6-2005

Rank-Based Methods for Repeated Measures Data Under Exchangeable Errors

John Kloke
Western Michigan University

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

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1039
Rank-Based Methods for Repeated Measures Data Under Exchangeable Errors

by

John Kloke

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Statistics

Dr. Joseph McKean

Western Michigan University
Kalamazoo, Michigan
June 2005
ACKNOWLEDGMENTS

First and foremost, I wish to thank my advisor Professor Joseph McKeen for his extraordinary patience and guidance. Next, I wish to thank the members of the committee members for their helpful comments. Also, I wish to thank my teachers over the years.

I also wish to thank those less involved with my education, however very supportive and very much involved with the other aspects of my life, my friends and family. A big thanks to my roommates over the years. Finally I would like to thank my good friend Travis Foust with whom I had many insightful discussions throughout my college career.

John Kloke
LIST OF TABLES

1. One-way Example Simulated Data .............................................. 65
2. One-way Example Tests .......................................................... 66
3. One-way Example Estimates ...................................................... 66
4. Simulation Results for One-way Normal Data .............................. 67
5. Simulation Results for One-way Cauchy Data ............................... 68
6. Simulation Results for One-way $t_3$ Data .................................. 69
7. Main Effects Example - Without Outlier .................................... 71
8. Main Effects Example - With Outlier ......................................... 71
9. Interaction Example - Without Outlier ....................................... 72
10. Interaction Example - With Outlier .......................................... 72
11. Main Effect Example without Outlier ....................................... 74
12. Main Effects Example with Outlier ......................................... 74
LIST OF FIGURES

1. Power Curve for Interaction, $H_{01} : \gamma_1 = \gamma_2$ ........................................ 76
2. Power Curves for Main Effects, $H_{01} : \theta_1 = \theta_2$ ........................................ 77
3. Power Curves for Interaction, $H_{01} : \gamma_1 = \gamma_2$ ........................................ 78
4. Power Curves for Main Effects, $H_{01} : \theta_1 = \theta_2$ ........................................ 79
5. Contaminated Normal Power Curves for $\sigma = 3$ ................................................. 80
6. Contaminated Normal Power Curves for $\sigma = 5$ ................................................. 81
7. Contaminated Normal Power Curves for $\sigma = 10$ ................................................. 82
CHAPTER I

INTRODUCTION

The simplest repeated measures design is the paired data problem. One example of this type of is where a pair of subjects (for example siblings) are chosen. One member of the pair is then assigned to one of the two possible treatments the other member is then assigned to the other treatment. A common solution, as discussed in introductory statistics courses, is the paired t-test. The use of the t-test is equivalent to using a least squares estimate. It is well known that this method decreases the variability. That is, by using a paired design we have increased efficiency.

Another example is in which a measurement is taken on an individual then the individual is given some treatment for some time and then finally another measurement is taken on each of the subjects, forming a pair of observations. That is, each subject serves as its own control. The types of problems we consider in this paper generalize this simple scenario. That is, instead of two measurements per subject we have several ($n$).

In fact, repeated measures encompass a wider range of problems. For instance, consider measurements taken on the same family or taken at the same research center (hospital). In the case where the measurements are taken on the same individual over time, these data are often called longitudinal data. Another term sometimes applied to these types of data is cluster-correlated. Further, mixed model methods may be applied to these data.

Our goal is to obtain rank-based solutions for a certain general repeated
measures model. These estimates are extensions of the usual rank-based regression estimates. So we begin by reviewing least squares and rank regression as well as briefly discussing the asymptotic theory of the corresponding estimates.

1.1 Regression Estimation

Recall the multiple regression problem, where we are interested in the relationship between some measurement $y$ and some other set of measurements, $x_1, \ldots, x_p$. When we take such measurements on several experimental units (people, for example) we may write the problem as

$$
y_i = \alpha + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + e_i = \alpha + x_i^T \beta + e_i.
$$

Where, $x_i = [x_{i1}, \ldots, x_{ip}]^T$ is a $p \times 1$ vector of predictors, $\beta = [\beta_1, \ldots, \beta_p]^T$ is a $p \times 1$ vector of parameters, $\alpha$ is the intercept parameter, $y_i$ is the response variable, and $e_i$ is the error term. It is common to assume $e_1, \ldots, e_n$ are independent with mean $E[e_i] = 0$ and common variance $\text{var}[e_i] = \sigma^2$.

Now stack everything so that $y = [y_1, \ldots, y_n]^T$ is an $n \times 1$ response vector, $1$ is an $n \times 1$ vector of ones, $e = [e_1, \ldots, e_n]^T$ is an $n \times 1$ error vector, and

$$
X = 
\begin{bmatrix}
x_1^T \\
\vdots \\
x_n^T
\end{bmatrix}
$$

is an $n \times p$ design matrix. Write the linear model as,

$$
y = \alpha 1 + X \beta + e.
$$
Thus $E[e] = 0$ and $\text{var}[e] = \sigma^2 I$ where $\mathbf{0}$ is an $n \times 1$ vector of zeros and $I$ is an $n \times n$ identity matrix. We are interested in estimating $\beta$ and performing inference on $\beta$. We now describe two such estimates.

1.1.1 Least Squares Estimates

The most often used estimate of $\beta$ is the least squares (LS) estimate. A least squares solution to the regression problem is defined as a vector, $\hat{y}_{LS} = X \hat{\beta}_{LS}$ that is closest in Euclidean distance to the vector $y$ (see for example Trefethen and Bau 1997). That is,

$$\hat{\beta}_{LS} = \text{Argmin} \| y - X \beta \|_2 \text{ where } \| \mathbf{v} \|_2 = \left( \sum_{i=1}^{n} v_i^2 \right)^{1/2}$$

Under regularity conditions it can be shown that

$$\hat{\beta}_{LS} \sim N_p(\beta, \sigma^2 (X^T X)^{-1})$$

(see for example Stapleton 1995). We are using the symbol $\sim$ to denote approximately distributed. That is the distribution is exact at the limit $m \to \infty$. However, in practice we will make use of the approximate distribution to perform approximate inference on the vector of parameters $\beta$. A similar result holds for the rank-based estimates, which we describe next.

1.1.2 Rank-Based Estimates

Proposed by Jaeckel (1972), rank-based (R) estimates are an alternative to least squares estimates for the regression problem. As shown in McKean and Schrader (1980) this is equivalent to replacing Euclidean distance, with a rank-
based measure of distance. See McKean (2004) for a recent discussion.

A rank-based solution to the regression problem is defined as a vector, \( \hat{y}_\phi = X\hat{\beta}_\phi \) that is closest to \( y \) in terms of a rank dispersion function. That is,

\[
\hat{\beta}_\phi = \text{Argmin} \|y - X\beta\|_\phi \quad \text{where} \quad \|v\|_\phi = \sum_i a(R(v_i))v_i,
\]

(1.1)

\[ a(i) = \varphi \left( \frac{i}{n+1} \right), \]  and \( \varphi \) is a nondecreasing function, defined on the interval \((0,1)\) and standardized so that \( \int \varphi(u) du = 0 \) and \( \int \varphi^2(u) du = 1 \). Such a function, \( \varphi \), is called a rank score function. Some examples of score functions are the Wilcoxon \( \varphi(u) = \sqrt{12}(u - \frac{1}{2}) \), the sign \( (L) \) \( \varphi(u) = \text{sgn}[u - (1/2)] \), and the normal scores \( \varphi(u) = \Phi^{-1}(u) \).

It can be shown that under regularity conditions,

\[ \hat{\beta}_\phi \sim N_p(\beta, \tau^2_\phi(X^TX)^{-1}), \]

where \( \tau_\phi = \left[ \int \varphi(u)\varphi_f(u) du \right]^{-1} \) is a scale parameter and \( \varphi_f(u) = \frac{f(F^{-1}(u))}{f(F^{-1}(u))} \). Note that if we choose \( \varphi_f(u) \) as our score function then our estimates will be asymptotically efficient; see Cheng and Hettmansperger (1983) as well as section (1.8.2) of Hettmansperger and McKean (1998).

1.1.3 Regularity Conditions

In this section we briefly discuss the usual assumptions imposed on the design matrix \( X \). First define the hat matrix as the following projection matrix

\[ H = X(X^TX)^{-1}X. \]
Notice $H \hat{y} = \hat{y}_{LS}$. Now consider a sequence of such hat matrices $H_n$. Denote $h_{iin}$ as the $ith$ diagonal entry of $H_n$. The following design conditions are usually assumed for the asymptotic theory

$$\text{(D.2)} \lim_{n \to \infty} \max_{i} h_{iin} = 0$$

$$\text{(D.3)} \lim_{n \to \infty} n^{-1} X^T X = \Sigma$$

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

Condition (D.2) is called Huber’s condition which implies the following condition:

$$\text{(N.1)} \max_{i} \frac{x_{ik}^2}{\sum_{j=1}^{n} x_{jk}^2} \to 0 \text{ for } k = 1, \ldots, p,$$

which is called Noether’s condition.
 Consider a longitudinal study where we observe \( n \) repeated measures on each of \( m \) subjects. For example one might be interested in the effect that a new drug has on lowering cholesterol. In this case, repeated measures are then taken over time, perhaps weekly, for some period of time. It is this type of problem we wish to explore and develop rank methods for.

It is convenient to represent the repeated measures for a subject as a vector. So for each of the \( m \) individuals we then have an \( n \times 1 \) response vector, \( \mathbf{y}_k = [y_{k1}, \ldots, y_{kn}]^T \). Conventionally this problem is framed as a linear model,

\[
y_k = \alpha_k \mathbf{1}_n \mathbf{X}_k \beta + \mathbf{e}_k \text{ for } k = 1, \ldots, m
\]

where, \( \alpha_k \) is an intercept parameter, \( \mathbf{X}_k \) is a \( nxp \) design matrix, \( \beta \) is a \( p \)-vector of unknown parameters, and \( \mathbf{e}_k \) is the vector of errors for the \( k \)th subject. As usual the we assume the errors between subjects are independent and have mean vector zero, \( E[\mathbf{e}_k] = \mathbf{0} \). If however we are fitting a medians model (as opposed to a means model) we assume that the medians of the residuals are zero. See Hettmansperger and McKean (1998, ch.4) for a discussion on means versus medians models. Since we have included a design matrix in this model, there is no loss in generality in assuming that \( \alpha_k = \alpha \) for \( k = 1, \ldots, m \). However fitting several intercepts will improve the interpretation and simplify the inference. These are obvious extensions of the univariate regression problem discussed in the previous
chapter. It is also common to make assumptions on the variance covariance structure of the error vectors. In our case we assume that the within subject errors are exchangeable. If we wish to assume finite variance, this is equivalent to assuming that the error vectors have compound symmetry variance covariance structure, \( \text{var}[e_k] = \sigma^2 A(\rho) \), where \( A(\rho) = (1 - \rho)I + \rho J \), \( J = 11^T \), \( 1 \) is a \( n \)-vector of ones, and \( I \) is an \( nxn \) identity matrix. That is,

\[
A(\rho) = \begin{bmatrix}
1 & \rho & \cdots & \rho \\
\rho & 1 & \cdots & \rho \\
\vdots & \vdots & \ddots & \vdots \\
\rho & \rho & \cdots & 1
\end{bmatrix} > 0.
\]

This model is what we call the general model. In this chapter we explore some sub-models of this general model. In the next chapter we develop the asymptotic theory of the rank estimates for these models.

Our approach is both simple to understand and easily implemented using existing software, for example RGLM (Kapenga, et. al. 1988) or using an iteratively reweighted least squares approach as in Cheng and Hettmansperger (1983). Sievers and Abebe (2004) suggest centering by using the median of the residuals ensuring positive weights. We make use of some of the software of Terpstra and McKean (2004).

