Rank Based Methods for Repeated Measurement Data

by

Yankun Gong

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2011

Keywords: Rank, Nonparametric, Predation Preference, Weighted Function, IRLS,
Repeated Measures, Sensitivity Curve, Relative Efficiency

Approved by

Asheber Abebe, Chair, Associate Professor of Mathematics and Statistics
Peng Zeng, Associate Professor of Mathematics and Statistics
Mark Carpenter, Professor of Mathematics and Statistics

Abstract

This dissertation considers rank based methods for one sample and two sample repeated measurement data. As a specific example, in Chapter 2, this dissertation considers nonparametric tests for selective predation. [1] proposed a nonparametric tests for selective predation using the linear score function. Motivated by this method, general rank tests are given for the case of one predatory species and prey characterized by a binary feature of interest and the case of two predatory species and prey characterized by either a continuous or a categorical feature of interest. The score functions used to construct the test statistics are monotone and hence the test is designed to detect simple ordered alternatives. The results based on the asymptotic Gaussian distribution of the test statistics show that the tests retain nominal Type-I error rates. The results also show that power of the asymptotic test depends on the chosen score function. In Chapter 2, we study the one sample and two sample repeated measurement data with random censoring and the simulation results show that we can take the asymptotic distribution as the underlying distribution of the test statistic even with a high censoring rate and a small sample size.

In Chapter 3, this dissertation considers using rank based methods to test the trend of the difference between two samples for two sample repeated measurement data. As a specific example, we consider nonparametric methods for testing whether the rate of prey feature change in the selection of one species is faster than that of another species. Although the Page test is used in conjunction with a single randomized complete block design, we extend it to the situation where we have two randomized complete block designs. We derive the asymptotic distribution of a general test statistic which includes the Page statistic as a special case. The results based on the asymptotic Gaussian distribution of the test statistics show that the tests retain nominal Type-I error rates.

In Chapter 4, the finite sample performance of the rank estimator of regression coefficients obtained using the iteratively reweighted least squares (IRLS) of Sievers and Abebe (2004) is evaluated. Efficiency comparisons show that the IRLS method does quite well in comparison to least squares or the traditional rank estimates in cases of moderate tailed error distributions; however, the IRLS method does not appear to be suitable for heavy tailed data. Moreover, the results show that the IRLS estimator will have an unbounded influence function even if we use an initial estimator with a bounded influence function.

In Chapter 5, this dissertation study how to test the trend of the difference in the two sample repeated measurement data using two generalized estimating equation (GEE) models: the likelihood based GEE model proposed by Liang and Zeger (1986) and the IRLS Rank-based GEE model proposed by [2]. The results show that the GEE model proposed by [2] is robust to outliers in response space and can be used to analyze data with small sample sizes compared to the GEE model proposed by Liang and Zeger (1986).

Acknowledgments

My greatest thanks extend to Dr. Asheber Abebe for his patience and guidance throughout this endeavor. Dr. Abebe took considerable time and energy to further the progress in my studies of Statistics. I also wish to express my appreciation to my parents and my brother Yanfei Gong who have given their love and constant support throughout my life. I would also like to thank Dr. Peng Zeng and Dr. Mark Carpenter on the committee for their contribution to this dissertation and serving as committee members.

Table of Contents

Abstract ii

Acknowledgments iv

List of Figures vii

List of Tables viii

1 Introduction 1

2 Rank Tests for Selective Predation 7

2.1 Introduction 7

2.2 Data and Model in [1] 9

2.2.1 Experiment Frame and Data 9

2.2.2 Test Statistic 10

2.3 General Rank Test for Preference: One-Sample 12

2.4 Two Sample Comparison of Preference Patterns 14

2.4.1 Continuous Prey Feature 15

2.4.2 Categorical Prey Feature 16

2.5 Censored Data 17

2.5.1 One-sample Selection Problem 18

2.5.2 Two Sample Selection Problem 19

2.6 Monte Carlo Simulation 21

2.6.1 Null Simulation 21

2.6.2 Power Simulation 24

2.6.3 Null Simulation for Two Sample Case with Censoring 26

2.7 Conclusion 27

3 Testing the Difference Trend in Two Sample Repeated Measurement Data . . . 28

3.1 Motivation of the Method 28
3.2 Test Statistic and Theory 30
3.3 Monte Carlo Simulation 35
3.3.1 Null Simulation 35
3.3.2 Results for the data in [48] 36
3.4 Conclusion 37
4 Iteratively Reweighted Rank Regression Estimator 39
4.1 Introduction 39
4.2 Reweighted Least Squares Formulation 41
4.3 Computational Details and Results 43
4.3.1 Algorithm for Estimating Regression Coefficients 43
4.3.2 Variance Estimator 46
4.3.3 Influence Function 48
4.4 Tests of Linear Hypotheses 51
4.5 Conclusion 52
5 IRLS Estimation of Rank Generalized Estimation Equations Model 53
5.1 Introduction 53
5.2 Ordinary LS based GEE models 54
5.3 Iterated Reweighted Rank-Based Estimates for GEE Models 56
5.4 Our Case for the Ordinary LS based GEE Model 59
5.5 Our Case for the Iterated Reweighted Rank-Based Estimators for GEE Models 61
5.6 Example 63
5.7 Conclusion 65
6 Summary 66
Bibliography 68
Appendix: Proofs 74

List of Figures

2.1 Figure for Null Rejection Rates in Table 2.2 23

3.1 Boxplots for CRP levels at time 0 and time 24 for the two groups HI and LO . 30

3.2 Mean and median profiles for the CRP levels in two groups HI and LO 37

4.1 Sensitivity Curves 50

5.1 Residul plots for The ordinary GEE and rank-based GEE 64

List of Tables

2.1 Selections of five large or small armyworms in bird trials 10

2.2 Simulated null rejection rates of W_ψ^+ using its asymptotic normal critical constant for continuous distributions, $\psi(t) \propto t,\ \alpha = .05$. 22

2.3 The simulated power of W_ψ^+ using its asymptotic normal distribution where $X_j \sim N(\mu_j, 1),\ Y_j \sim N(0,1),\ 1 \leq j \leq k,\ \alpha = 0.05$ 25

2.4 Simulated null rejection rates of W_ψ^+ using its asymptotic normal critical constant, $\psi(t) \propto t,\ \alpha = .05, P(\varepsilon_{rj} = 1) = P(\xi_{sj} = 1) = .5$. 26

3.1 CRP levels in [48] 29

3.2 Simulated null rejection rates of $\mathbf{W}_\psi$ using its asymptotic normal critical constant, $\psi(t) \propto t,\ \alpha = .05$ and $k = 5$. 36

4.1 Relative efficiencies of IRLS versus HM and LS when the errors follow a t_{df} distribution 45

4.2 Relative efficiencies of IRLS versus HM and LS when the errors follow a $CN(\epsilon, 3)$ distribution 45

4.3 REs of $\widehat{\mathbf{V}}$ of IRLS vs HM when $e_i \sim t_{df}$ 48

4.4 REs of $\widehat{\mathbf{V}}$ of IRLS vs HM when $e_i \sim CN(\epsilon, 3)$ 48

4.5 Empirical α levels for IRLS vs HM 51

Chapter 1

Introduction

In the 70+ years since their origin in the mid-1930s, nonparametric statistical methods have flourished and have emerged as valuable methodology for statisticians and other scientists performing data analysis. Roughly speaking, a nonparametric procedure is a statistical procedure that has certain desirable properties that hold under relatively mild assumptions regarding the underlying population from which the data are obtained. Nonparametric techniques have the following advantages compared to parametric techniques:

1. Nonparametric methods require few assumptions about the underlying population from which the data are drawn. In particular, nonparametric procedures forgo the traditional normality assumption.

2. Nonparametric procedures enable the user to obtain exact p-values for tests, exact coverage probabilities for confidence intervals, exact experimentwise error rates for multiple comparison procedures, and exact coverage probabilities for confidence bands without relying on assumptions that the underlying populations are normal.

3. Nonparametric techniques are often (although not always) easier to apply than their normal theory counterparts.

4. Usually the nonparametric procedures are only slightly less efficient than their normal theory competitors when the underlying populations are normal, and they can be mildly or wildly more efficient than these competitors when the underlying populations are not normal.

5. Nonparametric methods are widely used for studying populations that take on a ranked order (such as movie reviews receiving one to four stars). The use of nonparametric

methods may be necessary when data have a ranking but no clear numerical interpretation, such as when assessing preference; in terms of levels of measurement, for data on an ordinal scale.

6. Due to the reliance on fewer assumptions, nonparametric methods are relatively robust to outlying observations and other violations.

In this dissertation, we discuss the nonparametric techniques in repeated measurement data. Repeated measurement data, in which the same response variable is recorded on each observational unit on several different occasions, occur frequently in many different disciplines. One example is that sequential experiments for treatment comparisons are widely used in clinical (e.g., a new treatment and a control or placebo) or pharmaceutical trials and in many other applied contexts. There are two possible inferential goals in these studies. One goal is to establish a significant overall difference between the two treatments under study, the other goal is after we conclude that there does exist a difference between the two treatments, to check if the difference stays the same, decays, or expands over time.

Several approaches have been considered in the past to deal with comparison of two groups with respect to their repeated measures. A useful and common initial step in the analysis of repeated measures data is to graph the data in some way. A method often employed, particularly in medical publications, is to plot means by treatment group for every time point. An example of such a plot for the data from the trail of two treatments for the control of intestinal parasites in cattle can be found in [3].

Another commonly used method of analysis for repeated measures that involve a number of treatment groups, particularly in medical and related research, is to compare the groups at each time point, by using either t-test or some nonparametric equivalent. [4] suggested that this approach may be quite useful if the occasions are few and the intervals between them are large.

A more relevant, but still relatively straightforward, approach to the analysis of repeated measures data is that involving the use of summary measures, and sometimes known as

response feature analysis. Here the responses for each subject are used to construct a single number that summarizes some aspect of the subject's response profile. (In some cases more than a single summary measure may be used). The summary measure to be used needs to be chosen before the analysis of the data and should, of course, be relevant to the particular questions of interest in the study. [5] gave a list of potentially useful summary measures. [6] suggest three methods of analysis using summary statistics, which are post-treatment means (POST), mean changes (CHANGE) and analysis of covariance (ANCOVA).

The multivariate procedure involves testing on a set of transformed variables representing the within-subject difference of each within-subject factor and their interactions. The hypothesis that the means of a set of transformed variables representing a within-subject factor, or an interaction between within-subject factors, are zero can be tested by using Hotelling's T^2 statistics which is introduced in [7]. To deal with the cases when the variables are not normally distributed or the sample size is small, [8] proposed numerous nonparametric approaches.

The repeated measures data most appropriately analysed by the methods described above are those from designed experiments where all subjects have the same number of observations measured at equivalent time interval. Since these conditions often do not hold in clinical trials, methods to accommodate missing data have been developed ([9]). Many of these are likelihood-based methods ([10], [11]). When values are missing completely at random, non-parametric and semiparametric methods have been developed to analyze such data ([10], [12], [13]). When the missing values are not random, [14] propose a semiparametric estimation procedure for treatment differences over time based on repeated measurements of an outcome variable when the patient's follow-up time may depend on observed and/or missing measurements. They examine the effect of switching to didanosine(ddI) from zidovudine(AZT) for HIV-infected patients who had tolerated AZT for at least 16 weeks, 304 patients were randomly chosen to continue AZT therapy and 298 patients were assigned to

take ddI at a daily dose of 500 mg. The patient's CD4 cell count, a proxy for the progression of HIV infection, was obtained periodically during the trial. In addition to the usual clinical endpoint analysis, they also evaluate the group difference based on such repeated measurements.

Apart from clinical trials, another important application field of repeated measurements is to study the predation patterns in evolutionary Biology. In Chapter 2 of this dissertation, we use the Wilcoxon signed rank statistic to study the repeated measurement data model in [1]. Our intention is to develop nonparametric hypothesis tests for selective predation as a result of values taken by prey feature of interest. We do this by extending the method of [1] to allow for differing gradients in prey selection and random censoring. We also extend the method to study the two sample predation preference.

If we conclude that the two predator species do have different prey preference, in order to test whether the rate of prey feature change in the selections of one species is faster than that of another species, in Chapter 3, we need to use theory of long memory processes.

The phenomenon of long memory had been known long before suitable stochastic models were developed. Scientists in diverse fields of statistical applications observed empirically that correlations between observations that are far apart (in time or space) decay to zero at a slower rate than one would expect from independent data or data following classic Markov-type models. As a result of Mandelbrot's pioneering work, self-similar and related stationary processes with long memory were later introduced to statistics, to provide a sound mathematical basis for statistical inference. Since then, long memory (or long-range dependence) has become a rapidly developing subject. Because of the diversity of applications, the literature on the topic is broadly scattered in a large number of journals, including those in fields such as agronomy, astronomy, chemistry, economics, engineering, environmental sciences, mathematics, physics, geosciences, hydrology, and statistics. Best known is the occurrence of long-range dependence in geophysics and hydrology (for a review, see [15]). In particular, the so-called Hurst Effect ([16]) can be explained by slowly decaying correlations. However,

there are many other fields of application where this type of correlation occurs. As early as 1895 the astronomer Newcomb discussed the phenomenon of long-range dependence in astronomical data sets and called it "semi-systematic" errors. He also proposed a heuristic explanation by superposition of independent random errors and constant systematic errors. In 1902, Karl Pearson observed slowly decaying correlations in simulated astronomical observations. Further examples are discussed, for instance, by [17] and [18] for economical data, [19] for data from biology, geophysics, meteorology and hydrology, [20] for meteorological data, [21] for telecommunication data and [22] and [23] for agricultural data.

In Chapter 3, using stationary process with long-memory theory, we propose a procedure that is analogous to the Page test [24] to study the trend of the difference between two samples of repeated measurement data. Although the Page test is used in conjunction with a single randomized complete block design, we extend it to the situation where we have two randomized complete block designs. We derive the asymptotic distribution of a general test statistic that includes the Page statistic as a special case. This provides an approximate α level test for the changing rates in the simple ordering case.

For the repeated measurement data, when the outcome variable is approximately Gaussian, statistical methods are well developed, e.g. [25] and [26]. For non-Gaussian outcomes, however, less development has taken place. For binary data, repeated measures models in which observations for a subject are assumed to have exchangeable correlations have been proposed by [27] using a probit link, by [28] using a logit link and by [29] using log linear models. Only the model proposed by [28] allows for time-dependent covariates. [10] proposed an extension of generalized linear models to the analysis of longitudinal data. They introduce a class of estimating equations that give consistent estimates of the regression parameters and of their variance under mild assumptions about the time dependence. Using a working generalized linear model for the marginal distribution of the response, the estimating equations are derived without specifying the joint distribution of a subject's observations yet

they reduce to the score equations for multivariate Gaussian outcomes. Asymptotic theory is presented for the general class of estimators.

The methods, however, are not robust against outliers since they are based on score equations from the maximum likelihood method of estimation. A solution proposed by [30] is to use M-type estimation by involving downweighting schemes. Another solution is one given by [31] who proposed an adaptation of the Wilcoxon-Mann-Whitney method of estimating linear regression parameters for use in longitudinal data analysis under the working independence model. They used joint ranking (JR) of all observations in their development. [32] consider the same model as [31] but they use separate between-subject and within-subject ranks to specify their Wilcoxon-Mann-Whitney estimating equation. [2] provided a direct rank estimation analogue of the GEE model of [10] obtained via the minimization of rank dispersion function of [33]. This is a generalization of the work of [31]. Their development used the iterated reweighted least squares (IRLS) formulation of the rank dispersion function given in [34].

In Chapter 4 of this dissertation, we evaluate the finite sample performance of the IRLS estimate of $\widehat{\boldsymbol{\beta}}_R$ and its variance. We discuss the IRLS formulation, provide computational algorithms as well as simulation results pertaining to the estimator, its variance, and its influence function.

In Chapter 5, Using the LS based GEE model in [10] and the IRLS Rank-Based GEE model proposed by [2] (see also [35]), we study the problem of selective predation patterns.

Chapter 2

Rank Tests for Selective Predation

2.1 Introduction

Repeated measurement data, in which the same response variable is recorded on each observational unit on several different occasions, occur frequently in many different disciplines. One major field that generates a great deal of repeated measures data is in Biology, when people study the predatory preference. In this chapter we will study the predatory preference problem as a special case of repeated measure data. All the test statistics we used and the corresponding asymptotic properties can be used for similar repeated measures data.

Predation is necessary for life to exist. It is a force in evolution of species as natural selection is biased in favor of effective predators and elusive prey. We are interested in the case where predatory behavior is guided by certain features of the prey. Dating back to the time of Darwin, scientists have long recognized that signals used by animals to attract mates can have the unintended consequence of attracting predators and resulting in gender selective predation [36]. Conversely, as a recent study of [37] shows, in some species, signaling places the receiver of the signal at a higher risk of predation than the signaler. Several methods have been given in the past to quantify these types of selective predation [38]. As pointed out in [39] and [1], many of these methods have major shortcomings in that they ignore the order in which the predator selects its prey. To that end, [39] proposed a test procedure, based on the maximum likelihood principle, that accounts for the order of selection. The performance of this method was unsatisfactory for small sample sizes. As a remedy, [1] suggested to use a Wilcoxon signed rank statistic to test for predatory preference. Their approach takes into account not only the selection of prey but also the order of in which they are selected.

Further, as most nonparametric methods, the method provides satisfactory results for small samples. However, this method, by using a linear score function, assumes that changes in consecutive prey selections are equally important regardless of when the selection is made.

In this chapter, it will be assumed that there is only one feature of the prey that determines predatory preference. Hereafter, we will refer to this as "prey feature of interest." Our intention is to develop nonparametric hypothesis tests for selective predation as a result of values taken by prey feature of interest. We will do this by extending the method of [1] in three directions:

1. We provide a class of general rank tests for the experiment considered by [1]; that is, testing for selective predation of one random sample of predators where the prey feature of interest takes one of only two possible values. Examples of dichotomous prey features of interest are gender and size categorized as "large" and "small". General rank tests will allow one to place varied emphasis on selections at different stages of the selection experiment. This could be helpful in controlling the manner in which extraneous variables, such as the level of hunger of predators, can affect selection preference at the beginning versus at the end of the experiment.

2. We propose a class of general rank score tests for the difference in predation patterns of two random samples of predators (eg. two species). In this case, we treat the case where the prey feature of interest is continuous ("surface area", for example). We then give the test where prey feature is a categorical variable with m classes ($m \geq 2$). The test statistic for this case becomes equivalent to that of the continuous case when there are no ties in the data.

3. If we look at our data as a repeated measures data, X_{rj} denote the measurement of the rth experimental subject in population $\mathscr{X}$ at condition or time point j, similar for Y_{sj}, we may not have a complete data frame, i.e. some observations of the experimental subjects maybe missing, which is often the case in clinical trials. We propose the

test statistic for randomly missing data for the cases of one sample problem and two samples problem in repeated measures data.

In all the three cases, we study the asymptotic distributions of the test statistics as the number of predators increases while the number of allowed selections remains constant for all predators.

2.2 Data and Model in [1]

2.2.1 Experiment Frame and Data

The purpose of this chapter is to suggest a test statistic that can be useful to establish preference between two response in repeated measures experimental design settings. The context for this study was experimental work conducted at the University of Florida aimed at conservation of naturally occurring enemies of crop pests as a means of improving biological control of cropping systems. At issue was whether birds could be useful in crop pest control due to their consumption of pests that escape mortality from other agents of biological control. Specially, the research investigated aspects of pest insects that would make them more attractive food sources for birds. Trails were conducted using red-winged blackbirds with fall armyworms as their food source. The purpose of the trails was to determine whether the birds preferred larger versus smaller armyworms and whether the birds preferred non-parasitized armyworms to parasitized armyworms of the same size. The parasite used was larva of E.plathypenae, a species common in Florida and previously investigated for its biological control value. A description of the experiment can be found in [40].

In this experiment, a plastic chamber divided into two equal sized parts was used to present each bird with two food choices. The positions of the two foods in the subchambers (left or right) were varied over the trials and no directional preference was observed. In trials to determine preference for food size, for example, each bird was presented with five large armyworms in one subchamber and five small armyworms in the other. The researchers

recorded the first five choices made by the bird, including the order in which the armyworms were selected by the birds. One of the data sets from this experiments is presented in Table 2.1. It displays the selection by each bird: large (L) or small (S), in the order in which they were selected (left to right).

Table 2.1: Selections of five large or small armyworms in bird trials

Trial number	Choice:	1	2	3	4	5
	Score:	5	4	3	2	1
1		L	L	L	L	S
2		L	L	S	L	S
3		S	L	L	L	S
4		L	S	L	S	S
5		L	L	L	L	S
6		L	L	L	S	L
7		L	L	L	S	S
8		L	S	L	L	S
9		S	S	L	L	L
10		L	L	S	L	L
11		L	L	L	S	S
12		L	L	L	L	S
13		L	L	L	L	S
14		L	L	L	S	S
15		L	L	L	S	S
16		S	L	S	S	L
17		L	L	S	L	S
18		L	L	S	L	S
19		L	L	L	S	S
20		S	L	L	S	L
21		L	L	S	S	L

2.2.2 Test Statistic

When testing for preference, the null hypothesis is "no preference", i.e. that the bird randomly selects its food type (L or S) and the five selections made by each bird are independent. In other words, each no preference selection by a bird is like a coin toss with probability $p = \frac{1}{2}$ of selection L over S and with independence between selections. A Wilcoxon signed rank statistic can be used to describe the selections made in each trial. A "+" is assigned to

each large armyworm selected and a "$-$" is assigned to each small armyworm. The score of 5 is assigned to the first selection made by the bird, 4 to the second selection,..., and 1 to the fifth selection made by the bird. Using these scores means that early selection of a food type and more selections of that food type, will both contribute to increasing the indication of preference for that food type. Under the alternative that the birds exhibit a preference for one of the two food types, we would expect $p \neq \frac{1}{2}$. For example, $p > \frac{1}{2}$ would indicate a preference for L, and $p < \frac{1}{2}$ would indicate a preference for S. Because of the particular design of this experiment, abundance of the two types of food is not an issue. If the two food types were mixed together in one food chamber, abundance would play a more important role in the analysis. However, in this setting they were in separate subchambers and the researchers recorded only which subchamber was chosen by the bird in each selection.

In the jth bird trial we let

$$T_j^+ = \text{(sum of the scores assigned to +(L) food selections)}$$

and

$$T_j^- = \text{(sum of the scores assigned to -(S) food selections)}$$

Summing over the trials on different birds produces the test statistic

$$T^+ = \sum_{j=1}^{n} T_j^+$$

where n denotes the number of birds used in the experiment.

