRBD estimator.
CHAPTER 3
TEST PROCEDURES

In this Chapter, we construct testing procedures for a general linear hypothesis in model (2.1.2). We partition the $p \times 1$ dimensional regression parameter $\beta$ as $\beta^\top = (\beta_1^\top, \beta_2^\top)$, where $\beta_1^\top$ and $\beta_2^\top$ are $p-q$ and $q$ dimensional vectors, respectively. The linear model (2.1.2) then can be rewritten as

$$Y = \alpha 1_N + X_1 \beta_1 + X_2 \beta_2 + \epsilon,$$

(3.0.1)

where $X_1$ and $X_2$ are $N \times (p-q)$ and $N \times q$ dimensional matrices with $X = (X_1, X_2)$. In this model, our objective is to construct a test for the following hypotheses:

$$H_0 : \beta_2 = 0, \text{ with } \beta_1 \text{ unspecified, against } H_A : \beta_2 \neq 0, \text{ with } \beta_1 \text{ unspecified.}$$

(3.0.2)

In these hypotheses, the parameter vector $\beta_1$ plays the role of a nuisance parameter. The framework of this hypothesis is very general. It covers tests for main and interaction effects in ANOVA models, tests for covariance analysis, and tests for coefficients of regression parameters in regression models. We consider three tests: drop, score and Wald tests for the hypothesis (3.0.2).
3.1 Drop Test

The dispersion function $D_n(\cdot)$ is a convex function of $\beta$. Our estimator $\hat{\beta}$ minimizes this convex function. We can interpret the magnitude of $D_n(\hat{\beta})$ as a measure of the goodness-of-fit of the linear model to the ORRD data. In fact, $D_n(\hat{\beta})$ measures the minimum distance from the data vector to the subspace determined by model (3.0.1). This distance is defined by $D_n(\cdot)$. Thus the function $D_n(\cdot)$ would be a natural test statistic to use in a test for the hypothesis (3.0.2).

Let $\beta_r^T = (\beta_1^T, 0^T)$ be the parameter vector under the reduced model. Then the estimator $\hat{\beta}_r^T = (\hat{\beta}_1^T, 0^T)$ minimizes $\beta_r$ under the restriction that $\beta_2 = 0$. The dispersion function $D_n(\hat{\beta}_r)$ measures the goodness-of-fit of the reduced model to the data and it is always larger than $D_n(\hat{\beta})$. A natural test would reject the null hypothesis $H_0$ if the difference between the fits of the reduced and full models is too large

$$RD = D_n(\hat{\beta}_r) - D_n(\hat{\beta}),$$

where the statistic $RD$ is a goodness-of-fit statistic for the reduced model.

This test is similar to the generalized F-test based on extra sum of square residuals in a least squares analysis, where the null hypothesis is rejected for large values of

$$FL = \frac{\text{SSE}(\text{reduced}) - \text{SSE}(\text{full})}{\text{MSE}(\text{full})}.$$ 

In the above equation, SSE and MSE denote the sum of squared errors and mean squared errors, respectively.

The quantity $RD$, similar to $FL$, measures the amount of reduction in the dispersion of $D_n(\cdot)$ going from the reduced model to the full model. A large drop indicates that the reduced model is not capable of explaining the data adequately. Hence, the test should reject the null hypothesis for sufficiently large values of $RD$. 

70
The construction of the rejection region of the test statistic $RD$ requires one to establish its distributional properties. For finite sample sizes, this will depend on the underlying distribution and the judgement ranking process. Asymptotic results in Chapter 2 suggest that the limiting distribution of $RD$ may be distribution free.

To establish the limiting distribution of $RD$, we use a quadratic approximation. The uniform convergence result in Corollary 2.5.4 along with Theorem 2.5.6 yields that

$$D_n(\hat{\beta}_r) - D_n(\tilde{\beta}) = Q_n(\hat{\beta}_r) - Q_n(\tilde{\beta}) + o_p(1).$$

The proof of this expression relies only on the convexity of $D_n(\cdot)$, $Q_n(\cdot)$ and the uniform convergence result in Corollary 2.5.4. Since the proof is given in Hettmansperger and McKean (2011 page 191-194) in the context of a simple random sampling design, it is not provided here.

Let $\Sigma_{i,j}$ and $\Psi_{i,j}$, for $i, j = 1, 2$, be the partitions of the matrices $\Sigma$ and $\Psi$ in assumptions $A2$ and $A5$ based on the partition of the parameter vector $\beta^\top = (\beta_1^\top, \beta_2^\top)$,

$$\Psi = \lim_{N \to \infty} \frac{1}{N} X^\top \Omega X = \begin{bmatrix} \Psi_{1,1} & \Psi_{1,2} \\ \Psi_{2,1} & \Psi_{2,2} \end{bmatrix}$$

and

$$\Sigma = \lim_{N \to \infty} \frac{1}{N} X^\top \Gamma X = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix},$$

where $\Psi_{1,1}$ and $\Sigma_{1,1}$ are $(p-q) \times (p-q)$ matrices; $\Psi_{2,2}$ and $\Sigma_{2,2}$ are $q \times q$ matrices; and $\Psi_{2,1} = \Psi_{1,2}$, $\Sigma_{2,1} = \Sigma_{1,2}$. We further define

$$W^{-1} = \Psi_{2,2} - \Psi_{2,1} \Psi_{1,1}^{-1} \Psi_{1,2}$$

71
and
\[ V = \Psi_{2,1} \Psi_{1,1}^{-1} \Sigma_{1,1} \Psi_{1,2}^{-1} \Psi_{1,2} - 2 \Psi_{2,1} \Psi_{1,1}^{-1} \Sigma_{1,2} + \Sigma_{2,2}. \]

**Theorem 3.1.1.** Under model (2.1.1), and assumptions J1 and A1-A5, if \( \beta_{r0}^T = (\beta_{10}^T, 0^T) \) is the true value of \( \beta_r \), we have
\[
2[D_n(\beta_r) - D_n(\hat{\beta})] \overset{D}{\to} \sum_{i=1}^{q} \lambda_i \chi_i^2(1),
\]
where \( \lambda_1, \ldots, \lambda_q \) are the \( q \) positive eigen values of the matrix \( V^{1/2} W V^{1/2} \) and \( \chi_i(1), i = 1, \ldots, q \) are independent chi-squared random variables each with 1-degree of freedom.

**Proof.** Without loss of generality, we assume that \( \beta_{10} = 0 \). From the quadratic approximation in equation (2.5.7), we have
\[
Q_n(\beta) = D_n(0) - (\sqrt{N} \beta^T) S_n(0) \frac{1}{\sqrt{N}} + \frac{1}{2} (\sqrt{N} \beta)^T \frac{1}{N} X^T \Omega X (\sqrt{N} \beta),
\]
where \( ||\beta|| < c/\sqrt{N} \). Note that \( \sqrt{N} \beta_{10} \) is bounded in probability. Hence quadratic equation \( Q_n(\beta) \) holds for \( \beta_{10} \). We first express the reduced and full model parameter estimators as functions of the score function
\[
\sqrt{N} \beta = \Psi^{-1} \frac{1}{\sqrt{N}} S_n(0) + o_p(1),
\]
and
\[
\sqrt{N} \hat{\beta}_r = \begin{pmatrix} \Psi_{1,1}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \frac{1}{\sqrt{N}} S_n(0) + o_p(1).
\]
By inserting $\sqrt{N}\tilde{\beta}_r$ in $Q_n(\tilde{\beta}_r)$, we obtain

$$Q_n(\tilde{\beta}_r) = D_n(0) - \sqrt{N}\tilde{\beta}_r^\top \frac{1}{\sqrt{N}} S_n(0) + \frac{1}{2}(\sqrt{N}\tilde{\beta}_r)\Psi^{\top} \sqrt{N}\tilde{\beta}_r$$

$$= D_n(0) - \frac{1}{\sqrt{N}} S_n^\top(0) \left( \begin{array}{cc} \Psi_{1,1}^{-1} & 0 \\ 0 & 0 \end{array} \right) \frac{1}{\sqrt{N}} S_n(0)$$

$$+ \frac{1}{2} \frac{1}{\sqrt{N}} S_n^\top(0) \left( \begin{array}{cc} \Psi_{1,1}^{-1} & 0 \\ 0 & 0 \end{array} \right) \frac{1}{\sqrt{N}} S_n(0)$$

$$= D_n(0) - \frac{1}{2} \frac{1}{\sqrt{N}} S_n^\top(0) \left( \begin{array}{cc} \Psi_{1,1}^{-1} & 0 \\ 0 & 0 \end{array} \right) \frac{1}{\sqrt{N}} S_n(0).$$

In a similar fashion, under the full model, we write $Q_n(\tilde{\beta})$ as

$$Q_n(\tilde{\beta}) = D_n(0) - \frac{1}{2} \frac{1}{\sqrt{N}} S_n^\top(0) \Psi^{-1} \frac{1}{\sqrt{N}} S_n(0).$$

By combining $Q_n(\tilde{\beta}_r)$ and $Q_n(\tilde{\beta})$, we write

$$Q_n(\tilde{\beta}_r) - Q_n(\tilde{\beta}) = \frac{1}{2} \frac{1}{\sqrt{N}} S_n^\top(0) \left\{ \Psi^{-1} - \left( \begin{array}{cc} \Psi_{1,1}^{-1} & 0 \\ 0 & 0 \end{array} \right) \right\} \frac{1}{\sqrt{N}} S_n(0).$$

Using a well-known result for the inverse of a partitioned matrix (Manahan, 2008, page 268), we have:

$$\Psi^{-1} = \begin{bmatrix} \Psi_{1,1}^{-1} & 0_{(p-q)\times q} \\ 0_{q\times(p-q)} & 0_{q\times q} \end{bmatrix} + \begin{bmatrix} -\Psi_{1,1}^{-1}\Psi_{1,2} \\ -\Psi_{2,1}\Psi_{1,1}^{-1} I_q \end{bmatrix} W \begin{bmatrix} -\Psi_{2,1}\Psi_{1,1}^{-1} I_q \\ I_q \end{bmatrix},$$

where $I_q$ is $q \times q$ identity matrix and $W^{-1} = \Psi_{2,2} - \Psi_{2,1}\Psi_{1,1}^{-1}\Psi_{1,2}$. Using this matrix identity, we rewrite $Q_n(\tilde{\beta}_r) - Q_n(\tilde{\beta})$ as

$$Q_n(\tilde{\beta}_r) - Q_n(\tilde{\beta}) = \frac{1}{2} \frac{1}{\sqrt{N}} S_n(0)^\top \left[ \begin{array}{c} -\Psi_{1,1}^{-1}\Psi_{1,2} \\ I_q \end{array} \right] W \left[ \begin{array}{c} -\Psi_{2,1}\Psi_{1,1}^{-1} I_q \\ I_q \end{array} \right] \frac{1}{\sqrt{N}} S_n(0).$$
Let $T_n = \frac{1}{\sqrt{N}} \left[ -\Psi_{2,1} \Psi^{-1}_{1,1} I_q \right] S_n(0)$. From Theorem 2.4.4, we can establish that $T_n$ converges to a $q-$dimensional normal distribution with mean $0$ and variance $V$, where

$$V = \begin{bmatrix} -\Psi_{2,1} \Psi^{-1}_{1,1} I_q \end{bmatrix} \Sigma \begin{bmatrix} -\Psi^{-1}_{1,1} \Psi_{1,2} \\ I_q \end{bmatrix} = \Psi_{2,1} \Psi^{-1}_{1,1} \Sigma_{1,1} \Psi^{-1}_{1,1} \Psi_{1,2} - 2 \Psi_{2,1} \Psi^{-1}_{1,1} \Sigma_{1,2} + \Sigma_{2,2},$$

and $V$ is a $q \times q$ positive definite matrix. Let $Z^* = V^{-\frac{1}{2}} T_n$. Then the difference $Q_n(\tilde{\beta}_r) - Q_n(\tilde{\beta})$ can be written as

$$Q_n(\tilde{\beta}_r) - Q_n(\tilde{\beta}) = \frac{1}{2} Z^* \left( V^{\frac{1}{2}} W V^{\frac{1}{2}} \right) Z^*.$$

Note that $V^{\frac{1}{2}} W V^{\frac{1}{2}}$ is a positive definite matrix. Then there exists a $p \times p$ orthogonal matrix $\xi$ such that

$$V^{\frac{1}{2}} W V^{\frac{1}{2}} = \xi D \xi^T,$$

where $D$ is a $p \times p$ diagonal matrix of the $q$ positive eigenvalues of the matrix $V^{\frac{1}{2}} W V^{\frac{1}{2}}$

$$D = \text{diag}(0, \ldots, 0, \lambda_{p-q+1}, \ldots, \lambda_p).$$

With this eigenvalue representation, the expression $Q_n(\tilde{\beta}_r) - Q_n(\tilde{\beta})$ reduces to

$$Q_n(\tilde{\beta}_r) - Q_n(\tilde{\beta}) = \frac{1}{2} Z^* \left( V^{\frac{1}{2}} W V^{\frac{1}{2}} \right) Z^* = \frac{1}{2} Z^* \left( \xi D \xi^T \right) Z^* = \frac{1}{2} Z D Z^T,$$

where $Z = \xi^T Z^* \overset{D}{\rightarrow} N_p(0, I)$. This completes the proof of the theorem

$$2[Q_n(\tilde{\beta}_r) - Q_n(\tilde{\beta})] \overset{D}{\rightarrow} \sum_{i=1}^{q} \lambda_i \chi_i^2(1).$$

The asymptotic null distribution of $RD$ is in the form of a linear combination of
independent chi-squared random variables. The coefficient of this linear combination depends on the design matrix through eigenvalues of $V^{\frac{1}{2}} W V^{\frac{1}{2}}$. To compute these eigenvalues, we need to determine $\Omega$ and $\Gamma$. As we indicated in Chapter 2, values of these matrices may depend on the underlying distribution and judgment ranking process. The estimation of these matrices from the data will be discussed in Chapter 4. Once we estimate $\Omega$ and $\Gamma$, we can either bootstrap or simulate the null distribution of $RD$, and construct the critical values of the test.