2.1 One-Way Setup

In the one-way setup there is only one treatment level. The parameter of interest is a vector of means, \( X_k \beta = \mu \) for \( k = 1, \ldots, m \). The one-way repeated measures model is often written in an overparameterized fashion, \( \mu = \theta 1 + \gamma, k = \ldots \).
1, \ldots, m. Here, $\theta 1 = P_1 \mu$ and $\gamma = P_{1\perp}$ where, $P_1$ and $P_{1\perp}$ are the projections onto the space spanned by a column of ones and the space orthogonal to a column of ones, respectively. The main inference of interest is done in two parts. First, a test for effect $H_0 : \mu = \theta 1$, which is equivalent to testing $H_0 : \gamma = 0$. Second, confidence intervals and tests on contrasts $h^T \mu$. One such contrast is baseline versus the last time point, $h^T = [1, 0, \ldots, 0, -1]$. That is the stage one analysis is a test of effect $H_0 : \gamma = 0$. If we indeed reject the hypothesis of effect in the stage one analysis we may proceed to the stage two analysis, inference on $h^T \mu$. Where $h$ is a contrast, that is $h^T 1 = 0$.

There are numerous ways of looking at this problem. Often it is thought of as a randomized complete block or framed as a multivariate location problem. In the rest of this section we look at some nonparametric approaches to this problem. The first two are standard methods and finally the third is our new approach.

2.1.1 Friedman’s Test

The classical nonparametric approach to this problem is Friedman’s (1937) test. Randles and Wolfe (1979) point out that the test is distribution free given that the errors are exchangeable under the null. Here the model is written as a randomized complete block design, $y_{kj} = \theta + \gamma_j + \tau_k + e_{kj}$, where subject is the block. For the method, rank within subject (row) and form the test statistic

$$T_1 = \frac{12}{mn(n+1)} \sum_{j=1}^{n} \left( R_j - \frac{m(n+1)}{2} \right)^2,$$

where $R_j$ is the sum of the ranks for the $j$th column, $R_j = \sum_{k=1}^{m} R_{kj}$. The asymptotic distribution is $\chi^2_{(n-1)}$ (see Lehmann 1998). There are corrections for ties in the data, see for example Hollander and Wolfe (1999). Note this procedure
results only in a test. In particular, there is no estimate of $\mu$ nor any stage two analysis.

2.1.2 Multivariate

The multivariate solution is to model this as a one sample multivariate location problem,

$$
\mathbf{y} = \begin{bmatrix}
  y_1^T \\
  \vdots \\
  y_m^T
\end{bmatrix} = \mathbf{1}\mu^T + \mathbf{e} \quad \text{where} \quad \mathbf{e} = \begin{bmatrix}
  e_1^T \\
  \vdots \\
  e_m^T
\end{bmatrix}.
$$

For this model, the compound symmetry assumption is not required. The rank based approach we consider here is that of Davis and McKean (1993). Their estimate of $\mu$ is the solution of $S^+(\mathbf{y} - \mathbf{1}\mu^T) = \mathbf{0}$. Where,

$$
S^+(\mathbf{y} - \mathbf{1}\mu^T) = [S_1^+(\mathbf{y} - \mathbf{1}\mu^T), \ldots, S_n(\mathbf{y} - \mathbf{1}\mu^T)]
$$

and $S_j^+(\mathbf{y} - \mathbf{1}\mu^T) = \sum_{k=1}^n \text{sgn}(y_{kj} - \mu_j)\alpha^+(R_j(|Y_{kj} - \mu_j|))$. Wilcoxon scores result in componentwise Hodges-Lehmann estimates with the following asymptotic distribution

$$
\sqrt{n}(\hat{\mu} - \mu) \rightarrow N_n(0, \mathbf{T}\Sigma\mathbf{T}),
$$

where $\mathbf{T}$ is a diagonal matrix of scale parameters and $\Sigma$ is a correlation matrix on the ranks. Consistent estimates are $\hat{\mathbf{T}} = \text{diag}(\hat{\tau}_1, \ldots, \hat{\tau}_n)$ and $\hat{\Sigma} = [\sigma_{ij}]$ where

$$
\sigma_{ij} = \frac{12}{m-1} \sum_{k=1}^m \left[ \frac{R(e_{ik})}{m+1} - \frac{1}{2} \right] \left[ \frac{R(e_{jk})}{m+1} - \frac{1}{2} \right].
$$
We use the estimates of $\tau_j$ discussed in Davis and McKeown (1993). We compute these with with Terpstra and McKeown's (2004) function, \texttt{wilcoxonontau} on the residuals, $\hat{e}^{(j)} = y^{(j)} - \hat{\mu}_j 1$. Where $y^{(j)}$ is the $j$th column of $Y$. That is $\hat{\tau}_j = \tau(\hat{e}^{(j)})$.

The stage one test is written as a set of contrasts on the means, $H_0: I\mu'K = 0$. One such $K$ is $K = [-1_{n-1} I_{n-1}]$, that is $K$ is a $n-1 \times n$ contrast matrix. The resulting Wald type test is

$$T_2 = m[K \hat{\mu}]^T[K \hat{T} \hat{\Sigma} \hat{T} K^T]^{-1}[K \hat{\mu}]$$

which asymptotically has a $\chi^2_{(n-1)}$ distribution.

2.1.3 Arnold's Transformation

Our new method depends on some results due to Arnold (1979, 1981), who, working under the assumption of normality, developed the following simplifying transformation. Consider the orthogonal transformation applied to each subject,

$$y^*_k = \Gamma y_k = \begin{bmatrix} \frac{1}{\sqrt{n}} 1^T \\ C^T \end{bmatrix} y_k = \begin{bmatrix} y_{k1}^* \\ y_{k2}^* \end{bmatrix} = \mu^* + e_k^*,$$

where the columns of $C$ are an orthonormal basis for $1^\perp$ ($C^T 1 = 0$),

$$\mu^* = \begin{bmatrix} \theta^* \\ \gamma^* \end{bmatrix}, e_k^* = \begin{bmatrix} e_{k1}^* \\ e_{k2}^* \end{bmatrix},$$

$\theta^* = \sqrt{n} \theta$, and $\gamma^* = C^T \gamma$. The random variables $y_{11}^*, \cdots, y_{m1}^*$ are independent with mean, $E[y_{k1}] = \theta^* = \sqrt{n} \theta$ and variance $\text{var}[y_{k1}^*] = \lambda_1 = \sigma^2 (1 + (n-1) \rho)$. The
random $n-1$ vectors $\mathbf{y}_1^*, \ldots, \mathbf{y}_m^*$ are independent with mean, $E[\mathbf{y}_k^*] = \gamma^* = C^T \gamma$ and variance $\text{var}[\mathbf{y}_k^*] = \lambda_2 I_{(n-1)} = \sigma^2 (1-\rho) I_{(n-1)}$. Further, the covariance between the two pieces is zero, $\text{cov}[\mathbf{y}_k^*, \mathbf{y}_k^*] = 0$.

It will be convenient to represent the two pieces separately by

$$z_1 = [\mathbf{y}_{11}^*, \ldots, \mathbf{y}_{m1}^*]^T = \theta^* \mathbf{1}_m + \mathbf{e}_1$$

and

$$z_2 = [\mathbf{y}_{12}^*, \ldots, \mathbf{y}_{m2}^*]^T = \mathbf{W} \gamma^* + \mathbf{e}_2$$

where

$$\mathbf{W} = \begin{bmatrix} I_{(n-1)} \\ \vdots \\ I_{(n-1)} \end{bmatrix}$$

is a $m(n-1)\times(n-1)$ matrix and $I_{(n-1)}$ is an $(n-1)\times(n-1)$ identity matrix. So that $z_1$ is an $m$-vector and $z_2$ is an $m(n-1)$-vector. Notice that the first component is the mean and the second component is a set of orthogonal contrasts. Further, the elements of $z_1$ are independent; hence inference on the means (first component) is immediate. Under normal errors, all elements of $z$ are independent, so in particular both components are independent. So the elements of $z_2$ are iid, hence inference on the contrasts (second component) is immediate. However, as shown in the next chapter, the normality assumption is not required.

Since the transformation is one-to-one, estimation of the mean vector is straightforward,

$$\hat{\mu} = \Gamma^T \hat{\mu}^* = \begin{bmatrix} \frac{1}{\sqrt{n}} \mathbf{1} & C \end{bmatrix} \begin{bmatrix} \theta^* \\ \gamma^* \end{bmatrix} = \frac{1}{\sqrt{n}} \mathbf{1} \hat{\theta}^* + C \hat{\gamma}.$$
Recall the main inference of interest is two parts. First, test all the effects are zero, \( H_0 : \mu = \theta 1 \) which is equivalent to testing \( H_0 : \gamma^* = 0 \). Second, form confidence intervals and tests on contrasts, \( h^T \mu \). Since \( h \) is a contrast \( h^T \mu = h^T [\theta 1 + \gamma] = h^T \gamma \) Hence, for the one-way problem the main inference of interest is in the second component.

Since,

\[
\sigma^2 A(\rho) \frac{1}{\sqrt{n}} 1 = \sigma^2 [(1 - \rho) + n\rho] \frac{1}{\sqrt{n}} 1,
\]

\( \lambda_1 = \sigma^2 [(1 - \rho) + n\rho] \) is the first eigenvalue and \( v_1 = \frac{1}{\sqrt{n}} 1 \) is its corresponding eigenvector. Further, since,

\[
\sigma^2 A(\rho) C = \sigma^2 (1 - \rho) C,
\]

\( \lambda_2 = \sigma^2 (1 - \rho) \) is the second eigenvector with multiplicity \( n - 1 \) and the \( n - 1 \) columns of \( C \) are the corresponding eigenvectors.

Recall after we apply Arnold’s transformation all the information on the effect is in the second piece, \( z_{2k} = \gamma^* + \epsilon_{2k} \). Further, the elements of \( \epsilon_{2k} \) are uncorrelated, \( \text{var}[\epsilon_{2k}] = \lambda I \). This corresponds to working independence in GEE terminology (Liang and Zeger 1986). If we are willing to assume uncorrelated implies independence, we may use the usual rank estimates as in Chapter 3 of Hettmansperger and McKean (1998).

To fit we write \( z_2 = W \gamma^* + \epsilon_2 \) as a regression model, \( z_2^* = X \beta + \epsilon_2 \) where, \( \beta = E^{-1} \gamma = [\alpha, \beta_1, \cdots, \beta_{n-2}]^T \) and \( X = WE \). So that \( \gamma^* = E \hat{\beta} \) follows an asymptotic
$N(\gamma^*, \tau_0^2 \mathbf{E}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{E}^T)$ distribution. A Wald type test of $H_0 : \gamma^* = 0$ is

$$T_4 = \frac{\hat{\gamma}^T \left[ \tau_0^2 \mathbf{E}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{E}^T \right]^{-1} \hat{\gamma}^* / (n - 1)}{\tau_0^2}$$

$$= \frac{\hat{\beta}^T (\mathbf{X}^T \mathbf{X}) \hat{\beta} / (n - 1)}{\tau_0^2}$$

$$= \frac{\|\hat{\delta}_2^*\|^2 / (n - 1)}{\tau_0^2}$$

has an approximate $F_{\alpha, n-1, mn-m-n+1}$ distribution. In fact $(n - 1)T_4$ is asymptotically $\chi^2(n - 1)$. The $F$ standardization has been shown to have good small sample properties; see Hettmansperger and McKean (1998, ch.3).

2.2 Two Sample

Suppose instead of testing for a time effect of one treatment we are interested in comparing two treatments. Two common examples of these types of comparisons are comparing a new drug compound to a standard treatment or comparing the new drug to placebo. Maybe the effect a new drug has on lowering cholesterol compared to a standard treatment. In this section we explore some rank-based Arnold transformation solutions to these types of two-sample repeated measures problems. In these problems the subjects are assigned to one of two treatment categories and measurements are taken over time.