In general, suppose there are k selections made by each subject in the experimental situation and n subjects in the experiment. Then under the null hypothesis, T^+ has the distribution of the sum of n independent Wilcoxon signed rank statistics in which k is the

largest rank assigned in each of the signed rank statistics. It is clear that

$$E_{H_0}[T^+] = \frac{nk(k+1)}{4} \quad \text{and} \quad Var_{H_0}[T^+] = \frac{nk(k+1)(2k+1)}{24}$$

When n is large and k is fixed, the null distribution of T^+ can be approximately by a normal distribution. This follows from the usual Central Limit Theorem because under the null, T^+ is the sum of n independent and identically distributed statistics, T_j, each with finite mean and variance. From the data in Table 2.1, the exact p-value is 0.0001510.

2.3 General Rank Test for Preference: One-Sample

Suppose that we have an experiment in which there is a random sample of n predators ($n > 1$) and that each predator is allowed to select k preys ($k > 1$). We shall assume that relative abundance of prey remains unchanged during successive selections, or equivalently, that successive selections are made with replacement. Let ω_{ij} represent the feature of interest (eg. color, size, etc) of the jth prey selected by predator i, where $1 \leq i \leq n$ and $1 \leq j \leq k$. It will be assumed that predators make their selections independently of each other. As in [1], we will focus on the case where the prey feature of interest is categorized into two classes (2 sizes, 2 genders etc). In this case, we can define a random variable $X = 0$ or $X = 1$ depending on the category. We will write X_{ij} for $X(\omega_{ij})$ and $\{X = c\}$ for $\{\omega : X(\omega) = c\},\ c = 0, 1$. Then, under the null hypothesis that there is no preferential predation and selections are independent, $X_{i1}, \ldots, X_{ik}$ are independently and identically distributed according to the Bernoulli distribution with probability of success $p = 1/2$, $1 \leq i \leq n$. A value of $p > 1/2$ indicates preference for $\{X = 1\}$ prey. We will confine our attention to testing whether $\{X = 1\}$ is the preferred prey feature in that predators tend to select $\{X = 1\}$ prey more than $\{X = 0\}$ prey at the beginning of the selection experiment than at the end.

[1] gave a test statistic for detecting such selection patterns. In our notation, the Wilcoxon signed rank test statistic given by [1] is

$$T^+ = \sum_{i=1}^{n} \sum_{j=1}^{k} (k - j + 1) X_{ij} \ . \tag{2.3.1}$$

As intended, the test rejects the null if T^+ is large; that is, if there are more $\{X = 1\}$ prey selections at the beginning of selection than at the end. This is done by giving less importance to later $\{X = 1\}$ prey selections than earlier selections using the *score* function $\psi(j) \equiv k - j + 1$. However, the importance gradient remains constant since $\psi(j)$ decreases linearly as j goes from 1 to k. This does not allow investigators to factor in extraneous variables (such as time of day, predator's hunger level, location, etc) that can potentially influence the selection process. Using a general score function [35] in the definition of the test statistic provides investigators the flexibility to decide on a gradient. To that end, let ψ be a non-increasing function defined on the interval $(0, 1)$ such that $\sum_{j=1}^{k} \psi^2\{j/(k+1)\} < \infty$. We now define the generalized Wilcoxon signed rank test statistic for preference patterns as

$$T_\psi^+ = \sum_{i=1}^{n} \sum_{j=1}^{k} \psi_j X_{ij} \ , \tag{2.3.2}$$

where $\psi_j = \psi\{j/(k+1)\}$. The following theorem gives the first two moments and the asymptotic distribution of the test statistic $\mathbf{T}_\psi^+$. The proof is straightforward and may be found, for instance, in [35].

Theorem 2.1 *Assume that the null hypothesis is true. Then for $k < \infty$ fixed and $\sum_{j=1}^{k} \psi_j^2 < \infty$ we have*

$$\frac{T_\psi^+ - E_0(T_\psi^+)}{\{var_0(T_\psi^+)\}^{1/2}} \to N(0, 1) \text{ in distribution as } n \to \infty \ ,$$

where $E_0(T_\psi^+) = 2^{-1} n \sum_{j=1}^{k} \psi_j$ and $var_0(T_\psi^+) = 4^{-1} n \sum_{j=1}^{k} \psi_j^2$.

The score function ψ can be normalized such that $\sum_{j=1}^{k} \psi_j = 0$ and $\sum_{j=1}^{k} \psi_j^2 = 1$. One way to obtain normalized score functions is to take

$$\psi\{j/(k+1)\} = c\,[b - h\{j/(k+1)\}] \tag{2.3.3}$$

and to solve for b and c using the constraints $\sum_{j=1}^{k} \psi_j = 0$ and $\sum_{j=1}^{k} \psi_j^2 = 1$. Here b and c are scalars and $h(u)$ is a nondecreasing function defined on $[0, 1]$ depending on u alone. The score function leading to a test statistic equivalent to (2.3.1) of [1] may be obtained by taking $h(u) = u$. Other simple score functions use $h(u) = u^2$ and $h(u) = sgn(u)$.

Taking $h(u) = u$ in (2.3.3) gives $b = 1/2$ and $c = [\{12(k+1)\}/\{k(k-1)\}]^{1/2}$. For the data given in [1], one obtains a p-value of 3.14×10^{-5} for $h(u) = u$ and 1.22×10^{-5} for $h(u) = u^2$ using the asymptotic test in Theorem 2.1. [1] found a p-value of 1.51×10^{-5} using a permutation test.

2.4 Two Sample Comparison of Preference Patterns

Suppose now that we have two predatory species $\mathscr{X}$ and $\mathscr{Y}$ and record the first k choices made by these two species. As earlier, we will assume that there is only one prey feature of interest. However, in this case, we will not restrict our analyses to the case where this feature is categorized into two classes. Instead, we will consider a case where the feature is continuously measured and also the case of an arbitrary number of categories. To simplify our argument, we will assume that prey are infinitely abundant or that selection is made with replacement. Our goal is to test whether there is a significant difference in the predatory preferences of these two species.

The data from this experiment are presented as X_{rj} $(r = 1, \ldots, n_1; j = 1, \ldots, k)$ and Y_{sj} $(s = 1, \ldots, n_2; j = 1, \ldots, k)$. Here X_{rj} and Y_{sj} denote the feature of interest of the jth prey selected by the rth individual from species $\mathscr{X}$ and the sth individual from species $\mathscr{Y}$, respectively. When there is no confusion, we will use $\mathscr{X}$ and $\mathscr{Y}$ as generic random variables

denoting prey feature selected by species $\mathscr{X}$ and $\mathscr{Y}$, respectively. We are in particular interested in testing whether the $\mathscr{X}$ species prefer larger values of the particular feature than the $\mathscr{Y}$ species. This will be characterized by larger values of $pr(\mathscr{X} > \mathscr{Y})$ at the beginning of the experiment than at the end.

We propose the following test statistic for comparing two species selection patterns:

$$W_\psi^+ = \sum_{r=1}^{n_1}\sum_{s=1}^{n_2}\sum_{j=1}^{k}\psi_j\varphi(X_{rj} - Y_{sj}) \,, \tag{2.4.1}$$

where

$$\varphi\,(x) = \begin{cases} 1 & (x > 0) \,, \\ 1/2 & (x = 0) \,, \\ 0 & (x < 0) \,. \end{cases}$$

In the following we will study the null asymptotic behavior of W_ψ^+ for continuous and categorical prey feature, respectively. All proofs are given in the Appendix.

2.4.1 Continuous Prey Feature

Assume that $\mathscr{X}$ and $\mathscr{Y}$ are measured continuously. Under the null hypothesis that there is no preferential selection, $pr(\varphi(\mathscr{X} - \mathscr{Y}) = 1) = 0.5 = pr(\varphi(\mathscr{X} - \mathscr{Y}) = 0)$. The null asymptotic distribution of W_ψ^+ is given in the following theorem.

Theorem 2.2 *Under the null hypothesis that there is no difference in preferences, $k < \infty$ fixed, and $\sum_{j=1}^k \psi_j^2 < \infty$ we have*

$$\frac{W_\psi^+ - E_0(W_\psi^+)}{\left\{var_0(W_\psi^+)\right\}^{1/2}} \;\to\; N(0,1) \text{ in distribution as } \min(n_1, n_2) \to \infty \,,$$

where $E_0(W_\psi^+) = 2^{-1}n_1n_2\sum_{j=1}^k\psi_j$ and $var_0(W_\psi^+) = 12^{-1}n_1n_2(n_1+n_2+1)\sum_{j=1}^k\psi_j^2$.

Proof 2.3 *Write $W_\psi^+ = \sum_{j=1}^k \psi_j U_{n_1n_2j}$, where $U_{n_1n_2j} = \sum_{r=1}^{n_1}\sum_{s=1}^{n_2}\varphi\,(X_{rj} - Y_{sj})$ for $j = 1, \ldots, k$. For $j = 1, \ldots, k$ fixed, under the null hypothesis, $U_{n_1n_2j}$ is just the Mann-Whitney*

U-statistic and thus [41]

$$E_0(U_{n_1n_2j}) = \frac{n_1n_2}{2} \quad \textit{and} \quad var_0(U_{n_1n_2j}) = \frac{n_1n_2(n_1+n_2+1)}{12} \ .$$

Moreover, as $\min(n_1, n_2) \to \infty$

$$\frac{U_{n_1n_2j} - n_1n_2/2}{\{n_1n_2(n_1+n_2+1)/12\}^{1/2}} \ \to \ N(0,1)$$

in distribution whenever the null hypothesis is true. Thus

$$\frac{W_\psi^+ - (n_1n_2/2)\sum_{i=1}^k \psi_j}{\{n_1n_2(n_1+n_2+1)/12\}^{1/2}} = \sum_{j=1}^k \psi_j \left\{ \frac{U_{n_1n_2j} - n_1n_2/2}{\{n_1n_2(n_1+n_2+1)/12\}^{1/2}} \right\} \ \to \ N\left(0, \ \sum_{j=1}^k \psi_j^2\right)$$

in distribution since $k < \infty$ *and* $\sum_{j=1}^k \psi_j^2 < \infty$.

Once again, in applications, standardizing the score function ψ such that $\sum_{j=1}^k \psi_j = 0$ and $\sum_{j=1}^k \psi_j^2 = 1$ will simplify the expressions for mean and variance.

2.4.2 Categorical Prey Feature

Assume that the prey feature of interest is categorized into m disjoint classes. We wish to extend the two sample selective predation problem to this case where selections are being made with replacement. Under the null hypothesis of no selective predation, the combined sample $\{X_{1j}, \ldots, X_{n_1j}, Y_{1j}, \ldots, Y_{n_2j}\}$ has a multinomial distribution with mass function

$$P(T_1 = t_{j1}, \ldots, T_m = t_{jm}) = (n_1+n_2)! \prod_{l=1}^m \frac{p_l^{t_{jl}}}{t_{jl}!} \ ,$$

where $\sum_{l=1}^m p_l = 1$ and $\sum_{l=1}^m t_{jl} = n_1 + n_2$ $(j = 1, \ldots, k)$. In this case, we may have ties in our data that need to be taken into account. The following theorem gives the asymptotic normality of W_ψ^+.

Theorem 2.4 *Suppose prey feature is categorical with m possible outcomes ($m > 1$) and that the null hypothesis is true. Let t_{jl} be the observed frequencies for selection j ($j = 1, \ldots, k;\ l = 1, \ldots, m$). Then*

$$\frac{W_\psi^+ - E_0(W_\psi^+)}{\left\{var_0(W_\psi^+)\right\}^{1/2}} \ \rightarrow \ N(0,1) \text{ in distribution as } \min(n_1, n_2) \rightarrow \infty \ ,$$

where $E_0(W_\psi^+) = 2^{-1} n_1 n_2 \sum_{j=1}^k \psi_j$ and

$$var_0(W_\psi^+) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} \sum_{j=1}^k \psi_j^2 - \frac{n_1 n_2}{12(n_1 + n_2)(n_1 + n_2 - 1)} \sum_{j=1}^k \sum_{l=1}^m \psi_j^2 (t_{jl}^3 - t_{jl}) \ .$$

Proof 2.5 *The proof follows from the proof of Theorem 2.2. The only difference is in the computation of the null variance of $U_{n_1 n_2 j}$ which is given by [42] as*

$$var_0(U_{n_1 n_2 j}) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \frac{n_1 n_2}{12(n_1 + n_2)(n_1 + n_2 - 1)} \sum_{l=1}^m (t_{jl}^3 - t_{jl}) \ .$$

We remark here that the added complexity presented by a discrete prey feature is the introduction of ties in $\mathscr{X}$ and $\mathscr{X}$ that one needs to take into account. As in most nonparametric statistics, although ties affect the null variance, they do not usually affect the null expected value of the statistic [41]. It can be easily seen that the null variance of the discrete case is the same as the null variance of the continuous prey feature case when $t_{jl} = 1$ ($j = 1, \ldots, k;\ l = 1, \ldots, m$).

2.5 Censored Data

In repeated measurement data study, especially in clinical trials, if the repeated measurements are taken over time during the study, some experimental subjects may not have a complete vector of observations. some of the observations may be censored. In this section we assume all the censored observations are censored completely at random.

Define the $i.i.d$ censor indicators ε_{rj} and ξ_{sj} to be:

$$\varepsilon_{rj} = \begin{cases} 1, & \text{if } X_{rj} \text{ is observed,} \\ 0, & \text{otherwise.} \end{cases}$$

and

$$\xi_{sj} = \begin{cases} 1, & \text{if } Y_{Sj} \text{ is observed,} \\ 0, & \text{otherwise.} \end{cases}$$

for $r = 1, \ldots, n_1,\ s = 1, \ldots, n_2$ and $j = 1, \ldots, k$

we also assume:

$$P(\varepsilon_{rj} = 1) = p_j \text{ and } P(\xi_{sj} = 1) = q_j$$

2.5.1 One-sample Selection Problem

The problem in Section 2.3 can be considered as a one sample repeated measures problem. Let $X_r = (X_{r1}, \ldots, X_{rk})'$ denote independent random samples for the rth subject, and $j = 1, \ldots, k$ are indices representing a set of conditions or a series of prespecified conditions or time points for which repeated measurements from each experimental subject are obtained. Some experimental subjects may not have a complete vector of observations. We use ε_{rj} to indicate if the jth observation of rth experimental subject in group $\mathscr{X}$ is censored or not. Then the test statistic in (2.3.2) can be modified as:

$$T^c = \sum_{j=1}^{k} \sum_{r=1}^{n} \psi_j \varepsilon_{rj} X_{rj} \tag{2.5.1}$$

It is easy to see that

$$E_0(T^c) = \frac{n}{2} \sum_{j=1}^{k} \psi_j E(\varepsilon_j) = \frac{n}{2} \sum_{j=1}^{k} p_j \psi_j$$

$$
\begin{aligned}
Var_0(T^c) &= \sum_{j=1}^{k} \psi_j^2 Var_0(\sum_{r=1}^{n} \varepsilon_{rj} X_{rj}) \\
&= \sum_{j=1}^{k} \psi_j^2 n Var_0(\varepsilon_{rj} X_{rj}) \\
&= \sum_{j=1}^{k} (E_0(\varepsilon_{rj}^2 X_{rj}^2) - E_0^2(\varepsilon_{rj} X_{rj})) \\
&= \sum_{j=1}^{k} (E(\varepsilon_{rj}^2) E_0(X_{rj}^2) - E^2(\varepsilon_{rj}) E_0^2(X_{rj})) \\
&= n \sum_{j=1}^{k} (\frac{1}{2} E(\varepsilon_{rj}^2) - \frac{1}{4} E^2(\varepsilon_{rj})) \psi_j^2 \\
&= n \sum_{j=1}^{k} (\frac{1}{2} p_j - \frac{1}{4} p_j^2) \psi_j^2
\end{aligned}
$$

Then because $\sum_{r=1}^{n} \psi_j \varepsilon_{rj} X_{rj}$ are independently and identically distributed for $j = 1, \ldots, k$, using the Central limit theorem, we can get the asymptotic normality of T^c similar as the result in Theorem 2.1.

2.5.2 Two Sample Selection Problem

For the two sample predation preference problem, we also can look it as a two sample repeated measures problem. For example, in clinical trials, we have control group $\mathscr{X}$ and the new treatment group $\mathscr{Y}$. We record the data of the experimental subjects in both groups under some conditions or at some prespecified time points $j = 1, \ldots, k$. The investigator wishes to draw an overall conclusion whether the new treatment constitutes an "improvement" over the control for the entire study. As defined in Section 2.4, we use $\mathscr{X}$ and $\mathscr{Y}$ to denote the two populations for the two groups and we use ε_{rj} to denote if the jth observation of rth experimental subject in group $\mathscr{X}$ is censored or not, similar for ξ_{sj}. Then the test statistic in (2.4.1) can be rewritten as:

$$
W_\psi^c = \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} \sum_{j=1}^{k} \psi_j \varepsilon_{rj} \xi_{sj} \varphi(X_{rj} - Y_{sj}) \tag{2.5.2}
$$

We can calculate the null expectation and variance of W_ψ^c

$$\begin{aligned} E_0(W_\psi^c) &= \frac{n_1 n_2}{2} \sum_{j=1}^{k} \psi_j E(\varepsilon_{rj}) E(\xi_{sj}) \\ &= \frac{n_1 n_2}{2} \sum_{j=1}^{k} \psi_j p_j q_j \end{aligned}$$

$$Var_0(W_\psi^c) = \sum_{j=1}^{k} \psi_j^2 Var_0[\sum_{r=1}^{n_1} \sum_{s=1}^{n_2} \varepsilon_{rj} \xi_{sj} \varphi(X_{rj} - Y_{sj})]$$

if we denote $Q_{rsj} = \varepsilon_{rj} \xi_{sj} \varphi(X_{rj} - Y_{sj})$

$$\begin{aligned} Var_0(W_\psi^c) &= \sum_{j=1}^{k} \psi_j^2 \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} Var_0(Q_{rsj}) + \sum_{j=1}^{k} \psi_j^2 \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} \sum_{r_1=1}^{n_1} \sum_{s_1=1}^{n_2} Cov_0(Q_{rsj}, Q_{r_1 s_1 j}) \\ &= \sum_{j=1}^{k} \psi_j^2 [n_1 n_2 Var_0(Q_{11j}) + n_1 n_2 (n_2 - 1) Cov_0(Q_{11j}, Q_{12j}) \\ &+ n_1 n_2 (n_1 - 1) Cov_0(Q_{11j}, Q_{21j})] \end{aligned}$$

Where:

$$\begin{aligned} Var_0(Q_{11j}) &= E(Q_{rsj}^2) - E^2(Q_{rsj}) \\ &= E_0(\varepsilon_{1j} \xi_{1j} \varphi(X_{1j} - Y_{1j}))^2 - [E(\varepsilon_{1j}) E(\xi_{1j}) E_0(\varphi(X_{1j} - Y_{1j}))]^2 \\ &= \frac{1}{2} E(\varepsilon_{1j}^2) E(\xi_{1j}^2) - \frac{1}{4} E^2(\varepsilon_{1j}) E^2(\xi_{1j}) \\ &= \frac{1}{2} E(\varepsilon_{1j}) E(\xi_{1j}) - \frac{1}{4} E^2(\varepsilon_{1j}) E^2(\xi_{1j}) \\ &= \frac{1}{2} p_j q_j - \frac{1}{4} p_j^2 q_j^2 \end{aligned}$$

$$
\begin{aligned}
Cov_0(Q_{11j}, Q_{12j}) &= E_0(Q_{11j}Q_{12j}) - E_0(Q_{11j})E_0(Q_{12j}) \\
&= E_0(\varepsilon_{1j}\xi_{1j}\varphi(X_{1j} - Y_{1j})\varepsilon_{1j}\xi_{2j}\varphi(X_{1j} - Y_{2j})) - E_0^2(\varepsilon_{1j}\xi_{1j}\varphi(X_{1j} - Y_{2j})) \\
&= E(\varepsilon_{1j}^2)E(\xi_{1j}\xi_{2j})E_0(\varphi(X_{1j} - Y_{1j})\varphi(X_{1j} - Y_{2j})) \\
&- E^2(\varepsilon_{1j})E^2(\xi_{1j})E_0^2(\varphi(X_{1j} - Y_{1j})) \\
&= \frac{1}{3}E(\varepsilon_{1j}^2)E(\xi_{1j}\xi_{2j}) - \frac{1}{4}E^2(\varepsilon_{1j}\xi_{1j}) \\
&= \frac{1}{3}p_j q_j^2 - \frac{1}{4}p_j^2 q_j^2
\end{aligned}
$$

and similarly

$$
\begin{aligned}
Cov(Q_{11j}, Q_{21j}) &= \frac{1}{3}E(\varepsilon_{1j}\varepsilon_{2j})E(\xi_{1j}^2) - \frac{1}{4}E^2(\varepsilon_{1j}\xi_{1j}) \\
&= \frac{1}{3}p_j^2 q_j - \frac{1}{4}p_j^2 q_j^2
\end{aligned}
$$

Then because $\sum_{r=1}^{n_1}\sum_{s=1}^{n_2} Q_{rsj}$ are independently and identically distributed for $j = 1, \ldots, k$, using the Central limit theorem, we can get the asymptotic property of W_ψ^c following similar steps as in Theorem 2.2.

2.6 Monte Carlo Simulation

We performed a large scale simulation to evaluate the finite sample performance of the asymptotic distribution of the statistics W_ψ^+ and W_ψ^c.