3.2 Score Test

The score test is constructed based on the gradient function of $D_n(\beta)$, which is the score function $S_n(\beta)$, with some adjustment to reduce the impact of the unspecified nuisance parameter $\beta_1$ in hypotheses (3.0.2). To achieve this, we find the estimate of $\beta_1$ under the reduced model $H_0$, denoted by $\hat{\beta}_r^T = (\hat{\beta}_1^T, 0^T)$. From the reduced model residuals, $Y - X_1^T \hat{\beta}_{10}$, we compute the $q \times 1$ dimensional centered rank statistics $S_{2n}(\hat{\beta}_{10}) = S_{2n}(Y - X_1^T \hat{\beta}_{10})$, where $S_{2n}(\hat{\beta}_{10})$ is the score function that corresponds to parameter vector $\beta_2$. This centered rank statistic is called the vector of aligned rank statistics. The original data is adjusted (aligned) by removing the effect of nuisance parameter $\beta_1$ before constructing the test statistic for the null hypothesis, $H_0 : \beta_2 = 0$, with unspecified $\beta_1$. Because the effect of nuisance parameter $\beta_1$ is removed, if the null hypothesis, $H_0 : \beta_2 = 0$ is true, we expect that $S_{2n}(Y - X_1^T \hat{\beta}_{10}) = 0$. A sensible test then should assess the magnitude of $S_{2n}(Y - X_1^T \hat{\beta}_{10})$. Since $S_{2n}(\cdot)$ is a $q$-dimensional vector, we form a quadratic function to measure its magnitude

$$T_n^* = (\frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}_r))^\top M_n^{-1}(\frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}_r)),$$

75
where

\[ M_s = \left[ -\Psi_{2,1} \Psi_{1,1}^{-1} \ I_q \right] \Sigma \left[ -\Psi_{2,1} \Psi_{1,1}^{-1} \ I_q \right]^T \]

is used for an appropriate standardization of the test statistic \( T^*_n \). The test rejects the null hypothesis \( H_0 : \beta_2 = 0 \) in favor of \( H_A : \beta_2 \neq 0 \) for larger values of \( T^*_n \). The finite sample null distribution of \( T^*_n \) may depend on the underlying distribution \( F \) and the judgment ranking process. The next theorem, for large values of \( n \), shows that the limiting distribution of \( T^*_n \) is a central chi-squared distribution with \( q \)-degrees of freedom.

**Theorem 3.2.1.** Under model (2.1.1) and assumptions J1 and A1-A5, when \( \beta_2 = 0 \), the test statistic \( T^*_n \) converges to a chi-squared distribution with \( q \)-degrees of freedom.

\[ T^*_n \overset{D}{\rightarrow} \chi^2(q). \]

**Proof.** The score function can be partitioned in accordance with the partition of the parameter vector \( \beta \)

\[ S_n(\beta) = \left\{ \begin{array}{l} S_{1n}(\beta) \\ S_{2n}(\beta) \end{array} \right\}, \]

where \( S_{1n} \) and \( S_{2n} \) are the \( p-q \) and \( q \) dimensional score functions that correspond to the parameter vectors \( \beta_1 \) and \( \beta_2 \). From the uniform convergence result in equation (2.5.5), for \( \sqrt{N}||\beta - \beta_0|| \leq c \) we have

\[ \frac{1}{\sqrt{N}} S_n(\beta) = \frac{1}{\sqrt{N}} S_n(\beta_0) - \Psi \sqrt{N} (\beta - \beta_0) + o_p(1). \]

We insert the rank regression estimator \( \hat{\beta} \) in the above equation, and write

\[ \frac{1}{\sqrt{N}} S_n(\hat{\beta}) = \frac{1}{\sqrt{N}} S_n(\beta_0) - \Psi \sqrt{N} (\hat{\beta} - \beta_0) + o_p(1). \]

The above equation can be written in terms of the partitioned parameter vector \( \beta \) as
follows
\[
\frac{1}{\sqrt{N}} \begin{pmatrix} S_{1n}(\hat{\beta}) \\ S_{2n}(\hat{\beta}) \end{pmatrix} = \frac{1}{\sqrt{N}} \begin{pmatrix} S_{1n}(\beta_0) \\ S_{2n}(\beta_0) \end{pmatrix} - \begin{bmatrix} \Psi_{1,1} & \Psi_{1,2} \\ \Psi_{2,1} & \Psi_{2,2} \end{bmatrix} \begin{pmatrix} \hat{\beta}_1 - \beta_{10} \\ \hat{\beta}_2 - \beta_{20} \end{pmatrix} \sqrt{N} + o_p(1).
\]

From this partitioned score function, it is easy to observe that
\[
\frac{1}{\sqrt{N}} S_{1n}(\hat{\beta}) = \frac{1}{\sqrt{N}} S_{1n}(\beta_0) - \Psi_{1,1} \sqrt{N}(\hat{\beta}_1 - \beta_{10}) - \Psi_{1,2} \sqrt{N}(\hat{\beta}_2 - \beta_{20}) + o_p(1), \tag{3.2.1}
\]
and
\[
\frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}) = \frac{1}{\sqrt{N}} S_{2n}(\beta_0) - \Psi_{2,1} \sqrt{N}(\hat{\beta}_1 - \beta_{10}) - \Psi_{2,2} \sqrt{N}(\hat{\beta}_2 - \beta_{20}) + o_p(1). \tag{3.2.2}
\]

Under the null hypothesis \( \beta_2 = 0 \), the reduced model parameter estimate is \( \hat{\beta}_r^\top = (\hat{\beta}_1^\top, 0^\top) \). Since \( \hat{\beta}_r \) solves the estimating equation \( S_{1n}(\hat{\beta}_r) = 0 \), \( S_n(\hat{\beta}_r) \) reduces to
\[
S_n(\hat{\beta}_r) = \begin{cases} S_{1n}(\hat{\beta}_r) \\ S_{2n}(\hat{\beta}_r) \end{cases} = \begin{cases} 0 \\ S_{2n}(\hat{\beta}_r) \end{cases}.
\]

Under the reduced model, equation (3.2.1) simplified to
\[
\frac{1}{\sqrt{N}} S_{1n}(\beta_0) - \Psi_{1,1} \sqrt{N}(\hat{\beta}_1 - \beta_{10}) + o_p(1) = 0.
\]

Solving this expression for \( \sqrt{N}(\hat{\beta}_1 - \beta_{10}) \), we obtain
\[
\sqrt{N}(\hat{\beta}_1 - \beta_{10}) = \Psi_{1,1}^{-1} \frac{1}{\sqrt{N}} S_{1n}(\beta_0) + o_p(1). \tag{3.2.3}
\]

Under the reduced model, equation (3.2.2) can be written as
\[
\frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}_r) = \frac{1}{\sqrt{N}} S_{2n}(\beta_0) - \Psi_{2,1} \sqrt{N}(\hat{\beta}_1 - \beta_{10}) + o_p(1).
\]
We insert equation (3.2.3) in the above expression, to obtain

\[
\frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}_r) = \frac{1}{\sqrt{N}} S_{2n}(\beta_0) - \Psi_{2,1} \Psi_{1,1}^{-1} \frac{1}{\sqrt{N}} S_1(\beta_0) + o_p(1).
\]

This expression can be written in compact notation as,

\[
\frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}_r) = \left[ -\Psi_{2,1} \Psi_{1,1}^{-1} I_q \right] \frac{1}{\sqrt{N}} S_n(\beta_0) + o_p(1).
\]

We have already established in Theorem 2.4.4 that \( \frac{1}{\sqrt{N}} S_n(\beta_0) \) converges to a \( p \)-dimensional multivariate normal distribution with mean 0 and covariance matrix \( \Sigma \). A multivariate version of Slutsky’s theorem then yields that

\[
\frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}_r) \xrightarrow{D} N_q(0, M_s),
\]

where \( M_s = [-\Psi_{2,1} \Psi_{1,1}^{-1} I_q] \Sigma [-\Psi_{2,1} \Psi_{1,1}^{-1} I_q] \top \). Note that \( M_s \) is a positive definite matrix. Then there exists a square root matrix \( M_s^{-1/2} \) such that \( M_s^{-1/2} M_s^{-1/2} = M_s^{-1} \). Hence, the expression \( M_s^{-1/2} \frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}_r) \) converges to a \( q \)-dimensional multivariate normal distribution with the identity variance covariance matrix. We then conclude that \( T_n^* \) converges to a chi-squared distribution with \( q \)-degrees of freedom. This completes the proof.

We again note that the construction of the test statistic \( T_n^* \) requires the computation of matrices \( \Gamma \) and \( \Omega \). In practice, these matrices must be estimated from the data to reduce the dependence on the judgment ranking model and the underlying distribution.

We now consider the asymptotic power of the score test. Assuming that the score test rejects the hypothesis \( H_0 : \beta_2 = 0 \) in favor of \( H_A : \beta_2 \neq 0 \), we then evaluate the
power of the test in the sequence of alternative hypotheses,

\[ Y = \alpha 1 + X_1 \beta_1 + X_2 \theta_{2,n} + \epsilon, \tag{3.2.4} \]

\[ \theta_{2,n} = N^{-1/2} \theta_2, \text{ and } \theta_2 \text{ is a } q \times 1 \text{ vector. Under this sequence of alternatives, we assume that} \]

\[ \lim_{N \to \infty} N^{1/2} \theta_{2,n} = \theta_2. \]

**Theorem 3.2.2.** Under model (3.2.4) and assumptions A1-A5, we have that

\[ P_{\theta_{2,n}}(T_n^* \leq y) \xrightarrow{D} (\chi^2(\eta_2) \leq y), \]

where \( \eta_2 = \theta_2^T (\Psi_{2,2} - \Psi_{2,1} \Psi_{1,1} \Psi_{1,2}) M_s^{-1}(\Psi_{1,2} - \Psi_{2,1} \Psi_{1,1} \Psi_{1,2}) \theta_2. \)

**Proof.** Given the true parameter value \( \beta_0 \), from equation (2.5.5) in Corollary 2.5.4, we can write

\[ \frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}) = \frac{1}{\sqrt{N}} S_{2n}(\beta_0) - \Psi_{2,1}(\hat{\beta}_{10} - \beta_{10}) \sqrt{N} + \Psi_{2,2} \sqrt{N} \theta_{2,n} + o_p(1). \tag{3.2.5} \]

Under the reduced model, with the true parameter vector, the estimator \( \hat{\beta}_{10} \) can be expressed asymptotically as

\[ \sqrt{N}(\hat{\beta}_{10} - \beta_{10}) = \Psi_{1,1}^{-1} \frac{1}{\sqrt{N}} S_{1n}(\beta_0) + \Psi_{1,2}^{-1} \Psi_{1,2} \theta_{2,n} \sqrt{N} + o_p(1). \]
We insert this expression in equation (3.2.5) and write

\[
\frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}) = \left[ -\Psi_{2,1} \Psi_{1,1}^{-1}, \mathbb{I} \right] \left( \begin{array}{c} S_{1n}(\beta_0) \\ S_{2n}(\beta_0) \end{array} \right) \frac{1}{\sqrt{N}}
- \Psi_{2,1} \Psi_{1,1}^{-1} \Psi_{1,2} \theta_{2,n} \sqrt{N} + \Psi_{2,2} \sqrt{N} \theta_{2,n} + o_p(1)
= \left[ -\Psi_{2,1} \Psi_{1,1}^{-1}, \mathbb{I} \right] \frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}_0)
+ (\Psi_{2,2} - \Psi_{2,1} \Psi_{1,1}^{-1} \Psi_{1,2}) \theta_{2,n} \sqrt{N} + o_p(1).
\]

Note that \( \sqrt{N} \theta_{2,n} \) has limit \( \theta_2 \) and \( \left[ -\Psi_{2,1} \Psi_{1,1}^{-1}, \mathbb{I} \right] \frac{1}{\sqrt{N}} S_{2n}(\hat{\beta}_0) \) converges to a normal distribution with mean \( \theta_2 \) and covariance matrix \( M_s \). We then conclude that

\[
\frac{1}{\sqrt{N}} M_s^{-1/2} S_{2n}(\hat{\beta}_0) \text{ converges to a normal distribution with mean } \mu,
\]

\[
\mu = M_s^{-1/2} (\Psi_{2,2} - \Psi_{2,1} \Psi_{1,1}^{-1} \Psi_{1,2}) \theta_2,
\]

and covariance matrix \( \mathbb{I} \). The quadratic expression \( T_n^* \), under the sequence of alternative hypotheses \( \theta_{2,n} \), converges to a non-central chi-squared distribution with \( q \)-degrees of freedom and non-centrality parameter \( \eta_2 \) where

\[
\eta_2 = \theta_2^T (\Psi_{2,2} - \Psi_{2,1} \Psi_{1,1} \Psi_{1,2}) M_s^{-1} (\Psi_{1,2} - \Psi_{2,1} \Psi_{1,1} \Psi_{1,2}) \theta_2.
\]

\[
\square
\]

### 3.3 Wald Test

In this section we consider hypotheses of the form

\[
H_0 : G \beta = G \beta_0 \text{ versus } H_A : G \beta \neq G \beta_0.
\]
where $G$ is a $q \times p$ matrix with full row rank. The matrix $G$ allows us to construct a hypothesis for $q$ independent linear functions of the $p$-dimensional parameter vectors. The selection of $G$ depends on the hypothesis of interest. The first step in the construction of the Wald test is to estimate the regression parameter $\beta$ with $\hat{\beta}$ under the full model and construct a quadratic expression in $G\hat{\beta}$

$$T_{n}^{**} = N(\hat{\beta} - \beta_0)^\top G^\top M_w^{-1}G(\hat{\beta} - \beta_0),$$

where $M_w = G\Psi^{-1}\Sigma\Psi^{-1}G^\top$.

If the null hypothesis is true, $G\beta = G\beta_0$, we anticipate that $G\hat{\beta} = G\beta_0$ and the test statistic $T_{n}^{**}$ tends to be smaller. On the other hand, if the null hypothesis is not true, then $T_{n}^{**}$ tends to be larger. Hence large value of $T_{n}^{**}$ lead to rejection of the null hypothesis.

**Theorem 3.3.1.** Under the null hypothesis $G\beta = 0$ and assuming that assumptions $J1$ and $A1-A5$ hold, the asymptotic distribution of $T_{n}^{**}$, as $N$ goes to infinity, is a chi-squared distribution with $q$-degrees of freedom,

$$T_{n}^{**} \xrightarrow{D} \chi^2(q),$$

where $q$ is the row rank of the matrix $G$.

**Proof.** By Corollary 2.5.7, under model (2.1.1) and assumptions $A1-A5$, we have that

$$\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{D} N_p(0, \Psi^{-1}\Sigma\Psi^{-1}).$$

Using Slutsky’s Theorem we obtain

$$\sqrt{N}(G\hat{\beta} - G\beta_0) \xrightarrow{D} N_q(0, M_w),$$. 