Suppose there are $m_1$ subjects assigned to the first treatment and there are $m_2$ subjects assigned to the second treatment. Denote, $\mathbf{u}_1, \ldots, \mathbf{u}_{m_1}$, as the vectors of repeated measures for the subjects on the first treatment. Similarly denote, $\mathbf{v}_1, \ldots, \mathbf{v}_{m_2}$ as the vectors of repeated measures for the subjects on the second treatment.
Now, model each subject’s \( n \times 1 \) response vector as a linear model,

\[
\begin{align*}
\mathbf{u}_i &= \mathbf{\mu}_1 + \mathbf{r}_i = \theta_1 \mathbf{1} + \gamma_1 + \mathbf{r}_i \text{ for } i = 1, \ldots, m_1 \\
\mathbf{v}_j &= \mathbf{\mu}_2 + \mathbf{s}_j = \theta_2 \mathbf{1} + \gamma_2 + \mathbf{s}_j \text{ for } j = 1, \ldots, m_2
\end{align*}
\]

Assume, the errors vectors are independent, have zero mean vector, and common variance-covariance. That is \( \mathbf{r}_1, \ldots, \mathbf{r}_{m_1}, \mathbf{s}_1, \ldots, \mathbf{s}_{m_2} \) are independent

\[
E[\mathbf{r}_i] = E[\mathbf{s}_j] = \mathbf{0}, \text{ and } \text{var}[\mathbf{r}_i] = \text{var}[\mathbf{s}_j] = \mathbf{V}
\]

for \( i = 1, \ldots, m_1 \) and \( j = 1, \ldots, m_2 \).

Now write one common linear model,

\[
\mathbf{y} = \begin{bmatrix} 1_{nm_1} & 0 \\ 0 & 1_{nm_2} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \mathbf{W}_1 & 0 \\ 0 & \mathbf{W}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\gamma}_1 \\ \mathbf{\gamma}_2 \end{bmatrix} + \mathbf{e}
\]

Where

\[
\mathbf{y} = [\mathbf{u}_1^T, \ldots, \mathbf{u}_{m_1}^T, \mathbf{v}_1^T, \ldots, \mathbf{v}_{m_2}^T]^T \text{ and } \mathbf{e} = [\mathbf{r}_1^T, \ldots, \mathbf{r}_{m_1}^T, \mathbf{s}_1^T, \ldots, \mathbf{s}_{m_2}^T]^T
\]

are \( mn \times 1 \) vectors, \( \mathbf{W}_1 = 1_{m_1} \otimes \mathbf{I}_n \) is an \( m_1n \times n \) incidence matrix and \( \mathbf{W}_2 = 1_{m_2} \otimes \mathbf{I}_n \) is an \( m_2n \times n \) incidence matrix. Define \( m = m_1 + m_2 \) as the total sample size. So that \( \mathbf{\mu}_1 = \theta_1 \mathbf{1} + \mathbf{\gamma}_1 \) is the common mean vector for the first treatment and \( \mathbf{\mu}_2 = \theta_2 \mathbf{1} + \mathbf{\gamma}_2 \) is the common mean vector for the second treatment. That is we have written \( \mathbf{\mu}_1 \) and \( \mathbf{\mu}_2 \) in an over-parameterized fashion.

As with the oneway solution we make the assumption that the within
subject errors are exchangeable. Again, this is loosely equivalent to assuming that the error vectors have compound symmetry covariance structure. That is

\[ V = \text{var}[e_k] = \sigma^2 A(\rho) = \sigma^2 [(1 - \rho)I + \rho J] = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \ddots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} > 0 \]

where \( J = 11^T \), \( 1 \) is a \( n \times 1 \) vector of ones, and \( I \) is an \( n \times n \) identity matrix.

2.2.1 Profile Analysis

A typical analysis for this setup is a profile analysis; see for example, Seber (1984) or Fitzmaurice et. al. (2004). The three tests of interest are,

- \( H_{01} : \gamma_1 = \gamma_2 \) parallel profiles or no interaction
- \( H_{02} : \theta_1 = \theta_2 \) or no main effects
- \( H_{03} : \gamma_1 = \gamma_2 = 0 \) or flat profiles

Note that \( H_{01} \) and \( H_{02} \) results in coincident profiles. We will focus on tests of coincident profiles. We feel that the most often the primary inference of interest is \( H_{02} \). That is, is the active treatment performs better then placebo or the new treatment performs better than the standard treatment. In the presence of interaction, however, main effects may be difficult to interpret. So consider the following two scenarios.

Suppose we are comparing a new active treatment to placebo. Then we probably expect interaction. That is we expect the profile for placebo to be relatively flat where as the profile for the active treatment should be improving. So that we will probably reject parallel profiles. Therefore the test of main effects may
not be valid. However a contrast such as \( c = [1, 0, \ldots, 0, -1, 0, \ldots, 0, 1] \) is often of great interest. Suppose we are comparing to active treatments, one standard treatment and one new. One hopes that in addition to having fewer side effects and a better over all treatment effect the new treatment will take effect faster. Hence a contrast of interest in this case may be \( c = [1, -1, \ldots, 0] - 1, 1, \ldots, 0] \).

2.2.2 Arnold Transformation Solution

As in the one sample case we apply Arnold’s transformation to each subject vector,

\[
\begin{align*}
\mathbf{u}_i^* &= \begin{bmatrix} u_{1i}^* \\ u_{2i}^* \end{bmatrix} = \Gamma \mathbf{u}_i = \begin{bmatrix} \theta_1^* \\ \gamma_1^* \end{bmatrix} + r_i^* \text{ for } i = 1, \ldots, m_1 \\
\mathbf{v}_j^* &= \begin{bmatrix} v_{1j}^* \\ v_{2j}^* \end{bmatrix} = \Gamma \mathbf{v}_j = \begin{bmatrix} \theta_2^* \\ \gamma_2^* \end{bmatrix} + s_j^* \text{ for } j = 1, \ldots, m_2.
\end{align*}
\]

Next, break into two pieces, \( z_1 = [u_{11}^*, \ldots, u_{m_1}^*, v_{11}^*, \ldots, v_{m_2}^*]^T \) and \( z_2 = [u_{21}^T, \ldots, u_{m_1}^T, v_{12}^T, \ldots, v_{m_2}^T]^T \). Now write each as a linear model,

\[
\begin{align*}
z_1 &= \begin{bmatrix} 1_{m_1} & 0 \\ 0 & 1_{m_2} \end{bmatrix} \begin{bmatrix} \theta_1^* \\ \theta_2^* \end{bmatrix} + \varepsilon_1 \\
\text{and} \\
z_2 &= \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} \gamma_1^* \\ \gamma_2^* \end{bmatrix} + \varepsilon_2 = W^* + \varepsilon_2 \end{align*}
\]

where, \( W_1 = 1_{m_1} \otimes I_{n-1} \) and \( W_2 = 1_{m_2} \otimes I_{n-1} \). So that \( W_1 \) is \( m_1(n-1) \times (n-1) \), \( W_2 \) is \( m_2(n-1) \times (n-1) \), and \( W \) is \( m(n-1) \times 2(n-1) \).

As we have transformed to working independence, we may use the usual
regression methods. To estimate $\gamma^*$ we fit the linear model

$$z_2 = W \gamma^* + \epsilon_2.$$ 

2.2.3 Inference

We first consider the least squares solution of Arnold. The hypothesis of no interaction is $H_{01}: \gamma_1^* = \gamma_2^*$. Since,

$$\hat{\gamma}_{LS} \sim N_{2(n-1)}(\gamma^*, \sigma_2^2(W^T W)^{-1})$$

where

$$(W^T W)^{-1} = \begin{bmatrix} \frac{1}{m_1} I & 0 \\ 0 & \frac{1}{m_2} I \end{bmatrix}$$

the level $\alpha$ Wald-type least squares test of parallel profiles is to reject $H_{01}$ if

$$\frac{m_1 m_2}{m_1 + m_2} \frac{||\hat{\gamma}_{LS,2} - \hat{\gamma}_{LS,1}||^2}{\hat{\sigma}^2} > F_{\alpha, n-1, m(n-1)-2(n-1)}.$$ 

Since,

$$\hat{\gamma}_\varphi \sim N_{2(n-1)}(\gamma^*, \tau_\varphi^2(W^T W)^{-1})$$

an $\alpha$ level Wald-type test of parallel profiles, is to reject $H_{01}$ if

$$\frac{m_1 m_2}{m_1 + m_2} \frac{||\hat{\gamma}_\varphi,2 - \hat{\gamma}_\varphi,1||^2}{\hat{\tau}_\varphi^2} > F_{\alpha, n-1, m(n-1)-2(n-1)}.$$ 

A test for main effects may be based on the the transformed parameters, $H_{02}: \theta_1^* = \theta_2^*$. This results in a two sample location problem on $u_{11}^*, \ldots, u_{m1}^*$ and $v_{11}^*, \ldots, v_{m2}^*$. 

17
Let $\hat{\Delta}_{LS} = \bar{v}_1^* - \bar{u}_1^*$ then an level $\alpha$ least squares test of main effects is to reject $H_{02}$ if

$$\frac{\hat{\Delta}_{LS}}{\hat{\sigma}^* \sqrt{\frac{1}{m_1} + \frac{1}{m_2}}} > t_{\alpha, m_1 + m_2 - 2}.$$ 

We could use the Mann-Whitney-Wilcoxon rank sum statistic, $S^+_R = \#\{v_{j_1}^* - u_{i_1}^* > 0\}$ but instead we use the t-test described in chapter two of Hettmansperger and McKean (1998). Let

$$E = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad E^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad \text{and} \quad \beta^* = \begin{bmatrix} \alpha^* \\ \Delta^* \end{bmatrix}.$$ 

So that $\alpha^*$ is the intercept and $\Delta^*$ is the shift parameter.

The Wilcoxon rank-based estimate is,

$$\hat{\Delta}_\phi^* = \text{med}\{v_{j_1}^* - u_{i_1}^*\}.$$ 

and a level $\alpha$ test of main effects is to reject $H_{02}$ if,

$$\frac{\hat{\Delta}_\phi^*}{\hat{\tau}^*_\phi \sqrt{\frac{1}{m_1} + \frac{1}{m_2}}} > t_{\alpha, m_1 + m_2 - 2}.$$ 

To test for flat profiles we test if all the effect are zero, $H_{03} : \gamma_1 = \gamma_2 = 0$ which is equivalent to testing the transformed parameters are zero, $\gamma_1^* = \gamma_2^* = 0$. As such we end up with a test similar to the one sample problem.

2.2.4 Contrasts

We may estimate any general contrast. If we use the median the resulting inference may have less efficiency. If we are willing to assume symmetry we may
use a Hodges-Lehmann type estimate of intercept. However there are certain contrasts which we may estimate which are invariant to intercept. In this section we explore these types of contrasts.

To aid our discussion consider the following table representing the means of the two treatment levels.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>\cdots</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1</td>
<td>\mu_{11}</td>
<td>\mu_{12}</td>
<td>\cdots</td>
<td>\mu_{1n}</td>
</tr>
<tr>
<td>Treatment 2</td>
<td>\mu_{21}</td>
<td>\mu_{22}</td>
<td>\cdots</td>
<td>\mu_{2n}</td>
</tr>
</tbody>
</table>

Represent the treatment means as vectors,

\[ \begin{align*}
\mu_1 &= \theta_1 1 + \gamma_1 \\
\mu_2 &= \theta_2 1 + \gamma_2
\end{align*} \]

and stack the two mean vectors,

\[ \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \theta_1 1 + \gamma_1 \\ \theta_2 1 + \gamma_2 \end{bmatrix}. \]

Let \( h = [h_1 h_2]^T \). We are interested in estimating contrasts of the form

\[ \eta = h^T \mu = h_1^T \mu_1 + h_2^T \mu_2 = h_1^T (\theta_1 1 + \gamma_1) + h_2^T (\theta_2 1 + \gamma_2). \]

Clearly we may estimate contrasts within a treatment (row). For example if we are interested in a contrast on treatment one, \( h_2 = 0 \) so that \( \eta = h_1^T \gamma_1 \).

What about the same contrast within the rows? That is \( h_1 = -h_2 \) so that \( \eta = h_1^T \gamma_1 + h_2^T \gamma_2 = h_1^T (\gamma_1 - \gamma_2) \). One final contrast we consider is the difference between treatment one and treatment two within a column. That is \( \mu_{1j} - \mu_{2j} = \)
This contrast may be estimated without estimating a common intercept. However we do need to estimate an overall treatment effect, something we avoid in the previous contrasts.