2.6.1 Null Simulation

First we performed a null simulation for the two-sample continuous prey feature case under the nominal $\alpha = 0.05$. For this study, independent, identically distributed samples were generated from the standard Cauchy (C), central t with 5 degrees of freedom (t_5), standard normal (N), and the standard Laplace (L) distributions. We considered all $\binom{4}{2} = 10$ possible combinations of the distributions for the two samples. For the sample sizes we took

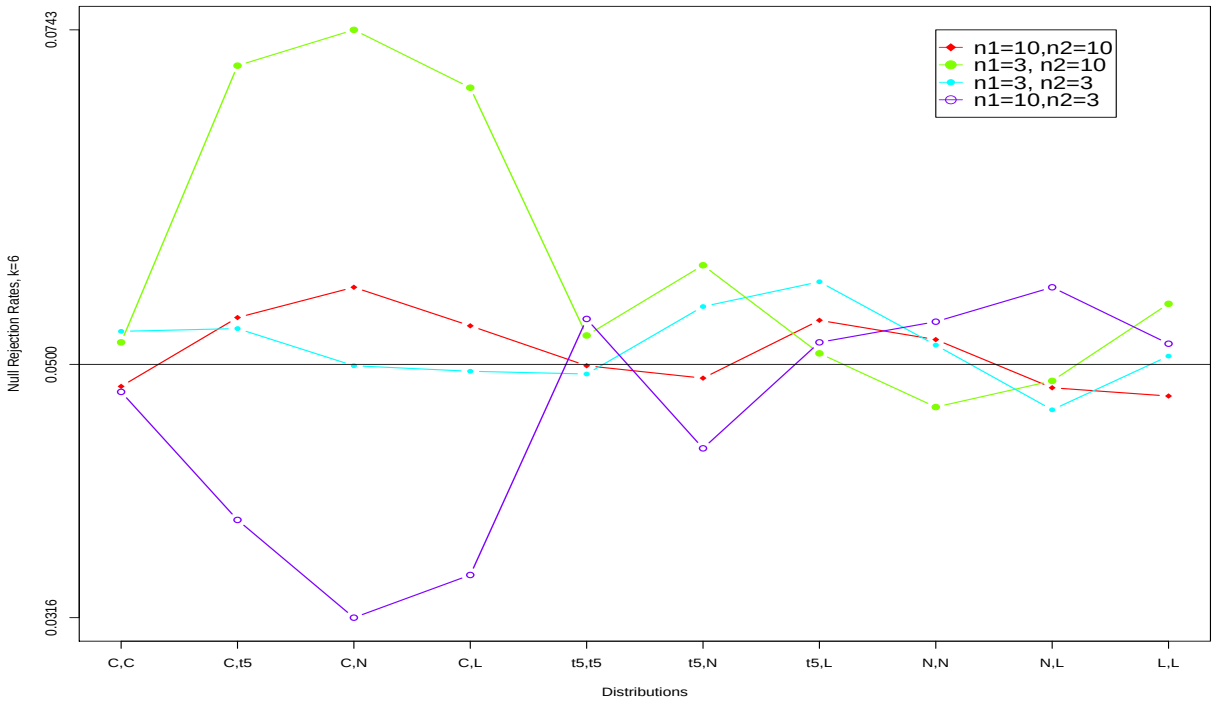
(n_1, n_2) combinations of $(3, 10)$, $(3, 3)$, $(10, 3)$, and $(10, 10)$. For the number of selections of each subject we took $k = 3, 6,$ and 12. The asymptotic distribution given in Theorem 2.2 is used to determine the null rejection rates. We performed 10^4 repetitions and found the proportion of cases for which $\{W_\psi^+ - E_0(W_\psi^+)\}/\{var_0(W_\psi^+)\}^{1/2} > 1.645$. The results for $\psi(t) = \Phi^{-1}(1-t)$ were quite similar to the results for $\psi(t) = 1 - t$ which we report in Table 2.2. We note that for most distribution combinations, the rejection rates were very

Table 2.2: Simulated null rejection rates of W_ψ^+ using its asymptotic normal critical constant for continuous distributions, $\psi(t) \propto t,\ \alpha = .05$.

k	N,N	C,N	t_5,N	N,L	C,C	C,t_5	C,L	t_5,t_5	t_5,L	L,L
					$n_1 = 10, n_2 = 10$					
3	0.0482	0.0513	0.0496	0.0484	0.0470	0.0559	0.0538	0.0479	0.0508	0.0457
6	0.0518	0.0556	0.0490	0.0483	0.0484	0.0534	0.0528	0.0499	0.0532	0.0477
12	0.0496	0.0545	0.0510	0.0456	0.0517	0.0557	0.0531	0.0478	0.0500	0.0493
					$n_1 = 3, n_2 = 10$					
3	0.0449	0.0785	0.0576	0.0445	0.0487	0.0661	0.0671	0.0502	0.0472	0.0507
6	0.0469	0.0743	0.0572	0.0488	0.0516	0.0717	0.0701	0.0521	0.0508	0.0544
12	0.0492	0.0695	0.0548	0.0452	0.0480	0.0679	0.0667	0.0487	0.0510	0.0509
					$n_1 = 3, n_2 = 3$					
3	0.0440	0.0471	0.0435	0.0468	0.0454	0.0454	0.0481	0.0481	0.0472	0.0479
6	0.0514	0.0499	0.0542	0.0467	0.0524	0.0526	0.0495	0.0493	0.0560	0.0506
12	0.0509	0.0507	0.0502	0.0490	0.0458	0.0503	0.0502	0.0508	0.0481	0.0480
					$n_1 = 10, n_2 = 3$					
3	0.0483	0.0281	0.0473	0.0575	0.0488	0.0361	0.0342	0.0505	0.0529	0.0499
6	0.0531	0.0316	0.0439	0.0556	0.0480	0.0387	0.0347	0.0533	0.0516	0.0515
12	0.0487	0.0296	0.0454	0.0566	0.0488	0.0355	0.0345	0.0520	0.0494	0.0517

close to the nominal $\alpha = 0.05$ despite using the asymptotic distribution for such small samples. In the hybrid distribution cases where the two samples were drawn from two different distributions, when the sample size from the heavier tailed distribution is smaller than the sample size from the lighter tailed one, we found that the rejection rates were higher than 0.05. The reverse phenomenon occurs when the sample sizes are reversed. It appears that the asymptotic variances are under- and over-estimated, respectively, for the two cases. We can see the pattern more clearly from Figure 2.1.

Figure 2.1: Figure for Null Rejection Rates in Table 2.2



In Figure 2.1, the horizontal axis denotes different distributional settings and the vertical axis denotes null rejection rate. Different color denotes different sample size combinations and the flat line in the middle denotes the nominal $\alpha = 0.05$. The Figure shows the null rejection rate for different distributional settings and different sample size combinations. From this Figure, we can see some clear patterns: First, for most distributional settings, the null rejection rates are fairly close to 0.05 despite using the asymptotic distribution for such small sample size. Second, the red line is the most preferred line since overall, it's the closed one to the flat line which denotes the nominal $\alpha = 0.05$. This indicates the test statistic performs the best for equal large sample size ($n_1 = n_2 = 10$) which meets our expectation. Third, there are some abnormality due to including the extremely heavy tailed Cauchy distribution to generate the sample. If the sample size generated using Cauchy distribution is relatively larger compared to the other sample size, the asymptotic variance tends to be over-estimated, which would lead to a smaller null rejection rate. Similarly, if the sample

size generated using Cauchy distribution is relatively smaller compared to the other sample size, the asymptotic variance tends to be under-estimated, which would lead to a larger null rejection rate.

2.6.2 Power Simulation

To determine the power of the test based on the asymptotic distribution in Theorem 2.2, we generated independent samples for the jth selection as $X_j \sim N(\mu_j, 1)$ and $Y_j \sim N(0, 1)$ $(j = 1, \ldots, k)$. We study the performance of the asymptotic test under simple ordering $\mu_1 \geq \cdots \geq \mu_k$ and umbrella ordering $\mu_1 \leq \cdots \leq \mu_q \geq \cdots \geq \mu_k$, $q \in \{1, \ldots, k\}$, of the X means. We used Wilcoxon scores $\psi(t) = 1 - t$ and normal scores $\psi(t) = \Phi^{-1}(1 - t)$, where Φ is the standard normal cdf. The proportion of cases out of 10^4 in which the standardized statistic exceeded 1.645 were recorded. In the first part of the simulation, we generated the X samples by drawing the values of the first approximately $k/2$ selections from the $N(0.5, 1)$ distribution and the remaining selections from the $N(0, 1)$ distribution. For the second part, the first $k/3$ X selections were drawn from the $N(1, 1)$ distribution, the middle $k/3$ were from the $N(0.5, 1)$ distribution, and the final $k/3$ were from the $N(0, 1)$ distribution. Finally, the third part used the first $k/3$ selections from $N(0, 1)$, the middle $k/3$ from $N(0.5, 1)$ and the final $k/3$ from $N(0, 1)$ creating an umbrella pattern. The results are given in Table 2.3. It is evident that the rejection rates were increasing with increasing k or increasing n_1 and n_2 for both score functions. It also appears that the test's ability to detect simple ordered mean patterns increased when the number of means used to generate the X sample was larger. This goes hand in hand with our intuition that the test statistic becomes better trained to detect patterns when there are more distinct classes. The test based Wilcoxon scores was more powerful than the test based on normal scores. This is expected since the means are arranged in a linearly decreasing pattern where linear scores are optimal. Normal scores are optimal for the case where the means are quickly decreasing at the beginning and at the end but slowly decreasing in the middle.

Table 2.3: The simulated power of W_{ψ}^{+} using its asymptotic normal distribution where $X_j \sim N(\mu_j, 1),\ Y_j \sim N(0,1),\ 1 \le j \le k,\ \alpha = 0.05$

				$\psi(t) = 1 - t$			$\psi(t) = \Phi^{-1}(1-t)$		
				$n_1 = n_2$			$n_1 = n_2$		
Mean pattern			k	2	5	7	2	5	7
{0.5	0.5	0}	3	0.111	0.138	0.164	0.046	0.112	0.134
			12	0.240	0.492	0.624	0.154	0.273	0.350
{1	0.5	0}	3	0.248	0.453	0.567	0.072	0.220	0.286
			12	0.569	0.939	0.984	0.270	0.562	0.711
{0	0.5	0}	3	0.099	0.109	0.112	0.026	0.046	0.047
			12	0.115	0.185	0.228	0.055	0.050	0.051

The power of the test was low in detecting an umbrella ordered pattern in the alternative. We note that the test is not designed to detect such patterns since the score function is monotone non-increasing with each successive selection. If one is interested in such patterns, then the test of [43] for umbrella alternatives may be adapted for this case in a straightforward manner.

If the populations that generated the samples are as in Table 2.3 but the X mean configuration is $\{1, 1/4, 0, 0\}$, then a test using linear scores will not be as powerful as one using a score function that is decreasing fast at first and slowly at the end. One such score function is $\psi(t) = G^{-1}(1-t)$, where G is the Gamma$(1/5, 4)$ distribution. Using the Wilcoxon score function, for $k = 4$ selections, 10^3 replications gave rejection rates of 29%, and 40% for $n_1 = n_2$ values of 5 and 7, respectively. The corresponding rejection rates for the gamma score function were 40% and 54%. Clearly, the choice of score functions is important. It is possible to select bent scores as discussed on p. 100 of [44] or Section 4 of [45]. Alternatively, nonparametric estimates of the optimal score function may be obtained using the methods of [46] and [47].

Finally, we performed a simulation study with the selections were correlated. This was done by generating the X sample from a trivariate normal distribution with mean $(1, 1/2, 0)$

and an order one autoregressive (AR(1)) covariance matrix with (i,j)th entry $\rho^{|i-j|}$ for $\rho = 0, 0.2, 0.5, 0.9$. The rejection rates were nearly the same as the ones in Table 2.3.

2.6.3 Null Simulation for Two Sample Case with Censoring

Similar to the null simulation for the two-sample continuous prey feature case, for this study, under the nominal $\alpha = 0.05$, independent, identically distributed samples were generated from the standard Cauchy (C), central t with 5 degrees of freedom (t_5), standard normal (N), and the standard Laplace (L) distributions. We considered all $\binom{4}{2} = 10$ possible combinations of the distributions for the two samples. For the sample sizes we took (n_1, n_2) combinations of $(3, 10)$, $(3, 3)$, and $(10, 3)$. For the number of repeated measurements of each subject we took $k = 3, 12$. For the censor probability, we took $(p_j = q_j = 0.5)$ The asymptotic distribution of W_ψ^c is used to determine the null rejection rates. We performed 10^4 repetitions and found the proportion of cases for which $\{W_\psi^c - E_0(W_\psi^c)\}/\{var_0(W_\psi^c)\}^{1/2} > 1.645$.

Table 2.4: Simulated null rejection rates of W_ψ^+ using its asymptotic normal critical constant, $\psi(t) \propto t,\ \alpha = .05, P(\varepsilon_{rj} = 1) = P(\xi_{sj} = 1) = .5$.

k	N,N	C,N	t_5,N	N,L	C,C	C,t_5	C,L	t_5,t_5	t_5,L	L,L
					$n_1 = 3, n_2 = 10$					
3	0.0436	0.0479	0.0427	0.0413	0.0431	0.0476	0.0521	0.0402	0.0439	0.0449
12	0.0459	0.0589	0.0460	0.0453	0.0481	0.0572	0.0526	0.0476	0.0469	0.0420
					$n_1 = 3, n_2 = 3$					
3	0.0504	0.0487	0.0509	0.0498	0.0512	0.0523	0.0451	0.0468	0.0524	0.0499
12	0.0469	0.0461	0.0460	0.0441	0.0437	0.0454	0.0431	0.0480	0.0444	0.0446
					$n_1 = 10, n_2 = 3$					
3	0.0445	0.0380	0.0401	0.0472	0.0421	0.0403	0.0404	0.0451	0.0447	0.0445
12	0.0482	0.0383	0.0449	0.0546	0.0460	0.0444	0.0365	0.0466	0.0493	0.0490

It is evident from Table 2.4, Although the same size is very small, $n_1,\ n_2$ just takes 3 and 10, for a very large censor probability $p_j = q_j = 0.5$, for all the distribution settings, the null rejection rate is still fairly close to the nominal $\alpha = 0.05$.

Finally, we performed simulations for different censoring probabilities ($p_j = 0.1,\ q_j = 0.1$), ($p_j = 0.1,\ q_j = 0.9$), ($p_j = 0.25,\ q_j = 0.25$), ($p_j = 0.25,\ q_j = 0.75$) and ($p_j = 0.9,\ q_j = 0.9$) . In all these cases, the null rejection rates are all close to the nomial 0.05, with less censoring probability corresponding to closer to 0.05, which meets our expectation.

2.7 Conclusion

This chapter gave nonparametric rank based test statistics for detecting preference patterns in selective predation by extending the method of [1]. This test statistics can be used broadly in repeated measures data study. We provided a class of general rank score tests for the one sample case where we only have one species of predators and prey have two features of interest. This gives the flexibility to place varied emphasis on consecutive selections at different stages of the selection experiment. This could be helpful in controlling the manner in which extraneous variables can affect selection preference patterns.

This chapter also proposes a class of general rank score tests for the difference in predation patterns of two predatory species, which is a special case for the two-sample repeated measures data. In this case, prey feature of interest can be continuous or categorical. It is shown that the test statistic for the categorical case becomes equivalent to the continuous case if the data have no ties. In both cases, the asymptotic distribution of the test statistic is Gaussian. We also study the test statistics and its asymptotic properties for the one-sample and two-sample repeated measures data with random censoring. The results of a simulation study using the asymptotic Gaussian distribution but small samples shows that the test has a satisfactory finite-sample performance. The null simulation shows that null rejection rates are close to nominal α values. The asymptotic test is powerful in detecting simple-ordered alternatives.

Chapter 3

Testing the Difference Trend in Two Sample Repeated Measurement Data

3.1 Motivation of the Method

Consider the predator-prey relationship discussed in the previous chapter. The manner in which the predator selects its prey is crucial for its survival. Several methods have been given in the past to quantify selective predation. As pointed out in [1], one drawback of most existing methods is that they do not take into account the order in which different preys are taken. [1] suggest to use a Wilcoxon statistic to test for food preference while taking into account the order in which different prey are taken. The use of the Wilcoxon statistic means that the gradient of selection remains the same throughout the experiment. In the previous chapter we gave a generalized form of the Wilcoxon statistic used by [1] to allow experimenters place varied emphasis on the selection gradient at different parts of the experiment. Further we proposed a method to test the difference in preference patterns of two different species of predators, where prey feature can be either continuous or categorical. In this chapter, we are in particular interested in testing whether the rate of prey feature change in the consecutive selections of one species is faster than that of another species.

Another problem we consider involves an experiment conducted by the Department of Kinesiology at Auburn University. Let us consider part of the study by [48] concerning the effect of a single session of high intensity aerobic exercise on inflammatory markers of subjects taken over time. Eighteen subjects were placed into two groups (High Fitness and Moderate Fitness) depending on their fitness levels, nine in each group. The response here is the C-Reactive protein (CRP). Elevated CRP levels are a marker of low-grade chronic inflammation and may predict a higher risk for cardiovascular disease. Our effect of interest is the trend of difference of CRP levels between the two groups (High Fitness - Moderate Fitness), we

are interested in testing whether the difference stays the same over time or decays to zero over time. Each individual in the two groups was required to finish an exercise session which consumes 500 kcal. CRP levels were obtained 24 hours and immediately prior to the acute bout of exercise and subsequently 24, 72 and 120 hours following exercise.

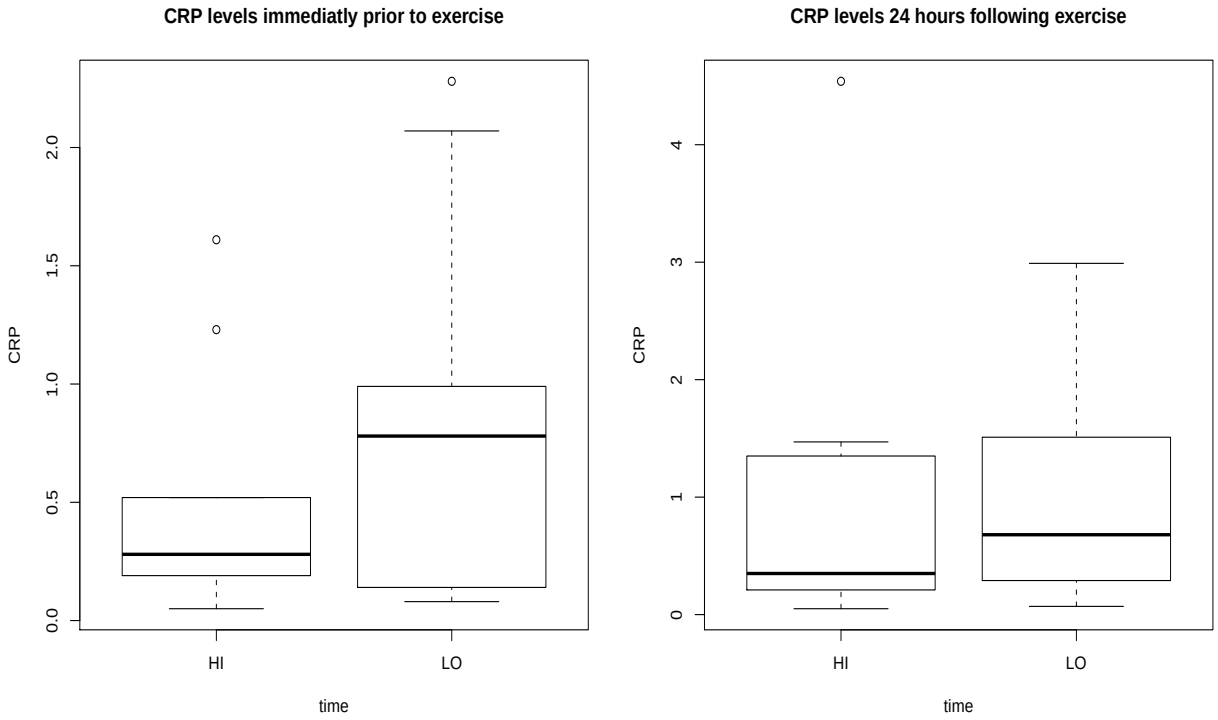
Table 3.1: CRP levels in [48]

id	group	-24	0	24	72	120
1	LO	1.79	0.78	0.68	0.86	0.83
2	LO	0.09	0.14	0.07	0.14	0.51
3	LO	1.13	0.99	1.06	0.91	0.92
4	LO	2.94	2.07	1.51	1.05	1.03
5	LO	1.82	2.28	2.99	1.95	1.78
6	LO	0.17	0.08	0.29	0.3	0.22
7	LO	0.38	0.19	0.29	0.49	0.26
8	LO	1.32	0.96	1.8	0.82	1.68
9	LO	0.4	0.11	0.25	0.2	0.31
10	HI	0.11	0.19	0.21	0.22	0.08
11	HI	0.15	0.17	0.15	0.11	0.21
12	HI	0.06	0.05	0.05	0.06	0.16
13	HI	0.31	0.2	0.28	0.14	0.24
14	HI	0.21	0.28	0.35	0.1	0.2
15	HI	0.36	0.37	4.54	1.8	1.1
16	HI	0.92	1.23	1.47	0.96	0.81
17	HI	0.52	0.52	0.51	0.31	0.35
18	HI	2.05	1.61	1.35	0.73	0.67

In this example, the data sample size is small, we only have 9 subjects in each group and each subject has only 5 repeated measurement data. The data are highly skewed and contain outliers, which we can see from Figure 3.1.

Both the two-sample selective predation problem and the CRP level problem are specific examples for two sample repeated measurement data, and the problem we are interested in can be generalized as studying the trend difference between the two samples over repeated measurements. In the current chapter, we propose a procedure that is analogous to the Page test [24] to study the trend of the difference between two samples in a repeated measurement data. Although the Page test is used for randomized complete block designs, we extend it to the situation where we have two randomized complete block designs. We derive the

Figure 3.1: Boxplots for CRP levels at time 0 and time 24 for the two groups HI and LO



asymptotic distribution of a general test statistic that includes the Page statistic as a special case. This provides an approximate α level test for the changing rates in the simple ordering case.

3.2 Test Statistic and Theory

Suppose now that we have two populations of repeated measurement data represented by $\mathscr{X}$ and $\mathscr{Y}$, and each subject in the two groups has k repeated measurements. We will assume that the repeated measurements for each subject are continuously measured. Suppose we have a random sample n_1 from $\mathscr{X}$ and $n_2 = N - n_1$ from $\mathscr{Y}$ to participate in the experiment. The data from this experiment are presented as an $n_1 \times k$ matrix $\mathbf{X}$ and $n_2 \times k$ matrix $\mathbf{Y}$. In both cases, rows represent individuals and columns represent repeated measurements. Let $\mathbf{U} = [\mathbf{X}' \quad \mathbf{Y}']'$ be the $N \times k$ matrix of responses. In the following $\mathbf{J}_{p\times q}$ denotes a $p \times q$ matrix of ones and $\mathbf{0}_{p\times q} \equiv \mathbf{J}_{p\times q} - \mathbf{J}_{p\times q}$ is a $p \times q$ matrix of zeros.

Consider the following multivariate linear model

$$\mathbf{U} = \alpha \mathbf{J}_{N\times k} \; + \; \mathbf{C}\boldsymbol{\Delta} \; + \; \mathbf{E}\,, \tag{3.2.1}$$

$\alpha \in \mathbb{R}$ is an unknown intercept parameter, $\mathbf{C} = [\mathbf{J}'_{n_1\times k} \quad \mathbf{0}'_{n_2\times k}]'$, $\boldsymbol{\Delta} = \text{diag}(\Delta_1, \ldots, \Delta_k)$ is a diagonal matrix of k real unknown parameters, and $\mathbf{E}$ is an $N \times k$ matrix of random error terms. In the model in (3.2.1), if we assume that $E[\mathbf{E}] = \mathbf{0}_{N\times k}$, then, for any $r = 1, \ldots, n_1$ and $s = 1, \ldots, n_2$, $E(X_{rj} - Y_{sj}) = \Delta_j$, for $j = 1, \ldots, k$. Thus $\boldsymbol{\Delta}$ is a shift parameter.