81
where $M_w = G \Psi^{-1} \Sigma \Psi^{-1} G^T$. Since the rows of $G$ are linearly independent, $M_w$ is a $q \times q$ invertible matrix. We then observe that

$$\sqrt{N} M_w^{-\frac{1}{2}} G (\hat{\beta} - \beta_0) \xrightarrow{D} N_q(0, I_q),$$

where $I_q$ is the $q \times q$ identity matrix. With the above result, it is easy to see that the quadratic expression $T_n^{**}$ converges to a chi-squared distribution with $q$-degrees of freedom:

$$T_n^{**} = N(\hat{\beta} - \beta_0)^\top G^\top M_w^{-1} G (\hat{\beta} - \beta_0) \xrightarrow{D} \chi^2(q).$$

It is relatively straightforward to implement the Wald test. If we wish to test a hypothesis of the form $H_0 : \beta_2 = \beta_{20}$ with $\beta_1$ unspecified against $H_A : \beta_2 \neq \beta_{20}$ with $\beta_1$ unspecified, we only need to determine an appropriate $G$ matrix to produce $G \beta = \beta_{20}$. It is clear that the Wald test rejects the null hypothesis $\beta_2 = \beta_{20}$ in favor of the alternative hypothesis $\beta_2 \neq \beta_{20}$ for large values of $T_n^{**}$. Again, the construction of the quadratic form $T_n^{**}$ requires the estimation of $\Gamma$ and $\Omega$ matrices, which we will discuss in Chapter 4.
CHAPTER 4
COMPUTATIONAL PROCEDURES

This chapter considers computational procedures necessary to carry out the statistical inference developed in the previous chapters. The first step is to construct a computationally efficient algorithm to compute $\hat{\beta}$, the minimizer of $D_n(\beta)$. The discussion in Chapter 2 indicates that the dispersion function $D_n(\beta)$ is a continuous and convex function of $\beta$. Then a natural choice for computing $\hat{\beta}$ is to use a gradient algorithm, such as steepest descent, but this is usually too slow for a practical use.

To solve the problem, one can use a version of a Newton algorithm based on the asymptotic quadratic nature of the dispersion function $D_n(\beta)$. The Newton algorithm requires initial estimates for $\beta$ and $\Omega$, which will be denoted as $\hat{\beta}^{(0)}$ and $\hat{\Omega}^{(0)}$. These initial values help to create a centered residual vector $\hat{e}^{(0)}$. This can be achieved in two steps. The first step computes $\epsilon^{*^{(0)}} = Y - X\hat{\beta}^{(0)}$ and then estimates the intercept parameter $\alpha$ with $\hat{\alpha} = \text{median}(\epsilon^{*^{(0)}})$. The initial residual vector $\hat{e}^{(0)}$ is then computed from $\hat{e}^{(0)} = \epsilon^{*^{(0)}} - \hat{\alpha}$. We now write the quadratic approximation to $D_n(\beta)$ as follows:

$$Q_n(\beta) = D_n(\hat{\beta}^{(0)}) + \frac{1}{2}(\beta - \hat{\beta}^{(0)})X^\top \hat{\Omega}^{(0)}X(\beta - \hat{\beta}^{(0)}) - (\beta - \hat{\beta}^{(0)})^\top S_n(\hat{\beta}^{(0)}).$$

The value of $\beta$ which minimizes $Q_n(\beta)$ is given by

$$\hat{\beta}^{(1)} = \hat{\beta}^{(0)} + (X^\top \hat{\Omega}^{(0)}X)^{-1}S_n(\hat{\beta}^{(0)}), \hspace{1cm} (4.0.1)$$

83
where
\[ S_n(\beta^{(0)}) = S_n(Y - X\hat{\beta}^{(0)}) = \frac{1}{N} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} X_{u[w]ij} [R(Y_{u[w]ij} - X_{u[w]ij}\beta^{(0)}) - \tau_w]. \]

Equation (4.0.1) suggests an iterative solution. We use the following algorithm to compute the rank regression estimator \( \hat{\beta} \).

**Algorithm:**

**Step 1.** Compute the initial value \( \hat{\beta}^{(0)} \) using some other procedure, such as least square regression, and set \( k = 1 \).

**Step 2.** Obtain the centered residual \( \hat{\epsilon}^{(k-1)} \) and \( \hat{\Omega}^{(k-1)} \).

**Step 3.** Compute \( \hat{\beta}^{(k)} \) from equation (4.0.1), and update \( \hat{\Omega}^{(k)} \) and \( \hat{\epsilon}^{(k)} \) based on \( \hat{\beta}^{(k)} \).

**Step 4.** If the convergence criteria is met for \( |\hat{\beta}^{(k)} - \hat{\beta}^{(k-1)}| \), take \( \hat{\beta}^{(k)} \) as the rank-regression estimator and go to Step 5. Otherwise, set \( \hat{\beta}^{(k-1)} = \hat{\beta}^{(k)}, \hat{\Omega}^{(k-1)} = \hat{\Omega}^{(k)}, \text{ and } \hat{\epsilon}^{(k-1)} = \hat{\epsilon}^{(k)} \), and then go to Step 3.

**Step 5.** Obtain the estimate \( \hat{\beta}^{(k)} \) of \( \beta \) and the final estimate of \( \Omega \).

One can alternatively compute \( \hat{\beta} \) by minimizing the dispersion function \( D_n(\beta) \) with a smooth optimization function. For example, minimization of \( D_n(\beta) \) can be achieved using the nlminb function in the R programming language. In our simulation study, we used the nlminb function to compute \( \hat{\beta} \).

The computational algorithm for obtaining \( \hat{\beta} \) and the construction of the test statistics in Chapter 3 require the estimation of \( \Omega \) and \( \Gamma \). In general, these matrices depend on the ranking process. To reduce the impact of ranking error, we construct estimators based on estimated rank residuals \( \hat{\epsilon} = Y - 1\hat{\alpha} - X\hat{\beta}_o \). Recall that \( \Gamma = \text{Var}(A_R(\epsilon)) \). Since observations from different sets are independent, \( \Gamma \) is a block
diagonal matrix. Each block is an $H \times H$ matrix, $\gamma$, where

$$
\gamma = \text{Var}(h_R(\epsilon_{11})) = \begin{bmatrix}
\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1H} \\
\gamma_{21} & \gamma_{22} & \cdots & \gamma_{2H} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{H1} & \gamma_{H2} & \cdots & \gamma_{HH}
\end{bmatrix},
$$

$$
\gamma_{wq} = \int (F(x) - \tau_w)(F(y) - \tau_q)dF_{[w,q]}(x,y), \ w = 1, \cdots, H, \text{ and } q = 1, \cdots, H.
$$

A naive estimator of $\gamma_{wq}$ can be obtained from the sample covariance $(\hat{\gamma}_{wq})$ between the $w$-th and $q$-th judgement class residuals, $\hat{\gamma}_{wq} = \text{Cov}(h_R^\top(\hat{e}_{[w]}), h_R^\top(\hat{e}_{[q]}))$, where

$$
\hat{h}_R^\top(\hat{e}_{[w]}) = (R(\hat{\epsilon}_{u_{w}[w]11}) - \tau_w, \cdots, R(\hat{\epsilon}_{u_{w}[w]K1}) - \tau_w, \cdots, R(\hat{\epsilon}_{u_{w}[w]K_n}) - \tau_w),
$$

and

$$
\hat{h}_R^\top(\hat{e}_{[q]}) = (R(\hat{\epsilon}_{u_{q}[q]11}) - \tau_q, \cdots, R(\hat{\epsilon}_{u_{q}[q]K1}) - \tau_q, \cdots, R(\hat{\epsilon}_{u_{q}[q]K_n}) - \tau_q)
$$

A naive estimator of $\Gamma$, $\hat{\Gamma}$, is then obtained by inserting $\hat{\gamma}$ in the block diagonal matrix. This estimator may be biased because of within-set correlation structure. We now propose an improved estimator (conditionally unbiased for a known error vector $\epsilon$ from distribution $F$). We note that the $w$-th and $q$-th entries of $\gamma$ can be written as

$$
\gamma_{wq} = \int (F(x) - \tau_w)(F(y) - \tau_q)dF_{[w,q]}(x,y)
$$

$$
= \int F(x)F(y)dF_{[\alpha,\beta]}(x,y) - \int F(y)dF_{[w]}(y) \int F(y)dF_{[q]}(y)
$$

$$
= \nu_{[w,q]} - \tau_w \tau_q,
$$

where $\nu_{[w,q]} = \int F(x)F(y)dF_{[w,q]}(x,y)$, and $\tau_q = \int F(y)dF_{[q]}(y)$. In this section,
unless stated otherwise, we assume that the error vectors $\epsilon$ are from $F$; they are not estimated residuals. Unbiased estimators of $\nu_{w,q}$ and $\tau_{w}$ are given by

$$\hat{\nu}_{w,q} = \frac{1}{A_n} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{q=1}^{K} \sum_{q'=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} \sum_{d'=1}^{K} \sum_{t'=1}^{n} I(\epsilon_{u[w]ij} \leq \epsilon_{\zeta[w]g[m]} I(\epsilon_{p[q]dt} \leq \epsilon_{\zeta[q]g[m]})$$

and

$$\hat{\tau}_{w} = \frac{1}{HKn(Kn-1)} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{q=1}^{K} \sum_{q'=1}^{n} \sum_{d=1}^{K} \sum_{t=1}^{n} \sum_{d'=1}^{K} \sum_{t'=1}^{n} I(\epsilon_{u[w]ij} \leq \epsilon_{\zeta[w]g[m]})$$

where

$$A_n = \frac{1}{H^2Kn(Kn-1)(Kn-2)}$$

and

$$I_{g,m}(i,j,d,t) = \{g = i, m = j\} \cap \{g = d, m = t\}.$$ 

The quantities $\hat{\nu}_{w,q}$ and $\hat{\tau}_{w}$ can not be computed in their current form since the error terms are not observable. In practice, we replace $\epsilon$ with the estimated residuals $\hat{\epsilon}$ to construct the estimator of $\gamma_{wq}$,

$$\hat{\gamma}_{wq}^* = \hat{\nu}_{w,q}^* - \hat{\tau}_{w}^* \hat{\tau}_{q}^*,$$

where $\hat{\nu}_{w,q}^*$ and $\hat{\tau}_{w}^*$ have the same expressions as in $\hat{\nu}_{w,q}$ and $\hat{\tau}_{w}$, but $\epsilon$ is replaced with the estimated residuals $\hat{\epsilon}$. Next, we consider the estimation of the dispersion matrix $\Omega$. We note that $\Omega$ is a diagonal matrix of $(\omega_1, \ldots, \omega_{nK})$, where $\omega_i = (\omega_1, \omega_2, \ldots, \omega_H)$, $i = 1, \ldots, nK$, and $\omega_h = \int f(y)f_{[h]}(y)dy < \infty$, for $h = 1, \ldots, H$.

The estimation of the matrix $\Omega$ requires the estimation of $\omega_h$, which can be accomplished using a kernel density estimation technique (Chen, 1999). Let $\hat{f}$ be a
kernel density estimator of $f$ based on centered error terms $\epsilon$

$$\hat{f}(x) = f_n(x) = \frac{F_n(x + \kappa_n/2) - F_n(x - \kappa_n/2)}{\kappa_n},$$

where $\kappa_n$ is a band-width and $F_n(x)$ is the empirical CDF of the residuals. We assume the usual regularity conditions hold: $\kappa_n \to 0$ and $n\kappa_n \to \infty$ as $n$ goes to infinity. The kernel density estimator, $\hat{f}(x)$, can be written as

$$\hat{f}(x) = \frac{1}{\kappa_n} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} \left(I(\epsilon_{u[w]ij} \leq x + \kappa_n/2) - I(\epsilon_{u[w]ij} \leq x - \kappa_n/2)\right)$$

$$= \frac{1}{\kappa_n} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} I(x - \kappa_n/2 \leq \epsilon_{u[w]ij} \leq x + \kappa_n/2)$$

$$= \frac{1}{\kappa_n} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} I(-\kappa_n/2 \leq \epsilon_{u[w]ij} - x \leq \kappa_n/2).$$

Using $\hat{f}(x)$, we define an unbiased estimator for $w_{[h]}$

$$\hat{w}_{[h]} = \frac{1}{C_{\kappa_n}} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{\{g=i,m=j\}} I(-\kappa_n/2 \leq \epsilon_{u[w]ij} - \epsilon_{[h]gm} \leq \kappa_n/2),$$

where $C_{\kappa_n} = HKn(Kn - 1)\kappa_n$ and the summation index excludes the indicator functions that contain two error terms from the same set. We now show that $\hat{w}_{[h]}$ is an unbiased estimator for $w_{[h]}$. We calculate the expected value of $\hat{w}_{[h]}$ as follows:

$$E(\hat{w}_{[h]}) = \frac{1}{C_{\kappa_n}} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{\{g=i,m=j\}} E[I(-\kappa_n/2 \leq \epsilon_{u[w]ij} - \epsilon_{[h]gm} \leq \kappa_n/2)]$$

$$= \frac{1}{C_{\kappa_n}} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{\{g=i,m=j\}} E[I(\epsilon_{u[w]ij} \leq \epsilon_{[h]gm} + \kappa_n/2)]$$

$$- \frac{1}{C_{\kappa_n}} \sum_{w=1}^{H} \sum_{i=1}^{K} \sum_{j=1}^{n} \sum_{\{g=i,m=j\}} E[I(\epsilon_{u[w]ij} \leq \epsilon_{[h]gm} - \kappa_n/2)].$$

87
The expression above can be simplified by conditioning on the error term $\epsilon_{\zeta[h]gm}$

\[
E(\hat{\omega}[h]) = \frac{1}{C_{\kappa_n}} \left( \sum_{w=1}^{H} \sum_{i=1}^{K_n} \sum_{j=1}^{n} \sum_{g=i,m=j}^{K_n} E(I(\epsilon_{u[w]ij} \leq \epsilon_{\zeta[h]gm} + \frac{\kappa_n}{2} | \epsilon_{\zeta[h]gm}) \right) \\
- \frac{1}{C_{\kappa_n}} \left( \sum_{w=1}^{H} \sum_{i=1}^{K_n} \sum_{j=1}^{n} \sum_{g=i,m=j}^{K_n} E(I(\epsilon_{u[w]ij} \leq \epsilon_{\zeta[h]gm} - \frac{\kappa_n}{2}) | \epsilon_{\zeta[h]gm}) \right) \\
= \frac{1}{C_{\kappa_n}} \left( \sum_{w=1}^{H} \sum_{i=1}^{K_n} \sum_{j=1}^{n} \sum_{g=i,m=j}^{K_n} (F_{w}(\epsilon_{\zeta[h]gm} + \frac{\kappa_n}{2}) - F_{w}(\epsilon_{\zeta[h]gm} - \frac{\kappa_n}{2})) \right).
\]

Under assumption $J1$, use of fundamental equation (2.3.1) yields that

\[
E(\hat{\omega}[h]) = \frac{1}{C_{\kappa_n}} \left( \sum_{i=1}^{K_n} \sum_{j=1}^{n} \sum_{g=i,m=j}^{K_n} H(F_{\epsilon_{\zeta[h]gm} + \frac{\kappa_n}{2}} - F_{\epsilon_{\zeta[h]gm} - \frac{\kappa_n}{2}}) \right) \\
= \frac{H}{C_{\kappa_n}} \left( \sum_{i=1}^{K_n} \sum_{j=1}^{n} \sum_{g=1}^{K_n} \sum_{m=1}^{n} (F(\epsilon_{\zeta[h]gm} + \frac{\kappa_n}{2}) - F(\epsilon_{\zeta[h]gm} - \frac{\kappa_n}{2})) \\
- \sum_{i=1}^{K_n} \sum_{j=1}^{n} (F(\epsilon_{\zeta[h]ij} + \frac{\kappa_n}{2}) - F(\epsilon_{\zeta[h]ij} - \frac{\kappa_n}{2})) \right).
\]

Note that the expected values in the above sums over indexes i and j are constant.