2.3 Analysis of Covariance

As a final example we consider the following simple analysis of covariance model,

\[ y_k = \theta \mathbf{1} + \gamma + \beta \mathbf{x}_k + e_k \quad \text{for } k = 1 \ldots, m \]

where \( \mathbf{x}_k = [x_{k1}, \ldots, x_{kn}]^T \) is an \( n \times 1 \) predictor vector for the \( k \)th subject.

As before apply Arnold’s transformation to each response vector,

\[ y_k^* = \Gamma y_k = \begin{bmatrix} y_{k1}^* \\ y_{k2}^* \end{bmatrix}. \]

Where,

\[ y_{k1}^* = \frac{1}{\sqrt{n}} \mathbf{1}^T y_k = \frac{1}{\sqrt{n}} \theta \mathbf{1} + \frac{1}{\sqrt{n}} \beta \mathbf{x}_k + e_{k1} \]

and

\[ y_{k2}^* = C^T y_k = C \gamma + \beta C^T \mathbf{x}_k + e_{k2}. \]

Consider two cases. First, if \( \mathbf{x}_k \) is centered, that is \( \bar{x}_k = 0 \). That is the design for the \( k \)th subject is centered. So all the information about \( \beta \) is in the second piece. Second, if \( \mathbf{x}_k \) only depends on \( k \), \( \mathbf{x}_k = x_k \mathbf{1} \). That is \( \mathbf{x}_k \) only depends on \( k \), the predictor is constant across measurements. So that all the information about \( \beta \) is in the first piece.
2.4 General Model

In this section we discuss the general model. Arnold (1979, 1981) develops the general theory by taking a coordinate-free approach. We begin with this coordinate free model, however, we take a regression approach to develop the asymptotic theory.

As with the examples discussed in the previous sections, let \( y_k = [y_{k1}, \ldots, y_{kn}]^T \) denote the vector of responses for the \( k \)th subject. Write each response vector as a means model,

\[
y_k = \mu_k + e_k = \alpha_k 1 + \eta_k + e_k \quad \text{for} \quad k = 1, \ldots, m.
\]

Let \( P_1 \) be a projection matrix onto the space spanned by one and let \( P_{1^\perp} \) be a projection matrix onto the space orthogonal to one. Then \( \alpha_k 1 = P_1 \mu_k \) and \( \eta_k = P_{1^\perp} \mu_k \). That is we have written \( \mu_k \) in an over-parameterized fashion. As before, assume that each error vector has mean zero and a compound symmetry variance covariance structure, that is,

\[
E[e_k] = 0 \quad \text{and} \quad \text{var}[e_k] = \sigma^2 A(\rho) \quad \text{for} \quad k = 1, \ldots, m,
\]

where \( A(\rho) = (1 - \rho)I + \rho J \) or

\[
A(\rho) = \\
\begin{bmatrix}
1 & \rho & \cdots & \rho \\
\rho & 1 & \cdots & \rho \\
\vdots & \ddots & \ddots & \vdots \\
\rho & \rho & \cdots & 1
\end{bmatrix}
\]

Since \( \sigma^2 A(\rho) \) is a covariance matrix, we assume that it is positive definite. This
amounts to assuming \((1 + (n - 1)p) > 0\) and \((1 - p) > 0\).

Define,

\[
\alpha = \begin{bmatrix}
\alpha_1 \\
\vdots \\
\alpha_m
\end{bmatrix}
\quad \text{and} \quad
\eta = \begin{bmatrix}
\eta_1 \\
\vdots \\
\eta_m
\end{bmatrix}
\]

where \(\alpha \in S\) and \(\eta \in T\). The subspace \(S\), is a proper subspace of \(\mathcal{R}^m\), that is \(\dim(S) = s < m\). The subspace \(T\), is a proper subspace of \(\mathcal{R}^m\). In fact \(\dim(T) = t < m(r - 1)\). In the one-sample location problem (section 2.2) \(\alpha_k = \theta\) and \(\gamma_k = \gamma\) for \(k = 1, \ldots, m\) so that, \(s = 1\) and \(t = n - 1\).

In the two-sample location problem (section 2.3) \(\alpha_k = \theta_1\) for \(k = 1, \ldots, m_1\), \(\alpha_k = \theta_2\) for \(k = m_1 + 1, \ldots, m\) \(\eta_k = \gamma_1\) for \(k = 1, \ldots, m_1\) and \(\eta_k = \gamma_2\) for \(k = m_1 + 1, \ldots, m\) so that, \(s = 2\) and \(t = 2 \times (n - 1)\).

For the asymptotic theory, write each response vector as an analysis of covariance model,

\[
y_k = \alpha_k 1 + X_k \beta + e_k \quad \text{for} \quad k = 1, \ldots, m.
\]

That is \(\eta_k = X_k \beta\) and \(X_k\) is a basis matrix for \(T_k\).

Now apply Arnold’s transformation to each response vector,

\[
y_k^* = \Gamma y_k = \begin{bmatrix} y_{k1}^* \\ y_{k2}^* \end{bmatrix} \quad \text{for} \quad k = 1, \ldots, m.
\]
2.4.1 Mean Component

The inference on the mean component is straight forward. For the $k$th subject we have the (univariate) random variable,

$$y_{k1} = \frac{1}{\sqrt{n}} 1^T y_k = \frac{1}{\sqrt{n}} 1^T [\alpha_k 1 + X_k \beta + e_k] = \sqrt{n} \alpha_k + e_{k1}^* = \alpha_k + \epsilon_{k1}^*$$

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

For this general model, observe that $y_{k1}, \ldots, y_{km}$ are independent random variables so that the inference on the first piece, $z_1 = [y_{k1}^*, \ldots, y_{km}^*]^T$, is simply a one-way ANOVA model. Write this as means (or medians) model, that is as a linear model,

$$z_1 = W \theta^* + \epsilon_1$$

where, $W \theta^* = \alpha^* = \sqrt{n} \alpha$. That is $W$ is a basis matrix for $T^*$. So that a least squares estimate has the following asymptotic distribution

$$\hat{\theta}_{LS}^* \sim N_s(\theta^*, \sigma_1^2 (W^T W)^{-1})$$

Similarly the rank estimate has the following asymptotic distribution

$$\hat{\theta}_\varphi^* \sim N_s(\theta^*, \tau_1^2 (W^T W)^{-1})$$

2.4.2 Contrast Component

In this section we describe the asymptotic theory for the contrast piece. For each subject we have the following vector of uncorrelated random variables,

$$y_{k2}^* = C^T y_k = C^T [\theta_k 1 + X_k \beta + e_k] = X_k^* \beta + e_{k2}^* \text{ for } k = 1, \ldots, m.$$
In terms of the coordinate-free approach \( \eta^* = C^T \eta \) so that \( X_k^* = C^T X_k \) is a basis matrix for \( V_k^* \).

The linear model we fit is

\[
Z_2 = X^* \beta + \epsilon_2.
\]

Where \( z_2 = [y_{12}, \ldots, y_{m2}]^T \), \( \epsilon_2 = [e_{12}, \ldots, e_{m2}]^T \), and

\[
X^* = \begin{bmatrix}
X_1^* \\
\vdots \\
X_m^*
\end{bmatrix}
\]

So that \( z_2 \) and \( \epsilon_2 \) are \( m(n-1) \times 1 \) vectors and \( X^* \) is an \( m(n-1) \times p \) design matrix. Further, \( E[\epsilon_2] = 0 \) and \( \text{var}[\epsilon_2] = \lambda_2 I \). Hence, not only are the transformed observations uncorrelated, they have the same variance.

The least squares estimate is defined as

\[
\hat{\beta}_{LS} = \text{Argmin} \| z_2 - X^* \beta \|_{LS} = \text{Argmin}(z_2 - X^* \beta)^T (z_2 - X^* \beta).
\]

So that \( \hat{\beta}_{LS} = (X^{*T}X^*)^{-1}X^{*T}z_2 \).

The least squares estimate has the following approximate distribution,

\[
\hat{\beta}_{LS} \sim N_p(\beta, \sigma^2 (X^{*T}X^*)^{-1}).
\]

However, for R estimates, we need to fit an intercept. This is no problem as there are two cases. Firstly, if \( 1 \) is in the span of \( X \) then we fit the model,

\[
z_2 = \alpha 1 + X_1^* \beta_1 + \epsilon_2 = X^* E E^{-1} \beta + \epsilon_2
\]

24
and use Wilcoxon scores to estimate the intercept. Secondly, if 1 is not in the span of $X$ then we fit the model,

$$z_2 = \alpha 1 + X^* \beta + \epsilon_2$$

and then project to the correct space (see Dixon and McKean 1996).

Define the rank-based estimates as

$$\hat{\beta}_\varphi = \text{Argmin} \| z_2 - X^* \beta \|_\varphi$$

If we are willing to assume that uncorrelated implies independence then the approximated distribution of the rank estimate is

$$\hat{\beta}_\varphi \sim N_p(\beta, \tau^2(X^* X)^{-1})$$

If we do not feel comfortable with the uncorrelated implies independence assumption, we may make use of the following correction

$$\hat{\beta}_\varphi \sim N_p(\beta, (X^* \Sigma \varphi X^*)^{-1})$$

Where

$$\Sigma \varphi = \text{var}[\varphi(\epsilon_2)] .$$

In the next chapter we develop this asymptotic theory.
2.5 Generalized Estimating Equations

In this section we set up the generalized estimating equations (GEE) approach for the general model. These estimates require an estimate of variance covariance. Arnold’s transformation gives us such a robust estimate for the compound symmetry variance covariance structure. We finish by looking at the one way submodel.

2.5.1 General Model

Consider the linear model,

\[ y_{kj} = \alpha + x_{kj}^T \beta + e_{kj} \quad \text{for} \quad k = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n_k. \]

Where \( e_{kj} \) has pdf \( f(t) \), \( E(e_{kj}) = 0 \), and \( \text{var}(e_{kj}) = \sigma^2 \). The random variables \( y_{k1} \ldots y_{kn_k} \) are related or possibly correlated. In a longitudinal study, these \( n_k \) random variables represent repeated measures on the \( k \)th subject. For each of the subjects \( (k = 1, \ldots, m) \), let \( \mathbf{y}_k = [y_{k1}, \ldots, y_{kn_k}]^T \) be an \( n_k \times 1 \) response vector and

\[
X_k = \begin{bmatrix}
    x_{k1}^T \\
    \vdots \\
    x_{kn_k}^T
\end{bmatrix}
\]

be an \( n_k \times p \) design matrix. So for each subject we have a regression model,

\[ \mathbf{y}_k = X_k \beta + e_k \quad \text{for} \quad k = 1, \ldots, m. \]
Let $V_k$ be an $n_k \times n_k$ working covariance matrix for the $k$th subject. That is $V_k$ may or may not be the covariance matrix of $y_k$. As we have not written this as a GLM extension, we do not need to specify a form of $V_k$.

We first write the GEE of Liang and Zeger (1986) as a least squares minimization problem. The least squares dispersion function is

$$D_{LS}(\beta) = \|y^\dagger - X\beta\|^2_{LS} = \sum_{k=1}^{m} (y_k^\dagger - X_k^\dagger \beta)^T (y_k^\dagger - X_k^\dagger \beta).$$

Where

$$y_k^\dagger = V^{-1/2} y_k$$

and

$$X_k^\dagger = V^{-1/2} X_k$$

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

Differentiating with respect to $\beta$ and equating to zero give the least squares GEE

$$\sum_{k=1}^{m} X_k^\dagger T (y_k^\dagger - X_k^\dagger T \beta) = \sum_{k=1}^{m} X_k^\dagger T V^{-1} (y_k - X_k \beta) = 0 \quad (2.1)$$

So that $\hat{\beta}_{LS}$ is a least squares GEE solution to (2.1).