We are interested in testing

$$H_0 : \Delta_1 = \cdots = \Delta_k$$
$$H_A : \Delta_1 \leq \cdots \leq \Delta_k\,, \quad \text{with at least one strict inequality.}$$

Rejecting H_0 in favor of H_A means that the gap between two groups is increasing over time. Let $N^* = n_1 n_2$ and define $N^* \times k$ matrix $\mathbf{Z} = \mathbf{X} \otimes \mathbf{J}_{n_2\times 1} - \mathbf{J}_{n_1\times 1} \otimes \mathbf{Y}$. Let $\mathbf{R}$ be the $N \times k$ matrix of row-ranks with (i, j)th element of $\mathbf{R}$ is $R_{ij} = \sum_{t=1}^{k} I\{Z_{it} \leq Z_{ij}\}$. Selection ranks are

$$\tilde{\mathbf{R}} = \mathbf{R}' \times \mathbf{J}_{N^*\times 1}\,.$$

[24] introduced a rank test of H_0 versus H_A for the case where the Δ's are parameters of a randomized complete block design by modifying the Friedman test. Extending this to our situation, we define an equivalent of the Page L statistic for testing H_0 versus H_1 as

$$W = A'_k \tilde{\mathbf{R}}\,,$$

where $A'_k = (1, \ldots, k)$. A level α test rejects H_0 in favor of H_A if $W > w_\alpha$ where w_α is chosen to satisfy $P_0(W > w_\alpha) = \alpha$. Here P_0 stands for the probability measure under the restriction imposed by H_0. The exact value of w_α can be determined; however, this is a very tedious process. So, we opt for deriving an asymptotic test based on the asymptotic distribution of

W.

We will derive the asymptotic distribution of a more general test statistic of which W is one special case. To that end, let $\Psi = (\psi_1, \ldots, \psi_k)'$ be a vector of contrast coefficients such that $\psi_1 \leq \cdots \leq \psi_k$ and normalized such that $\mathbf{J}'_{k\times 1}\Psi = 0$ and $\Psi'\Psi = 1$.

We then define our "generalized" Page test statistic as

$$\mathbf{W}_\Psi = \Psi'\tilde{\mathbf{R}} \,. \tag{3.2.2}$$

The following theorem gives the null expectation and variance of $\mathbf{W}_\Psi$. The proof is quite long and is found in the appendix.

Theorem 3.1 *Assume that the null hypothesis is true and the score function Ψ are normalized. Then $E_0(\mathbf{W}_\Psi) = 0$ and*

$$\begin{aligned} var_0(W_\Psi) &= \Psi'\mathbf{V}\Psi \\ &= n_1 n_2 \frac{k(k+1)}{12} + n_1 n_2 (n_1 + n_2 - 2)\{-(2k-4)A \\ &\quad - (k-2)B + (k-1)(k-2)C + \frac{3k^2 - 11k + 16}{12}\} \end{aligned}$$

where

$$\begin{aligned} A &= \int P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} > t)\{1 - P_0(U_{11} - U_{11} + U_{(n_1+1)2} > t)\}dF(t);, \\ B &= \int \{P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} > t)\}^2 dF(t) \;, \text{ and} \\ C &= \int \{P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} < t)\}^2 dF(t) \;. \end{aligned}$$

For practical purposes, the quantities A, B, and C in the expression for $var_0(\mathbf{W}_\psi)$ have to be estimated from available data. One can use the nonparametric estimators $\hat{A}_{n_1 n_2}$, $\hat{B}_{n_1 n_2}$, and $\hat{C}_{n_1 n_2}$ by using the empirical distribution function in the defining integrals. This is equivalent to using the Riemann sum of the integrals.

Let $Z_i, \quad i = 1, \ldots, n$ be a stationary time series, let $\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Assuming the variance of $\overline{Z}_n$ decay to zero proportional to $n^{-\alpha}$ for some $0 < \alpha \leq 1$, that is, assuming $Var(\overline{Z}_n) \approx C_{var} n^{-\alpha}$ for some constant $C_{var} > 0$. We have short-range dependence in the case $\alpha = 1$ and long-range dependence (or long memory) in the case $\alpha < 1$. It is usual to characterize a long-memory process by the number $H = 1 - \alpha/2$, the so called *Hurst parameter*. In the following, we will show that test statistic $\mathbf{W}_\Psi$ is a sum of a long-memory process. We can write

$$\begin{aligned} Var_0(\mathbf{W}_\Psi) &= n_1 n_2 C_1(k) + n_1 n_2 (n_1 + n_2 - 2) C_2(k) \\ Var_0(\overline{\mathbf{W}}_\Psi) &= \frac{1}{N^*} C_1(k) + \frac{N-2}{N^*} C_2(k) \end{aligned}$$

where $C_1(k)$ and $C_2(k)$ are functions of k and $N^* = n_1 n_2$. When N^* tends to ∞, we can rewrite $Var_0(\overline{\mathbf{W}}_\Psi)$ as:

$$Var_0(\overline{\mathbf{W}}_\Psi) \approx C_2(k) N^{*(2H-2)}$$

where the *Hurst parameter* $H = \frac{1}{2}\left(\frac{\log(n_1+n_2)}{\log(n_1)+\log(n_2)} + 1\right)$ with $\alpha = 2 - 2H < 1$ and

$$H \to 3/4 \quad \text{as} \quad \min(n_1, n_2) \to \infty \, .$$

If we rewrite the test statistic (3.2.2) as

$$\mathbf{W}_\Psi = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \Psi' R_{ij} = \sum_{i=1}^{n_1} T_i$$

then

$$\overline{\mathbf{W}}_\Psi = \frac{1}{N*} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \Psi' R_{ij} = \sum_{i=1}^{n_1 n_2} \frac{T_i}{n_1 n_2}$$

where we assume $T_i, i = 1, \ldots, n_1$ is a stationary process. So we conclude $\mathbf{W}_\Psi$ is a sum of long-memory process.

To continue our discussion, we need to assume that there is a function $G \in L^2(\mathbb{R})$, such that $T_i/n_2 = G(X_i)$, where X_i is a stationary Gaussian process with long memory. The *Hurst parameter* is now $m = \inf\{k > 0 : c_k \neq 0\}$, where k is the coefficient in the Hermite polynomial expansion of G as $G(x) = \sum_{k=0}^{\infty} \frac{c_k}{k!} H_k(x)$.

Theorem 3.2 *Let $\mathscr{G}_m$ be the class of all functions in $L^2(\mathbb{R})$ with Hermite rank m. Let $G \in \mathscr{G}_m$ and X_t is a stationary Gaussian process with long-range correlation and $S_n = \sum_{t=1}^{n} G(X_t)$. Then the following holds:*

- *If $1/2 < H < 1 - 1/(2m)$, then*

$$\sigma_S^2 = \lim_{n\to\infty} n^{-1} \sum_{j,l=1}^{n} E[G(X_j)G(X_l)]$$

exists and

$$S_n^* = n^{-\frac{1}{2}} S_n \to S \text{ in distribution as } n \to \infty$$

where S is a normal random variable with zero mean and variance σ_S^2

- *If $1 - 1/(2m) < H < 1$, then*

$$S_n^* = n^{-1-m(H-1)} S_n \to \sqrt{c_m} \frac{a_m}{m!} Z_m \text{ in distribution,}$$

where

$$c_m = \frac{2c_\gamma^m m!}{(1 - m(2-2H))(2 - m(2-2H))}$$

c_γ^m is a dealing with X_t and Z_m is a nondegenerate random variable.

Let us discuss the possible values of m.

1. If $m > 2$, $1/2 < 3/4 = H < 1 - 1/(2m)$, which falls under case 1 of Theorem 3.2, in which $\sigma_S^2 = \lim_{n\to\infty} n^{-1} \sum_{j,l=1}^{n} E[G(X_j)G(X_l)]$ exists. In our case,

$$\lim_{n\to\infty} n^{-1} \sum_{j,l=1}^{n} E[G(X_j)G(X_l)] = \lim_{n_1\to\infty} n_1^{-1} \sum_{j,l=1}^{n_1} E[\frac{T_{jn}}{n}\frac{T_{ln}}{n}] \approx \lim_{n_1\to\infty} (n_1 - 1)G_k = \infty$$

where G_k is a function of k. So we conclude that this can not be the case, $m > 2$ is not true.

2. If $m = 2$, which falls under case 2 of Theorem 3.2, but in this case, the distribution of Z_m is complicated.

3. If $m = 1$, which still falls in case 2 of Theorem 3.2. In our case, we can get that

$$N^{*H}\mathbf{W}_\Psi \xrightarrow{\mathcal{D}} \sqrt{c_m}a_m N(0,1)$$

where $\sqrt{c_m}a_m$ is a constant we need to estimate. Since $Var_0(\overline{\mathbf{W}}_\Psi) \approx C_2(k)N^{*(2H-2)}$, we get $Var_0(\mathbf{W}) = C_2(k)N^{*2H}$, we can use $\sqrt{C_2(k)}$ as an estimator of the constant $\sqrt{c_m}a_m$.

We will use simulation to show this is the case.

3.3 Monte Carlo Simulation

3.3.1 Null Simulation

First we perform a null simulation for the two-sample continuous prey feature case under the nominal $\alpha = 0.05$. For this study, independent, identically distributed samples are generated from the standard Cauchy (C), central t with 5 degrees of freedom (t_5), standard normal (N), and the standard Laplace (L) distributions. For the sample sizes we take combinations (n_1, n_2) as $(2,2)$, $(2,4)$,$(2,7)$,$(2,10)$,$(4,4)$,$(4,7)$,$(4,10)$,$(7,7)$,$(7,10)$, and $(10,10)$. For the number of selections of each subject we take $k = 5$. The asymptotic

distribution given in Theorem 3.2 is used to determine the null rejection rates. We perform 10^4 repetitions and find the proportion of cases for which $(n_1 n_2)^H \mathbf{W}_\psi / \sqrt{C_2(k)} > 1.645$. We note that for all the distribution combinations, the null rejection rate are all close to

Table 3.2: Simulated null rejection rates of $\mathbf{W}_\psi$ using its asymptotic normal critical constant, $\psi(t) \propto t$, $\alpha = .05$ and $k = 5$.

(n_1, n_2)	N, N	C, C	t_5, t_5	L, L
2,2	0.0555	0.0590	0.0537	0.0574
2,4	0.0550	0.0613	0.0539	0.0580
2,7	0.0577	0.0554	0.0612	0.0553
2,10	0.0601	0.0586	0.0575	0.0616
4,4	0.0611	0.0558	0.0569	0.0569
4,7	0.0507	0.0522	0.0548	0.0571
4,10	0.0512	0.0529	0.0549	0.0548
7,7	0.0551	0.0536	0.0520	0.0548
7,10	0.0524	0.0591	0.0547	0.0509
10,10	0.0545	0.0524	0.0521	0.0522

and a little above the nominal $\alpha = 0.05$ despite using the asymptotic distribution as the underlying distribution for such small samples. We can conclude that in Theorem 3.2, we should take $m = 1$, and $\sqrt{C_2(k)}$ is a fine although not perfect estimate of the constant $\sqrt{c_m} a_m$. To improve the simulation result, we need to find a better estimate than $\sqrt{C_2(k)}$ for the constant $\sqrt{c_m} a_m$.

3.3.2 Results for the data in [48]

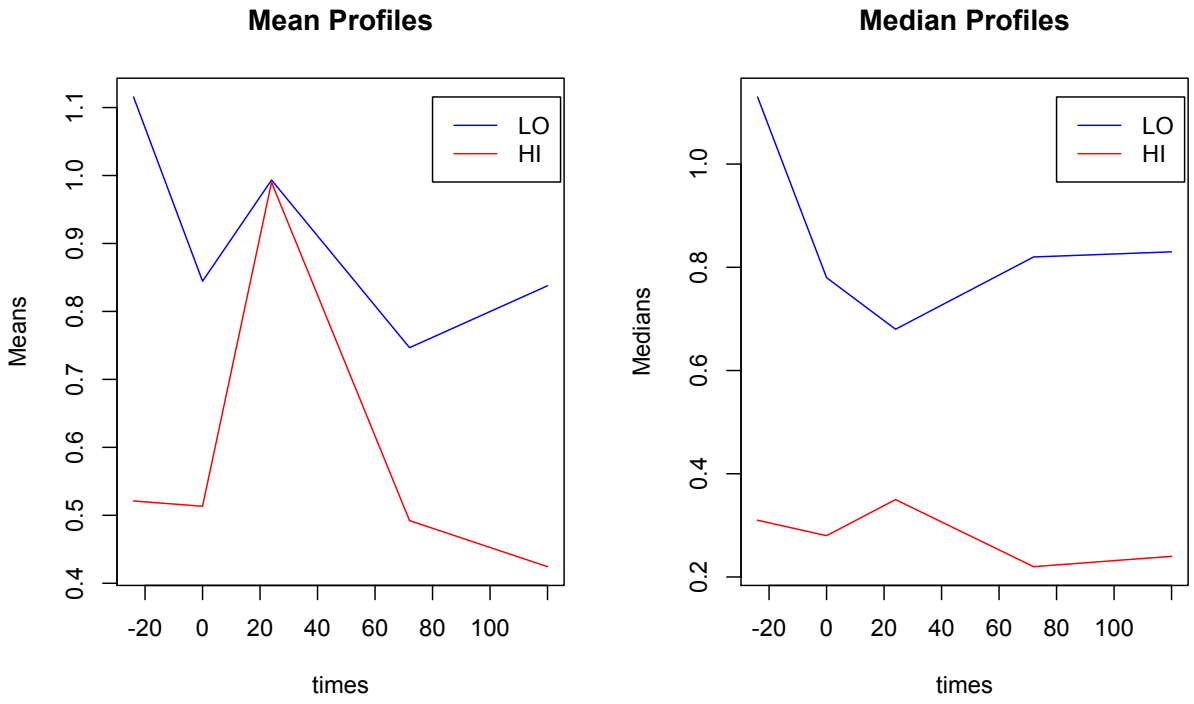
For the data set in [48], we are interested in testing the following hypothesis:

- H_0 : The difference between two groups stays the same.
- H_a : The difference between two groups decays to zero.

Using the nondecreasing normalized linear score function and the asymptotic Gaussian distribution of the test statistic $\mathbf{W}_\psi$, we find that the $p - value$=.466. We would not reject H_0 using this $p - value$ and we conclude that the difference between the two groups (Moderate Fitness and High Fitness) does not decay to zero, which we can also see from the Figure 3.2

.

Figure 3.2: Mean and median profiles for the CRP levels in two groups HI and LO



In both the two figures, the horizontal axis denotes the times at which we obtain the CRP values for each individual. The first figure shows the mean CRP levels in each group (Moderate Fitness and High Fitness) at the five time points. The second figure shows the median CRP levels in each group (Moderate Fitness and High Fitness) at the five time points. Since from figure 3.1 we know the data is highly skewed and there are outliers in this data, we prefer to use the median profile, which clearly verify our result that the difference of CRP levels between the two groups does not decay to zero over time.

3.4 Conclusion

If we already have enough evidence to conclude that two species of predator tend to choose different prey, then we may want to study the difference further, for example, to compare the trends of the two selection patterns. In this chapter, we proposed a rank based method to test the trend between two samples in repeated measurement data. We

establish the asymptotic normality of the test statistic. Our simulation study showed that the asymptotic test retains the nominal Type I error rate. If we want to test the trend of the difference of the opposite direction, that is, $H_a : \Delta_1 \geq \Delta_2 \geq \ldots \geq \Delta_k$ with at least one strict inequality, then we can just multiply the generalized normalized non-decreasing weight function by -1. This will not affect the asymptotic properties of the test statistic.

Chapter 4

Iteratively Reweighted Rank Regression Estimator

4.1 Introduction

In this chapter, we introduce and study IRLS rank estimators. This method of estimating rank regression parameters is essential in developing a general rank approach to generalized estimating equations (GEE). We shall reconsider testing in the two-sample repeated measures problems using rank GEE methods in the next chapter. Moreover, using the rank GEE approach we will develop a program for rank testing and estimation of treatment effects for repeated measurement models with the response distribution assumed to be continuous and in the exponential family of distributions.

Consider the general linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}\ , \tag{4.1.1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown parameters, $\mathbf{Y} \in \mathbb{R}^n$ denotes the response vector, and $\mathbf{X}$ denotes the $n \times p$ design matrix of predictors. The random error vector $\mathbf{e} = (e_1, \ldots, e_n)^T$ is such that $e_1, \ldots, e_n$ are independent and identically distributed according to a distribution function F with density f.

Least squares (LS) procedures are widely used for the analysis of linear models such as the one given in (4.1.1). In addition to their mathematical clarity, they offer the user a unified methodology with which to attack many diverse problems. The LS estimator of $\boldsymbol{\beta}$ in (4.1.1) is given by

$$\widehat{\boldsymbol{\beta}}_{LS} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\text{Argmin}} \, \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2\ ,$$

where $\|\cdot\|^2$ is the squared Euclidean norm; that is $\|\mathbf{v}\|_{LS}^2 = \sum_{i=1}^n v_i^2$ for $\mathbf{v} \in \mathbb{R}^n$. The LS estimator is given by $\widehat{\boldsymbol{\beta}}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Under regularity conditions, the large sample distribution is given by

$$\widehat{\boldsymbol{\beta}}_{LS} \xrightarrow{\mathcal{D}} N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \ .$$

Since the LS estimator minimizes the Euclidean distance between the response vector and the space defined by the linear model, the LS estimator is not robust to outliers and skewness.

There are several classes of robust estimators that are analogous to the traditional LS estimators. They are usually obtained by replacing the Euclidean norm with another norm. Furthermore, depending on the selected norm, the analysis can be made robust and highly efficient compared to the LS analysis. A popular rank-based estimator uses the rank semi-norm instead of the Euclidean norm. The rank semi-norm on $\mathbb{R}^n$ is given by

$$\|\mathbf{v}\|_R = \sum_{i=1}^n a(R(v_i))v_i \ , \qquad \mathbf{v} \in \mathbb{R}^n$$

where $R(v_i)$ denotes the rank of v_i among $v_1, \ldots, v_n$, $a(i) = \phi(i/(n+1))$ and $\phi(u)$ is the score function, which can be any nondecreasing function on $(0, 1)$. The fact that this is a semi-norm is shown in [49]. The rank estimate of $\boldsymbol{\beta}$ is a vector $\widehat{\boldsymbol{\beta}}_R$ such that

$$\widehat{\boldsymbol{\beta}}_R = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\text{Argmin}} \, \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_R \ .$$

This method was proposed by [33]. Under regularity conditions found in Chapter 3 of [44], the asymptotic distribution of the rank estimator is given by

$$\widehat{\boldsymbol{\beta}}_R \xrightarrow{\mathcal{D}} N(\boldsymbol{\beta}, \mathbf{V}_R) \ , \tag{4.1.2}$$

where $\mathbf{V}_R = \tau^2(\mathbf{X}'\mathbf{X})^{-1}$ with

$$\tau^{-1} = \int_0^1 \phi(u) \left\{ -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \right\} du \ . \tag{4.1.3}$$

A common rank estimator is the Wilcoxon estimator that uses $\phi(u) = \sqrt{12}(u - 0.5)$. In this case, $\tau^{-1} = \sqrt{12}\int_{-\infty}^{\infty} f^2(t)dt$.

It can be shown that the rank estimate solves the estimating equations $\mathbf{X}'\mathbf{a}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$, where $\mathbf{a}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ denotes the n-vector with ith component $a[R(Y_i - \mathbf{x}_i'\boldsymbol{\beta})]$. The solution cannot be obtained in closed form, but there are some algorithms available for obtaining an approximate solution [44]. One approach proposed by [34] uses iteratively reweighted least squares (IRLS) to estimate $\widehat{\boldsymbol{\beta}}_R$. The appeal of this method is its simplicity and that it can be obtained using any package that can compute LS estimates.

The purpose of this chapter is to evaluate the finite sample performance of the IRLS estimate of $\widehat{\boldsymbol{\beta}}_R$ and its variance. We provide computational algorithms as well as simulation results pertaining to the estimator, its variance, and its influence function.

4.2 Reweighted Least Squares Formulation

For any $\boldsymbol{\beta} \in \mathbb{R}^p$, let $r_i(\boldsymbol{\beta})$ denote the ith residual $Y_i - \mathbf{x}_i^T\boldsymbol{\beta}$ and $\mathbf{r}(\boldsymbol{\beta}) = (r_1(\boldsymbol{\beta}), \ldots, r_n(\boldsymbol{\beta}))^T$. Now $\widehat{\boldsymbol{\beta}}_n$ minimizes the dispersion function

$$D_n(\boldsymbol{\beta}) := \|\mathbf{r}(\boldsymbol{\beta})\|_R = \sum_{i=1}^{n} \phi\left\{\frac{R(r_i(\boldsymbol{\beta}))}{n+1}\right\} r_i(\boldsymbol{\beta}) \,. \tag{4.2.1}$$

We will assume that the score generating function ϕ is standardized so that $\int_0^1 \phi(u)du = 0$ and $\int_0^1 \phi^2(u)du = 1$. Since ϕ is nondecreasing, $\phi(\nu) = 0$ for some $\nu \in (0, 1)$ and ν is such that $\phi(u) \leq 0$ (≥ 0) whenever $u \leq \nu$ $(u \geq \nu)$. Let $m_\nu(\boldsymbol{\beta})$ be the νth quantile of the $r_i(\boldsymbol{\beta})$, $i = 1, \ldots, n$. The dispersion function in (4.2.1) may be written as

$$\begin{aligned} D_n(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \phi\left\{\frac{R(r_i(\boldsymbol{\beta}))}{n+1}\right\} [r_i(\boldsymbol{\beta}) - m_\nu(\boldsymbol{\beta})] \\ &= \sum_{i=1}^{n} w_i(\boldsymbol{\beta})[r_i(\boldsymbol{\beta}) - m_\nu(\boldsymbol{\beta})]^2 \end{aligned}$$

where

$$w_i(\boldsymbol{\beta}) = \begin{cases} \frac{\phi\left\{\frac{R(r_i(\boldsymbol{\beta}))}{n+1}\right\}}{r_i(\boldsymbol{\beta}) - m_\nu(\boldsymbol{\beta})} , & \text{if } |r_i(\boldsymbol{\beta}) - m_\nu(\boldsymbol{\beta})| \neq 0 , \\ 0 , & \text{if } |r_i(\boldsymbol{\beta}) - m_\nu(\boldsymbol{\beta})| = 0 . \end{cases}$$

Due to the centering of the residuals by $m_\nu(\boldsymbol{\beta})$, the weights, $w_i(\boldsymbol{\beta})$, are nonnegative. The expression above suggests an iterative scheme. Given the kth step estimate, $\widehat{\boldsymbol{\beta}}_k$, the $(k+1)$th step estimate $\widehat{\boldsymbol{\beta}}_{k+1}$ minimizes the $(k+1)$th step dispersion given by

$$D_n^*(\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}}_k) = \sum_{i=1}^{n} w_i(\widehat{\boldsymbol{\beta}}_k)((Y_i - \mathbf{x}_i^T\boldsymbol{\beta}) - m_\nu(\widehat{\boldsymbol{\beta}}_k))^2, \quad k = 0, 1, \ldots , \tag{4.2.2}$$

that is

$$\widehat{\boldsymbol{\beta}}_{k+1} = \underset{\boldsymbol{\beta}}{\text{Argmin}}\, D_n^*(\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}}_k) .$$

Such algorithms have been considered in the past for computing estimators. [50] used such a method to obtain least absolute deviation estimators. [51] used a similar strategy to obtain M-estimates of regression coefficients. [52] develop an IRLS algorithm for the dispersion function given by (4.2.1) but centering using the mean of the residuals at every step.