Then the expected value of $\hat{\omega}[h]$ reduces to

\[
E(\hat{\omega}[h]) = \frac{1}{Kn(Kn - 1)\kappa_n} \left( Kn \sum_{g=1}^{K_n} \sum_{m=1}^{n} (F(\epsilon_{\zeta[h]gm} + \frac{\kappa_n}{2}) - F(\epsilon_{\zeta[h]gm} - \frac{\kappa_n}{2})) \\
- \sum_{i=1}^{K_n} \sum_{j=1}^{n} (F(\epsilon_{\zeta[h]ij} + \frac{\kappa_n}{2}) - F(\epsilon_{\zeta[h]ij} - \frac{\kappa_n}{2})) \right).
\]

88
After further simplification, we write

\[
E(\hat{\omega}_{[h]})
= \frac{1}{Kn(Kn-1)\kappa_n} E \left( (Kn-1) \sum_{g=1}^{K} \sum_{m=1}^{n} \left( F(\epsilon_{[h]gm} + \frac{\kappa_n}{2}) - F(\epsilon_{[h]gm} - \frac{\kappa_n}{2}) \right) \right)
\]

\[
= \frac{1}{Kn} \kappa_n E \left( \sum_{g=1}^{K} \sum_{m=1}^{n} \left( F(\epsilon_{[h]gm} + \frac{\kappa_n}{2}) - F(\epsilon_{[h]gm} - \frac{\kappa_n}{2}) \right) \right)
\]

\[
= \frac{1}{Kn} E \left( \sum_{g=1}^{K} \sum_{m=1}^{n} \left( \frac{F(\epsilon_{[h]gm} + \frac{\kappa_n}{2}) - F(\epsilon_{[h]gm} - \frac{\kappa_n}{2})}{\kappa_n} \right) \right)
\]

\[
= \frac{1}{Kn} \kappa_n \int [F(x + \frac{\kappa_n}{2}) - F(x - \frac{\kappa_n}{2})] dF_{[h](x)}
\]

\[
= \int \frac{F(x + \frac{\kappa_n}{2}) - F(x - \frac{\kappa_n}{2})}{\kappa_n} dF_{[h](x)}
\]

As \(n\) goes to infinity, \(\kappa_n\) converges to zero. Hence, \(E(\hat{\omega}_{[h]})\) becomes asymptotically unbiased for \(\omega_{[h]}\), as \(E(\hat{\omega}_h) \rightarrow \int f(x) dF_{[h](x)} = \int f(x) f_{[h]}(x) dx\).

The construction of the estimator \(\hat{\omega}_{[h]}\) requires a band-width selection, \(\kappa_n\). In this dissertation, we use the band-width selection procedure developed for rank regression estimator in Hettmansperger (1984, page 249)

\[
\kappa_n = \frac{4.11}{N^{1/2}} 4\hat{\sigma},
\]

where \(\hat{\sigma}\) is the estimated interquartile range of the estimated residuals. We again note that \(\hat{\omega}_{[h]}\) can not be computed in its current form since the errors \((\varepsilon)\) are not observable. In practice, we estimate \(\hat{\omega}_{[h]}\) with \(\hat{\omega}^*_{[h]}\), where \(\hat{\omega}^*_{[h]}\) has the same expression as \(\hat{\omega}_{[h]}\), but \(\varepsilon\) is replaced with estimated error vector \(\hat{\varepsilon}\).
CHAPTER 5
SIMULATION RESULTS

5.1 Simulation Settings

In this chapter, we investigate the empirical properties of the proposed procedures under a wide range of simulation settings. The simulation parameters include different set sizes ($H$), replication sizes ($n$), varying quality of ranking information and some common underlying distributions. The simulation study has three different objectives. The first objective is to investigate finite sample behavior of the Type I error rates. The second objective is to determine the sensitivity of the testing procedures against ranking error. The last objective is to evaluate the empirical powers of the tests for finite sample sizes.

The ranking process in ORRDs uses internal variation among the within-set experimental units to create judgement classes. Good quality ranking information creates a significant amount of separation among judgment classes, leading to reduction in the variance of the regression estimator and test statistics. Poor quality ranking information, however, generally increases the variances of the estimators and tests, and may cause a loss in the efficiency of an ORRD. Thus, modeling the quality of ranking information is necessary to evaluate the performance of the ORRDs.

In the literature, there are three classes of judgment ranking models. The first class of models was originally introduced by Bohn and Wolfe (1994). Later, Frey
(2007) extended the model to a bigger class. These models express the distribution of the judgment ranked order statistics as a mixture distribution of the order statistics in a set of size $H$. Under these models, the $j$-th judgment class distribution has a CDF

$$F_{[j]}(\epsilon) = \sum_{i=1}^{H} p_{ij}F_{(i)}(\epsilon),$$

where $p_{ij}$ is the probability of assigning the $i$-th order statistic to the $j$-th judgment class. To ensure that $F_{[j]}(\cdot)$ is a valid probability distribution, the matrix of the judgment ranking probabilities $P = (p_{ij})$ must be a doubly stochastic matrix. The quality of ranking information is modeled by the matrix $P$. An identity matrix corresponds to a perfect ranking model, whereas a matrix with entries $p_{ij} = \frac{1}{H}$ corresponds to a completely random ranking model. Intermediate quality of ranking information can be modeled by an appropriately chosen $P$ matrix.

The second class of models, called the monotone likelihood ratio (MLR) ranking model, was introduced by Fligner and MacEachern (2006). Under this model, the variable of interest $T$ have a distribution that depends on the parameter $\theta$, $T|\theta \sim F_\theta$. The perceived values $U$ of $T$ are modeled as a function of the actual values $T$, $U|T \sim G_T$, where the distribution $G_T$ depends on the parameter $\theta$ only through $T$. The MLR-ranking model has two basic assumptions: (1) $F_\theta$ is MLR in $T$ and (2) $G_T$ is MLR in $U$. Under these assumptions, Fligner and MacEachern (2006) showed that $U|\theta$ is stochastically increasing in $\theta$. This leads to a stronger result that in a set of size $H$, the distribution of $T_{[i]}$ is stochastically smaller than that of $T_{[j]}$ for $i < j$.

The third class of models uses assumption that the rank order of units in a small set can be estimated from an additive model (Dell and Clutter, 1972). This model can be written as, $\mu_i = \epsilon_i + w_i$, where $\mu_i$ is the estimated residual, $\epsilon_i$ is the true error term from the underlying CDF $F$, and $w_i$ represents a random noise that is added to the model to make the ranking process harder. The random error $w_i$ is normally
distributed with mean 0 and variance $\tau^2$, $w_i \overset{i.i.d.}{\sim} N(0, \tau^2)$. In this model, $\epsilon_i$ and $w_i$ are assumed to be independent. It is also assumed that $\epsilon_i$ is standardized so that it has mean zero and variance 1.

In this dissertation, we use the Dell and Clutter judgment ranking model to generate the error terms in an ORRD. To construct a within-set ORRD error vector $\epsilon$ in model (2.1.2), a vector of $H$ independent error terms $(\epsilon_1, \cdots, \epsilon_H)$ is generated from the underlying distribution $F$. These residuals are scaled so that they all have variance 1. Another vector of $H$ independent normal errors, $w = (w_1, \cdots, w_H)$, is generated with mean 0 and variance $\tau^2$. These two vectors are added to obtain $\mu = \epsilon + w$. The vector $\mu$ is sorted, and the corresponding $\epsilon$ values are taken as within-set error terms in an ORRD. The quality of ranking information is controlled by the correlation coefficient between $\epsilon$ and $\mu$, $\rho = \frac{1}{\sqrt{1 + \tau^2}}$, or equivalently by the variance of the random variable $w$, $\tau^2$. For example, if $\rho = 1$ (or $w$ is a degenerate random variable), the ranking mechanism leads to a perfect ranking model. On the other hand, if $\rho = 0$ (or $\tau^2$ is large), the Dell and Clutter model produces ranking at random. Other intermediate values of ranking information can be determined by selecting an appropriate correlation coefficient $\rho$ (or $\tau^2$). For example, if we want ranking information with correlation coefficient $\rho = 0.80$, we would select the random variable $w$ with $\tau^2 = \frac{1 - \rho^2}{\rho^2} = \frac{9}{16}$.

5.2 Type I Error Rates

In this section, we provide empirical evidence to evaluate the performance of the tests we developed in Chapter 3. The simulation settings consist of different set sizes ($H$), replication sizes ($n$), treatment levels ($L$), degrees of ranking errors and different underlying distributions.

In the first part of the simulation study, we investigate the finite sample behavior of
Type I error rates under various simulation settings. The simulation study considers two ORRDs, Design 1 and Design 2. Design 1 uses a set size $H = 2$, number of sets per replication $K = 2 \binom{L}{2}$ and treatment levels $L = 2, 3, 4$. Replication size $n$ varies between 3 and 12 depending on the total sample size $N$. Design 2 uses the set sizes $H = 3, 4, 5$, number of sets per replication $K = H$, treatment levels $L = H$ and replication sizes $n = 4, 6, 8$. The empirical size and power of the tests are computed based on 5000 data sets generated from model (2.1.2). The quality of the ranking information varies with $\rho$ ($\rho = 1, 0.90, 0.75, 0.50$). The ORRD error terms are generated from three different underlying distributions. These underlying distributions include the standard normal distribution, Student’s $t$-distribution with 3-degrees of freedom, and log-normal distribution, representing symmetric, heavy tailed and skewed distributions, respectively.

Throughout the simulation study, the response vector $Y_{ij}$ in the $i$-th set and $j$-th replication is generated in two steps. In the first step, we construct the deterministic part of model (2.1.2), $X_{ij}\beta$, for given values of $\beta$ and design matrix $X$. Without loss of generality, we use $\beta = 0$ in this part of the simulation. In the second step, we generate the within-set ORRD error terms $\epsilon_{ij}$ from the Dell and Clutter model. The response vector $Y_{ij}$ is then constructed by adding $X_{ij}\beta$ and $\epsilon_{ij}$, $Y_{ij} = X_{ij}\beta + \epsilon_{ij}$, for $i = 1, \cdots, K$, and $j = 1, \cdots, n$.

The empirical Type I error rates of the score test for Design 1 and Design 2 are given in Tables 5.1-5.6 for selected simulation parameters. The Type I error rates are calculated using two different approximations to critical values: a chi-squared distribution approximation with $q$-degrees of freedom and an $F$-distribution approximation with $q$ and $(nK - q + 1)$ degrees of freedom. We set the nominal level to be 0.05. The critical values are taken to be the 95-th percentile of these two approximating distributions.
Simulation results in Tables 5.1-5.6 indicate that the estimated Type I error rates converge to the nominal value 0.05 at different rates depending on the approximating distribution of the asymptotic null distribution of the score test statistic. For small sample sizes, empirical Type I error rates based on the F-approximation are much close to 0.05 than those of the chi-squared approximation. For moderately large sample sizes, however, the estimated Type I error rates based on the chi-squared approximation provide better convergence results than those based on the F-approximation. For large sample sizes, it appears that both approximations generally yield Type I error rates reasonably close to the nominal value 0.05. The empirical Type I error rates in Design 2 converge to the nominal value 0.05 more quickly than those in Design 1.

In Tables 5.1-5.6, it appears that the quality of ranking information (\( \rho \)) does not have a big impact on the estimated Type I error rates. For a given sample size, the estimated Type I error rates are relatively constant for \( \rho = 1, 0.90, 0.75, 0.50 \). This clearly indicates that the calibrated tests are robust against these departures from the perfect ranking in the ranking process.

Another observation in Tables 5.1-5.6 is that the testing procedures work reasonably well for the selected symmetric, skewed and heavy-tailed distributions. For example, there is not much difference among the estimated Type I error rates when the ORRD error terms are generated from standard normal, Student’s t and log-normal distributions.

The empirical Type I error rates of the Wald test for Design 1 and Design 2 are given in Tables 5.7-5.12 for selected simulation parameters. Again, Type I error rates are computed with two approximations: the chi-squared and F distribution approximations. Results similar to those discussed in Tables 5.1-5.6 hold for the Wald test with a noticeable difference. In the Wald test, the empirical Type I error rates based on an F-approximation provide slightly better convergence rates than
those for the score test. For example, the estimated Type I error rates (based on an F-approximation) in Tables 5.7-5.12 are closer to the nominal value 0.05 than those in Tables 5.1-5.6 for matched sample sizes and correlation coefficients.

The empirical Type I error rates of the drop test are given in Tables 5.13-5.18 for selected simulation parameters. The drop test requires knowing the distribution of a linear combination of \( q \) independent chi-squared distributions with 1 degree of freedom, \( \sum_{i=1}^{q} \lambda_i \chi_i^2(1) \), where \( \lambda_i \) are the eigenvalues of the matrix \( \mathbf{V}^{1/2} \mathbf{W} \mathbf{V}^{1/2} \) in Theorem 3.1.1. Our simulation study takes two approaches to estimate the null distribution of the drop test.

The first approach uses a single critical value in all iterations. It first estimates the eigenvalues based on a simulation of 5000 replications

\[
\hat{\lambda}_i = \frac{1}{5000} \sum_{j=1}^{5000} \hat{\lambda}_{j,i}, \quad i = 1, \cdots, q.
\]

These estimated eigenvalues are then used to construct the empirical distribution of \( \sum_{i=1}^{q} \hat{\lambda}_i \chi_i^2(1) \) based on 10,000 replications. The 95-th percentile of this distribution is taken to be the critical value of the drop test. In this case, the estimated Type I error rates are presented under the heading “Avg” in Tables 5.13-5.18.

The second approach estimates the eigenvalues \( \lambda_{j,i} \) with \( \hat{\lambda}_{j,i}, \ i = 1, \cdots, q \), from the \( j \)-th simulated data set, \( j = 1, \cdots, 5000 \). For each \( j \), these estimated eigenvalues are then used to generate the simulated distribution of \( \sum_{i=1}^{q} \hat{\lambda}_{j,i} \chi_i^2(1) \) based on 10,000 replications. The critical value of the drop test for the \( j \)-th data set is then computed from the 95-th percentile of the simulated distribution of \( \sum_{i=1}^{q} \hat{\lambda}_{j,i} \chi_i^2(1) \). The estimated Type I error rates for this approximation are represented under the heading “Each” in Tables 5.13-5.18. We note that in this approach, the critical value of the drop test is estimated separately at each iteration of the simulation.
The empirical Type I error rates for the drop test have similar patterns as for the previous two tests. It appears that the way we compute the critical values of the drop test does not make a significant difference on the estimated Type I error rate, when the sample size is large. The critical values from approach 1 and approach 2 yield similar estimated Type I error rates for a given sample size, set size and correlation coefficient. On the other hand, for moderate sample sizes, approach 2 yields Type I error rates closer to nominal value 0.05 than those generated by approach 1. When the sample size increases, the Type I error rates become much closer to the nominal level 0.05 in both approaches. The empirical Type I error rates for Design 2 converges to 0.05 more quickly than those for Design 1.