Define the rank dispersion function as

$$D_R(\beta) = \|y^\dagger - X^\dagger \beta\|^2_\varphi = \sum_{k=1}^{m} \sum_{i=1}^{n_k} \varphi \left( \frac{R(y_{ki}^\dagger - x_{ki}^\dagger T \beta)}{n + 1} \right) (y_{ki}^\dagger - x_{ki}^\dagger T \beta) \quad (2.2)$$

We now take a weighted least squares approach similar to Sievers and Abebe (2004). That is we write this rank dispersion function as a weighted least squares minimization problem, where the weights depend on $\hat{b}$, a consistent estimate of...
Rewrite (2.2) as the following weight least squares dispersion function

\[
D_R(\beta) = \sum_{k=1}^{m} \sum_{j=1}^{n_k} \varphi \left( \frac{R(y_{kj}^\dagger - x_{kj}^T \beta)}{n + 1} \right) (y_{kj}^\dagger - x_{kj}^T \beta)^2.
\]

Where \( m(\beta) = \text{med}(y_{kj}^\dagger - x_{kj}^T \beta) \) and

\[
w_i(\beta) = \begin{cases} 
\frac{a(R(y_i - x_i^T \beta - m(\beta)))}{y_i - x_i^T \beta - m(\beta)} & \text{if } y_i - x_i^T \beta \neq m(\beta) \\
0 & \text{otherwise.}
\end{cases}
\]

Next define a dispersion function similar to (2.3) such that the weights depend on \( \hat{b} \)

\[
D_R(\beta|\hat{b}) = \sum_{k=1}^{m} \sum_{j=1}^{n_k} w_{kj}(\hat{b})(y_{kj}^\dagger - x_{kj}^T \beta - m(\hat{b}))^2
\]

One possibility is \( \hat{b} = \hat{\beta}^{(k)} = \text{Argmin}D(\hat{\beta}^{(k)}|\hat{\beta}^{(k-1)}) \). That is an iteratively reweighted least squares estimate.

As with the least squares we differentiate (2.4) with respect to \( \beta \) to obtain rank-weighted GEE

\[
\sum_{k=1}^{m} X_k^T W_k(\hat{b})(y_k^\dagger - X_k \beta - m(\hat{b})1) = 0
\]

or

\[
\sum_{k=1}^{m} X_k V_{k}^{-1/2} W_k(\hat{b})V_{k}^{-1/2}(y_k - X_k \beta - m(\hat{b})V_{k}^{1/2}1) = 0.
\]
So that $\beta_{RW}$ is a solution to (2.4).

2.5.2 Compound Symmetry

We now look at the one-way model under the compound symmetry assumption. Under this model $V_{k}^{-1/2} = \Gamma^T \Lambda^{-1/2} \Gamma$ for $k = 1, \ldots, m$. Recall, $\lambda_1 = \text{var}(z_{i1})$ for $i = 1, \ldots, m$ and $\lambda_2 = \text{var}(e_{2j})$ for $j = 1, \ldots, m(n-1)$ and $i = 1, \ldots, (n-1)$ so that $\Lambda^{-1/2} = \text{diag}[\lambda_1^{-1/2}, \lambda_2^{-1/2}, \ldots, \lambda_2^{-1/2}]$. We will use our estimates $\hat{\lambda}_1^{1/2} = \text{mad}(z_1)$ and $\hat{\lambda}_2^{1/2} = \text{mad}(e_2)$, where $e_2$ are the residuals from the fit in the next section.

We now write the model,

$$y_k^T = \hat{V}^{-1/2} y_k = \hat{V}^{-1/2} [\mu + e_k] = \hat{V}^{-1/2} \mu + e_k^T$$

To fit however, we need to center the design matrix, $X_c = \hat{V}^{-1/2} - P_1 \hat{V}^{-1/2}$ ($P_1$ is a projection matrix onto the space spanned by a column of ones). Thus we fit the model

$$y_k^T = X_c \mu + e_k^T = X_c \hat{E} \hat{\mu} + e_k^T = [1X_{1c}] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + e_k^T.$$

Observe that $X_{1c}$ is the centered last $(n-1)$ columns of $\hat{V}^{-1/2}$. The estimated mean vector is

$$\hat{\mu} = \hat{E} \hat{\beta} = \begin{bmatrix} 1 & 0^T \\ 1 & I \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \hat{\alpha} \\ \hat{\alpha} 1_{n-1} + \hat{\beta}_1 \end{bmatrix}$$

which because, $E(W^T W)^{-1} E^T = (X^T X)^{-1} = \frac{1}{m} \hat{V}$, has an asymptopic $N_n(\mu, \tau^2 \frac{1}{m} \hat{V})$
distribution.

Before we write the test statistic, notice that the test is invariant to intercept,

\[
K\hat{\mu} = [-1_{n-1}I_{n-1}] \begin{bmatrix} \hat{\alpha} \\ \hat{\alpha}1_{n-1} + \hat{\beta}1_{n-1} \end{bmatrix} = -\hat{\alpha}1_{n-1} + \hat{\alpha}1_{n-1} + \hat{\beta} = \hat{\beta}.
\]

The Wald type test for no time effect, \( H_0 : K\mu = 0 \) (\( K \) is a contrast matrix), is

\[
T_3 = \frac{(K\hat{\mu})^T(X_{1c}^TX_{1c})(K\hat{\mu})}{\frac{\hat{\beta}^T(X_{1c}^TX_{1c})\hat{\beta}}{(n-1)}}\frac{(n-1)}{\hat{\beta}^2}
\]

which has an asymptotic \( F_{(n-1, mn-n)} \) distribution.
CHAPTER III

ASYMPTOTIC THEORY

In this chapter we develop asymptotic theory for rank-based estimates for the general repeated measures linear models under exchangeable errors. We take an approach similar to the one discussed in the monograph by Hettmansperger and McKean (1998) for the linear model. That is we first obtain the asymptotic distribution of the gradient, which is due to a result of Thompson (1990). Then to obtain the asymptotic distribution of the regression estimates we obtain contiguity and linearity results.

3.1 Model

The general repeated measures model we are considering is

\[ y_k = \alpha 1_n + X_k \beta + e_k \text{ for } k = 1, \ldots, m \]

where the within subject errors are exchangeable. In other words the distribution of any permutation of the errors is the same,

\[ \mathcal{L}(e_1, \ldots, e_n) = \mathcal{L}(e_{\alpha_1}, \ldots, e_{\alpha_n}) \]

where \( \alpha_1, \ldots, \alpha_n \) is a permutation of the integers 1, \ldots, n. From which we get that the errors have the same marginal distributions,

\[ \mathcal{L}(e_i) = \mathcal{L}(e_j). \]
Under this model \( \mathbf{y}_k = [y_{k1}, \ldots, y_{kn}]^T \) is an \( n \times 1 \) response vector and

\[
X_k = \begin{bmatrix}
    x_{k1}^T \\
    \vdots \\
    x_{kn}^T
\end{bmatrix}
\]

is an \( n \times p \) design matrix. So that \( \beta \) is a \( p \times 1 \) vector of slope parameters, \( \alpha \) is an intercept parameter, \( \mathbf{1}_n \) is an \( n \times 1 \) vector of ones.

Now stack everything,

\[
y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \quad \text{and } e = \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix}.
\]

So that \( \mathbf{y} \) is an \( mn \times 1 \) response vector, \( X \) is an \( mn \times p \) design matrix, \( e \) is an \( mn \times 1 \) vector of errors. Hence we have the linear model,

\[
y = \alpha \mathbf{1}_{mn} + X\beta + e.
\]

Without loss of generality assume \( X \) is centered and the true vector of slope parameters is \( 0 \). Since we fit the intercept separately, there is also no loss in generality in assuming \( \alpha = 0 \). We consider local alternatives such that \( \sqrt{n}\beta = O(1) \).

As in the usual regression problem (ch.1), define the rank estimate as the value which minimizes a rank dispersion function,

\[
\hat{\beta}_\varphi = \text{Argmin} \| \mathbf{y} - X\beta \|_\varphi.
\]

32
Similarly, define the gradient as

\[ S(\beta) = X^T a(R(y - X\beta)) = \sum_{k=1}^{m} X_k^T a(R(y_k - X_k\beta)). \]  (3.2)

Following Hettmansperger and McKean (1998, A.3) let,

\[ \Delta = (X^TX)^{1/2} \beta \]
\[ C = X(X^TX)^{-1/2}. \]

So that without loss of generality we are modeling,

\[ y_{kj} = x_{kj}^T \beta + \varepsilon_{kj} = c_{kj}^T \Delta + \varepsilon_{kj} \text{ for } j = 1, \ldots, n \text{ and } k = 1, \ldots, m. \]

Notice,

\[ \Delta = O(1). \]  (3.3)

Let \( d_{kj} = -c_{kj}^T \Delta \). Note that the design conditions (A.3.4)-(A.3.6) on page 410 of Hettmansperger and McKean (1998) hold.

That is,

\[ d_{..} = \frac{1}{mn} \sum_{k=1}^{m} \sum_{j=1}^{n} d_{kj} \]
\[ = -\frac{1}{mn} \sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{i=1}^{p} c_{kji} \Delta_i \]
\[ = -\frac{1}{mn} \sum_{i=1}^{p} \Delta_i \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji} \]
\[ = 0, \]  (3.4)
\[
\sum_{k=1}^{m} \sum_{j=1}^{n} d_{kj}^2 = \sum_{k=1}^{m} \sum_{j=1}^{n} (\Delta^T c_{kj})^2 \\
\leq \sum_{k=1}^{m} \sum_{j=1}^{n} \|\Delta\|^2 \|c_{kj}\|^2 \\
= \|\Delta\|^2 \sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{i=1}^{p} c_{kji} \\
= p\|\Delta\|^2, \quad (3.5)
\]

and

\[
\max_i d_i^2 \leq \|\Delta\|^2 \max_{kj} \|c_{kj}\|^2 = \|\Delta\|^2 \max_{kj} h_{kj}^2 \to 0 \text{ as } n \to \infty. \quad (3.6)
\]

The last condition is due to the fact that \( \|\Delta\| \) is bounded and that \( \|c_{kj}\| = h_{kj} \) the leverage value of \( H_X \) the hat matrix for \( X \). Finally we invoke Noether's condition (N.1) see section 1.1.3.

The following shift theorem is similar Theorem 1.3.1 of Hettmansperger and McKean (1998) and will be useful in developing the asymptotic theory in the next sections.

**Theorem 3.1.1.** Suppose we can write \( S(\Delta) = S(y - \Delta) \) then \( E_{\Delta}[S(0)] = E_{0}[S(-\Delta)]. \)

**Proof.** First note that if \( y \) has location \( \Delta \) and \( e \) has location 0. That is \( L(y) = L(e + \Delta) \) and \( F_\Delta(t) = F_0(t - \Delta). \) Now, notice that,

\[
E_{\Delta}[S(0)] = E_{\Delta}[S(y - 0)] = \int S(y - 0) dF_\Delta(y) = \int S(y - 0) dF_0(y - \Delta).
\]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Finally let $e = y - \Delta$ which gives the result,

$$E_\Delta[S(0)] = \int S(e - (-\Delta))dF_\Delta(e) = E_\Delta[S(e - (-\Delta))] = E_\Delta[S(-\Delta)].$$

\[\square\]

3.2 Contiguity

In this section, we consider the model, (as defined in expression (3.1)),

$$y_{kj} = x_{kj}^T \beta + e_{kj} = c_{kj}^T \Delta + e_{kj}, \quad (3.7)$$

for $j = 1, \ldots, n$; $k = 1, \ldots, m$. As in Section 3.1, let $d_{kj} = c_{kj}^T \Delta$. Although $d_{kj}$ depends on $m$, we suppress the $m$ in its subscript. Let $d_k = (d_{k1}, \ldots, d_{kn})^T$. For Model (3.7), recall that the conditions (3.4) - (3.6) are true. Let $e_k = (e_{k1}, \ldots, e_{kn})^T$ denote the vector of errors for the $k$th subject. Recall that $e_1, \ldots, e_m$ are iid random vectors. Let $g(t) = g(t_1, \ldots, t_n)$ denote the common pdf. In this section, other than the regularity conditions given next, we make no further assumptions about $g(t)$. In particular, we will not assume that the error distribution is exchangeable.