It is shown in [34] that $\sqrt{n}\|\widehat{\boldsymbol{\beta}}_k - \widehat{\boldsymbol{\beta}}_n\| \to 0$ in probability as $n \to \infty$ for every $k \geq 1$ under some regularity conditions including the consistency of the initial estimator $\widehat{\boldsymbol{\beta}}_0$. They argued that since the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$ are known, the asymptotic properties of $\widehat{\boldsymbol{\beta}}_k$ are also known and equivalent to those of $\widehat{\boldsymbol{\beta}}_n$ for every $k \geq 1$. What are not known, however, are the finite sample properties of $\widehat{\boldsymbol{\beta}}_k$.

It is well known that the LS estimator is more efficient than the rank estimator when the error distribution is normal and the rank estimator has higher asymptotic efficiency than the LS estimator when the error distribution has long tails [44]. For finite samples, however, we expect the IRLS estimator discussed in this chapter to borrow some of its properties from LS. As such, we hypothesize that the IRLS estimator will have higher efficiency than both the LS and the rank estimators for error distributions with moderate tail thicknesses.

4.3 Computational Details and Results

4.3.1 Algorithm for Estimating Regression Coefficients

Let $Y_i^*(\boldsymbol{\beta}) = Y_i - m_\nu(\boldsymbol{\beta})$. Taking the derivative of equation (4.2.2) with respect to $\boldsymbol{\beta}$ and setting the resulting expression equal to 0 we obtain

$$\sum_i w_i(\widehat{\boldsymbol{\beta}}_k)[Y_i^*(\widehat{\boldsymbol{\beta}}_k) - \mathbf{x}_i^T\boldsymbol{\beta}]\mathbf{x}_i = 0 \tag{4.3.1}$$

which is the weighted LS estimating equation for the model $Y_i^*(\widehat{\boldsymbol{\beta}}_k) = \mathbf{x}_i^T\boldsymbol{\beta}+\varepsilon_i\ , \quad 1 \leq i \leq n.$

Given an initial estimate $\widehat{\boldsymbol{\beta}}_0$, we find the $(k+1)$th step IRLS estimate as

$$\widehat{\boldsymbol{\beta}}_{k+1} = (\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_k)\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_k)(\mathbf{Y} - m_\nu(\widehat{\boldsymbol{\beta}}_k)) \tag{4.3.2}$$

for $k = 0, 1, \ldots$.

Here we describe the algorithm used for computing the rank estimate of $\boldsymbol{\beta}$ using IRLS.

Algorithm 4.1

1. *Let the initial estimator $\widehat{\boldsymbol{\beta}}_0$ be given. Set c_w and c_β to small positive numbers (usually 10^{-5}) and set $k = 1$.*

2. *If $|r_i(\widehat{\boldsymbol{\beta}}_{k-1}) - m_\nu(\widehat{\boldsymbol{\beta}}_{k-1})| < c_w$, then set $w_i(\widehat{\boldsymbol{\beta}}_{k-1}) = 0$ for $i = 1, \ldots, n$. Set $\mathbf{W}(\widehat{\boldsymbol{\beta}}_{k-1}) = diag(w_1(\widehat{\boldsymbol{\beta}}_{k-1}), \ldots, w_n(\widehat{\boldsymbol{\beta}}_{k-1}))$.*

3. *Calculate $\widehat{\boldsymbol{\beta}}_k = (\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_{k-1})\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\boldsymbol{\beta}_{k-1})(\mathbf{Y} - m_\nu(\widehat{\boldsymbol{\beta}}_{k-1}))$*

4. *If $\|\widehat{\boldsymbol{\beta}}_k - \widehat{\boldsymbol{\beta}}_{k-1}\| > c_\beta\|\widehat{\boldsymbol{\beta}}_{k-1}\|$, then set $k = k+1$ and return to step 2; else, stop.*

We compare our approach to that of the traditional approach of obtaining the rank estimate of $\boldsymbol{\beta}$. The traditional approach uses the quasi-Newton algorithm proposed by [53]. It performs a Newton step starting with an initial estimate. It then checks if the step is successful in reducing the dispersion function. If the step is successful, then it will continue Newton

step iterations. Otherwise, it performs a linear search to bracket the solution. We will use the Wilcoxon semi-norm and the R code given by [54] that uses quantile regression methods of [55] to compute the estimate using the traditional approach. This code is available at `http://www.stat.wmich.edu/mckean/HMC/Rcode`. We will refer to the estimate obtained using this method the HM estimate.

The simulation considers estimating $\boldsymbol{\beta} = (\beta_2, \beta_3, \beta_4)^T$ in the model

$$Y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + e_i \ , \qquad i = 1, \ldots, n \ .$$

We generated x_{ji} from the Uniform$(0, 10j)$ distribution, $j = 2, 3, 4$. The errors e_i were randomly generated from the t distribution or the contaminated normal distribution in each repetition. We shall set the true parameter $\boldsymbol{\beta} = (0, 0, 0)^T$; that is we take $Y_i = e_i$.

In repetition b, let the estimates of $\boldsymbol{\beta}$ obtained using IRLS, HM, and LS be $\widehat{\beta}_{Ij}^{(b)}$, $\widehat{\beta}_{HMj}^{(b)}$, and $\widehat{\beta}_{LSj}^{(b)}$, respectively, for $j = 2, 3, 4$ and $b = 1, \ldots, B$. The average mean squared error estimates IRLS, HM, and LS are found as

$$\widehat{\text{MSE}}_I = \frac{1}{3B} \sum_{j=1}^{3} \sum_{b=1}^{B} (\widehat{\beta}_{Ij}^{(b)} - \beta)^2 \ ,$$
$$\widehat{\text{MSE}}_{HM} = \frac{1}{3B} \sum_{j=1}^{3} \sum_{b=1}^{B} (\widehat{\beta}_{HMj}^{(b)} - \beta)^2 \ , \text{ and}$$
$$\widehat{\text{MSE}}_{LS} = \frac{1}{3B} \sum_{j=1}^{3} \sum_{b=1}^{B} (\widehat{\beta}_{LSj}^{(b)} - \beta)^2 \ ,$$

respectively. The estimated relative efficiencies of IRLS with respect to HM and LS are

$$RE(I, HM) = \frac{\widehat{\text{MSE}}_{HM}}{\widehat{\text{MSE}}_I} \quad \text{and} \quad RE(I, LS) = \frac{\widehat{\text{MSE}}_{LS}}{\widehat{\text{MSE}}_I} \ ,$$

where $RE > 1$ indicates that IRLS is more efficient than the corresponding procedure.

The t distribution was chosen as an error distribution since it covers a wide range of tail thicknesses starting with the Cauchy (t_1) to the normal (t_∞). The contaminated normal distribution is ideal for studying the effect of outliers on estimates. We will consider a scale contamination of the $N(0,1)$ distribution which is created by sampling a proportion $1-\epsilon$ of the time from the $N(0,1)$ distribution and a proportion ϵ of the time from the $N(0,\sigma^2)$ distribution. We will use $CN(\epsilon,\sigma)$ to represent the contaminated normal distribution whose distribution function is given by

$$F_{\epsilon,\sigma}(x) = (1-\epsilon)\Phi(x) + \epsilon\Phi(x/\sigma) \ ,$$

where Φ is the standard normal distribution function. Table 4.1 and 4.2 give the results of our simulation study for $B = 5000$ repetitions. As can be seen in Table 4.1, when the errors are

Table 4.1: Relative efficiencies of IRLS versus HM and LS when the errors follow a t_{df} distribution

	$df = 3$		$df = 6$		$df = 12$	
n	HM	LS	HM	LS	HM	LS
10	0.981	1.267	1.051	1.042	1.057	.987
15	0.965	1.474	1.024	1.040	1.046	.993
20	0.945	1.433	1.011	1.065	1.039	1.005

generated from the t_{df} distribution, HM is more efficient than LS and IRLS when the error distribution is heavy tailed ($df = 3$). LS is more efficient than HM and IRLS when the tail of the distribution approaches the tails of the normal distribution ($df = 12$). IRLS is more efficient than HM and LS for a moderate-tailed distribution ($df = 6$). Table 4.2 contains

Table 4.2: Relative efficiencies of IRLS versus HM and LS when the errors follow a $CN(\epsilon,3)$ distribution

	$\epsilon = 0$		$\epsilon = .05$		$\epsilon = .10$	
n	HM	LS	HM	LS	HM	LS
10	1.096	.970	1.040	1.062	1.041	1.022
15	1.059	.968	1.011	1.094	.984	1.146
20	1.057	.974	.996	1.117	.957	1.182

results of the case where the error distribution is $N(\epsilon,3)$. The results for $CN(\epsilon,5)$ were

similar; hence omitted. It is clear that LS is the most efficient of the three when the error distribution is normal ($\epsilon = 0$). For large contamination ($\epsilon = .10$), HM is more efficient than LS and IRLS except for very small sample sizes. IRLS seems to have superior performance for such cases. IRLS is the most efficient of the three for moderate contamination ($\epsilon = .05$).

4.3.2 Variance Estimator

Suppose $\widehat{\boldsymbol{\beta}}_k$ is the consistent kth step IRLS estimator of $\boldsymbol{\beta}$. Given $\widehat{\boldsymbol{\beta}}_k$, the variance of $\widehat{\boldsymbol{\beta}}_{k+1}$ given in (4.3.2) is

$$\widehat{\mathbf{V}} = E[Var(\widehat{\boldsymbol{\beta}}_{k+1}|\widehat{\boldsymbol{\beta}}_k)] + Var[E(\widehat{\boldsymbol{\beta}}_{k+1}|\widehat{\boldsymbol{\beta}}_k)] \ ,$$

where

$$Var(\widehat{\boldsymbol{\beta}}_{k+1}|\widehat{\boldsymbol{\beta}}_k) = [(\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_k)\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_k)]Var(\mathbf{Y})[(\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_k)\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_k)]' \quad (4.3.3)$$

and

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}_{k+1}|\widehat{\boldsymbol{\beta}}_k) &= (\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_k)\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_k)(\mathbf{X}'\boldsymbol{\beta} - m_\nu(\widehat{\boldsymbol{\beta}}_k)) \\ &= \boldsymbol{\beta} - (\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_k)\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\widehat{\boldsymbol{\beta}}_k)m_\nu(\widehat{\boldsymbol{\beta}}_k) \end{aligned} \quad (4.3.4)$$

We will use the leave-one-out procedure described below to obtain estimates of (4.3.3) and (4.3.4).

Algorithm 4.2

1. *For $i = 1, \ldots, n$, leave the ith observation out, calculate the regression coefficient $\widehat{\boldsymbol{\beta}}^{(i)}$ using the $(n-1)$ observations in* Algorithm 4.1. *Let the number of steps required to converge be $k^{(i)} + 1$.*

2. *Use* (4.3.3) *and* (4.3.4) *to calculate $Var(\widehat{\boldsymbol{\beta}}^{(i)}_{k^{(i)}+1}|\widehat{\boldsymbol{\beta}}^{(i)}_{k^{(i)}})$ and $E(\widehat{\boldsymbol{\beta}}^{(i)}_{k^{(i)}+1}|\widehat{\boldsymbol{\beta}}^{(i)}_{k^{(i)}})$*

3. Set $\widehat{\mathbf{V}}_1 = \frac{1}{n}\sum_{i=1}^{n} Var(\widehat{\boldsymbol{\beta}}^{(i)}_{k^{(i)}+1}|\widehat{\boldsymbol{\beta}}^{(i)}_{k^{(i)}})$ *and* $\widehat{\mathbf{V}}_2 = Var(\frac{1}{n}\sum_{i=1}^{n} E(\widehat{\boldsymbol{\beta}}^{(i)}_{k^{(i)}+1}|\widehat{\boldsymbol{\beta}}^{(i)}_{k^{(i)}}))$.

4. Take $\widehat{\mathbf{V}} = \widehat{\mathbf{V}}_1 + \widehat{\mathbf{V}}_2$ *as estimate of* $\mathbf{V}$.

In the literature, there are a number of proposed methods to estimate $\mathbf{V} = \tau^2(\mathbf{X}'\mathbf{X})^{-1}$. All of them are based on estimating τ from the set of residuals via either considering lengths of nonparametric intervals based on the Walsh averages of the residuals [53] or nonparametric density estimation [56]. We shall compare our estimates of $\mathbf{V}$ obtained through Algorithm 4.2 to that of [56]. The algorithm of [56], described briefly here, is implemented in the R code of [54].

The method of [56] uses kernel density estimates with rectangular kernels. Let $\widehat{e}_1, \ldots, \widehat{e}_n$ be the estimated residuals and let

$$\widehat{H}_n(y) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}[\phi^*(j/n) - \phi^*(i/n)]I(|\widehat{e}_{(i)} - \widehat{e}_{(j)}| \leq y),$$

where

$$\phi^*(u) = \frac{\phi(u) - \phi(0)}{\phi(1) - \phi(0)}\,.$$

A consistent estimator of τ is given by

$$\hat{\tau} = \sqrt{\frac{n}{n-p-1}}\left\{\frac{[\phi(1) - \phi(0)]\widehat{H}_n(\widehat{H}_n^{-1}(\delta))/\sqrt{n}}{2\widehat{H}_n^{-1}(\delta)/\sqrt{n}}\right\}^{-1}\,.$$

The recommended values of δ are 0.8 if $n > 5p$ and 0.9 otherwise. Similar to the case of coefficients, we will refer to the estimate of the variance obtained this way as HM.

In the following, we perform a Monte Carlo comparison of the variances of the IRLS method to HM. The setting of the study is the same as the one used in Subsection 4.3.1 and the MSEs are computed in a similar fashion. Since the distribution of the errors is specified in the simulation, the true value of τ is found via exact or numerical integration using $\tau^{-1} = \sqrt{12}\int_{-\infty}^{\infty} f^2(t)dt$. The relative efficiencies of IRLS versus HM using the average MSEs are given in Tables 4.3 and 4.4. We can see from Table 4.3 that when the errors

Table 4.3: REs of $\widehat{\mathbf{V}}$ of IRLS vs HM when $e_i \sim t_{df}$

n	$df = 3$	$df = 6$	$df = 12$
10	.481	6.608	3.674
15	.391	2.200	2.482
20	.044	1.413	1.710

follow the t_{df} distribution, HM is more efficient for heavier tailed distributions ($df = 3$) and IRLS is more efficient for moderate and light tailed distributions ($df = 6, 12$). From

Table 4.4: REs of $\widehat{\mathbf{V}}$ of IRLS vs HM when $e_i \sim CN(\epsilon, 3)$

n	$\epsilon = 0$	$\epsilon = .05$	$\epsilon = .10$
10	9.907	2.827	3.809
15	3.871	1.775	1.916
20	2.431	.855	.347

Table 4.4, we observe that for the $CN(\epsilon, 3)$ distribution, HM is more efficient for larger contamination ($\epsilon = .05, .10$) when the sample size is relatively large ($n = 20$). Generally IRLS gives more efficient estimates when there is no contamination $\epsilon = 0$ or the sample size is small ($n = 10, 15$).

4.3.3 Influence Function

One measure of robustness is the influence function [57] which measures the amount of change in the estimator that can be brought about by an infinitesimal local contamination. Following the approach of [58], we can derive the influence function of $\widehat{\boldsymbol{\beta}}_{k+1}$ given in (4.3.2) at a given point $(\mathbf{x}_0, y_0)$. Let $r = y - \mathbf{x}'\boldsymbol{\beta}$ and define $A_w = \Gamma_w^{-1}\Lambda_w$ and $z(\mathbf{x}_0, y_0) = \Gamma_w^{-1} w(y_0 - \mathbf{x}_0'\boldsymbol{\beta})(y_0 - \mathbf{x}_0'\boldsymbol{\beta})\mathbf{x}_0$, where

$$\Lambda_w = -E\left[r\frac{\partial w}{\partial r}\mathbf{x}\mathbf{x}'\right] \quad \text{and} \quad \Gamma_w = E[w\mathbf{x}\mathbf{x}'] \,.$$

Now the influence function of $\widehat{\boldsymbol{\beta}}_{k+1}$ at $(\mathbf{x}_0, y_0)$ is

$$\mathrm{IF}(\widehat{\boldsymbol{\beta}}_{k+1}; \mathbf{x}_0, y_0) = \sum_{i=0}^{k} A_w^i z(\mathbf{x}_0, y_0) + A_w^{k+1} \mathrm{IF}(\widehat{\boldsymbol{\beta}}_0; \mathbf{x}_0, y_0) \ .$$

Note that for our weights

$$z(\mathbf{x}_0, y_0) = \Gamma_w^{-1} \phi(F(y_0 - \mathbf{x}_0' \boldsymbol{\beta}))\mathbf{x}_0 \ .$$

So, $\mathrm{IF}(\widehat{\boldsymbol{\beta}}_{k+1}; \mathbf{x}_0, y_0)$ will be bounded in the y-direction if we use a bounded score function ϕ and $\mathrm{IF}(\widehat{\boldsymbol{\beta}}_0; \mathbf{x}_0, y_0)$ is bounded in the y-direction. However, $\mathrm{IF}(\widehat{\boldsymbol{\beta}}_{k+1}; \mathbf{x}_0, y_0)$ remains unbounded in the $\mathbf{x}$-direction regardless of the boundedness of $\mathrm{IF}(\widehat{\boldsymbol{\beta}}_0; \mathbf{x}_0, y_0)$. This is in fact true for rank estimates. Rank estimates that achieve bounded influence have been given by [47] and [59]. These use the so-called weighted Wilcoxon dispersion function proposed by [60]. The resulting dispersion function is not the same as the one in (4.2.1).

To evaluate the robustness of the IRLS estimator in finite sample cases, we performed a Monte Carlo analysis using the sensitivity curve of the estimator. The sensitivity curve is the empirical influence function defined by [61, 62]

$$\mathrm{SC}_n(\mathbf{z}) = \frac{\widehat{\boldsymbol{\beta}}(\mathbf{z}, \mathbf{Z}_1, \ldots, \mathbf{Z}_n) - \widehat{\boldsymbol{\beta}}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)}{1/(n+1)} \ ,$$

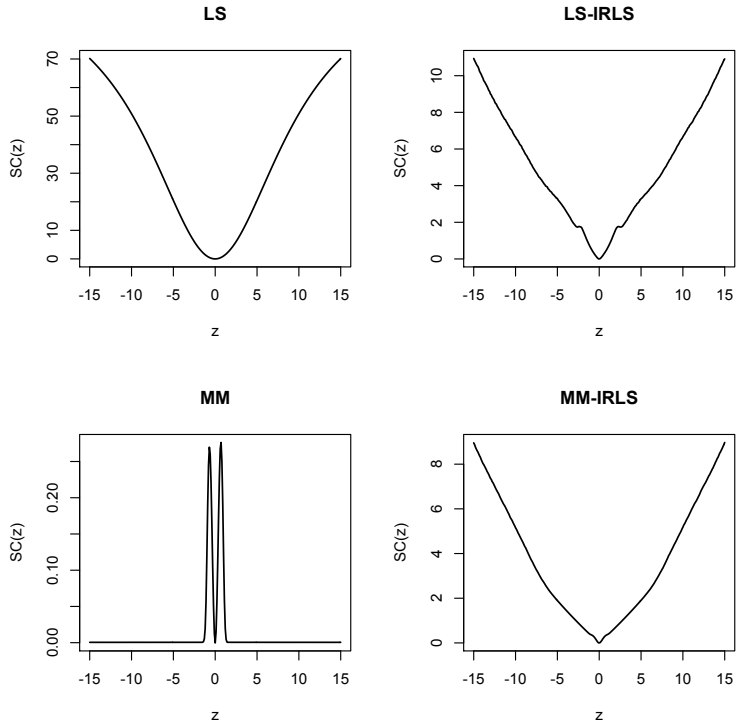
where $\mathbf{Z}_i = (\mathbf{X}_i, Y_i)$ for $i = 1, \ldots, n$, $\mathbf{z} \in \mathbb{R}^{p+1}$, and $\widehat{\boldsymbol{\beta}}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ represents the estimator of $\boldsymbol{\beta}$ based on the observations $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. This is used to get an idea of the sensitivity of $\widehat{\boldsymbol{\beta}}$ to local changes.

We performed a Monte Carlo analysis of the sensitivity of the IRLS estimator. In our simulation, we considered a simple linear model through the origin given by

$$Y_i = \beta X_i + \varepsilon_i \ , \qquad i = 1, \ldots, 100 \ ,$$

where $X_i \sim N(0,1)$, $\varepsilon_i \sim N(0, 16^{-1})$, and the true value of β was taken to be zero. We obtained 1000 IRLS estimates of β, each time using the generated data and after a point $(x, y) \in \{(-15, -15), (-14.9, -14.9), \ldots, (15, 15)\}$ was added. As initial values, we used the least squares estimate of β as well as the high-breakdown and bounded influence MM estimate of [63]. The average sensitivity curve of the initial (left panel) and the fully iterated IRLS estimate (right panel) are given in Figure 4.1. It is clear from the figure that MM is

Figure 4.1: Sensitivity Curves



a robust estimator while LS has an unbounded sensitivity curve. It is also clear that, even when we start with a highly robust estimator, the IRLS estimate of β remains sensitive to outlying values in the data.

4.4 Tests of Linear Hypotheses

Consider testing a general linear hypothesis of the form

$$H_0 : A\boldsymbol{\beta} = 0 \quad \text{versus} \quad H_1 : A\boldsymbol{\beta} \neq 0 \tag{4.4.1}$$

where A is a $q \times p$ hypothesis matrix with rank q. For example, when A is the $p \times p$ identity matrix, H_1 corresponds to regression significance. The Wald test rejects H_0 if

$$T^2 = (A\widehat{\boldsymbol{\beta}})'(A\widehat{\mathbf{V}}A')^{-1}(A\widehat{\boldsymbol{\beta}}) \tag{4.4.2}$$

is larger than $\chi^2_{1-\alpha}(q)$. However, as shown in [64], for finite samples, a better test rejects H_0 if $T^2/q > F_{1-\alpha}(q, n-p-1)$, where $F_{1-\alpha}(q, n-p-1)$ corresponds to the $(1-\alpha)100\%$ percentile of the F distribution with q and $n-p-1$ degrees of freedom.

We now perform a small Monte Carlo simulation to compare the performance of T^2 obtained via IRLS and HM. To this end, we consider the model used in Subsection 4.3.1 with $n = 30$. We are interested in testing

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \quad \text{versus} \quad H_0 : \boldsymbol{\beta} \neq \mathbf{0} \; .$$

This is equivalent to using $A = I_3$, the 3×3 identity matrix, in the general linear hypothesis (4.4.1). Table 4.5 contains the proportion of rejections under H_0 in 1000 repetitions. Table 4.5 shows that the Wald test based on IRLS appears to overestimate the nominal α

Table 4.5: Empirical α levels for IRLS vs HM

	$N(0,1)$			$CN(.1,3)$		
	.01	.05	.10	.01	.05	.10
IRLS	.024	.072	.129	.012	.063	.083
HM	.004	.035	.074	.005	.037	.070

and the Wald test based on HM underestimates α. IRLS comes closer to the nominal α in the case of contaminated normal errors.