Among the three tests, the drop test appears to produce better convergence results for the Type I error rates. For example, the Type I error rates in Tables 5.13-5.18 are much closer to nominal size 0.05 than those values for the Wald and score tests in Tables 5.1-5.12 for moderate sample sizes.
<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>ρ</th>
<th>Underlying distribution</th>
<th>Approximated distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>Log-normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>χ²</td>
<td>F₁,nK−1</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>1.00</td>
<td>0.078</td>
<td>0.051</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>0.90</td>
<td>0.090</td>
<td>0.059</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>0.75</td>
<td>0.088</td>
<td>0.060</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>0.50</td>
<td>0.082</td>
<td>0.054</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>1.00</td>
<td>0.065</td>
<td>0.044</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>0.90</td>
<td>0.070</td>
<td>0.051</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>0.75</td>
<td>0.073</td>
<td>0.055</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>0.50</td>
<td>0.066</td>
<td>0.046</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>1.00</td>
<td>0.060</td>
<td>0.045</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>0.90</td>
<td>0.071</td>
<td>0.053</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>0.75</td>
<td>0.070</td>
<td>0.054</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>0.50</td>
<td>0.076</td>
<td>0.059</td>
</tr>
<tr>
<td>12</td>
<td>48</td>
<td>1.00</td>
<td>0.053</td>
<td>0.040</td>
</tr>
<tr>
<td>12</td>
<td>48</td>
<td>0.90</td>
<td>0.068</td>
<td>0.058</td>
</tr>
<tr>
<td>12</td>
<td>48</td>
<td>0.75</td>
<td>0.063</td>
<td>0.051</td>
</tr>
<tr>
<td>12</td>
<td>48</td>
<td>0.50</td>
<td>0.062</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Table 5.1: Empirical Type I error rates for the score test, Design 1, H=2, L=2, K=2, and simulation size 5000.
<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>ρ</th>
<th>Underlying distribution</th>
<th>Approximated distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>Log-normal</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>1.00</td>
<td>0.052 0.025</td>
<td>0.050 0.025</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>0.90</td>
<td>0.055 0.026</td>
<td>0.052 0.025</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>0.75</td>
<td>0.049 0.025</td>
<td>0.051 0.028</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>0.50</td>
<td>0.050 0.023</td>
<td>0.048 0.023</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>1.00</td>
<td>0.047 0.027</td>
<td>0.047 0.029</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>0.90</td>
<td>0.052 0.033</td>
<td>0.049 0.033</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>0.75</td>
<td>0.052 0.033</td>
<td>0.053 0.032</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>0.50</td>
<td>0.046 0.025</td>
<td>0.056 0.032</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>1.00</td>
<td>0.057 0.044</td>
<td>0.050 0.040</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>0.90</td>
<td>0.051 0.039</td>
<td>0.051 0.039</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>0.75</td>
<td>0.047 0.035</td>
<td>0.049 0.036</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>0.50</td>
<td>0.047 0.034</td>
<td>0.041 0.028</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
<td>1.00</td>
<td>0.048 0.043</td>
<td>0.052 0.043</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
<td>0.90</td>
<td>0.047 0.038</td>
<td>0.049 0.038</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
<td>0.75</td>
<td>0.052 0.040</td>
<td>0.045 0.037</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
<td>0.50</td>
<td>0.054 0.042</td>
<td>0.048 0.039</td>
</tr>
</tbody>
</table>

Table 5.2: Empirical Type I error rates for the score test, Design 1, H=2, L=3, K=6 and simulation size 5000.
<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>ρ</th>
<th>Normal</th>
<th>Log-normal</th>
<th>t3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\chi^2_3$  $F_{3,nK-3}$</td>
<td>$\chi^2_3$  $F_{3,nK-3}$</td>
<td>$\chi^2_3$  $F_{3,nK-3}$</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>1.00</td>
<td>0.047 0.031</td>
<td>0.052 0.034</td>
<td>0.047 0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td>0.048 0.033</td>
<td>0.051 0.035</td>
<td>0.049 0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td>0.047 0.031</td>
<td>0.054 0.034</td>
<td>0.048 0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>0.049 0.031</td>
<td>0.050 0.032</td>
<td>0.045 0.030</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>1.00</td>
<td>0.050 0.037</td>
<td>0.049 0.037</td>
<td>0.058 0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td>0.050 0.038</td>
<td>0.053 0.039</td>
<td>0.054 0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td>0.044 0.032</td>
<td>0.047 0.037</td>
<td>0.048 0.035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>0.045 0.035</td>
<td>0.053 0.040</td>
<td>0.053 0.039</td>
</tr>
<tr>
<td>6</td>
<td>144</td>
<td>1.00</td>
<td>0.050 0.042</td>
<td>0.050 0.042</td>
<td>0.048 0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td>0.053 0.042</td>
<td>0.056 0.047</td>
<td>0.044 0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td>0.047 0.039</td>
<td>0.052 0.045</td>
<td>0.051 0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>0.049 0.042</td>
<td>0.049 0.039</td>
<td>0.048 0.039</td>
</tr>
<tr>
<td>8</td>
<td>192</td>
<td>1.00</td>
<td>0.050 0.042</td>
<td>0.056 0.046</td>
<td>0.048 0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td>0.044 0.036</td>
<td>0.053 0.047</td>
<td>0.052 0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td>0.050 0.043</td>
<td>0.047 0.041</td>
<td>0.053 0.047</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>0.050 0.044</td>
<td>0.058 0.049</td>
<td>0.048 0.042</td>
</tr>
</tbody>
</table>

Table 5.3: Empirical Type I error rates for the score test, Design 1, H=2, L=4, K=12 and simulation size 5000.
<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>( \rho )</th>
<th>Underlying distribution</th>
<th>Approximated distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>( \chi_2^2 )</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>1.00</td>
<td>0.122</td>
<td>0.070</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>0.90</td>
<td>0.163</td>
<td>0.092</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>0.75</td>
<td>0.168</td>
<td>0.093</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>0.50</td>
<td>0.165</td>
<td>0.094</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>1.00</td>
<td>0.065</td>
<td>0.044</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>0.90</td>
<td>0.070</td>
<td>0.051</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>0.75</td>
<td>0.073</td>
<td>0.055</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>0.50</td>
<td>0.066</td>
<td>0.046</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>1.00</td>
<td>0.052</td>
<td>0.025</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>0.90</td>
<td>0.055</td>
<td>0.026</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>0.75</td>
<td>0.049</td>
<td>0.025</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>0.50</td>
<td>0.050</td>
<td>0.023</td>
</tr>
</tbody>
</table>

Table 5.4: Empirical Type I error rates for the score test, Design 2, H=3, L=3, K=3 and simulation size 5000.
Table 5.5: Empirical Type I error rates for the score test, Design 2, when H=4, L=4, K=4 and simulation size 5000.
Table 5.6: Empirical Type I error rates for the score test, Design 2, H=5, L=5, K=5 and simulation size 5000.
<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>( \rho )</th>
<th>Underlying distribution</th>
<th>Approximated distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>( \chi^2 ) F_{1,nK-1}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Log-normal</td>
<td>( \chi^2 ) F_{1,nK-1}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>t3</td>
<td>( \chi^2 ) F_{1,nK-1}</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>1.00</td>
<td>0.124 0.079</td>
<td>0.101 0.069</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>0.90</td>
<td>0.134 0.086</td>
<td>0.112 0.075</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>0.75</td>
<td>0.120 0.078</td>
<td>0.113 0.072</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>0.50</td>
<td>0.117 0.068</td>
<td>0.112 0.073</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>1.00</td>
<td>0.082 0.056</td>
<td>0.075 0.053</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>0.90</td>
<td>0.091 0.065</td>
<td>0.077 0.054</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>0.75</td>
<td>0.085 0.055</td>
<td>0.078 0.053</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>0.50</td>
<td>0.078 0.051</td>
<td>0.074 0.052</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>1.00</td>
<td>0.068 0.048</td>
<td>0.060 0.045</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>0.90</td>
<td>0.066 0.047</td>
<td>0.066 0.050</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>0.75</td>
<td>0.077 0.057</td>
<td>0.064 0.048</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>0.50</td>
<td>0.061 0.043</td>
<td>0.061 0.047</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>1.00</td>
<td>0.065 0.050</td>
<td>0.051 0.041</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>0.90</td>
<td>0.066 0.050</td>
<td>0.059 0.047</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>0.75</td>
<td>0.065 0.047</td>
<td>0.056 0.045</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>0.50</td>
<td>0.064 0.050</td>
<td>0.054 0.042</td>
</tr>
<tr>
<td>12</td>
<td>48</td>
<td>1.00</td>
<td>0.067 0.054</td>
<td>0.049 0.038</td>
</tr>
<tr>
<td>12</td>
<td>48</td>
<td>0.90</td>
<td>0.060 0.048</td>
<td>0.056 0.045</td>
</tr>
<tr>
<td>12</td>
<td>48</td>
<td>0.75</td>
<td>0.064 0.053</td>
<td>0.056 0.044</td>
</tr>
<tr>
<td>12</td>
<td>48</td>
<td>0.50</td>
<td>0.063 0.051</td>
<td>0.055 0.044</td>
</tr>
</tbody>
</table>

Table 5.7: Empirical Type I error rates for the Wald test, Design 1, H=2, L=2, K=2 and simulation size 5000.
<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>ρ</th>
<th>Underlying distribution</th>
<th>Approximated distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>χ²</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>1.00</td>
<td>0.046 0.023</td>
<td>0.078</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>0.90</td>
<td>0.044 0.022</td>
<td>0.063</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>0.75</td>
<td>0.046 0.023</td>
<td>0.070</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>0.50</td>
<td>0.048 0.025</td>
<td>0.072</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>1.00</td>
<td>0.045 0.027</td>
<td>0.058</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>0.90</td>
<td>0.048 0.031</td>
<td>0.059</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>0.75</td>
<td>0.045 0.030</td>
<td>0.060</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>0.50</td>
<td>0.043 0.026</td>
<td>0.064</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>1.00</td>
<td>0.046 0.031</td>
<td>0.055</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>0.90</td>
<td>0.049 0.036</td>
<td>0.060</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>0.75</td>
<td>0.048 0.034</td>
<td>0.057</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>0.50</td>
<td>0.043 0.032</td>
<td>0.062</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
<td>1.00</td>
<td>0.049 0.040</td>
<td>0.052</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
<td>0.90</td>
<td>0.049 0.038</td>
<td>0.049</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
<td>0.75</td>
<td>0.048 0.039</td>
<td>0.058</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
<td>0.50</td>
<td>0.044 0.036</td>
<td>0.061</td>
</tr>
</tbody>
</table>

Table 5.8: Empirical Type I error rates for the Wald test, Design 1, H=2, L=3, K=6 and simulation size 5000.
<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>ρ</th>
<th>Underlying distribution</th>
<th>Approximated distribution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>Log-normal</td>
<td>t3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>1.00</td>
<td>0.051 0.032</td>
<td>0.060 0.047</td>
<td>0.055 0.038</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>0.90</td>
<td>0.044 0.030</td>
<td>0.059 0.044</td>
<td>0.045 0.030</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>0.75</td>
<td>0.049 0.034</td>
<td>0.064 0.048</td>
<td>0.053 0.033</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>0.50</td>
<td>0.047 0.029</td>
<td>0.069 0.048</td>
<td>0.047 0.033</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>1.00</td>
<td>0.044 0.033</td>
<td>0.055 0.042</td>
<td>0.057 0.043</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>0.90</td>
<td>0.049 0.039</td>
<td>0.059 0.047</td>
<td>0.055 0.040</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>0.75</td>
<td>0.049 0.035</td>
<td>0.055 0.045</td>
<td>0.054 0.040</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>0.50</td>
<td>0.045 0.032</td>
<td>0.069 0.056</td>
<td>0.051 0.038</td>
</tr>
<tr>
<td>6</td>
<td>144</td>
<td>1.00</td>
<td>0.048 0.037</td>
<td>0.058 0.051</td>
<td>0.053 0.044</td>
</tr>
<tr>
<td>6</td>
<td>144</td>
<td>0.90</td>
<td>0.050 0.041</td>
<td>0.063 0.056</td>
<td>0.049 0.041</td>
</tr>
<tr>
<td>6</td>
<td>144</td>
<td>0.75</td>
<td>0.051 0.045</td>
<td>0.066 0.056</td>
<td>0.048 0.042</td>
</tr>
<tr>
<td>6</td>
<td>144</td>
<td>0.50</td>
<td>0.047 0.038</td>
<td>0.060 0.052</td>
<td>0.055 0.047</td>
</tr>
<tr>
<td>8</td>
<td>192</td>
<td>1.00</td>
<td>0.045 0.038</td>
<td>0.061 0.053</td>
<td>0.051 0.045</td>
</tr>
<tr>
<td>8</td>
<td>192</td>
<td>0.90</td>
<td>0.052 0.045</td>
<td>0.049 0.042</td>
<td>0.053 0.046</td>
</tr>
<tr>
<td>8</td>
<td>192</td>
<td>0.75</td>
<td>0.043 0.038</td>
<td>0.054 0.048</td>
<td>0.054 0.048</td>
</tr>
<tr>
<td>8</td>
<td>192</td>
<td>0.50</td>
<td>0.048 0.043</td>
<td>0.068 0.059</td>
<td>0.049 0.044</td>
</tr>
</tbody>
</table>