We now list the regularity conditions we require for $g(t)$. These are the usual maximum likelihood conditions for a location problem. Let $g(t - \theta) = g(t_1 - \theta_1, \ldots, t_n - \theta_n)$, for $\theta \in \mathbb{R}^n$. The three assumptions are:

A. All third partials of $g(t - \theta)$ with respect to $\theta$ exist.

B. The operations of partial derivatives and integration can be interchanged and the expectations of all partials of order less than or equal to three exist.
For $i, j = 1, \ldots, n$, define

$$I_{g,ij} = E \left[ \frac{\partial \log g(e)}{\partial \theta_i} \frac{\partial \log g(e)}{\partial \theta_j} \right].$$

(3.8)

Define the $n \times n$ matrix $I_g = [I_{g,ij}]$ and assume that it is positive definite. This is the (location) information matrix. Notice because the support of $g(t)$ is $R^n$, the information is free of $\theta$.

C. Suppose there exists functions $M_{ijl}$ such that

$$\left| \frac{\partial^3 \log g(e)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right| \leq M_{ijl}(e), \quad \text{for all } \theta \in R^n,$$

(3.9)

where $E[M_{ijl}(e)] < \infty$, for all $i, j, l = 1, \ldots, n$.

In this section, we will call these assumptions the regularity conditions.

We call Model (3.7) with $\Delta = 0$ the null model. The likelihood function for this model is given by

$$p_m = \prod_{k=1}^{m} g(e_k).$$

(3.10)

We refer to the model when $\Delta \neq 0$ as the alternative model. The likelihood function for this model is given by

$$q_m = \prod_{k=1}^{m} g(e_k - d_k),$$

(3.11)

where we have used the abbreviation $g(e_k - d_k) = g(e_{k1} - d_{k1}, \ldots, e_{kn} - d_{kn})$.

In this section, we establish that the sequence of alternative $\{q_m\}$ is contiguous to $\{p_m\}$. We basically follow the discussion in A.3 of Hettmansperger and McKean (1998), who established this result for the univariate linear model with iid errors. In turn, their discussion made use of Chapter 6 in the 1967 monograph
by Hájek and Šidák.

For \( \eta = (\eta_1, \ldots, \eta_n)^T \in \mathbb{R}^n \), define the function

\[
l(\eta) = \sum_{k=1}^{m} \log \frac{g(e_{k1} - \eta_1d_{k1}, \ldots, e_{kn} - \eta_nd_{kn})}{g(e_k)}.
\] (3.12)

This is of course the log of the ratio of the likelihoods when \( \eta = 1_n \), where \( 1_n \) denotes the vector of \( n \) ones, i.e., \((1, \ldots, 1)^T\). As above, we will use the abbreviation \( g(e_k - \eta d_k) = g(e_{k1} - \eta_1d_{k1}, \ldots, e_{kn} - \eta_nd_{kn}) \). Recall, (3.3), that \( \Delta = O(1) \) and that \( \max_k \|d_k\| \to 0 \), as \( m \to \infty \). Hence, this is a sequence of local alternatives.

Expand each of the \( m \)-terms in (3.12) in a Taylor series of length three plus a remainder about 0 and then evaluate the expansion at \( \eta = 1_n \). Denote the expansion as

\[
l = l(1) = \sum_{k=1}^{m} [H_{k,1} + H_{k,2} + H_{k,3} + R_k] = H_{,1} + H_{,2} + H_{,3} + R.
\] (3.13)

For each \( k \), the constant term \( H_{k,1} = 0 \); hence, \( H_{,1} = 0 \).

We will often drop the subscript \( k \) on the random vector \( e_k \). For the linear term, denote the \( j \)th partial derivative of \( g \) by

\[
g^{(j)} = \frac{\partial g}{\partial \eta_j}.
\] (3.14)

We need the vector of first partials of \( l(\eta) \). For the \( k \)th term of \( l(\eta) \), the \( j \)th term of this vector is given by

\[
\frac{\partial}{\partial \eta_j} l(\eta)_k = \frac{\partial \log g(e - \eta d)}{\partial \eta_j} = -d_{kj} g^{(j)}(e - \eta d) \cdot \frac{g(e - \eta d)}{g(e)},
\] (3.15)
In the expansion this last expression is evaluated at \( \eta = 0 \). The expansion itself is then the inner product of this vector and the vector \( 1 - 0 \). That is,

\[
H_{k,2} = \sum_{j=1}^{n} (-d_{kj}) \frac{g^{(j)}(e_k)}{g(e_k)}.
\] (3.16)

For the quadratic term, we need the \( n \times n \) matrix of mixed second partials for \( i, j = 1, \ldots, n \). Denote this matrix by \( H \). From the expression of first partials, (3.15), for \( i \neq j \), we have

\[
H_{ij}(e - \eta d) = \frac{\partial}{\partial \eta_i} \left[ \frac{\partial}{\partial \eta_j} - l(\eta) \right] = -d_{kj} \left[ g(e - \eta d)g^{(i,j)}(e - \eta d)(-d_{ki}) - g^{(j)}(e - \eta d)(-d_{ki})g^{(i)}(e - \eta d) \right] \frac{g^2(e - \eta d)}{g^2(e - \eta d)}
\]

The term that goes into the expansion is this last expression evaluated at \( \eta = 0 \); that is,

\[
H_{ij} = H_{ij}(e) = d_{ki}d_{kj} \left[ \frac{g^{(i,j)}(e)}{g(e)} - \left( \frac{\partial \log g(e)}{\partial \eta_i} \right) \left( \frac{\partial \log g(e)}{\partial \eta_j} \right) \right].
\] (3.17)

In the same way, it can be shown that this last expression for the case \( i = j \) is

\[
H_{ii} = H_{ii}(e) = d_{ki}^2 \left[ \frac{g^{(i,i)}(e)}{g(e)} - \left( \frac{\partial \log g(e)}{\partial \eta_i} \right)^2 \right].
\] (3.18)

Therefore, the quadratic term is

\[
H_{k,3} = \frac{1}{2} 1^T H 1.
\] (3.19)
We will discuss the remainder term later.

We next obtain the asymptotic properties of the linear and quadratic terms which require their expectations and variances. As assumed in the regularity conditions, we can interchange the order of differentiation and expectation. It follows that the expectation of linear term is

$$E(H_{k,2}) = \sum_{k=1}^{m} \sum_{j=1}^{n} (-d_{kj}) E \left( \frac{\partial \log g(e_k)}{\partial \theta_j} \right) = 0. \tag{3.20}$$

While the variance is

$$\text{Var}(H_{k,2}) = E \left[ \sum_{j=1}^{n} (-d_{kj}) \frac{\partial \log g(e_k)}{\partial \theta_j} \sum_{s=1}^{n} (-d_{ks}) \frac{\partial \log g(e_k)}{\partial \theta_s} \right]$$

$$= \sum_{j=1}^{n} \sum_{s=1}^{n} d_{kj}d_{ks} E \left[ \frac{\partial \log g(e_k) \partial \log g(e_k)}{\partial \theta_j \partial \theta_s} \right]$$

$$= \sum_{j=1}^{n} \sum_{s=1}^{n} d_{kj}d_{ks} I_{g,js} = d_k^T I_g d_k,$$

where $I_{g,js}$ is the location information defined in (3.8). Hence, by independence

$$\text{Var}(H_{k,2}) = \sum_{k=1}^{m} d_k^T I_g d_k. \tag{3.21}$$

For the mean of the quadratic term, the expectation of the first term of (3.17) is given by

$$E \left[ d_{ki}d_{kj} \frac{g^{(k)}(e)}{g(e)} \right] = d_{ki}d_{kj} \cdot 0 = 0.$$  

Further, the mean of the second term is

$$-E \left[ d_{ki}d_{kj} \left( \frac{\partial \log g(e)}{\eta_i} \right) \left( \frac{\partial \log g(e)}{\eta_j} \right) \right] = -d_{ki}d_{kj} I_{g,ij}.$$
Likewise, the expectation of (3.18) is $-d_{kj}^2 I_{jj}$. Therefore,

$$E [H_{ij}] = -\frac{1}{2} \sum_{k=1}^{m} d_k^T I_g d_k.$$  \hspace{1cm} (3.22)

Our first theorem obtains the asymptotic distribution of $H_{ij}$. The next lemma helps with the proof.

**Lemma 3.2.1.** Assume the regularity conditions and the design conditions hold. Let

$$B^2 = \lim_{m \to \infty} \sum_{k=1}^{m} d_k^T I_g d_k.$$ \hspace{1cm} (3.23)

Then the limit exists and $0 < B^2 < \infty$.

**Proof.** Let

$$B_m^2 = \sum_{k=1}^{m} d_k^T I_g d_k.$$ \hspace{1cm} (3.24)

Denote the spectral decomposition of the matrix $I_g$ by

$$I_g = \Gamma^T \Lambda \Gamma = \sum_{i=1}^{n} \lambda_i v_i v_i^T,$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ are the eigenvalues of $I_g$ and $v_1, v_2, \ldots, v_n$ are the
corresponding eigenvectors. Then

\[
B_m^2 = \sum_{k=1}^{m} \sum_{i=1}^{n} \lambda_i |d_k^T v_i|^2 \\
\leq \sum_{k=1}^{m} \sum_{i=1}^{n} \lambda_i \|d_k\|^2 \\
\leq n\lambda_1 \sum_{k=1}^{m} \|d_k\|^2 \\
= n\lambda_1 \sum_{k=1}^{m} \sum_{j=1}^{n} d_{kj}^2 \\
\leq n\lambda_1 p \|\Delta\|^2,
\]

where the first inequality follows from the Cauchy-Schwartz inequality and the last inequality follows from condition (3.5). Note that since \(I_g\) is positive definite, \(B_m > 0\). Further, by (3.3), \(\|\Delta\| = O(1)\). Thus \(B_m^2\) is a positive increasing series which is bounded above; hence, the limit exists and satisfies \(0 < B < \infty\). \(\Box\)

This leads to the first of our main results.

**Theorem 3.2.1.** Assume the regularity conditions and the design conditions hold. Then

\[
H_{1,2} \overset{\mathcal{D}}{\rightarrow} N(0, B^2),
\]

as \(m \rightarrow \infty\).

**Proof.** As above,

\[
H_{1,2} = \sum_{k=1}^{m} H_{k,2}.
\]

From (3.20) and (3.21), \(E(H_{k,2}) = 0\) and \(\text{Var}(H_{k,2}) = d_k^T I_g d_k\). However, while \(H_{1,2}, \ldots, H_{m,2}\) are independent they are not identically distributed. Hence, to obtain the result we will use the version of the Lindeberg-Feller Central Limit
Theorem which is found on page 397 of Hettmansperger and McKean (1998). By Lemma 3.2.1, \( \text{Var}(H_{k,2}) = B_m^2 \to B^2 \), where \( 0 < B^2 < \infty \). Further, using the notation of the lemma and condition (3.6), we have

\[
\max_{1 \leq k \leq m} \text{Var}(H_{k,2}) = \max_{1 \leq k \leq m} \sum_{i_1}^{n} \lambda_i d_k^T I_g d_k \\
\leq n\lambda_1 \|d_k\|^2 \to 0.
\]

(3.26)

Let \( \epsilon > 0 \) be given. To satisfy the Lindeberg-Feller conditions and hence obtain the result, we need to show that

\[
\lim_{m \to \infty} \sum_{k=1}^{m} E \left[ H_{k,2}^2 I_\epsilon(H_{k,2}) \right] = 0,
\]

(3.27)

where \( I_\epsilon(x) = 1 \) or 0 if and only if \( |x| > \epsilon \) or \( |x| \leq \epsilon \), respectively. Recall that

\[
H_{k,2} = \sum_{j=1}^{n} (-d_{k,j}) \frac{g^{(j)}(e_k)}{g(e_k)}.
\]

Hence,

\[
|H_{k,2}| \leq \sum_{j=1}^{n} |d_{j,k}| \left| \frac{g^{(j)}(e_k)}{g(e_k)} \right| \\
\leq \max_{1 \leq k \leq m, 1 \leq j \leq n} |d_{kj}| \left\{ \sum_{j=1}^{n} \left| \frac{g^{(j)}(e_k)}{g(e_k)} \right| \right\}.
\]

\[
= \text{def} \ d_m^* \left\{ \sum_{j=1}^{n} \left| \frac{g^{(j)}(e_k)}{g(e_k)} \right| \right\}.
\]