4.5 Conclusion

For linear models, the least squares estimator of the regression coefficient is optimal if the error distribution is normal. For distributions with longer tails than the normal, however, rank estimators are more efficient than the least squares estimator. [34] proposed an IRLS method for estimating rank estimators. In this chapter, we study the finite sample performance of the IRLS method. For Wilcoxon scores, using the LS estimate as an initial estimate, we find that the IRLS algorithm leads to estimates whose efficiency is between those of the LS and rank estimates obtained using the algorithm of [54]. The LS estimates are efficient when the distribution of the error is short-tailed. The algorithm of [54] gives efficient estimates when the distribution of the error is long-tailed while the IRLS algorithm gives efficient estimates when the distribution of the error is moderate-tailed. It is also observed that there is a need for an IRLS formulation of bounded influence rank estimators [47]. The boundedness of the initial estimator is a necessary condition for the boundedness of the influence function of IRLS estimator, but it is not sufficient.

Chapter 5

IRLS Estimation of Rank Generalized Estimation Equations Model

5.1 Introduction

The analysis of data resulting from repeated measurement designs such designs often becomes complicated because of the non-zero within-subject correlation. To avoid this problem [10] proposed the Generalized Estimation Equations (GEE) Model, an extension of generalized linear models to the analysis of longitudinal data. They introduced a class of estimating equations that give consistent estimates of the regression parameters and of their variance under mild assumptions about the time dependence. The estimating equations are derived without specifying the joint distribution of a subject's observations yet they reduce to the score equations for multivariate Gaussian outcomes. The GEE specify only the relationship between the marginal mean of the response variable and covariates. Within-subject correlation is then accounted for through a 'working' correlation matrix.

Since then, the GEE model has been used to deal with many specific data models, for example, for binary models [65] [66], in nonlinear regression models [67], in proportional odds model [68], in Binomial model [69] [70], in logistic model [71], in contingency tables [72], in mixed logistic models [73], in normally distributed models [74].

However, the solution of GEEs proposed by [10] is obtained using an iterated reweighted least squares fitting which is not robust. One solution proposed by [30] is by introducing a diagonal weight matrix for the within-subject correlation into the estimating equations. Another solution is one given by [31] who proposed an adaption of the Wilcoxon-Mann-Whitney method of estimating linear regression parameters for use in longitudinal data analysis under the working independence model. They used joint ranking (JR) of all observations in their development. [32] consider the same model as [31] but they use separate between-subject

and within-subject ranks to specify their Wilcoxon-Mann-Whitney estimating equation. [2] proposed a rank-based fitting procedure that only involves substituting a norm based on a score function for the Euclidean norm used by [10]. Their subsequent fitting, while also an iterated reweighted least squares solution to GEEs, is robust to outliers in response space and further the weights can easily be adapted for robustness in factor space.

In this Chapter, we will first introduce the ordinary Least Square based GEE models proposed by [10] and the Iterated Reweighted Rank-Based Estimates for GEE models. Then under similar assumptions, we introduce our selective predation pattern problem as a special case of the two different GEEs.

5.2 Ordinary LS based GEE models

Consider a longitudinal set of observations over K subjects, let y_{ij} denote the jth response for the ith subject for $j = 1, 2 \ldots, n_i$ and $i = 1, 2, \ldots, K$. Assume that x_{ij} is a $p \times 1$ vector of corresponding covariates. Let $N = \sum_{i=1}^{K} n_i$ denote the total sample size. Assume that the marginal distribution of y_{ij} is of the exponential class of distributions and is given by

$$f(y_{ij}) = exp[y_{ij}\theta_{ij} - a(\theta_{ij}) + b(y_{ij})]\phi \tag{5.2.1}$$

where $\phi > 0$, $\theta_{ij} = h(\eta_{ij})$ and $\eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}$. Thus the mean and variance are given by

$$E(y_{ij}) = a^{'}(\theta_{ij}) \;\; and \;\; Var(y_{ij}) = a^{''}(\theta_{ij})/\phi$$

In this notation, the link function is $h^{-1} \circ (a^{'})^{-1}$.

Let $\mathbf{Y}_i = (y_{i1}, \ldots, y_{in_i})$ and $\mathbf{X}_i = (x_{i1}, \ldots, x_{in_i})$ denote the $n_i \times 1$ vector of response and the $n_i \times p$ matrix of covariates, respectively, for the ith individual. We consider the general case where the components of the vector of responses for the ith subject, $\mathbf{Y}_i$, are dependent. Let $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{in_i})^T$, so that $E(\mathbf{Y}_i) = \mathbf{a}'(\boldsymbol{\theta}_i) = (a'(\theta_{i1}), \ldots, a'(\theta_{in_i}))^T$. For a $s \times 1$ vector of unknown parameters $\boldsymbol{\alpha}$, let $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\alpha})$ denote a $n_i \times n_i$ correlation matrix. Define

the matrix

$$\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}_i(\alpha)\mathbf{A}_i^{1/2}/\phi \tag{5.2.2}$$

where $\mathbf{A}_i = diag(a''(\theta_{i1}), \ldots, a''(\theta_{in_i}))$. We refer to $\mathbf{R}_i$ as the working correlation matrix. $\mathbf{V}_i$ will be equal to $cov(\mathbf{Y}_i)$ if $\mathbf{R}_i(\boldsymbol{\alpha})$ is the true correlation matrix for the $\mathbf{Y}_i$. For estimation, let $\widehat{\mathbf{V}}_i$ be an estimate of $\mathbf{V}_i$ and $\widehat{\mathbf{R}}_i$ be an estimate of $\mathbf{R}_i$, which, in general, requires estimation of $\boldsymbol{\alpha}$ and often an initial estimate of $\boldsymbol{\beta}$. In general, we will denote the estimator of $\boldsymbol{\alpha}$ by $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)$ to reflect its dependence on $\boldsymbol{\beta}$ and ϕ.

Liang and Zeger (1986) defined their estimate in terms of general estimation equations (GEE). Define the $n_i \times p$ Hessian matrix,

$$\mathbf{D}_i = \frac{\partial \mathbf{a}'(\theta_i)}{\partial \boldsymbol{\beta}}, i = 1, \ldots, K$$

Then their GEE estimator $\widehat{\boldsymbol{\beta}}_{LS}$ is the solution to the equations

$$\sum\nolimits_{i=1}^{K} \mathbf{D}_i^T \widehat{\mathbf{V}}_i^{-1}[\mathbf{Y}_i - \mathbf{a}'(\boldsymbol{\theta}_i)] = 0. \tag{5.2.3}$$

which we denoted by

$$\sum\nolimits_{i=1}^{K} \mathbf{U}_i[\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\}] = 0$$

We can define the dispersion function in terms of the Euclidean norm.

$$\begin{aligned}
\mathbf{D}_{LS}(\boldsymbol{\beta}) &= \sum\nolimits_{i=1}^{K} [\mathbf{Y}_i - \mathbf{a}'(\boldsymbol{\theta}_i)]^T \widehat{\mathbf{V}}_i^{-1} [\mathbf{Y}_i - \mathbf{a}'(\boldsymbol{\theta}_i)] \\
&= \sum\nolimits_{i=1}^{K} [\widehat{\mathbf{V}}_i^{-\frac{1}{2}} \mathbf{Y}_i - \widehat{\mathbf{V}}_i^{-\frac{1}{2}} \mathbf{a}'(\boldsymbol{\theta}_i)]^T [\widehat{\mathbf{V}}_i^{-\frac{1}{2}} \mathbf{Y}_i - \mathbf{V}_i^{-\frac{1}{2}} \mathbf{a}'(\boldsymbol{\theta}_i)] \\
&= \sum\nolimits_{i=1}^{K} \sum\nolimits_{j=1}^{n_i} [y_{ij}^* - d_{ij}(\boldsymbol{\beta})]^2
\end{aligned}$$

where $\mathbf{Y}_i^* = \widehat{\mathbf{V}}_i^{-\frac{1}{2}}\mathbf{Y}_i = (y_{i1}^*, \ldots, y_{ik}^*)^T$, $d_{ij}(\boldsymbol{\beta}) = \mathbf{c}_j^T\mathbf{a}'(\boldsymbol{\theta}_i)$, and c_j^T is the jth row of $\widehat{\mathbf{V}}_i^{-\frac{1}{2}}$. The gradient of $\mathbf{D}_{LS}(\beta)$ is

$$\nabla\mathbf{D}_{LS}(\boldsymbol{\beta}) = -\sum_{i=1}^{K}\mathbf{D}_i^T\widehat{\mathbf{V}}_i^{-1}[\mathbf{Y}_i - a'(\boldsymbol{\theta}_i)]. \tag{5.2.4}$$

Thus the solution to (5.2.3) also can be expressed as

$$\hat{\boldsymbol{\beta}}_{LS} = Argmin\mathbf{D}_{LS}(\boldsymbol{\beta}) \tag{5.2.5}$$

From this point of view, $\hat{\boldsymbol{\beta}}_{LS}$ is a nonlinear least squares (LS) estimator. The following theorem given by [10] establishes the asymptotic normality of $\hat{\boldsymbol{\beta}}_{LS}$.

Theorem 5.1 *Under mild regularity conditions and given that:*

A.1 $\hat{\alpha}$ *is* $K^{\frac{1}{2}}$*-consistent given* $\boldsymbol{\beta}$ *and* ϕ*.*

A.2 $\hat{\phi}$ *is* $K^{\frac{1}{2}}$*-consistent given* $\boldsymbol{\beta}$*; and*

A.3 $|\partial\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi)/\partial\phi| \leq H(\mathbf{Y},\beta)$ *which is* $O_p(1)$*.*

Then $K^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta})$ *is asymptotically multivariate Gaussian with zero mean and covariance matrix* $\mathbf{V}_{LS}$ *given by*

$$\mathbf{V}_{LS} = \lim_{K\to\infty} K(\sum_{i=1}^{K}\mathbf{D}_i^T\mathbf{V}_i^{-1}\mathbf{D}_i)^{-1}\{\sum_{i=1}^{K}\mathbf{D}_i^T\mathbf{V}_i^{-1}cov(\mathbf{Y}_i\mathbf{V}_i^{-1}\mathbf{D}_i)\}(\sum_{i=1}^{K}\mathbf{D}_i^T\mathbf{V}_i^{-1}\mathbf{D}_i)^{-1}$$

5.3 Iterated Reweighted Rank-Based Estimates for GEE Models

[75] developed a class of nonlinear robust estimators. Similar to nonlinear LS estimators, these estimators minimize a norm of the residuals where, for a vector $v \in \mathbf{R}^n$, the norm is defined by

$$\|v\| = \sum_{i=1}^{n}\psi[r(v_i)/(n+1)]v_i$$

where $r(v_i)$ denote the rank of v_i among $v_1, \ldots, v_n$ and the score function $\psi(u)$ is a nondecreasing, square-integrable function defined on the interval $(0, 1)$, Without loss of generality, we standardized ψ so that

$$\int \psi(u)du = 0 \; and \; \int \psi(u)^2 du = 1$$

For nonnegative weights, we need one other assumption on the score function. For discussion, we also assume that the score function is odd about 1/2; that is

$$\psi(1-u) = -\psi(u)$$

let $\mathbf{Y}_i^* = \widehat{\mathbf{V}}_i^{\frac{1}{2}}\mathbf{Y}_i = (y_{i1}^*, \ldots, y_{ik}^*)^T$,$g_{ij}(\boldsymbol{\beta}) = \mathbf{c}_j^T\mathbf{a}'(\boldsymbol{\theta}_i)$, where $\mathbf{c}_j^T$ is the jth row of $\widehat{\mathbf{V}}_i^{\frac{1}{2}}$, and let $\mathbf{G}_i^* = [g_{ij}]$. The rank-based dispersion function is given by

$$\mathbf{D}_R(\beta) = \sum\nolimits_{i=1}^{K}\sum\nolimits_{j=1}^{n_i}\psi[r(y_{ij}^* - g_{ij}(\beta))/(n+1)][y_{ij}^* - g_{ij}(\beta)] \tag{5.3.1}$$

We next write the R estimator as weighted LS estimator. From this representation the asymptotic theory of the R estimator can be derived. Furthermore, it naturally suggests an IRLS algorithm. Let $e_{ij} = y_{ij}^* - g_{ij}\boldsymbol{\beta}$ denote the (i,j)th residual and let $m(\boldsymbol{\beta}) = med_{(i,j)}\{e_{ij}(\boldsymbol{\beta})\}$ denote the median of all the residuals. Then because the scores sum to 0, we have the identity,

$$\begin{aligned}
\mathbf{D}_R(\boldsymbol{\beta}) &= \sum\nolimits_{i=1}^{K}\sum\nolimits_{j=1}^{n_i}\psi[r(e_{ij}(\boldsymbol{\beta}))/(n+1)][e_{ij}(\boldsymbol{\beta}) - m(\boldsymbol{\beta})] \\
&= \sum\nolimits_{i=1}^{K}\sum\nolimits_{j=1}^{n_i}\frac{\psi[r(e_{ij}(\boldsymbol{\beta}))/(n+1)]}{e_{ij}(\boldsymbol{\beta}) - m(\boldsymbol{\beta})}[e_{ij}(\boldsymbol{\beta}) - m(\boldsymbol{\beta})] \\
&= \sum\nolimits_{i=1}^{K}\sum\nolimits_{j=1}^{n_i} w_{ij}(\boldsymbol{\beta})[e_{ij}(\boldsymbol{\beta}) - m(\boldsymbol{\beta})]^2
\end{aligned}$$

where $w_{ij}(\boldsymbol{\beta}) = \psi[r(e_{ij}(\boldsymbol{\beta}))/(n+1)]/[e_{ij}(\boldsymbol{\beta}) - m(\boldsymbol{\beta})]$ is a weight function. As usual, we take $w_{ij}(\boldsymbol{\beta}) = 0$ if $e_{ij}(\boldsymbol{\beta}) - m(\boldsymbol{\beta}) = 0$. Note that the weights are nonnegative.

Now let $\widehat{\boldsymbol{\beta}}_R^{(0)}$ denote an initial estimator of $\boldsymbol{\beta}$. As estimates of the weights, we use $\widehat{w}_{ij}(\widehat{\boldsymbol{\beta}}_R^{(0)})$; i.e., the weight function evaluated at $\widehat{\boldsymbol{\beta}}^{(0)}$, we have the dispersion function

$$\begin{aligned}\mathbf{D}_R^*(\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}}_R^{(0)}) &= \sum\nolimits_{i=1}^{K}\sum\nolimits_{j=1}^{n_i}\widehat{w}_{ij}(\widehat{\boldsymbol{\beta}}_R^{(0)})[e_{ij}(\boldsymbol{\beta}) - m(\widehat{\boldsymbol{\beta}}_R^{(0)})]^2 \\ &= \sum\nolimits_{i=1}^{K}\sum\nolimits_{j=1}^{n_i}[\sqrt{\hat{w}_{ij}(\widehat{\boldsymbol{\beta}}_R^{(0)})}e_{ij}(\boldsymbol{\beta}) - \sqrt{\widehat{w}_{ij}(\widehat{\boldsymbol{\beta}}_R^{(0)})}m(\widehat{\boldsymbol{\beta}}_R^{(0)})]^2\end{aligned}$$

Let

$$\widehat{\boldsymbol{\beta}}_R^{(1)} = Argmin\mathbf{D} * (\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}}_R^{(0)})$$

This establishes a sequence of IRLS estimates, $\widehat{\boldsymbol{\beta}}_R^{(k)}\}, k = 1, 2, \ldots$

After some algebraic simplification, we can obtain the gradient

$$\nabla\mathbf{D}_R^*(\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}}_R^{(k)}) = -2\sum\nolimits_{i=1}^{K}\mathbf{D}_i^T\widehat{\mathbf{V}}_i^{-\frac{1}{2}}\widehat{\mathbf{W}}_i\widehat{\mathbf{V}}_i^{-\frac{1}{2}}[\mathbf{Y}_i - \mathbf{a}'(\boldsymbol{\theta}) - \mathbf{m}^*(\widehat{\boldsymbol{\beta}}_R^{(k)})] \tag{5.3.2}$$

where $\mathbf{m}^*(\widehat{\boldsymbol{\beta}}_R^{(k)}) = \widehat{\mathbf{V}}_i^{\frac{1}{2}}m(\widehat{\boldsymbol{\beta}}_R^{(k)})\mathbf{1}$,$\mathbf{1}$ denotes a $n_i \times 1$ vector all of whose elements are 1, and $\widehat{\mathbf{W}} = diag\{\widehat{w}_{i1}, \ldots, \widehat{w}_{in_i}\}$ is the diagonal matrix of weights for the ith subject. Hence $\widehat{\boldsymbol{\beta}}_R^{(k+1)}$ satisfies the general estimating equations (GEE) given by,

$$\sum\nolimits_{i=1}^{K}\mathbf{D}_i^T\widehat{\mathbf{V}}_i^{-\frac{1}{2}}\widehat{\mathbf{W}}_i\widehat{\mathbf{V}}_i^{-\frac{1}{2}}[\mathbf{Y}_i - \mathbf{a}^{'}(\boldsymbol{\theta}) - \mathbf{m}^*(\boldsymbol{\beta}_R^{(k)})] = 0 \tag{5.3.3}$$

Which we denoted by

$$\sum\nolimits_{i=1}^{K}\mathbf{Z}_i(\boldsymbol{\beta}, \boldsymbol{\alpha} * (\boldsymbol{\beta})) = 0$$

Theorem 5.2 *Under these assumptions,*

A.1 $\sqrt{K}|\hat{\phi}(\boldsymbol{\beta}) - \phi| = O_p(1)$*, as* $K \to \infty$*, when* $\boldsymbol{\beta}$ *is known.*

A.2 $\sqrt{K}|\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi) - \boldsymbol{\alpha}| = O_p(1)$ *when* $\boldsymbol{\beta}$ *and* ϕ *are known.*

A.3 $|\partial\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)/\partial\phi| \leq H(\mathbf{Y}, \boldsymbol{\beta})$ *which is* $O_p(1)$

A.4 (Lindeberg-Feller Conditions): For $i = 1, \ldots, K$, *let* $\mathbf{U}_i = \mathbf{V}_i^{-\frac{1}{2}}$ *and* $\mathbf{U}_N = [\mathbf{U}_1^T \mathbf{U}_2^T \ldots \mathbf{U}_K^T]^T$. *Denote the* (l, j)*th entry of* $\mathbf{U}_N$ *by* u_{lj},$l = 1, 2, \ldots, n; j = 1, 2, \ldots, n_i$. *Then*

$$\max_{1 \le l \le K} \frac{u_{lj}^2}{\sum_{m=1}^{n}} u_{mj}^2 \to 0, \textit{for all } j = 1, \ldots, n_i$$

and

$$\lim_{K \to \infty} \frac{1}{n} \mathbf{U}_N^T \mathbf{U}_N \textit{ exists and is positive definite.}$$

A.5 The score function $\varphi(u)$ *is bounded and satisfies the standardizing conditions.*

A.6 The marginal pdf of $e_{ij}^{\dagger} = y_{ij} - g_{ij}(\boldsymbol{\beta})$ *is continuous and variance-covariance matrix given in the following is positive definite.*

Assume that the initial estimate satisfies $\sqrt{K}(\widehat{\boldsymbol{\beta}}_R^0 - \boldsymbol{\beta}) = O_p(1)$. *Then under the above assumptions, for* $k \ge 1$, $\sqrt{K}(\widehat{\boldsymbol{\beta}}_R^{(k)} - \boldsymbol{\beta})$ *has an asymptotic normal distribution with mean* 0 *and covariance matrix,*

$$\lim_{K \to \infty} K\{\sum\nolimits_{i=1}^{K} \mathbf{D}_i^T \mathbf{V}_i^{-\frac{1}{2}} \mathbf{W}_i \mathbf{V}_i^{-\frac{1}{2}} \mathbf{D}_i\}^{-1} \{\sum\nolimits_{i=1}^{K} \mathbf{D}_i^T \mathbf{V}_i^{-\frac{1}{2}} Var(\varphi_i^{\dagger}) \mathbf{V}_i^{-\frac{1}{2}} \mathbf{D}_i\} \{\sum\nolimits_{i=1}^{K} \mathbf{D}_i^T \mathbf{V}_i^{-\frac{1}{2}} \mathbf{W}_i \mathbf{V}_i^{-\frac{1}{2}} \mathbf{D}_i\}$$

where φ_i *denotes the* $n_i \times 1$ *vector* $(\varphi[r(e_{i1}^{\dagger}/(n+1))], \ldots, \varphi[r(e_{in_i}^{\dagger})/(n+1)])^T$ *and where*

$$\mathbf{Y}_i^{\dagger} = \mathbf{V}_i^{-1/2} \mathbf{Y}_i = (y_{i1}^{\dagger}, \ldots, y_{ik}^{\dagger})^T$$

$$\mathbf{G}_i^{\dagger}(\boldsymbol{\beta}) = V_i^{-1/2} \mathbf{a}_i'(\boldsymbol{\theta}) = [g_{ij}^{\dagger}]$$

$$e_{ij}^{\dagger} = y_{ij}^{\dagger} - g_{ij}^{\dagger}(\boldsymbol{\beta})$$

5.4 Our Case for the Ordinary LS based GEE Model

In our case of studying two sample predation preference, we are considering a longitudinal set of observations over $n_1 + n_2$ subjects, Using the notation in Section 5.2, $K = n_1 + n_2$

and $n_i = k$. let u_{ij}denote the jth response for ith subject for $j = 1, 2 \ldots, k$ and $i = 1, 2, \ldots, K$ then we can write the model as:

$$u_{ij} = \alpha^* + x_{ij}^T \boldsymbol{\beta}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)^T$, α^* is a known intercept. If α^* is unknown here, use the median of x_{ij}, where $n_1 < i \leq K$. Here $x_{ij} = I(i \leq n_1)I_j$, where I_j is a $k \times 1$ vector with the jth item equal to 1, and all the other items equal to 0. If we denote $u_{ij} - \alpha$ by y_{ij} then our model becomes:

$$y_{ij} = x_{ij}^T \boldsymbol{\beta} \tag{5.4.1}$$

Let $N = (n_1 + n_2)k$ denote the total sample size. Assume that the marginal distribution of y_{ij} is of the exponential class of distributions and is given by (5.2.1). In our case, $\theta_{ij} = \eta_{ij} = x_{ij}^T \boldsymbol{\beta}$. the link function $h^{-1} \circ (a^{'})^{-1} = I$ and the mean and variance are given by

$$E(y_{ij}) = a^{'}(\theta_{ij}) = x_{ij}^T \boldsymbol{\beta} \ \ and \ \ Var(y_{ij}) = a^{''}(\theta_{ij})/\phi = \frac{1}{\phi}$$

$$\mathbf{A}_i = diag\{a''(\theta_{i1}), \ldots, a''(\theta_{ik})\} = I$$

$$\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}/\phi = \mathbf{R}_i(\boldsymbol{\alpha})/\phi$$

$\mathbf{a}^{'}(\theta_i) = \mathbf{X}_i^T \boldsymbol{\beta}$, which means

$$\mathbf{D}_i = \begin{cases} I, & \text{if } i \leq n_1 \ , \\ 0, & \text{if } n_1 + 1 \leq i \leq K \ . \end{cases}$$

So the GEE estimator $\widehat{\boldsymbol{\beta}}_{LS}$ actually is the solution to the equations

$$\sum\nolimits_{i=1}^{n_1} \widehat{\mathbf{V}}_i^{-1} [\mathbf{Y}_i - \mathbf{X}_i^T \boldsymbol{\beta}] = 0. \tag{5.4.2}$$

which we denote by

$$\sum_{i=1}^{n_1} \mathbf{U}_i[\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\}] = 0$$

and the gradient of $\mathbf{D}_{LS}(\boldsymbol{\beta})$ is

$$\nabla \mathbf{D}_{Ls}(\boldsymbol{\beta}) = -\sum_{i=1}^{n_1} \widehat{\mathbf{V}}_i^{-1}[Y_i - \mathbf{X}_i^T\boldsymbol{\beta}]$$

We have the similar theorem as Theorem 5.1. The proof is given in the appendix.