Table 5.9: Empirical Type I error rates for the Wald test, Design 1, H=2, L=4, K=12 and simulation size 5000.
### Table 5.10: Empirical Type I error rates for the Wald test, Design 2, H=3, L=3, K=3, and simulation size 5000.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>ρ</th>
<th>Underlying distribution</th>
<th>Approximated distribution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>Log-normal</td>
<td>t3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\chi^2$</td>
<td>$F_{2,nK-2}$</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>1.00</td>
<td>0.087 0.036</td>
<td>0.066 0.031</td>
<td>0.077</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>0.90</td>
<td>0.080 0.032</td>
<td>0.072 0.032</td>
<td>0.080</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>0.75</td>
<td>0.076 0.030</td>
<td>0.079 0.037</td>
<td>0.082</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>0.50</td>
<td>0.075 0.030</td>
<td>0.094 0.047</td>
<td>0.076</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>1.00</td>
<td>0.071 0.038</td>
<td>0.057 0.033</td>
<td>0.064</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>0.90</td>
<td>0.073 0.044</td>
<td>0.057 0.032</td>
<td>0.067</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>0.75</td>
<td>0.070 0.038</td>
<td>0.064 0.037</td>
<td>0.063</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>0.50</td>
<td>0.060 0.035</td>
<td>0.068 0.042</td>
<td>0.068</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>1.00</td>
<td>0.061 0.040</td>
<td>0.052 0.035</td>
<td>0.049</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>0.90</td>
<td>0.061 0.041</td>
<td>0.053 0.036</td>
<td>0.063</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>0.75</td>
<td>0.064 0.042</td>
<td>0.059 0.041</td>
<td>0.060</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>0.50</td>
<td>0.060 0.040</td>
<td>0.063 0.044</td>
<td>0.061</td>
</tr>
<tr>
<td>n</td>
<td>N</td>
<td>ρ</td>
<td>Underlying distribution</td>
<td>Approximated distribution</td>
<td>t3</td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
<td>-------</td>
<td>-------------------------</td>
<td>--------------------------</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>F_{3,nK-3}</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>χ^2_3</td>
<td>χ^2_3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>1.00</td>
<td>0.085 0.034</td>
<td>0.073 0.029</td>
<td>0.085 0.038</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>0.90</td>
<td>0.094 0.038</td>
<td>0.075 0.035</td>
<td>0.083 0.035</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>0.75</td>
<td>0.087 0.037</td>
<td>0.092 0.041</td>
<td>0.095 0.039</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>0.50</td>
<td>0.083 0.031</td>
<td>0.094 0.046</td>
<td>0.084 0.034</td>
</tr>
<tr>
<td>6</td>
<td>96</td>
<td>1.00</td>
<td>0.069 0.043</td>
<td>0.059 0.033</td>
<td>0.067 0.038</td>
</tr>
<tr>
<td>6</td>
<td>96</td>
<td>0.90</td>
<td>0.074 0.041</td>
<td>0.076 0.043</td>
<td>0.074 0.045</td>
</tr>
<tr>
<td>6</td>
<td>96</td>
<td>0.75</td>
<td>0.071 0.041</td>
<td>0.067 0.040</td>
<td>0.079 0.047</td>
</tr>
<tr>
<td>6</td>
<td>96</td>
<td>0.50</td>
<td>0.065 0.035</td>
<td>0.076 0.048</td>
<td>0.068 0.041</td>
</tr>
<tr>
<td>8</td>
<td>128</td>
<td>1.00</td>
<td>0.067 0.043</td>
<td>0.051 0.036</td>
<td>0.064 0.041</td>
</tr>
<tr>
<td>8</td>
<td>128</td>
<td>0.90</td>
<td>0.067 0.043</td>
<td>0.065 0.041</td>
<td>0.067 0.045</td>
</tr>
<tr>
<td>8</td>
<td>128</td>
<td>0.75</td>
<td>0.066 0.043</td>
<td>0.068 0.047</td>
<td>0.063 0.043</td>
</tr>
<tr>
<td>8</td>
<td>128</td>
<td>0.50</td>
<td>0.060 0.041</td>
<td>0.073 0.051</td>
<td>0.067 0.045</td>
</tr>
</tbody>
</table>

Table 5.11: Empirical Type I error rates for the Wald test, Design 2, H=4, L=4, K=4, and simulation size 5000.
### Table 5.12: Empirical Type I error rates for the Wald test, Design 2, H=5, L=5, K=5, and simulation size 5000.

<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>ρ</th>
<th>Underlying distribution</th>
<th>Approximated distribution</th>
<th>t3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>χ²</td>
<td>F_{4,nK-4}</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>1.00</td>
<td>0.073</td>
<td>0.025</td>
<td>0.062</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0.90</td>
<td>0.083</td>
<td>0.034</td>
<td>0.078</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0.75</td>
<td>0.074</td>
<td>0.031</td>
<td>0.094</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0.50</td>
<td>0.071</td>
<td>0.025</td>
<td>0.096</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>1.00</td>
<td>0.064</td>
<td>0.034</td>
<td>0.058</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>0.90</td>
<td>0.071</td>
<td>0.040</td>
<td>0.069</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>0.75</td>
<td>0.063</td>
<td>0.034</td>
<td>0.075</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>0.50</td>
<td>0.058</td>
<td>0.033</td>
<td>0.081</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>1.00</td>
<td>0.055</td>
<td>0.037</td>
<td>0.050</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>0.90</td>
<td>0.066</td>
<td>0.043</td>
<td>0.059</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>0.75</td>
<td>0.069</td>
<td>0.044</td>
<td>0.062</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>0.50</td>
<td>0.060</td>
<td>0.036</td>
<td>0.077</td>
</tr>
</tbody>
</table>
Underlying distribution

<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>ρ</th>
<th>Normal</th>
<th>Log-normal</th>
<th>t3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg</td>
<td>Each</td>
<td>Avg</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>1.00</td>
<td>0.072</td>
<td>0.139</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td>0.081</td>
<td>0.149</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td>0.081</td>
<td>0.141</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>0.085</td>
<td>0.138</td>
<td>0.080</td>
</tr>
</tbody>
</table>

| 6 | 24| 1.00 | 0.058  | 0.085       | 0.050| 0.077 |
|   |   | 0.90 | 0.072  | 0.098       | 0.065| 0.094 |
|   |   | 0.75 | 0.067  | 0.094       | 0.064| 0.095 |
|   |   | 0.50 | 0.067  | 0.091       | 0.060| 0.085 |

| 8 | 32| 1.00 | 0.057  | 0.074       | 0.051| 0.066 |
|   |   | 0.90 | 0.062  | 0.075       | 0.053| 0.064 |
|   |   | 0.75 | 0.063  | 0.078       | 0.054| 0.071 |
|   |   | 0.50 | 0.070  | 0.090       | 0.063| 0.079 |

| 10| 40| 1.00 | 0.052  | 0.065       | 0.048| 0.061 |
|   |   | 0.90 | 0.058  | 0.069       | 0.047| 0.063 |
|   |   | 0.75 | 0.065  | 0.072       | 0.062| 0.067 |
|   |   | 0.50 | 0.064  | 0.074       | 0.063| 0.072 |

| 12| 48| 1.00 | 0.053  | 0.060       | 0.045| 0.051 |
|   |   | 0.90 | 0.056  | 0.066       | 0.052| 0.065 |
|   |   | 0.75 | 0.060  | 0.072       | 0.052| 0.064 |
|   |   | 0.50 | 0.061  | 0.064       | 0.059| 0.060 |

Table 5.13: Empirical Type I error rates for the drop test, Design 1, H=2, L=2, K=2 and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration.
| n | N  | $\rho$ | Underlying distribution | Approximated distribution | | | | | | | Normal | Log-normal | t3 | Avg | Each | Avg | Each | Avg | Each |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 36 | 1.00 | | 0.053 | 0.054 | 0.053 | 0.053 | 0.056 | 0.054 |
| 3 | 36 | 0.90 | | 0.058 | 0.065 | 0.051 | 0.056 | 0.056 | 0.061 |
| 3 | 36 | 0.75 | | 0.058 | 0.060 | 0.057 | 0.065 | 0.058 | 0.062 |
| 3 | 36 | 0.50 | | 0.061 | 0.065 | 0.059 | 0.067 | 0.055 | 0.059 |
| 4 | 48 | 1.00 | | 0.048 | 0.052 | 0.047 | 0.051 | 0.050 | 0.052 |
| 4 | 48 | 0.90 | | 0.050 | 0.055 | 0.051 | 0.057 | 0.052 | 0.056 |
| 4 | 48 | 0.75 | | 0.050 | 0.060 | 0.053 | 0.061 | 0.056 | 0.057 |
| 4 | 48 | 0.50 | | 0.050 | 0.055 | 0.055 | 0.061 | 0.055 | 0.060 |
| 6 | 72 | 1.00 | | 0.053 | 0.052 | 0.048 | 0.049 | 0.050 | 0.053 |
| 6 | 72 | 0.90 | | 0.052 | 0.051 | 0.056 | 0.062 | 0.054 | 0.053 |
| 6 | 72 | 0.75 | | 0.058 | 0.059 | 0.051 | 0.054 | 0.057 | 0.058 |
| 6 | 72 | 0.50 | | 0.061 | 0.059 | 0.059 | 0.063 | 0.051 | 0.054 |
| 8 | 96 | 1.00 | | 0.051 | 0.051 | 0.042 | 0.044 | 0.052 | 0.052 |
| 8 | 96 | 0.90 | | 0.055 | 0.055 | 0.044 | 0.047 | 0.047 | 0.048 |
| 8 | 96 | 0.75 | | 0.057 | 0.057 | 0.049 | 0.054 | 0.054 | 0.055 |
| 8 | 96 | 0.50 | | 0.054 | 0.054 | 0.054 | 0.059 | 0.056 | 0.054 |

Table 5.14: Empirical Type I error rates for the drop test, Design 1, H=2, L=3, K=6 and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration.
### Table 5.15: Empirical Type I error rates for the drop test, Design 1, H=2, L=4, K=12 and simulation size 5000.

<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>ρ</th>
<th>Underlying distribution</th>
<th>Approximated distribution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>Log-normal</td>
<td>t3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Avg</td>
<td>Each</td>
<td>Avg</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>1.00</td>
<td>0.052</td>
<td>0.055</td>
<td>0.054</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>0.90</td>
<td>0.052</td>
<td>0.054</td>
<td>0.057</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>0.75</td>
<td>0.053</td>
<td>0.058</td>
<td>0.060</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>0.50</td>
<td>0.049</td>
<td>0.054</td>
<td>0.062</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>1.00</td>
<td>0.048</td>
<td>0.052</td>
<td>0.060</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>0.90</td>
<td>0.055</td>
<td>0.053</td>
<td>0.049</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>0.75</td>
<td>0.054</td>
<td>0.051</td>
<td>0.057</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>0.50</td>
<td>0.055</td>
<td>0.054</td>
<td>0.062</td>
</tr>
<tr>
<td>6</td>
<td>144</td>
<td>1.00</td>
<td>0.050</td>
<td>0.052</td>
<td>0.048</td>
</tr>
<tr>
<td>6</td>
<td>144</td>
<td>0.90</td>
<td>0.045</td>
<td>0.052</td>
<td>0.054</td>
</tr>
<tr>
<td>6</td>
<td>144</td>
<td>0.75</td>
<td>0.051</td>
<td>0.054</td>
<td>0.055</td>
</tr>
<tr>
<td>6</td>
<td>144</td>
<td>0.50</td>
<td>0.055</td>
<td>0.052</td>
<td>0.060</td>
</tr>
<tr>
<td>8</td>
<td>192</td>
<td>1.00</td>
<td>0.051</td>
<td>0.053</td>
<td>0.052</td>
</tr>
<tr>
<td>8</td>
<td>192</td>
<td>0.90</td>
<td>0.049</td>
<td>0.049</td>
<td>0.055</td>
</tr>
<tr>
<td>8</td>
<td>192</td>
<td>0.75</td>
<td>0.051</td>
<td>0.048</td>
<td>0.061</td>
</tr>
<tr>
<td>8</td>
<td>192</td>
<td>0.50</td>
<td>0.047</td>
<td>0.045</td>
<td>0.048</td>
</tr>
</tbody>
</table>

*“Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration.*
<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>ρ</th>
<th>Normal Avg</th>
<th>Each</th>
<th>Log-normal Avg</th>
<th>Each</th>
<th>t3 Avg</th>
<th>Each</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>36</td>
<td>1.00</td>
<td>0.065</td>
<td>0.082</td>
<td>0.057</td>
<td>0.073</td>
<td>0.062</td>
<td>0.080</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>0.90</td>
<td>0.080</td>
<td>0.103</td>
<td>0.066</td>
<td>0.089</td>
<td>0.077</td>
<td>0.095</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>0.75</td>
<td>0.076</td>
<td>0.100</td>
<td>0.072</td>
<td>0.091</td>
<td>0.079</td>
<td>0.094</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>0.50</td>
<td>0.087</td>
<td>0.093</td>
<td>0.072</td>
<td>0.090</td>
<td>0.063</td>
<td>0.085</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>1.00</td>
<td>0.057</td>
<td>0.066</td>
<td>0.058</td>
<td>0.067</td>
<td>0.058</td>
<td>0.068</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>0.90</td>
<td>0.066</td>
<td>0.072</td>
<td>0.061</td>
<td>0.073</td>
<td>0.062</td>
<td>0.069</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>0.75</td>
<td>0.066</td>
<td>0.077</td>
<td>0.058</td>
<td>0.075</td>
<td>0.065</td>
<td>0.077</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>0.50</td>
<td>0.071</td>
<td>0.075</td>
<td>0.065</td>
<td>0.072</td>
<td>0.066</td>
<td>0.076</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>1.00</td>
<td>0.061</td>
<td>0.064</td>
<td>0.055</td>
<td>0.062</td>
<td>0.059</td>
<td>0.065</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>0.90</td>
<td>0.060</td>
<td>0.067</td>
<td>0.054</td>
<td>0.067</td>
<td>0.062</td>
<td>0.069</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>0.75</td>
<td>0.071</td>
<td>0.077</td>
<td>0.067</td>
<td>0.069</td>
<td>0.071</td>
<td>0.075</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>0.50</td>
<td>0.060</td>
<td>0.068</td>
<td>0.063</td>
<td>0.066</td>
<td>0.064</td>
<td>0.067</td>
</tr>
</tbody>
</table>

Table 5.16: Empirical Type I error rates for the drop test, Design 2, H=3, L=3, K=3, and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration.
Table 5.17: Empirical Type I error rates for the drop test, Design 2, H=4, L=4, K=4, and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration.
<table>
<thead>
<tr>
<th>n</th>
<th>N</th>
<th>ρ</th>
<th>Underlying distribution</th>
<th>Approximated distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>Avg</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>1.00</td>
<td>0.057 0.064</td>
<td>0.054 0.059</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0.90</td>
<td>0.070 0.076</td>
<td>0.067 0.078</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0.75</td>
<td>0.066 0.072</td>
<td>0.062 0.068</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0.50</td>
<td>0.069 0.067</td>
<td>0.067 0.071</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>1.00</td>
<td>0.059 0.064</td>
<td>0.051 0.055</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>0.90</td>
<td>0.071 0.067</td>
<td>0.062 0.069</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>0.75</td>
<td>0.057 0.064</td>
<td>0.067 0.066</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>0.50</td>
<td>0.060 0.062</td>
<td>0.057 0.063</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>1.00</td>
<td>0.055 0.056</td>
<td>0.050 0.057</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>0.90</td>
<td>0.063 0.066</td>
<td>0.059 0.060</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>0.75</td>
<td>0.059 0.060</td>
<td>0.054 0.058</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>0.50</td>
<td>0.062 0.061</td>
<td>0.058 0.063</td>
</tr>
</tbody>
</table>

Table 5.18: Empirical Type I error rates for the drop test, Design 2, H=5, L=5, K=5, and simulation size 5000. “Avg” stands for the approach using a single critical value in all iterations. “Each” stands for the approach using a particular critical value in each iteration.
5.3 Empirical Power

In this section we present results of a simulation study conducted to investigate the empirical power of the proposed tests in Chapter 3 for Design 1, Design 2 and CRD. In this part of the simulation, we only consider the case when the number of treatments is $L = 3$. The simulation study contains four additional simulation parameters, replication size ($n$), the correlation coefficient ($\rho$), location shift parameter ($\Delta$), and the underlying distribution ($F$). To make a fair comparison among these designs, we match the sample sizes. The simulation parameters in Design 1 are set to be $n = 3, 6, 9$, $H = 2$ and $K = 6$, while the same parameters in Design 2 are assigned to be $n = 4, 8, 12$, $K = 3$ for set size $H = 3$ and $n = 3, 6, 9$, $K = 6$ for set size $H = 2$. As a result, both designs have the same total sample sizes $N = 36, 72, 108$, respectively. As in the previous section, the correlation coefficient $\rho$ in the Dell-Clutter model is selected to be $\rho = 1.0, 0.9, 0.75$, or 0.5.