Note that the random variables within braces are identically distributed for \( k = 1, 2, \ldots, m \) and by our regularity conditions it is integrable. Call the random
variable in braces \( g_{m,1}^*(e_k) \). So,

\[
I_c(|H_{k,2}|) \leq I_c(d_m^* g_{m,1}^*(e_k)), \quad k = 1, 2, \ldots, m. \tag{3.28}
\]

Note that \( d_m^* \to 0 \) as \( m \to \infty \). Next we need a bound on \( H_{k,2}^2 \) given by,

\[
H_{k,2}^2 = \sum_{j=1}^{n} \sum_{j'=1}^{n} d_{k,j} d_{k,j'}^2 \frac{g^{(j)}(e_k) g^{(j')}(e_k)}{g(e_k) g(e_k)} \\
\leq \left\{ \max_{1 \leq j \leq n, \ 1 \leq j' \leq n} \left| \frac{g^{(j)}(e_k) g^{(j')}(e_k)}{g(e_k) g(e_k)} \right| \right\} \left\{ \sum_{j=1}^{n} \sum_{j'=1}^{n} |d_{k,j}||d_{k,j'}| \right\}.
\]

Call the random variable in the first set of braces \( g_{m,2}^*(e_k) \). Note that it is identically distributed for \( k = 1, \ldots, m \) and that by the regularity conditions it is integrable. For the quantity in the second set of braces we have

\[
\sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{j'=1}^{n} |d_{k,j}||d_{k,j'}| \leq \sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{j'=1}^{n} [d_{k,j}^2 + d_{k,j'}^2] \\
= 2n \sum_{k=1}^{m} \sum_{j=1}^{n} d_{k,j}^2 = 2np\|\Delta\|^2. \tag{3.29}
\]

Thus,

\[
\lim_{m \to \infty} \sum_{k=1}^{m} E[H_{k,2}^2 I_c(|H_{1,2}|)] \leq 2np\|\Delta\|^2 \lim_{m \to \infty} E \left[ g_{m,2}^*(e_1) I_c(d_m^* g_{m,1}^*(e_1)) \right]. \tag{3.30}
\]

Because \( g_{m,2}^*(e_1) I_c(d_m^* g_{m,1}^*(e_1)) \leq g_{m,1}^*(e_1) \), \( g_{m,1}^*(e_1) \) is integrable, and \( I_c(d_m^* g_{m,1}^*(e_1)) \to 0 \) as \( m \to \infty \), an application of the Lebesgue Dominated Convergence Theorem gives the result.

For the quadratic term, we have the result

**Theorem 3.2.2.** Assume the regularity conditions and the design conditions hold.
Then
\[ H_{,3} \xrightarrow{P} -\frac{1}{2} B^2, \quad \text{as } m \to \infty, \] (3.31)

where \( B^2 \) is given by (3.23).

Proof. Write \( H_{,3} \) as
\[
H_{,3} = \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} \sum_{k=1}^{m} d_{kj} d_{ki} \left[ \frac{g(e_k)g^{(ij)}(e_k)}{g^2(e_k)} - \frac{g^{(i)}(e_k)g^{(j)}(e_k)}{g^2(e_k)} \right].
\]

For fixed \( i \) and \( j \), define \( W_{mij}^{(1)} \) and \( W_{mij}^{(2)} \) as
\[
W_{mij}^{(1)} = \sum_{k=1}^{m} d_{kj} d_{ki} \frac{g(e_k)g^{(ij)}(e_k)}{g^2(e_k)}
\]
\[
W_{mij}^{(2)} = \sum_{k=1}^{m} d_{kj} d_{ki} \frac{g^{(i)}(e_k)g^{(j)}(e_k)}{g^2(e_k)}
\]

The random variables in these expressions are identically distributed, for \( k = 1, \ldots, m \), and by the regularity conditions have finite variances.

Note that \( E \left( W_{mij}^{(1)} \right) = 0 \). Denote the common variance of \( g(e_k)g^{(ij)}(e_k)/g^2(e_k) \) by \( \sigma^{(1)} \). Then
\[
\text{Var} \left( W_{mij}^{(1)} \right) = \sum_{k=1}^{m} d_{kj}^2 d_{ki}^2 \sigma^{(1)}
\]
\[
\leq \max_{1 \leq k \leq m} d_{kt}^2 \sigma^{(1)} \sum_{k=1}^{m} d_{ki}^2
\]
\[
\leq \max_{1 \leq k \leq m} d_{ki}^2 \sigma^{(1)} p \| \Delta \|^2 \to 0,
\]
as \( m \to \infty \). Thus, by Chebyshev's Theorem,
\[
W_{mij}^{(1)} \xrightarrow{P} 0, \quad \text{as } m \to \infty,
\] (3.32)
for all $i$ and $j$.

For $W_{m ij}^{(2)}$, we have by (3.22)

$$E \left[ W_{m ij}^{(2)} \right] = \sum_{k=1}^{m} d_{kj} d_{ki} E \left[ W_{ij}^{(2)} \right] = \sum_{k=1}^{m} d_{kj} d_{ki} I_{g,ij}.$$ 

It follows in the same way as with the sequence of random variables $W_{m ij}^{(1)}$, that $\text{Var}(W_{m ij}^{(2)}) \to 0$, as $m \to \infty$. Hence, by Chebyshev's Theorem,

$$W_{m ij}^{(2)} - \sum_{k=1}^{m} d_{kj} d_{ki} I_{g,ij} \overset{p}{\to} 0, \quad \text{as } m \to \infty,$$  \hspace{1cm} (3.33)

for all $i$ and $j$. Because $n$ is fixed, by results (3.32) and (3.33), we have

$$H_s + \frac{1}{2} B_m^2 \overset{p}{\to} 0, \quad \text{as } m \to \infty,$$

The result follows because $B_m^2 \to B^2$. \hfill \Box

Finally, as the following theorem shows, the remainder term goes to zero in probability.

**Theorem 3.2.3.** Assume the regularity conditions and the design conditions hold. Then

$$R \overset{p}{\to} 0, \quad \text{as } m \to \infty.$$  \hspace{1cm} (3.34)

**Proof.** The remainder term is composed of a finite ($n$ is finite) number of mixed third partial derivatives. From an examination of the partial differentiation, around expression (3.17), each such term has as a multiplying factor a triple of the form $d_{ki} d_{kj} d_{kl}$. Two of these $d$s will suffice to make these terms $O(1)$ while the remaining one can be used to majorize the term as $\max_{jk} |d_{jk}|$ which

45

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
goes to zero as $m \to \infty$. Based on this and (3.9), the result follows.

Based on these results, we easily determine the asymptotic distribution of $l$, the log of the likelihood, which we state in this theorem. Likelihood.

**Theorem 3.2.4.** Assume the regularity conditions and the design conditions hold. Then

$$l \xrightarrow{D} N\left(-\frac{1}{2}B^2, B^2\right),$$

(3.35)

where $B^2$ is defined in expression (3.23).

The contiguity result follows directly from this theorem as shown in Hájek and Šidák (1967). We state it as a corollary.

**Corollary 3.2.1.** The densities $q_m$ are contiguous to $p_m$.

### 3.3 Asymptotic Theory of $S_{m,i}$

The asymptotic distribution of $S_{m,i}$ under the null hypothesis and under Pitman alternatives was obtained by Thompson (1990).

Following Thompson (1990) define,

$$S_{m,i} = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji}a(R(y_{kj}))$$

as the $i$th component of the gradient vector, $S_m = [S_{m,1}, \ldots, S_{m,n}]^T$. We only require the null model theory ($\Delta = 0$). So that $y_{kj} = e_{kj}$ for $k = 1, \ldots, m$ and $j = 1, \ldots, n$.

Let

$$z_{k,i} = \frac{1}{mn + 1} \sum_{k'=1}^{m} \sum_{j'=1}^{n} \sum_{j=1}^{n} (c_{k'j'i} - c_{kji}) \int [I(t \geq y_{kj}) - F_{k'j}(t)] \phi'(H(t)) dF_{k'j}(t)$$

46
be the projection where,

\[ H(t) = \frac{1}{mn} \sum_{k=1}^{m} \sum_{j=1}^{n} F_{kj}(t). \]

Note that \( z_{k,i} = z_{k,i}(y_{k1}, \ldots, y_{kn}) \). Hence \( z_{1,i}, \ldots, z_{m,i} \) are independent random variables having mean zero. Define

\[ \sigma_{m,i}^2 = \text{var} \left[ \sum_{k=1}^{m} z_{k,i} \right] \]

and

\[ \mu_{m,i} = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji} \int \phi(H(t)) dF_{kj}(t). \]

### 3.3.1 Null Models

Under the null hypothesis (\( \Delta = 0 \)) and the errors having the same marginal distributions, we have \( F_{kj}(t) = F(t) \) for \( k = 1, \ldots, m \) and \( j = 1, \ldots, n \). So that

\[ H(t) = \frac{1}{mn} \sum_{k=1}^{m} \sum_{j=1}^{n} F_{kj}(t) = F(t). \]

Then the projection becomes,

\[ z_{k,i} = \frac{1}{mn+1} \sum_{k'=1}^{m} \sum_{j'=1}^{n} \sum_{j=1}^{n} (c_{k'j'i} - c_{kji}) \int [I(t \geq y_{kj}) - F(t)] \phi'(F(t)) dF(t). \]

\[ = -\frac{mn}{mn+1} \sum_{j=1}^{n} c_{kji} w_{kj} \]
Where

\[ w_{kj} = \int [I(t \geq y_{kj}) - F(t)] \varphi'(F(t)) \, dF(t) \]
\[ = \int_{y_{kj}}^{\infty} \varphi'(F(t)) \, dF(t) - \int_{-\infty}^{y_{kj}} F(t) \varphi'(F(t)) \, dF(t) \]
\[ = -\varphi(F(y_{kj})) \]

since,

\[ \int_{y_{kj}}^{\infty} \varphi'(F(t)) \, dF(t) = \varphi(F(t)) \bigg|_{y_{kj}}^{\infty} = \varphi(1) - \varphi(F(y_{kj})) \]

and, by using integration by parts,

\[ \int_{-\infty}^{\infty} F(t) \varphi'(F(t)) \, dF(t) = \varphi(F(t))F(t) \bigg|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \varphi(F(t)) \, dF(t) \]
\[ = \varphi(1) \cdot 1 - \varphi(0) \cdot 0 - 0 \]
\[ = \varphi(1) \]

(note that the last term is \( \int \varphi(F(t)) \, dF(t) = E[\varphi(u)] = 0 \) \((u \sim U(0,1))\)). Hence, under the null model, the projection comes down to

\[ z_{k,i} = \frac{mn}{mn + 1} \sum_{j=1}^{n} c_{kji} \varphi(F(y_{kj})) \]

which we may write in vector notation

\[ z_{k,i} = \frac{mn}{mn + 1} c^{(i)T} \varphi(F(y_{k})) \]
where \( \varphi(F(y_k)) = [\varphi(F(y_{k1}), \ldots, \varphi(F(y_{kn}))]^T \) and \( c_k^{(i)} = [c_{k1i}, \ldots, c_{kni}]^T \) is the \( i \)-th column of \( C_k \). The null variance of \( z_{ki} \) is

\[
\text{var}_0[z_{ki}] = \left( \frac{mn}{mn + 1} \right)^2 c_k^{(i)T} \text{var}_0[\varphi(F(y_k))]c_k^{(i)} = c_k^{(i)} \Sigma_\varphi c_k^{(i)}
\]

and

\[
\sigma_{m,i}^2(0) = \sum_{k=1}^{m} \text{var}_0[z_{ki}]
\]

\[
= \left( \frac{mn}{mn + 1} \right)^2 \sum_{k=1}^{m} c_k^{(i)T} \text{var}_0[\varphi(F(e_k))]c_k^{(i)}
\]

\[
= \left( \frac{mn}{mn + 1} \right)^2 \sum_{k=1}^{m} c_k^{(i)T} \Sigma_\varphi c_k^{(i)}.
\]

Where \( \Sigma_\varphi = \text{var}_0[\varphi(F(e_1))] \).