Theorem 5.3 *Under mild regularity conditions and given that:*

A.1 $\widehat{\boldsymbol{\alpha}}$ *is* $K^{\frac{1}{2}}$*-consistent given* $\boldsymbol{\beta}$ *and* ϕ*.*

A.2 $\hat{\phi}$ *is* $K^{\frac{1}{2}}$*-consistent given* $\boldsymbol{\beta}$*; and*

A.3 $|\partial\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)/\partial\phi| \leq H(\mathbf{Y}, \boldsymbol{\beta})$ *which is* $O_p(1)$*.*

Then $K^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta})$ *is asymptotically multivariate Gaussian with zero mean and covariance matrix* $\mathbf{V}_{LS}$ *given by*

$$\mathbf{V}_{LS} = \lim_{n_1 \to \infty} n_1 (\sum_{i=1}^{n_1} \mathbf{V}_i^{-1})^{-1} \{\sum_{i=1}^{n_1} \mathbf{V}_i^{-1} cov(\mathbf{Y}_i) \mathbf{V}_i^{-1}\} (\sum_{i=1}^{n_1} \mathbf{V}_i^{-1})^{-1}$$

5.5 Our Case for the Iterated Reweighted Rank-Based Estimators for GEE Models

In our case, first we need an initial intercept $\hat{\alpha}^*$ which is given by $median(u_{ij} - x_{ij}^T\boldsymbol{\beta}_R^0)$, where $\boldsymbol{\beta}_R^0$ is the initial $\boldsymbol{\beta}_R$ satisfying $\sqrt{K}(\widehat{\boldsymbol{\beta}}_R^0 - \boldsymbol{\beta}) = O_p(1)$ and $n_1 < i \leq K$.
the gradient is given by

$$D_R^*(\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}}_{\mathbf{V}}^{(k)}) = -2\sum_{i=1}^{n_1} \widehat{\mathbf{V}}_i^{-\frac{1}{2}} \widehat{\mathbf{W}}_i \widehat{\mathbf{V}}_i^{-\frac{1}{2}} [Y_i - \mathbf{X}'\boldsymbol{\beta} - m^*(\widehat{\boldsymbol{\beta}}_R^{(k)})]$$

where $\mathbf{W}_i$ is the same as defined in Section 5.3, $\mathbf{Y}_i = \mathbf{U}_i^k - \boldsymbol{\alpha}^{*(k)}$ and $\mathbf{V}_i = \mathbf{R}_i(\boldsymbol{\alpha})/\phi$
The GEE for $\widehat{\boldsymbol{\beta}}_R^{(k+1)}$ is given by

$$\sum_{i=1}^{n_1} \widehat{\mathbf{V}}_i^{-\frac{1}{2}} \widehat{\mathbf{W}}_i \widehat{\mathbf{V}}_i^{-\frac{1}{2}} [\mathbf{Y}_i - \mathbf{a}'(\boldsymbol{\theta}) - m^*(\boldsymbol{\beta}_R^{(k)})] = 0 \tag{5.5.1}$$

which we denote by

$$\sum_{i=1}^{n_1} Z_i(\boldsymbol{\beta}, \alpha * (\boldsymbol{\beta})) = 0 \ .$$

The following theorem gives the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_R^{(k)}$. The proof is given in the appendix.

Theorem 5.4 *Under these assumptions,*

A.1 $\sqrt{K}|\widehat{\phi}(\boldsymbol{\beta}) - \phi| = O_p(1)$, *as* $K \to \infty$, *when* $\boldsymbol{\beta}$ *is known.*

A.2 $\sqrt{K}|\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi) - \boldsymbol{\alpha}| = O_p(1)$ *when* $\boldsymbol{\beta}$ *and* ϕ *are known.*

A.3 $|\partial\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)/\partial\phi| \leq H(\mathbf{Y}, \boldsymbol{\beta})$ *which is* $O_p(1)$

A.4 (Lindeberg-Feller Conditions): For $i = 1, \ldots, K$, *let* $\mathbf{U}_i = \mathbf{V}_i^{-\frac{1}{2}}$ *and* $\mathbf{U}_n = [\mathbf{U}_1^T \mathbf{U}_2^T \ldots \mathbf{U}_{n_1}^T]^T$. *Denote the* (l, j)*th entry of* $\mathbf{U}_n$ *by* u_{lj},$l = 1, 2, \ldots, K; j = 1, 2, \ldots, k$. *Then*

$$\max_{1 \leq l \leq K} \frac{u_{lj}^2}{\sum_{m=1}^K} u_{mj}^2 \to 0, \textit{for all } j = 1, \ldots, k$$

and

$$\lim_{K \to \infty} \frac{1}{K} \mathbf{U}_N^T \mathbf{U}_N \textit{ exists and is positive definite.}$$

A.5 The score function $\varphi(u)$ *is bounded and satisfies the standardizing conditions.*

A.6 The marginal pdf of $e_{ij}^\dagger = y_{ij} - g_{ij}(\boldsymbol{\beta})$ *is continuous and variance-covariance matrix given in the following is positive definite.*

Assume that the initial estimate satisfies $\sqrt{K}(\widehat{\boldsymbol{\beta}}_R^0 - \boldsymbol{\beta}) = O_p(1)$. *Then under the above assumptions, for* $k \geq 1$, $\sqrt{K}(\widehat{\boldsymbol{\beta}}_R^{(k)} - \boldsymbol{\beta})$ *has an asymptotic normal distribution with mean* 0 *and covariance matrix,*

$$\lim_{n_1 \to \infty} n_1 \{\sum_{i=1}^{n_1} \mathbf{V}_i^{-\frac{1}{2}} \mathbf{W}_i \mathbf{V}_i^{-\frac{1}{2}}\}^{-1} \{\sum_{i=1}^{n_1} \mathbf{V}_i^{-\frac{1}{2}} Var(\varphi_i^{\dagger}) \mathbf{V}_i^{-\frac{1}{2}}\} \{\sum_{i=1}^{n_1} \mathbf{V}_i^{-\frac{1}{2}} \mathbf{W}_i \mathbf{V}_i^{-\frac{1}{2}}\}$$

where φ_i *denotes the* $k \times 1$ *vector* $(\varphi[r(e_{i1}^{\dagger}/(n+1))], \ldots, \varphi[r(e_{ik}^{\dagger})/(n+1)])^T$ *and where*

$$\mathbf{Y}_i^{\dagger} = \mathbf{V}_i^{-1/2} \mathbf{Y}_i = (y_{i1}^{\dagger}, \ldots, y_{ik}^{\dagger})^T$$

$$G_i^{\dagger}(\boldsymbol{\beta}) = \mathbf{V}_i^{-1/2} \mathbf{a}_i'(\boldsymbol{\theta}) = [g_{ij}^{\dagger}]$$

$$e_{ij}^{\dagger} = y_{ij}^{\dagger} - g_{ij}^{\dagger}(\boldsymbol{\beta})$$

5.6 Example

Let us consider once again the study of [48] concerning the effect of a single session of high intensity aerobic exercise on inflammatory markers of subjects taken over time. Recall that 18 subjects were placed into two groups (High Fitness and Moderate Fitness) depending on their fitness levels, 9 in each group, and the response of interest here is the C-Reactive protein (CRP). Our effect of interest is the trend of difference in CRP beween the two groups (High Fitness - Moderate Fitness), we are intereted in testing will the difference stay the same over time or decay to zero over time. In particular, we are interested in testing the following hypothesis:

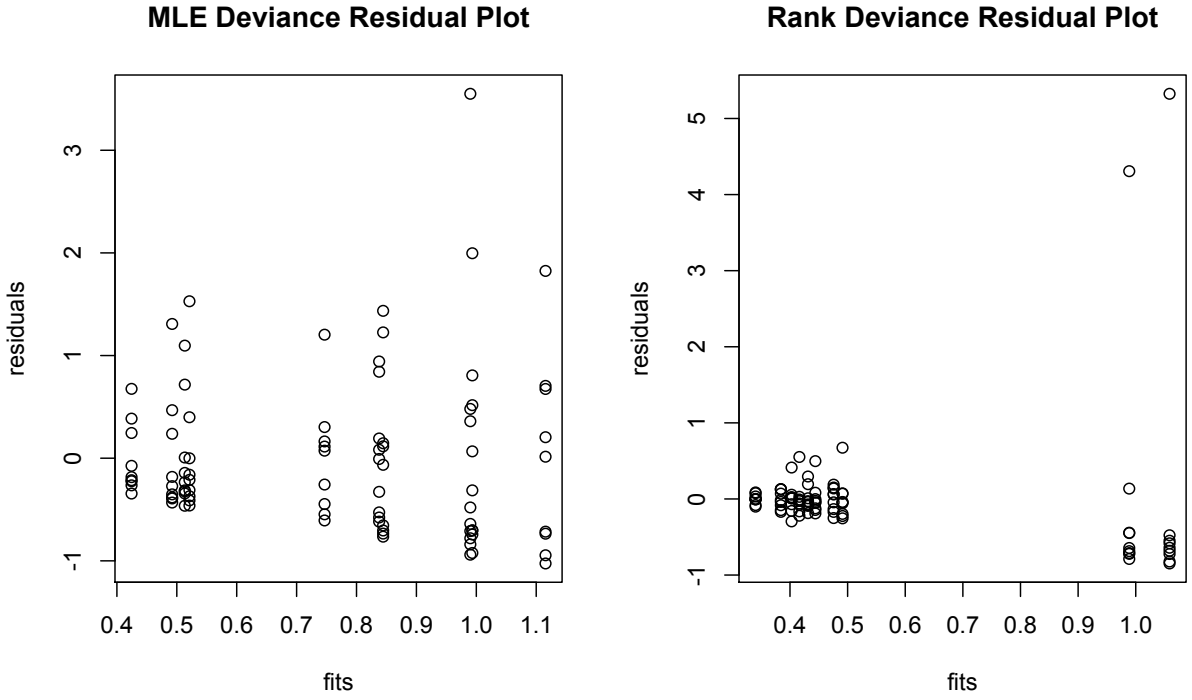
- H_0 : The difference between two groups remains the same.
- H_a : The difference between two groups is not the same.

Using the ordinary LS based GEE model, we get $p - value = 0.0087$ and using the iterated reweighted rank-based estimators for GEE model, we get $p - value = 0.443$ and the estimate

of difference is $(0.594, 0.331, 0.003, 0.254, 0.413)$. We draw two different conclusions using these two $p-values$. We tend to believe in the $p-value$ of the iterated reweighted rank-based estimators for GEE model because from Figure 3.2, it is clear that from either the mean profile or the median profile, the difference between the two groups HI and LO is not zero over time.

For the data we used in [48], Figure 5.1 shows the residual values versus fitted response

Figure 5.1: Residul plots for The ordinary GEE and rank-based GEE



values for the two different GEE models: The LS based GEE model and the Rank based GEE model. It is clear that the LS fit is adversely affected by the outlying observations. We find that overall the residuals are closer to 0 for the Rank Deviance Residual plot and it is easier to identify the two outliers in this plot. This is not the case with the LS deviance plot.

5.7 Conclusion

Repeated measurement designs occur in many areas of statistical research. The analysis of data resulting from such designs often becomes complicated because of the non-zero within-subject correlation. To avoid this problem [10] proposed the Generalized Estimation Equations Model (GEE). However, the solution of GEEs proposed by igns often becomes complicated because of the non-zero within-subject correlation, To avoid this problem [10] is not robust because it is based on iterated reweighted least squares fitting. [2] proposed a rank-based fitting procedure which is robust to outliers in response space. In this chapter, we used these two GEE models to analyze the selective predation pattern problem, as a special case of the two sample repeated measurement data. We were able to simplify some assumptions and derive the asymptotic normality of our estimates.

Chapter 6

Summary

In Chapter 2, this dissertation gave nonparametric test statistics for detecting preference patterns in selective predation by extending the method of [1]. We provided a class of general rank score tests for the one sample case where we only have one species of predators and prey have two features of interest. This gives the flexibility to place varied emphasis on consecutive selections at different stages of the selection experiment. This could be helpful in controlling the manner in which extraneous variables can affect selection preference patterns.

This dissertation also proposes a class of general rank score tests for the difference in predation patterns of two predatory species. In this case, prey feature of interest can be continuous or categorical. It is shown that the test statistic for the categorical case becomes equivalent to the continuous case if the data have no ties. In both cases, the asymptotic distribution of the test statistic is Gaussian. The results of a simulation study using the asymptotic Gaussian distribution but small samples shows that the test has a satisfactory finite-sample performance. The null simulation shows that null rejection rates are close to nominal α values. We also demonstrate that the asymptotic test is powerful in detecting simple-ordered alternatives.

If we already have evidence to conclude that two species of predator tend to choose different prey, then one might be interested in comparing the change in speed and the change in direction of the two overall patterns. We proposed a method to compare the change in the trends of prey selection patterns of the two species. The trend in the opposite direction can be tested by multiplying the generalized normalized weight function by -1 in the test statistic.

For linear models, the least squares estimator of the regression coefficient is optimal if the error distribution is normal. For distributions with longer tails than the normal, however, rank estimators are more efficient than the least squares estimator. [34] proposed an IRLS method for estimating rank estimators. In Chapter 4, we studied the finite sample performance of the IRLS method. For Wilcoxon scores, using the LS estimate as an initial estimate, we found that the IRLS algorithm leads to estimates whose efficiency is between those of the LS and rank estimates obtained using the algorithm of [54]. The LS estimates are efficient when the distribution of the error is short-tailed. The algorithm of [54] gives efficient estimates when the distribution of the error is long-tailed while the IRLS algorithm gives efficient estimates when the distribution of the error is moderate-tailed.

It is also observed that there is a need for an IRLS formulation of bounded influence rank estimators [47]. The boundedness of the initial estimator is a necessary condition for the boundedness of the influence function of IRLS estimator, but it is not sufficient.

In Chapter 5, we used the IRLS representation of rank estimation to propose a rank analogue of the GEE model proposed by Liang and Zegar (1986). We used these two GEE methods to study the changing trend problem studied in Chapter 3. The results show that the rank based GEE method gives superior performance than the GEE method of [10] for small datasets with outliers.

Bibliography

[1] Marinela Capanu, Gregory A. Jones, and Ronald H. Randles. Testing for preference using a sum of Wilcoxon signed rank statistics. *Computational Statistics & Data Analysis*, 51(2):793–796, NOV 15 2006.

[2] A. Abebe, J. W. McKean, and J. D. Kloke. Iterated reweighted rank-based estimates for gee models. *Submitted*, 2010.

[3] Michael G. Kenward. A method for comparing profiles of repeated measurements. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36(3):pp. 296–308, 1987.

[4] Dj Finney. Repeated Measurements-What is Measured and What Repeats. *Statistics in Medicine*, 9(6):639–644, JUN 1990.

[5] Jns Matthews, D Altman, Mj Campbell, and P Royston. Analysis of Serial Measurements in Medical-Research - Reply. *British Medical Journal*, 300(6725):680, MAR 10 1990.

[6] L Frison and Sj Pocock. Repeated Measures In Clinical-Trials - Analysis Using Mean Summary Statistics And Its Implications For Design. *Ststistics In Medicine*, 11(13):1685–1704, SEP 30 1992.

[7] T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley Publications in Statistics. John Wiley & Sons Inc., New York, 1958.

[8] Peter C. O'Brien. Procedures for comparing samples with multiple endpoints. *Biometrics*, 40(4):1079–1087, 1984.

[9] Roderick J. A. Little and Donald B. Rubin. *Statistical analysis with missing data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1987.

[10] Kung Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[11] MC Wu and KR Bailey. Estimation and Comparison of Changes in the Presence of Informative Right Censoring-Conditional Linear-Model. *Biometrics*, 45(3):939–955, SEP 1989.

[12] L. J. Wei and Wayne E. Johnson. Combining dependent tests with incomplete repeated measurements. *Biometrika*, 72(2):pp. 359–364, 1985.

[13] L. J. Wei and J. M. Lachin. Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *Journal of the American Statistical Association*, 79(387):pp. 653–661, 1984.

[14] D. V. Glidden and L. J. Wei. Rank estimation of treatment differences based on repeated measurements subject to dependent censoring. *J. Amer. Statist. Assoc.*, 94(447):888–895, 1999.

[15] A. J. Lawrance and N. T. Kottegoda. Stochastic modelling of riverflow time series. *Journal of the Royal Statistical Society. Series A (General)*, 140(1):pp. 1–47, 1977.

[16] HE Hurst. Long-term Storage Capacity of Reservoirs. *Transactions of The American Society of Civil Engineers*, 116:770–799, 1951.

[17] J. B. Carlin and A. P. Dempster. Sensitivity analysis of seasonal adjustments: Empirical case studies. *Journal of the American Statistical Association*, 84(405):pp. 6–20, 1989.

[18] S Porter-Hudak. An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates. *Journal of the American Statistical Association*, 85(410):338–344, JUN 1990.

[19] Benoit Mandelbrot. Long-run linearity, locally gaussian process, h-spectra and infinite variances. *International Economic Review*, 10(1):pp. 82–111, 1969.

[20] John Haslett and Adrian E. Raftery. Space-time modelling with long-memory dependence: Assessing ireland's wind power resource. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 38(1):pp. 1–50, 1989.

[21] Jan Beran. A goodness-of-fit test for time series with long range dependence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(3):pp. 749–760, 1992.

[22] P. Whittle. On the variation of yield variance with plot size. *Biometrika*, 43(3/4):pp. 337–343, 1956.

[23] P. Whittle. Topographic correlation, power-law covariance functions, and diffusion. *Biometrika*, 49(3/4):pp. 305–314, 1962.

[24] Ellis Batten Page. Ordered hypotheses for multiple treatments: a significance test for linear ranks. *J. Amer. Statist. Assoc.*, 58:216–230, 1963.

[25] NM Laird and JH Ware. Random-Effects Models For Longitudinal Data. *Biometrics*, 38(4):963–974, 1982.

[26] JH Ware. Linear-Models For The Analysis Of Longitudinal-Studies. *American Statistician*, 39(2):95–101, 1985.

[27] Y Ochi and RL Prentice. Likelihood Inference In A Correlated Probit Regression-Model. *Biometrika*, 71(3):531–543, 1984.

[28] R Stiratelli, N Laird, and JH Ware. Random-Effects Models For Serial Observations With Binary Response. *Biometrics*, 40(4):961–971, 1984.

[29] Gary G. Koch, J. Richard Landis, Jean L. Freeman, Jr. Freeman, Daniel H., and Robert G. Lehnen. A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, 33(1):pp. 133–158, 1977.

[30] Bahjat F. Qaqish and John S. Preisser. Resistant fits for regression with correlated outcomes: an estimating equations approach. *J. Statist. Plann. Inference*, 75(2):415–431, 1999. The Seventh Eugene Lukacs Conference (Bowling Green, OH, 1997).

[31] Sin-Ho Jung and Zhiliang Ying. Rank-based regression with repeated measurements data. *Biometrika*, 90(3):732–740, 2003.

[32] You-Gan Wang and Min Zhu. Rank-based regression for analysis of repeated measures. *Biometrika*, 93(2):459–464, 2006.

[33] Louis A. Jaeckel. Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.*, 43:1449–1458, 1972.

[34] Gerald L. Sievers and Asheber Abebe. Rank estimation of regression coefficients using iterated reweighted least squares. *J. Stat. Comput. Simul.*, 74(11):821–831, 2004.

[35] T. P. Hettmansperger and J. W. McKean. *Robust nonparametric statistical methods, 2nd ed.* Monographs on Statistics & Applied Probability. CRC Press, Boca Raton, 2011.

[36] M Zuk and GR Kolluru. Exploitation of sexual signals by predators and parasitoids. *Quarterly Review of Biology*, 73(4):415–438, DEC 1998.

[37] Nelika K. Hughes, Jennifer L. Kelley, and Peter B. Banks. Receiving behaviour is sensitive to risks from eavesdropping predators. *Oecologia*, 160(3):609–617, JUN 2009.

[38] J Chesson. The Estimation and Analysis of Preference and its Relationship to Foraging Models. *Ecology*, 64(5):1297–1304, 1983.

[39] B. F. J. Manly. A note on a model for selection experiments. *Biometrics*, 36(1):pp. 9–18, 1980.

[40] GA Jones, KE Sieving, ML Avery, and RL Meagher. Parasitized and non-parasitized prey selectivity by an insectivorous bird. *CROP PROTECTION*, 24(2):185–189, FEB 2005.

[41] Myles Hollander and Douglas A. Wolfe. *Nonparametric statistical methods*. Wiley Series in Probability and Statistics: Texts and References Section. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[42] J. H. Klotz. The Wilcoxon, ties, and the computer. *J. Amer. Statist. Assoc.*, 61:772–787, 1966.

[43] Gregory A. Mack and Douglas A. Wolfe. k-sample rank tests for umbrella alternatives. *J. Amer. Statist. Assoc.*, 76(373):175–181, 1981.

[44] T. P. Hettmansperger and J. W. McKean. *Robust nonparametric statistical methods*, volume 5 of *Kendall's Library of Statistics*. Edward Arnold, London, 1998.

[45] Joseph W. McKean, Thomas J. Vidmar, and Gerald L. Sievers. A robust two-stage multiple comparison procedure with application to a random drug screen. *Biometrics*, 45(4):pp. 1281–1297, 1989.

[46] Rudolf Beran. Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.*, 2:63–74, 1974.

[47] Joshua D. Naranjo and Joseph W. McKean. Rank regression with estimated scores. *Statist. Probab. Lett.*, 33(2):209–216, 1997.