The power of a test needs to be evaluated under an alternative model. We use a directional alternative model in which we fix $\beta_1 = 0$ and shift the parameter $\beta_2$ from 0 to 1 in increments of 0.1. As in section 5.1, the within-set residuals in ORRD are generated from the standard normal distribution, Student’s t-distribution with 3 degrees of freedom and log-normal distribution.

For each of the combinations of the simulation parameters, $n, \rho, \Delta$ and underlying distribution $F$, 5000 data sets are generated. Power curves are constructed based on a 5% test. The empirical power is computed from the percent of tests rejecting the null hypothesis among the 5000 replications. The critical values are constructed from the 95-th percentile of the 5000 test statistics when $\Delta = 0$.

Figures 5.1-5.9 illustrate the empirical power curves for 5% tests when the total sample size is $N = 36, 72$, and 108. In each figure, power curves are grouped in separate panels for each $\rho = 1, 0.9, 0.75$, and 0.5. The three different designs are
marked in different line styles in each panel. The solid, dashed and dashed-dotted lines represent the power curves of CRD, Design 1 and Design 2, respectively. The score, Wald, and drop tests are indicated by different symbols: dot, square, and triangle, respectively.

The discussion in Section 5.2 indicates that the tests do not achieve the exact value of the nominal size of the test at an $\alpha = 0.05$ significance level. To have a fair comparison among the empirical powers of all the tests, they must have the same size (Type I error rate). To achieve this objective, we ran a simulation study to determine the critical value of each test for each one of the simulation parameter combinations. The 95-th percentile of the test statistics calculated from the 5,000 simulation runs under the null hypothesis ($\Delta = 0$) was selected to be the critical value for that particular test. The empirical powers in Figures 5.1-5.9 are computed based on these empirical cut-off points, so that we have all the power curves aligned at 0.05 when the shift is 0.

Figures 5.1-5.9 reveal several noticeable differences among the different testing procedures and sampling designs. One can easily see that there is a significant difference between the power curves of the tests based on ORRDs and CRD. The power curves of the tests based on ORRDs always lie at or above the power curves of the tests based on CRD, as long as $\rho > 0.5$. If $\rho = 0.5$, it appears that ORRD and CRD tests perform equally well.

The power plots of the tests based on Design 1 and Design 2 appear to be different when the quality of ranking information is good ($\rho \geq 0.75$). Design 2 is at least as powerful as Design 1, and achieves the largest power gain under a perfect ranking model ($\rho = 1$). The efficiency gain decreases as the quality of ranking information becomes poor ($\rho$ decreases). Both designs have almost identical power curves at $\rho = 0.5$. 

116
We also observe that the drop, Wald and score tests, have comparable power curves for a given design. Even though the drop test appears to have a power curve that lies slightly above the other two power curves, the differences are not very noticeable.

Figures 5.1-5.9 also indicate that the power curves appear to be affected by the shape of the underlying distribution. For heavy and long tail distributions, such as Student’s t-distribution with 3 degrees of freedom, the slope of the power curves is not as steep as the slope of the power curves of the tests when the underlying distribution has lighter tails.

Based on the theoretical developments in Chapters 2 and 3, as well as the empirical evidence in this Chapter 5, we conclude that the tests we constructed in Chapter 3 outperform the comparable tests constructed based on a CRD. It is also clear that Design 2 provides better efficiency results than Design 1.
Figure 5.1: Empirical power curves of the score, Wald and drop tests. The underlying distribution is normal, and total sample size is N=36.
Figure 5.2: Empirical power curves of the score, Wald and drop tests. The underlying distribution is normal, and total sample size is N=72.
Figure 5.3: Empirical power curves of the score, Wald and drop tests. The underlying distribution is normal, and total sample size is N=108.
Figure 5.4: Empirical power curves of the score, Wald and drop tests. The underlying distribution is log-normal, and total sample size is N=36.
Figure 5.5: Empirical power curves of the score, Wald and drop tests. The underlying distribution is log-normal, and total sample size is N=72.
Figure 5.6: Empirical power curves of the score, Wald and drop tests. The underlying distribution is log-normal, and total sample size is N=108.
Figure 5.7: Empirical power curves of the score, Wald and drop tests. The underlying distribution is student’s t with 3 degrees of freedom, and total sample size is N=36.
Figure 5.8: Empirical power curves of the score, Wald and drop tests. The underlying distribution is student’s t with 3 degrees of freedom, and total sample size is N=72.
Figure 5.9: Empirical power curves of the score, Wald and drop tests. The underlying distribution is student's t with 3 degrees of freedom, and total sample size is N=108.
CHAPTER 6
DATA STUDY: AIDS CLINICAL TRIALS GROUP
PROTOCOL 116B/117 STUDY

6.1 Background

In this chapter, the testing procedures developed in Chapter 3 are applied to a large scale clinical trial data set. An ideal example to illustrate the use of ORRD should come from a data set generated from an ORRD. In the absence of such a data set, we rely on a large scale clinical trial to create a data set that resembles an ORRD as closely as possible.

One such clinical trial, ACTG 116B/117, is available from the US Department of Commerces National Technical Information Service at web site http://www.ntis.gov. This clinical trial compares three types of therapies, 750mg didanosine (treatment A), 500mg zidovudine (treatment B), and 500mg didanosine (Treatment C) on patients with the human immunodeficiency virus Type I (HIV-1). The experiment was conducted as a mulitcenter, randomized and double-blind clinical trail. It compared clinical efficacy and safety of switching from zidovudine to didanosine for patients who had already tolerated at least 16 weeks of zidovudine. The initial health status of a patient is accounted for by CD4 cell counts. In molecular biology, immune cells, such as T helper cells, monocytes, macrophages and dendritic cells which have a type of glycoprotein (cluster of differentiation 4) on their surface areas, are called...
CD4 cells. The CD4 cell count for a healthy person varies between 500 to 1500 cells per cubic millimeter. CD4 cell counts for people who have acquired immunodeficiency syndrome (AIDS) are much lower than a CD4 cell counts of a healthy person. Thus patients are stratified based on their initial CD4 cell counts prior to the experiment.

A total 913 patients were recruited for the study. All patients met the established criteria in age, the CD4 baseline level, and the length of the previous zidovudine therapy. Of those patients, 82% were white, 79% were homosexual or bisexual, and 96% were male. The median age was 36 years.

Any individual who had a new AIDS-defining event or died during the experiment was dropped from the study. The study found significantly fewer new AIDS-defining events and deaths among the patients who were assigned to treatment C (500 mg didanosine per day) than among the patients who were assigned to treatment B (500 mg zidovudine per day). Based on this clinical trial, researchers suggested that changing treatment from zidovudine to didanosine can be a helpful therapy to slow the progress of HIV disease.

6.2 Illustrative Example

To construct a data set that resembles an ORRD, we use CD4 cell counts at week 24 as the response variable. A within-set correlation structure is constructed by ranking the patients based on initial CD4 cell counts. The initial and the 24-th week CD4 cell counts are positively correlated with correlation coefficient 0.817. Out of 913 subjects, there were only 469 patients with no missing CD4 cell counts. We treat these 469 patients as our population. Among these 469 patients, 133 patients were in the control group treated with 500 mg zidovudine per day (Treatment B), 160 patients were in the group treated with 500 mg per day of didanosine (Treatment C), and 176 patients were in the group treated with 750 mg per day of didanosine.
(Treatment A). From this population, we sampled three data sets, one for Design 1 (Table 6.1), one for the Design 2 (Table 6.2), and one for a CRD (Table 6.4).

For Design 1, we generated a data set with $H = 2$, $K = 6$, and $n = 3$. This design yields a total of $N = 36$ patients. In this setting, Design 1 needs six sets for each replication, two sets for each contrasts, AB, AC and BC. We explain these contrasts in detail below.

For contrast AB, we need two sets, Set 1 and Set 2:

\[
\text{Set 1: } [AR_1, BR_2] \quad \text{and} \quad \text{Set 2: } [BR_1, AR_2].
\]

where the Set 1 indicates that the first patient has the smaller initial CD4 cell count ($R_1$) and is assigned to treatment A, and the second patient has the larger initial CD4 cell count ($R_2$) and is assigned to treatment B. The second set can be interpreted in a similar fashion. To construct these sets for our study, two patients, one from each treatment group A and B, are selected at random. The selected patients are ranked based on their initial CD4 cell counts. If the treatment and rank configuration matches either one of Set 1 or Set 2, these patients are selected for the contrast AB. Otherwise, the patients are returned back to the population, and the process is repeated until matches are found for both Set 1 and Set 2. The construction of the sets for the other contrast designs, AC and BC, is completed in a similar fashion. The entire process needs to be repeated 3 times to achieve the total sample size $N = 36$. The data set for this design is given in Table 6.1.

The Design 2 data is generated with the set size $H = 3$, number of sets $K = 3$ and the number of replications $n = 4$. This design has a Latin-square structure with three treatments (A, B, C), three rows (Set 1, Set 2, Set 3), and three columns ($R_1$, $R_2$, $R_3$). Notice that in a Latin-square design, each row (set) and column (ranks)
must have each treatment level only once. An example of a Latin-square allocation can be illustrated as follows:

Set 1: \([AR_1 , BR_2 , CR_3]\),

Set 2: \([BR_1 , CR_2 , AR_3]\),

and

Set 3: \([CR_1 , AR_2 , BR_3]\),

where Set 1 indicates that the three patients in the set are assigned to the treatments A, B, C, and had ranks \(R_1, R_2, R_3\) for their initial CD4 cell counts, respectively. The other sets can be interpreted in a similar fashion.

For this design, the ORRD data are collected in the following way. A potential set of 3 patients is chosen by selecting one patient at random from each of the three treatment groups. This potential set of patients is then ranked according to their initial CD4 cell counts. If the rank of treatment assignments matches the Set 1 configuration for this design, this set would be officially sampled, and the selected patients are deleted from the population. If the configuration does not match, the potential set is rejected and the patients are returned to the population. We continue to obtain another potential set until we have a match for Set 1. We perform the same process for Set 2 and Set 3. We repeat this entire process 4 times to have the needed sample size \(N = 36\). The generated data are given in Table 6.2.

In these data sets, we are interested in testing if the median treatment effects are significantly different. Let \(\theta_A, \theta_B, \) and \(\theta_C\) be the medians of treatments A, B, and C,
respectively. We wish to test

\[ H_0 : \theta_A = \theta_B = \theta_C \text{ versus } H_A : \text{ At least one median is different from the others.} \]

Both data sets in Tables 6.1 and 6.2 are analyzed using the tests we developed in Chapter 3. The same data sets were also analyzed by treating them as data sets generated from a CRD with the initial CD4 cell counts as a covariate. Observed test statistics along with their p-values are given in Table 6.3. Analyses of the data sets based on the score, Wald, and drop tests yield slightly different results. The p-values in Table 6.3 indicate that Wald and drop tests reject the null hypothesis while the score test fails to reject the null hypothesis at a 5% significance level. If the structure in ORRD is ignored and the data set is treated as a data set from CRD, all three tests fail to reject the null hypothesis at a 5% significance level.

We also selected a data set using the CRD. This data set contains 12 randomly selected patients from each treatment group and is presented in Table 6.4. This data is analyzed based on the score, Wald, and drop tests. The observed test statistics and p-values are presented in Table 6.5. It is clear that none of these three tests reject the null hypothesis at a 5% significance level.
<table>
<thead>
<tr>
<th>ID</th>
<th>TRT</th>
<th>RNK</th>
<th>CDBEG</th>
<th>CDEND</th>
<th>ID</th>
<th>TRT</th>
<th>RNK</th>
<th>CDBEG</th>
<th>CDEND</th>
</tr>
</thead>
<tbody>
<tr>
<td>361</td>
<td>A</td>
<td>1</td>
<td>27</td>
<td>14</td>
<td>877</td>
<td>C</td>
<td>1</td>
<td>74</td>
<td>52</td>
</tr>
<tr>
<td>730</td>
<td>B</td>
<td>2</td>
<td>145</td>
<td>210</td>
<td>308</td>
<td>A</td>
<td>2</td>
<td>258</td>
<td>285</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>1</td>
<td>184</td>
<td>74</td>
<td>358</td>
<td>B</td>
<td>1</td>
<td>100</td>
<td>56</td>
</tr>
<tr>
<td>705</td>
<td>A</td>
<td>2</td>
<td>192</td>
<td>47</td>
<td>206</td>
<td>C</td>
<td>2</td>
<td>215</td>
<td>20</td>
</tr>
<tr>
<td>503</td>
<td>A</td>
<td>1</td>
<td>31</td>
<td>45</td>
<td>158</td>
<td>C</td>
<td>1</td>
<td>56</td>
<td>59</td>
</tr>
<tr>
<td>890</td>
<td>C</td>
<td>2</td>
<td>119</td>
<td>41</td>
<td>572</td>
<td>B</td>
<td>2</td>
<td>227</td>
<td>198</td>
</tr>
<tr>
<td>911</td>
<td>C</td>
<td>1</td>
<td>37</td>
<td>6</td>
<td>317</td>
<td>A</td>
<td>1</td>
<td>161</td>
<td>130</td>
</tr>
<tr>
<td>520</td>
<td>A</td>
<td>2</td>
<td>195</td>
<td>112</td>
<td>230</td>
<td>B</td>
<td>2</td>
<td>171</td>
<td>252</td>
</tr>
<tr>
<td>544</td>
<td>B</td>
<td>1</td>
<td>15</td>
<td>43</td>
<td>420</td>
<td>B</td>
<td>1</td>
<td>120</td>
<td>50</td>
</tr>
<tr>
<td>278</td>
<td>C</td>
<td>2</td>
<td>268</td>
<td>440</td>
<td>467</td>
<td>A</td>
<td>2</td>
<td>213</td>
<td>201</td>
</tr>
<tr>
<td>707</td>
<td>C</td>
<td>1</td>
<td>51</td>
<td>34</td>
<td>374</td>
<td>A</td>
<td>1</td>
<td>163</td>
<td>248</td>
</tr>
<tr>
<td>183</td>
<td>B</td>
<td>2</td>
<td>165</td>
<td>122</td>
<td>90</td>
<td>C</td>
<td>2</td>
<td>166</td>
<td>110</td>
</tr>
<tr>
<td>124</td>
<td>A</td>
<td>1</td>
<td>135</td>
<td>76</td>
<td>428</td>
<td>C</td>
<td>1</td>
<td>50</td>
<td>37</td>
</tr>
<tr>
<td>743</td>
<td>B</td>
<td>2</td>
<td>151</td>
<td>148</td>
<td>342</td>
<td>A</td>
<td>2</td>
<td>74</td>
<td>48</td>
</tr>
<tr>
<td>363</td>
<td>B</td>
<td>1</td>
<td>60</td>
<td>26</td>
<td>902</td>
<td>B</td>
<td>1</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>81</td>
<td>A</td>
<td>2</td>
<td>357</td>
<td>420</td>
<td>216</td>
<td>C</td>
<td>2</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>689</td>
<td>A</td>
<td>1</td>
<td>11</td>
<td>6</td>
<td>826</td>
<td>C</td>
<td>1</td>
<td>29</td>
<td>9</td>
</tr>
<tr>
<td>384</td>
<td>C</td>
<td>2</td>
<td>51</td>
<td>45</td>
<td>221</td>
<td>B</td>
<td>2</td>
<td>203</td>
<td>136</td>
</tr>
</tbody>
</table>