[48] Eric P. Plaisance, J. Kyle Taylor, Sofiya Alhassan, Asheber Abebe, Michael L. Mestek, and Peter W. Grandjean. Cardiovascular fitness and vascular inflammatory markers after acute aerobic exercise. *International Journal of Sport Nutrition and Exercise Metabolism*, 17(2):152–162, APR 2007.

[49] Joseph W. McKean and Ronald M. Schrader. The geometry of robust procedures in linear models. *J. Roy. Statist. Soc. Ser. B*, 42(3):366–371, 1980.

[50] E. J. Schlossmacher. An iterative technique for absolute deviations curve fitting. *Journal of the American Statistical Association*, 68(344):pp. 857–859, 1973.

[51] Peter J. Huber. *Robust statistics*. John Wiley & Sons Inc., New York, 1981. Wiley Series in Probability and Mathematical Statistics.

[52] Kuo Sheuan Cheng and Thomas P. Hettmansperger. Weighted least-squares rank estimates. *Comm. Statist. A—Theory Methods*, 12(9):1069–1086, 1983.

[53] Joseph W. McKean and Thomas P. Hettmansperger. A robust analysis of the general linear model based on one step R-estimates. *Biometrika*, 65(3):571–579, 1978.

[54] JT Terpstra and JW McKean. Rank-based analyses of linear models using R. *Journal Of Statistical Software*, 14(7), JUL 2005.

[55] Roger Koenker and Gilbert Bassett, Jr. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

[56] Hira L. Koul, Gerald L. Sievers, and Joseph McKean. An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scand. J. Statist.*, 14(2):131–141, 1987.

[57] Frank R. Hampel. The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.*, 69:383–393, 1974.

[58] Michael B. Dollinger and Robert G. Staudte. Influence functions of iteratively reweighted least squares estimators. *J. Amer. Statist. Assoc.*, 86(415):709–716, 1991.

[59] William H. Chang, Joseph W. McKean, Joshua D. Naranjo, and Simon J. Sheather. High-breakdown rank regression. *J. Amer. Statist. Assoc.*, 94(445):205–219, 1999.

[60] Gerald L. Sievers. A weighted dispersion function for estimation in linear models. *Comm. Statist. A—Theory Methods*, 12(10):1161–1179, 1983.

[61] Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. *Robust statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. The approach based on influence functions.

[62] Antonio Cuevas and Juan Romo. On the estimation of the influence curve. *Canad. J. Statist.*, 23(1):1–9, 1995.

[63] Víctor J. Yohai. High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.*, 15(2):642–656, 1987.

[64] Thomas P. Hettmansperger and Joseph W. McKean. A robust alternative based on ranks to least squares in analyzing linear models. *Technometrics*, 19(3):275–284, 1977.

[65] Kuo-Chin Lin, Yi-Ju Chen, and Yu Shyr. A nonparametric smoothing method for assessing GEE models with longitudinal binary data. *Stat. Med.*, 27(22):4428–4439, 2008.

[66] Sin-Ho Jung and Chul W. Ahn. Sample size for a two-group comparison of repeated binary measurements using GEE. *Stat. Med.*, 24(17):2583–2596, 2005.

[67] Zhongwen Tang. *LOF of logistic GEE models and cost efficient Bayesian optimal designs for nonlinear combinations of parameters in nonlinear regression models*. ProQuest LLC, Ann Arbor, MI, 2008. Thesis (Ph.D.)–Kansas State University.

[68] Junxiang Luo. *Goodness-of-fit tests for proportional odds model with GEE for ordinal categorical responses and estimating sampling frequency in pollen exposure assessment over time*. ProQuest LLC, Ann Arbor, MI, 2006. Thesis (Ph.D.)–University of Cincinnati.

[69] Hui-Yi Lin and Leann Myers. Power and type I error rates of goodness-of-fit statistics for binomial generalized estimating equations (GEE) models. *Comput. Statist. Data Anal.*, 50(12):3432–3448, 2006.

[70] Hui-Yi Lin. *A comparison of goodness-of-fit tests for binomial generalized estimating equations (GEE) models*. ProQuest LLC, Ann Arbor, MI, 2004. Thesis (Ph.D.)–Tulane University.

[71] Scott Evans and Lingling Li. A comparison of goodness of fit tests for the logistic GEE model. *Stat. Med.*, 24(8):1245–1261, 2005.

[72] Ji-Hyun Lee and Bahjat F. Qaqish. Modified GEE and goodness of the marginal fit (GOMF) test with correlated bindary responses for contingency tables. *Biom. J.*, 46(6):675–686, 2004.

[73] Mohand-Larbi Feddag, Ion Grama, and Mounir Mesbah. Generalized estimating equations (GEE) for mixed logistic models. *Comm. Statist. Theory Methods*, 32(4):851–874, 2003.

[74] Chi-tsung Wu, Marcia L. Gumpertz, and Dennis D. Boos. Comparison of GEE, MINQUE, ML, and REML estimating equations for normally distributed data. *Amer. Statist.*, 55(2):125–130, 2001.

[75] A. Abebe and J. W. McKean. Highly efficient nonlinear regression based on the wilcoxon norm. *Festschrift in Honor of Mir Masoom Ali*, pages 340–357, 2007.

Appendix: Proofs

Proof of Theorem 3.1

Under H_0, $\Delta_1 = \cdots = \Delta_k = \Delta$. So, the expectations of all row ranks are equal and equal to $(k+1)/2$. Then $E_0(\tilde{\mathbf{R}}) = (N^*(k+1)/2)\mathbf{J}_{k\times 1}$. So, $E_0(\mathbf{W}_\Psi) = \Psi' E_0(\tilde{\mathbf{R}}) = (N^*(k+1)/2)\Psi'\mathbf{J}_{k\times 1} = 0$. Now $var_0(\mathbf{W}_\Psi) = \Psi' cov_0(\tilde{\mathbf{R}})\Psi$. But the (t,s)th element of $cov_0(\tilde{\mathbf{R}})$ is

$$v_{ts} = \sum_{i=1}^{N^*}\sum_{j=1}^{N^*} cov_0(R_{it}, R_{js})$$

When $t = s$, we have

$$\begin{aligned} v_{ts} &= N^* var(R_{it}) + N^*(N-2)\{\frac{1}{12}(k-1) + (C - \frac{1}{4})(k-1)(k-2)\} \\ &= N^*\frac{k^2-1}{12} + N^*(N-2)\{\frac{1}{12}(k-1) + (C - \frac{1}{4})(k-1)(k-2)\} \end{aligned}$$

When$t \neq s$, we have

$$\begin{aligned} v_{ts} &= N^* cov(R_{1t}, R_{1s}) + N^*(N-2)\{-\frac{1}{12} + 2(k-2)(A - \frac{1}{4}) + (B - \frac{1}{4})(k-2)\} \\ &= N^*(-\frac{k+1}{12}) + N^*(N-2)\{-\frac{1}{12} + 2(k-2)(A - \frac{1}{4}) + (B - \frac{1}{4})(k-2)\} \end{aligned}$$

where

$$\begin{aligned} A &= \int P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} > t)1 - P_0(U_{11} - U_{11} + U_{(n_1+1)2} > t)dF(t);, \\ B &= \int P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} > t)^2 dF(t)\ , \text{ and} \\ C &= \int \{P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} < t)\}^2 dF(t)\ . \end{aligned}$$

Denote $\sum_{s=1}^{k} \psi_s(\frac{s}{k+1})R_{is} = Q_i$, where $\psi_s = \psi(\frac{s}{k+1})$, we have:

$$\begin{aligned} Var(\mathbf{W}_\psi) &= \sum_{i=1}^{n_1 n_2} var(Q_i) + \sum_{i=1}^{n_1 n_2}\sum_{l=1}^{n_1 n_2} Cov[Q_i, Q_l] \\ &= n_1 n_2 var(Q_1) + n_1 n_2 (n_1 + n_2 - 2) Cov[Q_1, Q_2] \end{aligned}$$

We obtain

$$\begin{aligned} Cov(Q_{1,}Q_2) &= Cov(\sum_{s=1}^{k} \psi_s R_{1s}, \sum_{s=1}^{k} \psi_s R_{2s},) \\ &= \sum_{u=1}^{k}\sum_{v=1}^{k} \psi_u \psi_v Cov(R_{1u}, R_{2v}) \\ &= \sum_{u=1}^{k}\sum_{v=1}^{k} \psi_u \psi_v Cov(\sum_{p=1,p\neq u}^{k} I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u)} + 1, \\ &\quad \sum_{q=1,q\neq v}^{k} I(U_{1q} - U_{(n_1+2)q} < U_{1v} - U_{(n_1+1)v}) + 1) \\ &= \sum_{u=1}^{k}\sum_{v=1}^{k} \psi_u \psi_v \sum_{p=1,p\neq u}^{k}\sum_{q=1,q\neq v}^{k} E[I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u}) \\ &\quad I(U_{1q} - U_{(n_1+2)q} < U_{1v} - U_{(n_1+2)v})] \\ &\quad - \sum_{u=1}^{k}\sum_{v=1}^{k} \psi_u \psi_v \sum_{p=1,p\neq u}^{k}\sum_{q=1,q\neq v}^{k} E^2(I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u})) \end{aligned}$$

The calculation of the first term in the above equation is as the following four cases:

Case 1. $p \neq q, u \neq v$,

(1) $p \neq v, q \neq u$,

$$E^2(I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u})) = \frac{1}{4}$$

(2) $p = v, q = u$

$$\begin{aligned} & E[I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u})I(U_{1u} - U_{(n_1+2)u} < U_{1p} - U_{(n_1+2)p}] \\ = & \int\int P_0(U_{(n_1+1)u} - U_{(n_1+1)p} < t_1 - t_2)(1 - P(U_{(n_1+1)u} - U_{(n_1+1)p} < t_1 - t_2))dFt_1 dFt_2 \\ = & \frac{1}{6} \end{aligned}$$

(3) $p = v, q \neq u$

$$\begin{aligned} & E[I(U_{1p} - U_{(n_1+1)p} < X_{1u} - U_{(n_1+1)u})I(U_{1u} - _{(n_1+2)u} < U_{1p} - U_{(n_1+2)p}] \\ = & \int P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} > t)1 - P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} > t)dF(t) \\ = & A \end{aligned}$$

(4) $p \neq v, q = u$

$$\begin{aligned} & E[I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u})I(U_{1u} - U_{(n_1+2)u} < U_{1p} - U_{(n_1+2)p}] \\ = & \int P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} > t)1 - P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} > t)dF(t) \\ = & A \end{aligned}$$

Thus

$$\begin{aligned} & \sum_{u=1}^{k} \sum_{v=1,u\neq v}^{k} \psi_u \psi_v \sum_{p=1,p\neq u}^{k} \sum_{q=1,q\neq v}^{k} E[I(X_{1p} - Y_{1p} < X_{1u} - Y_{1u})I(X_{1q} - Y_{2q} < X_{1v} - Y_{2v})] \\ = & -[(2k-4)A + (k-2)(k-3)\frac{1}{4} + \frac{1}{6}] \end{aligned}$$

Case 2. $p = q, u = v$

$$\begin{aligned} & E[I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u})I(U_{1q} - U_{(n_1+2)q} < U_{1v} - U_{(n_1+2)v})] \\ = & \int\int P_0^2(U_{(n_1+1)u} - U_{(n_1+1)p} < t_1 - t_2)dFt_1dFt_2 \\ = & \tfrac{1}{3} \end{aligned}$$

Thus

$$\begin{aligned} & \sum_{u=1}^{k}\sum_{v=1}^{k}\psi_u\psi_u \sum_{p=1,p\neq u}^{k}\sum_{p=1,p\neq v}^{k} E[I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u}) \\ & I(U_{1q} - U_{(n_1+2)q} < U_{1v} - U_{(n_1+2)v})] \\ = & \frac{1}{3}(k-1) \end{aligned}$$

Case 3. $p = q, u \neq v$

$$\begin{aligned} & E[I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u})I(U_{1q} - U_{(n_1+2)q} < U_{1v} - U_{(n_1+2)v})] \\ = & \int P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} > t)^2 dF(t) \\ = & B \end{aligned}$$

Thus

$$\begin{aligned} & \sum_{u=1}^{k}\sum_{v=1}^{k}\psi_u\psi_v \sum_{p=1,p\neq u}^{k}\sum_{q=1,q\neq v}^{k} E[I(X_{1p} - Y_{1p} < X_{1u} - Y_{1u})I(X_{1q} - Y_{2q} < X_{1v} - Y_{2v})] \\ = & -(k-2)B \end{aligned}$$

Case 4. $p \neq q, u = v$

$$\begin{aligned} & E[I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u}) I(U_{1q} - U_{(n_1+2)q} < U_{1v} - U_{(n_1+2)v})] \\ = & \int \{P_0(U_{11} - U_{(n_1+1)1} + U_{(n_1+1)2} < t)\}^2 dF(t) \\ = & C \end{aligned}$$

Thus

$$\begin{aligned} & \sum_{u=1}^{k} \sum_{v=1}^{k} \psi_u \psi_u \sum_{p=1, p\neq u}^{k} \sum_{q=1, q\neq v}^{k} E[I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u}) \\ & I(U_{1q} - U_{(n_1+2)q} < U_{1v} - Y_{(n_1+2)v})] \\ = & (k-1)(k-2)C \end{aligned}$$

The second term

$$\begin{aligned} & \sum_{u=1}^{k} \sum_{v=1}^{k} \psi_u \psi_v \sum_{p=1, p\neq u}^{k} \sum_{q=1, q\neq v}^{k} E^2(I(U_{1p} - U_{(n_1+1)p} < U_{1u} - U_{(n_1+1)u})) \\ = & \sum_{u=1}^{k} \sum_{v=1}^{k} \psi_u \psi_v \sum_{p=1, p\neq u}^{k} \sum_{q=1, q\neq v}^{k} \frac{1}{4}(k-1)^2 \end{aligned}$$

Plug the result of $Cov(Q_1, Q_2)$ in $Var(\mathbf{W}_\psi)$, we proved the first part in Theorem 3.1, then using the Central Limit Theorem, we proved the second part in Theorem 3.1.

Proof of Theorem 5.3

Write $\alpha^*(\boldsymbol{\beta}) = \hat{\alpha}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\}$. We first expand $\sum_{i=1}^{K} \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta}))$ in a Taylor series about the true parameter $\boldsymbol{\beta}$ and evaluated at $\widehat{\boldsymbol{\beta}}_{LS}$. By the chain rule, the gradient in this expansion

is give by

$$\begin{aligned}\nabla_i &= \frac{\delta \mathbf{U}_i\{\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta})\}}{\delta \boldsymbol{\beta}} \\ &= \frac{\partial \mathbf{U}_i\{\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta})\}}{\partial \boldsymbol{\beta}} + \frac{\partial U_i\{\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta})\}}{\partial \boldsymbol{\alpha}^*} \frac{\partial \boldsymbol{\alpha}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ &= \mathbf{A}_i + \mathbf{B}_i \mathbf{C}\end{aligned}$$

Because $\widehat{\boldsymbol{\beta}_{LS}}$ solves equation $\sum_{i=1}^{K} \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta})) = 0$, we have

$$0 = \sum_{i=1}^{K} \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta})) + \sum_{i=1}^{K} \nabla_i(\widehat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta})$$

Solving for $\sqrt{K}(\widehat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta})$, we obtain

$$\sqrt{K}(\widehat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta}) = [-\frac{1}{K}\sum_{i=1}^{K} \nabla_i]^{-1}[\frac{1}{\sqrt{K}}\sum_{i=1}^{K} U_i\{\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}^*(\boldsymbol{\beta})}\}]$$

Secondly, we fix $\boldsymbol{\beta}$ and expand $\frac{1}{\sqrt{K}}\sum_{i=1}^{K} \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta}))$ about the true parameter $\boldsymbol{\alpha}$ and evaluated at $\boldsymbol{\alpha}^*$ to get

$$\begin{aligned}\frac{1}{\sqrt{K}}\sum_{i=1}^{K} \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta})) &= \frac{1}{\sqrt{K}}\sum_{i=1}^{K} U_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) \\ &\quad + \frac{1}{K}\sum_{i=1}^{n} \frac{\partial U_i(\boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \sqrt{K}(\boldsymbol{\alpha}^* - \boldsymbol{\alpha}) + o_p(1) \\ &= \mathbf{A}^* + \mathbf{B}^*\mathbf{C}^* + o_p(1)\end{aligned}$$

Now, $B^* = o_p(1)$, Using Law of Large Number and $\partial \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})/\partial \boldsymbol{\alpha}$ are linear functions of $Y_i - \mathbf{X}_i^T \boldsymbol{\beta}$, whose means are zeros, and condition (A.1) to (A.3) give

$$\begin{aligned} C^* &= \sqrt{n_1}(\boldsymbol{\alpha}^* - \boldsymbol{\alpha}) \\ &= \sqrt{K}[\hat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\} - \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi) + \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi) - \boldsymbol{\alpha}] \\ &= \sqrt{K}[\frac{\partial \hat{\boldsymbol{\alpha}}}{\partial \phi}(\boldsymbol{\beta}, \phi)(\hat{\phi} - \phi) + \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi) - \boldsymbol{\alpha}] \\ &= O_p(1) \end{aligned}$$

Consequently, $\frac{1}{\sqrt{n_1}} \sum_{i=1}^{K} \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta}))$ is asymptotically equivalent to $\mathbf{A}^*$ whose asymptotic distribution is multivariate Gaussian with zero mean and covariance matrix

$$\lim_{K \to \infty} \{\sum_{i=1}^{K} \mathbf{V}^{-1} cov\{Y_i\} \mathbf{V}^{-1}/K\}$$

Finally, $\sum_{i=1}^{K} = o_p(K)$ because $\partial \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*)/\partial \boldsymbol{\alpha}^*$ are also linear functions of $Y_i - \mathbf{X}_i^T \boldsymbol{\beta}$'s whose means are zero. because

$$\frac{\partial \boldsymbol{\alpha}^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{\alpha}^*(\boldsymbol{\beta})}{\partial \phi} \frac{\partial \phi}{\partial \boldsymbol{\beta}}$$

and then using condition (A.3), we can get $\mathbf{C} = O_p(1)$, and that $\sum_{i=1}^{K} \frac{\mathbf{A}_i}{K}$ converges as $K \to \infty$ to $-\sum_{i=1}^{K} \mathbf{V}^{-1}/K$. This completes the proof.

Proof of Theorem 5.4

This proof is quite similar to the previous one.

Let $\boldsymbol{\alpha}^*(\boldsymbol{\beta}) = \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta}))$. Let $k \geq 1$ be arbitrary but fixed. For $i = 1, \ldots, K$, let

$$\begin{aligned} Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta})) &= \widehat{\mathbf{V}}_i^{-\frac{1}{2}} \mathbf{W}_i \widehat{\mathbf{V}}_i^{-\frac{1}{2}} [Y_i - \mathbf{X}_i^T \boldsymbol{\beta} - m^*(\boldsymbol{\beta})] \\ &= \widehat{\mathbf{V}}_i^{-\frac{1}{2}} \mathbf{W}_i \widehat{\mathbf{V}}_i^{\frac{1}{2}} [Y_i - \mathbf{X}_i^T \boldsymbol{\beta} - m^*(\boldsymbol{\beta})] \\ &= \widehat{\mathbf{V}}_i^{-\frac{1}{2}} \mathbf{W}_i [Y_i^* - G_i^*(\boldsymbol{\beta}) - ml] \end{aligned}$$

We first expand $K^{-1/2}\sum_{i=1}^{K} Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha} * (\boldsymbol{\beta}))$ in a Taylor series about the true parameter $\boldsymbol{\beta}$ and evaluated at $\widehat{\boldsymbol{\beta}}_R^{(k)}$, we can get that

$$\sqrt{K}(\widehat{\boldsymbol{\beta}}_R^{(k)} - \boldsymbol{\beta}) = \{-\frac{1}{K}\sum_{i=1}^{K} \nabla_i\}^{-1}[\frac{1}{\sqrt{K}}\sum_{i=1}^{K} Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta}))]$$

where

$$\nabla_i = \frac{\partial Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha} * (\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} + \frac{\partial Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha} * (\boldsymbol{\beta}))}{\partial \boldsymbol{\alpha}}\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\beta}} = \mathbf{A}_i + \mathbf{B}_i\mathbf{C}$$

Secondly, we fixed $\boldsymbol{\beta}$ and expand $K^{-1/2}\sum_{i=1}^{K} Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta}))$ about the true parameter $\boldsymbol{\alpha}$ and evaluated at $\boldsymbol{\alpha}^*$ to get

$$\begin{aligned}\frac{1}{\sqrt{K}}\sum_{i=1}^{K} Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha}^*(\boldsymbol{\beta})) &= \frac{1}{\sqrt{K}}\sum_{i=1}^{K} Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) + \frac{1}{K}\sum_{i=1}^{K}\frac{\partial}{\partial\alpha} Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha})\sqrt{K}(\boldsymbol{\alpha}^* - \boldsymbol{\alpha}) + o_p(1) \\ &= \frac{1}{\sqrt{K}}\sum_{i=1}^{K} Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) + B^*C^* + o_p(1)\end{aligned}$$

Where the $o_p(1)$ term is due to regularity conditions which imply that the remainder term is $\frac{1}{K}O_p(1)$. Note that the weights are evaluated at the true parameters in this expansion too. Because $Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is evaluated at the true parameters, letting h_{ij}^T be the jth row of the $\mathbf{V}_i^{-1/2}$, we then have

$$\begin{aligned}\frac{1}{\sqrt{K}}\sum_{i=1}^{K} Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \frac{1}{\sqrt{K}}\sum_{i=1}^{K}\sum_{j=1}^{k} h_{ij}^T w_{ij}[y_{ij}^\dagger - g_{ij}^\dagger - m(\beta)] \\ &= \frac{1}{\sqrt{K}}\sum_{i=1}^{K}\sum_{j=1}^{k} h_{ij}^T a[r(y_{ij}^\dagger - g_{ij}^\dagger(\beta))] \\ &= \frac{1}{\sqrt{K}}\sum_{i=1}^{K} V_i^{-1/2} a[r(Y_i^\dagger - G_i^\dagger(\boldsymbol{\beta}))]\end{aligned}$$

The second equality holds because the weights are evaluated at the true parameters.

By Assumption [A.5] and [A.6], it follows from Theorem 3.1 of Brunner and Denker (1994) and the usual Cramer-Wold device that $\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1} Z_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is asymptotically normal with

mean 0 and variance-covariance matrix

$$M = \frac{1}{K}\sum_{i=1}^{K} \mathbf{V}_i^{-1/2} Var(\varphi_i^{\dagger}) \mathbf{V}_i^{-1/2}$$

The remainder of the proof follows from the previous proof. In particular, the results that $\mathbf{B}* = o_p(1)$ and $\mathbf{C}* = O_p(1)$ hold here, the proof of Theorem 5.4 is complete.