Table 6.1: ORRD data set for Design 1 with $n = 3$, $H = 2$, and $K = 6$. TRT = treatment, RNK = rank, CDBEG = initial CD4 cell counts, CDEND = 24th week CD4 cell counts, and ID = patient identification number.
<table>
<thead>
<tr>
<th>ID</th>
<th>TRT</th>
<th>RNK</th>
<th>CDBEG</th>
<th>CDEND</th>
<th>ID</th>
<th>TRT</th>
<th>RNK</th>
<th>CDBEG</th>
<th>CDEND</th>
</tr>
</thead>
<tbody>
<tr>
<td>725</td>
<td>A</td>
<td>1</td>
<td>25</td>
<td>80</td>
<td>256</td>
<td>A</td>
<td>1</td>
<td>85</td>
<td>92</td>
</tr>
<tr>
<td>668</td>
<td>B</td>
<td>2</td>
<td>64</td>
<td>34</td>
<td>497</td>
<td>B</td>
<td>2</td>
<td>150</td>
<td>201</td>
</tr>
<tr>
<td>325</td>
<td>C</td>
<td>3</td>
<td>263</td>
<td>357</td>
<td>525</td>
<td>C</td>
<td>3</td>
<td>291</td>
<td>257</td>
</tr>
<tr>
<td>384</td>
<td>C</td>
<td>1</td>
<td>51</td>
<td>45</td>
<td>130</td>
<td>C</td>
<td>1</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>870</td>
<td>A</td>
<td>2</td>
<td>105</td>
<td>82</td>
<td>139</td>
<td>A</td>
<td>2</td>
<td>58</td>
<td>31</td>
</tr>
<tr>
<td>893</td>
<td>B</td>
<td>3</td>
<td>203</td>
<td>169</td>
<td>119</td>
<td>B</td>
<td>3</td>
<td>235</td>
<td>230</td>
</tr>
<tr>
<td>722</td>
<td>B</td>
<td>1</td>
<td>45</td>
<td>90</td>
<td>896</td>
<td>B</td>
<td>1</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>827</td>
<td>C</td>
<td>2</td>
<td>156</td>
<td>119</td>
<td>100</td>
<td>C</td>
<td>2</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>897</td>
<td>A</td>
<td>3</td>
<td>226</td>
<td>194</td>
<td>520</td>
<td>A</td>
<td>3</td>
<td>195</td>
<td>112</td>
</tr>
<tr>
<td>774</td>
<td>A</td>
<td>1</td>
<td>15</td>
<td>10</td>
<td>503</td>
<td>A</td>
<td>1</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>14</td>
<td>B</td>
<td>2</td>
<td>42</td>
<td>8</td>
<td>721</td>
<td>B</td>
<td>2</td>
<td>90</td>
<td>60</td>
</tr>
<tr>
<td>1</td>
<td>C</td>
<td>3</td>
<td>209</td>
<td>156</td>
<td>303</td>
<td>C</td>
<td>3</td>
<td>158</td>
<td>120</td>
</tr>
<tr>
<td>346</td>
<td>C</td>
<td>1</td>
<td>48</td>
<td>14</td>
<td>450</td>
<td>C</td>
<td>1</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>148</td>
<td>A</td>
<td>2</td>
<td>72</td>
<td>97</td>
<td>40</td>
<td>A</td>
<td>2</td>
<td>87</td>
<td>80</td>
</tr>
<tr>
<td>750</td>
<td>B</td>
<td>3</td>
<td>264</td>
<td>339</td>
<td>881</td>
<td>B</td>
<td>3</td>
<td>245</td>
<td>198</td>
</tr>
<tr>
<td>604</td>
<td>B</td>
<td>1</td>
<td>25</td>
<td>17</td>
<td>578</td>
<td>B</td>
<td>1</td>
<td>63</td>
<td>73</td>
</tr>
<tr>
<td>36</td>
<td>C</td>
<td>2</td>
<td>106</td>
<td>90</td>
<td>211</td>
<td>C</td>
<td>2</td>
<td>116</td>
<td>78</td>
</tr>
<tr>
<td>874</td>
<td>A</td>
<td>3</td>
<td>183</td>
<td>109</td>
<td>350</td>
<td>A</td>
<td>3</td>
<td>188</td>
<td>138</td>
</tr>
</tbody>
</table>

Table 6.2: ORRD data set for the Design 2 with \( n = 4, H = 3, \) and \( K = 3. \) TRT = treatment, RNK = rank, CDBEG = initial CD4 cell counts, CDEND = 24th week CD4 cell counts, and ID = patient identification number.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Design 1</td>
<td>Score</td>
<td>3.353</td>
<td>0.063</td>
<td>CRD</td>
<td>Score</td>
<td>0.166</td>
<td>0.848</td>
</tr>
<tr>
<td>Design 1</td>
<td>Wald</td>
<td>5.491</td>
<td>0.016</td>
<td>CRD</td>
<td>Wald</td>
<td>0.170</td>
<td>0.844</td>
</tr>
<tr>
<td>Design 1</td>
<td>Drop</td>
<td>64.556</td>
<td>0.015</td>
<td>CRD</td>
<td>Drop</td>
<td>0.415</td>
<td>0.813</td>
</tr>
<tr>
<td>Design 2</td>
<td>Score</td>
<td>3.224</td>
<td>0.088</td>
<td>CRD</td>
<td>Score</td>
<td>0.826</td>
<td>0.447</td>
</tr>
<tr>
<td>Design 2</td>
<td>Wald</td>
<td>5.456</td>
<td>0.028</td>
<td>CRD</td>
<td>Wald</td>
<td>0.693</td>
<td>0.507</td>
</tr>
<tr>
<td>Design 2</td>
<td>Drop</td>
<td>30.389</td>
<td>0.013</td>
<td>CRD</td>
<td>Drop</td>
<td>1.552</td>
<td>0.460</td>
</tr>
</tbody>
</table>

Table 6.3: Observed statistics and p-values of the score, Wald and drop tests for the ORRD data sets in Tables 6.1 and 6.2. The p-values of Wald and score tests are computed based on an F-approximation. The p-values of the drop tests are computed based on a \( \chi^2 \) approximation. In the right half, the structure in ORRD is ignored and the data sets are treated as data sets from CRD.
Table 6.4: CRD data set with 3 treatments, each of which has sample size 12. TRT = treatment, CDBEG = initial CD4 cell counts, CDEND = 24th week CD4 cell counts, and ID = patient identification number.

<table>
<thead>
<tr>
<th>ID</th>
<th>TRT</th>
<th>CDBEG</th>
<th>CDEND</th>
<th>ID</th>
<th>TRT</th>
<th>CDBEG</th>
<th>CDEND</th>
</tr>
</thead>
<tbody>
<tr>
<td>59</td>
<td>A</td>
<td>52</td>
<td>48</td>
<td>538</td>
<td>B</td>
<td>215</td>
<td>12</td>
</tr>
<tr>
<td>761</td>
<td>A</td>
<td>139</td>
<td>261</td>
<td>156</td>
<td>B</td>
<td>103</td>
<td>38</td>
</tr>
<tr>
<td>689</td>
<td>A</td>
<td>11</td>
<td>6</td>
<td>881</td>
<td>B</td>
<td>245</td>
<td>198</td>
</tr>
<tr>
<td>375</td>
<td>A</td>
<td>67</td>
<td>41</td>
<td>438</td>
<td>B</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>556</td>
<td>A</td>
<td>224</td>
<td>362</td>
<td>261</td>
<td>B</td>
<td>34</td>
<td>9</td>
</tr>
<tr>
<td>256</td>
<td>A</td>
<td>85</td>
<td>92</td>
<td>836</td>
<td>B</td>
<td>250</td>
<td>188</td>
</tr>
<tr>
<td>717</td>
<td>A</td>
<td>165</td>
<td>320</td>
<td>404</td>
<td>C</td>
<td>25</td>
<td>6</td>
</tr>
<tr>
<td>643</td>
<td>A</td>
<td>122</td>
<td>68</td>
<td>713</td>
<td>C</td>
<td>216</td>
<td>231</td>
</tr>
<tr>
<td>728</td>
<td>A</td>
<td>9</td>
<td>30</td>
<td>12</td>
<td>C</td>
<td>21</td>
<td>24</td>
</tr>
<tr>
<td>20</td>
<td>A</td>
<td>224</td>
<td>216</td>
<td>742</td>
<td>C</td>
<td>99</td>
<td>88</td>
</tr>
<tr>
<td>365</td>
<td>A</td>
<td>81</td>
<td>34</td>
<td>105</td>
<td>C</td>
<td>205</td>
<td>210</td>
</tr>
<tr>
<td>440</td>
<td>A</td>
<td>195</td>
<td>70</td>
<td>861</td>
<td>C</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>269</td>
<td>B</td>
<td>114</td>
<td>99</td>
<td>859</td>
<td>C</td>
<td>178</td>
<td>126</td>
</tr>
<tr>
<td>357</td>
<td>B</td>
<td>251</td>
<td>270</td>
<td>27</td>
<td>C</td>
<td>94</td>
<td>123</td>
</tr>
<tr>
<td>638</td>
<td>B</td>
<td>21</td>
<td>17</td>
<td>792</td>
<td>C</td>
<td>153</td>
<td>169</td>
</tr>
<tr>
<td>547</td>
<td>B</td>
<td>232</td>
<td>159</td>
<td>472</td>
<td>C</td>
<td>265</td>
<td>126</td>
</tr>
<tr>
<td>315</td>
<td>B</td>
<td>170</td>
<td>144</td>
<td>499</td>
<td>C</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>273</td>
<td>B</td>
<td>145</td>
<td>141</td>
<td>839</td>
<td>C</td>
<td>75</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 6.5: Observed statistics and p-values of the score, Wald and drop tests for the CRD data set in Table 6.4. The p-values of the Wald and score tests are computed based on an F-approximation. The p-values of the drop test are computed based on a $\chi^2$ approximation.

<table>
<thead>
<tr>
<th>Design</th>
<th>Test</th>
<th>Obs. Stat.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRD</td>
<td>Score</td>
<td>1.850</td>
<td>0.173</td>
</tr>
<tr>
<td>CRD</td>
<td>Wald</td>
<td>1.825</td>
<td>0.177</td>
</tr>
<tr>
<td>CRD</td>
<td>Drop</td>
<td>3.966</td>
<td>0.138</td>
</tr>
</tbody>
</table>
7.1 Conclusions

One of the main objectives in the construction of a design of an experiment is to create homogeneous groups of experimental units to reduce the total variation. This can be achieved using all available information on the experimental units. The traditional designs, such as block and Latin square designs, create blocks only when the information can be translated into a small number of blocks. If the information is not precisely defined or is subjective in nature, it is usually ignored. Even this kind of subjective information, however, can be very helpful in reducing the total variation in an experiment if it is used properly. Ranked set sampling designs use subjective information through the ranking of a small set of experimental units. The ranking process creates a stratified sample and leads to reduction in the total variation of the experiment.

The use of a standard RSS procedure poses some practical challenges. One of the difficulties is that an RSS design requires a larger number of EUs to conduct the experiment. This is a challenge in settings where available experimental units are limited. Another concern is that it is not clear how the treatment allocation should be arranged to increase the efficiency of the experiments. This dissertation uses ORRD
designs to address these challenges. The ORRD uses all available EUs in a set along with a randomization technique to allocate the treatment levels to EUs.

Chapter 1 provided a literature review on ORRD and related designs. Chapter 2 introduced rank regression estimators to fit an additive model to ORRD data. We estimate the parameters of the regression model by projecting the response vector onto a subspace with respect to a rank-based dispersion function. Under some mild regularity conditions, we showed that the asymptotic distribution of the rank regression estimator is multivariate normal.

Chapter 3 developed three tests, the drop, score, and Wald tests, for generalized linear hypotheses. We showed that the asymptotic null distribution of the drop test is a linear combination of independent chi-squared random variables with 1 degree of freedom. The asymptotic null distribution of the other two tests is a chi-squared distribution with \( q \) degrees of freedom, where \( q \) is the dimension of the parameter of interest under the null hypothesis.

Chapters 4 and 5 addressed some computational issues and provided empirical evidence for the performance of the tests. The simulation results indicate that the Type I error rates of the calibrated tests are close to the nominal level of 0.05 for a wide range of judgment ranking information quality. Another simulation study of power comparisons showed that the proposed tests under ORRD outperform their competitors in CRD. This simulation also indicated that Design 2 is more powerful than Design 1.

Chapter 6 illustrated the use of the proposed tests on a clinical trial data set.

7.2 Future Research

This dissertation addressed only two particular ORRDs: Design 1 (All possible pairwise contrast design) and Design 2 (All contrasts design). For these designs, the
set sizes are selected to be $H = 2$ in Design 1 and $H = L$ in Design 2. One can relax these restrictions and increase the set sizes. In this case, the number of possible designs increases with the set size. Hence, the selection of an optimal design needs to be investigated.

In this research, we assumed that the response variable was an absolutely continuous random variable. Discrete responses are also quite common in many experimental settings. Thus, statistical inference for discrete responses from an ORRD design needs to be developed.

The performance of tests and estimators under ORRDs depends heavily on the quality of the ranking process. One way to improve ranking quality is to use all available ranking information, including multiple rankers.

MacEachern, Stasny and Wolfe (2004) introduced a multi-ranker model in JPS settings. A similar model can be implemented in ORRD to improve the ranking quality.
BIBLIOGRAPHY


Ozturk, O. and Deshpande J. V. (2006). Ranked-set sample nonparametric

Ozturk, O. and MacEachern, S. N. (2007). Order restricted randomized designs


Ozturk, O. and Sun, Y. (2009). Rank-sum test based on order restricted ran-

ranking errors in ranked set sampling. *Journal of Nonparametric Statistics*,
22(7), 823-840.

Ozturk, O. (2013). Combining multi-observer information in partially rank-
ordered judgment post-stratified and ranked set samples. *The Canadian

Ozturk, O. and MacEachern, S. N. (2013). Inference Based on General Linear
Models for Order Restricted Randomization. *Communications in Statistics