Further

\[
E[S_{m,i}] = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji} E[a(R(y_{kj}))]
\]

\[
= \sum_{k=1}^{m} \sum_{j=1}^{n} \frac{1}{mn} \left[ \sum_{i=1}^{mn} a(i) \right]
\]

\[= 0\]
and

\[
\mu_{m,i} = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kj} \int \varphi(H(t))dF_{kj}(t)
\]

\[
= \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kj} \int \varphi(F(t))dF(t)
\]

\[
= \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kj} \cdot 0
\]

\[
= 0.
\]

Thompson’s (1990) theorem 2.3 gives us the asymptotic normality,

\[
\frac{S_{m,i} - E[S_{m,i}]}{\sqrt{\text{var}[S_{m,i}]}} \rightarrow N(0,1)
\]

and

\[
\frac{S_{m,i} - \mu_{m,i}}{\sqrt{\sigma_{m,i}^2}} \rightarrow N(0,1).
\]

### 3.4 Linearity

In this section we obtain the asymptotic linearity of \( S_m \). We do so in a manner similar to that of Hettmansperger and McKean (1998) for the linear model (sections A.2-A.3). That is we first obtain a linearity result for \( T_{m,i} \), an approximation to \( S_{m,i} \). Next we show that \( S_{m,i} \) and \( T_{m,i} \) are asymptotically equivalent. Finally we use convexity to obtain the linearity result for the vector \( S_m \).

Define

\[
T_{m,i} = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kj}(\varphi(F(y_{kj}))).
\]
Now consider the approximation,

\[ S_{m,i} = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kj} a(R(y_{kj})) \]

\[ = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kj} \varphi \left( \frac{mn}{mn+1} F_n(y_{kj}) \right) \]

\[ = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kj} \varphi(F(y_{kj})). \]

where

\[ F_n(y_{kj}) = \sum_{k'=1}^{m} \sum_{j'=1}^{n} I(y_{kj} \geq y_{k'j'}) \]

is the empirical CDF of \( y_{kj} \).

Under the null hypothesis the mean of \( T_{m,i} \) is

\[ E[T_{m,i}] = E \left[ \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kj} \varphi(F(e_{kj})) \right] \]

\[ = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kj} E[\varphi(F(e_{kj}))] \]

\[ = 0 \]
and the variance is

$$\text{var}[T_{m,i}] = \text{var} \left[ \frac{mn + 1}{mn} \sum_{k=1}^{m} z_{k,i} \right]$$

$$= \left( \frac{mn + 1}{mn} \right)^2 \sum_{k=1}^{m} \text{var}[z_{k,i}]$$

$$= \left( \frac{mn + 1}{mn} \right)^2 \sum_{k=1}^{m} \left( \frac{mn}{mn + 1} \right)^2 c_k^{(i)T} \Sigma \varphi c_k^{(i)}$$

$$= \sum_{k=1}^{m} c_k^{(i)T} \Sigma \varphi c_k^{(i)} .$$

The next theorem is an extension of theorem A.2.5 of Hettmansperger and McKean (1998).

**Theorem 3.4.1.** Under $p_m$

$$T_{m,i}(0) - (T_{m,i}(\Delta) - E[T_{m,i}(\Delta)]) \xrightarrow{P} 0 .$$

**Proof.** Since $E[T_{m,i}(0)] = 0$ we only need to show $\text{var}[T_{m,i}(0) - T_{m,i}(\Delta)] \to 0 .

$$\text{var}[T_{m,i}(0) - T_{m,i}(\Delta)] = \text{var} \left[ \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kj} \varphi(F(e_{kj})) - \varphi(F(e_{kj} + d_{kj})) \right]$$

$$= \sum_{k=1}^{m} \text{var} \left[ \sum_{j=1}^{n} c_{kj} \varphi(F(e_{kj})) - \varphi(F(e_{kj} + d_{kj})) \right]$$

$$= \sum_{k=1}^{m} \text{var}[c_k^{(i)T} [\varphi(F(e_k)) - \varphi(F(e_k + d_k))] c_k^{(i)}]$$

$$= \sum_{k=1}^{m} c_k^{(i)T} \text{var}[\varphi(F(e_k)) - \varphi(F(e_k + d_k))] c_k^{(i)}$$

$$= \sum_{k=1}^{m} c_k^{(i)T} \Sigma_d c_k^{(i)}$$

52

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Where $\Sigma_d = \text{var}[\varphi(F(e_k) - \varphi(F(e_k + d_k)))]$ for $k = 1 \ldots m$. That is $\Sigma_d$ is the common variance covariance matrix for the difference between the score vectors. So that $\sigma^2_{d,j} = \text{var}[\varphi(F(e_{kj})) - \varphi(F(e_{kj} + d_{kj}))]$ is the jth diagonal element of $\Sigma_d$.

Let

$$\Sigma_d = \sum_{j=1}^{n} \lambda_j v_j v_j^T$$

be an eigenvalue decomposition of the matrix $\Sigma_d$. Since $v_1, \ldots, v_n$ is an orthonormal basis for $\mathbb{R}^n$ we can write $c_{k}^{(i)} = \sum_{j=1}^{n} a_{ki,j} v_j$ for some $a_1, \ldots a_n$. So that

$$\Sigma_d^{1/2} c_{k}^{(i)T} = \sum_{j=1}^{n} a_j \lambda_j v_j$$

and

$$c_{k}^{(i)} \Sigma_d c_{k}^{(i)} = \|\Sigma_d^{1/2} c_{k}^{(i)}\|^2$$

$$= \sum_{j=1}^{n} \lambda_j^2 a_j^2 v_j^T v_j$$

$$\leq \lambda_1^2 \sum_{j=1}^{n} a_j^2 v_j^T v_j$$

$$= \lambda_1^2 \|c_{k}^{(i)}\|^2.$$

The sum becomes

$$\sum_{k=1}^{m} c_{k}^{(i)T} \Sigma_d c_{k}^{(i)} = \lambda_1^2 \sum_{k=1}^{n} \|c_{k}^{(i)}\|^2 = \lambda_1^2 \cdot 1.$$ 

Since,

$$\lambda_1 \leq \text{trace}(\Sigma_d) \leq \sum_{j=1}^{n} \sigma_j^2 \leq n \cdot \max \sigma_j^2.$$ 

53
we only need to show that each of the diagonal elements of $\Sigma_d$ are arbitrarily small. Hettmansperger and McKean (1998, p.405) gives us the following result,

$$\sigma_j^2 = \text{var}[\varphi(F(e_{kj})) - \varphi(F(e_{kj} + d_{kj}))] \leq E[(\varphi(F(e_{kj})) - \varphi(F(e_{kj} + d_{kj}))^2] \leq \varepsilon.$$

Define

$$\gamma_f = \int_0^1 \varphi(u) \left( -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \right) du \quad (3.36)$$

and

$$\sigma_{cd,i} = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji}d_{kj} \quad (3.37)$$

Theorem 3.4.2. Under $p_m$

$$E_0[T_{m,i}(\Delta)] \rightarrow \gamma_f \sigma_{cd,i}$$

Proof. Recall that $\int \varphi(F(t))f(t)dt = 0$ so that

$$E_0[T_{m,i}(\Delta)] = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji}E_0[\varphi(F(e_{kj} + d_{kj}})]$$

$$= \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji} \int \varphi(F(t + d_{kj}))f(t)dt$$

$$= \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji} \int [\varphi(F(t + d_{kj})) - \varphi(F(t))]f(t)dt$$

$$= \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji} \int \varphi'(F(t))f(t)d_{kj}f(t)dt + o_p(1)$$

by the mean value theorem. Using integration by parts we get the following
representation,

\[ \int \varphi'(F(t)) f(t)^2 dt = - \int \varphi(F(t)) f'(t) dt. \]

Using substitution we get the result,

\[
- \int \varphi(F(t)) f'(t) dt = - \int \varphi(F(t)) \frac{f'(t)}{f(t)} f(t) dt \\
= \int \varphi(u) \left(- \frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \right) du \\
= \gamma_f
\]

\[ \square \]

The last two theorems gives us the linearity of \( T_{m,i} \)

\[ T_{m,i}(\Delta) = T_{m,i}(0) + \gamma_f \sigma_{cd} + o_p(1). \]

The following theorem is similar to theorem A.2.1 of Hettmansperger and McKean (1998). Since \( T_{m,i} \) is asymptotically equivalent to \( \sum_{k=1}^m Z_{m,i} \), the result is from Thompson's theorem (2.2).

**Theorem 3.4.3.** Under \( p_m \)

\[ T_{m,i}(0) - S_{m,i}(0) \xrightarrow{p} 0. \]

The following theorem, similar to theorem A.2.7 of Hettmansperger and McKean (1998), gives us pointwise linearity of \( S_{m,i} \).
Theorem 3.4.4. Under $p_m$,

\[
S_{m,i}(\Delta) = S_{m,i}(0) + \gamma_f \sigma_{cd,i} + a_p(1)
\]

\[
S_{m,i}(\Delta) = T_{m,i}(0) + \gamma_f \sigma_{cd,i} + a_p(1).
\]

Proof. By theorem (3.4.3) we have

\[
T_{m,i}(0) - S_{m,i}(0) \xrightarrow{p} 0
\]

under $p_m$. Since $p_m$ is contiguous to $q_m$ (Corollary 3.2.1) we know that

\[
T_{m,i}(0) - S_{m,i}(0) \xrightarrow{p} 0
\]

under $q_m$. So that

\[
T_{m,i}(\Delta) - S_{m,i}(\Delta) \xrightarrow{p} 0
\]

under $p_m$. Which gives the second expression. Which in turn, along with theorem (3.4.3), gives the first expression. □

Note we can rewrite $\sigma_{cd,i}$ as

\[
\sigma_{cd,i} = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji} d_{kj} = \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji} (-c_{kj}^T \Delta)
\]

\[
= -\sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji} \sum_{i'=1}^{p} c_{kji'} \Delta_{i'}
\]

\[
= -\sum_{i'=1}^{p} \Delta_{i'} \sum_{k=1}^{m} \sum_{j=1}^{n} c_{kji'} c_{kji} = -\Delta_i.
\]
So that we have

\[ S_{m,i}(\Delta) = S_{m,i}(0) - \gamma_f \Delta_i + o_p(1) \]
\[ S_{m,i}(\Delta) = T_{m,i}(0) - \gamma_f \Delta_i + o_p(1). \]

Let \( S_m(\Delta) = [S_{m,1}(\Delta), \ldots, S_{m,n}(\Delta)]^T \). Since componentwise convergence in probability implies convergence in probability of the corresponding vector we have the following theorem.

**Theorem 3.4.5.**

\[
\lim_{m \to \infty} P(||S_m(\Delta) - (S_m(0) - \gamma_f \Delta)|| \geq \varepsilon) = 0.
\]

Define

\[ Q_m(\Delta) = \gamma_f \Delta^T \Delta / 2 - \Delta^T S_m(0) + D_m(0). \]

So that \( Q_m(\Delta) \) is a quadratic approximation to \( D_m(\Delta) \). Denote

\[ \tilde{\Delta} = \text{Argmin} Q(\Delta). \]

That is, by differentiation,

\[ \tilde{\Delta} = \frac{1}{\gamma_f} S_m(0). \]

From which we immediately get the asymptotic distribution of \( \tilde{\Delta} \),

\[ \tilde{\Delta} \to N_p \left( 0, \frac{1}{\gamma_f^2} \Sigma \right). \]
Where
\[ \Sigma = \lim_{m \to \infty} \sum_{k=1}^{m} C_k^T \Sigma \phi C_k. \]

Since,
\[ S_m(0) \xrightarrow{D} N_p(0, \Sigma \phi). \]

See also Brown (1984) for a general result. From theorem (3.4.5) above, asymptotic linearity and quadraticity follow from convexity by the usual diagonalization argument (see A.3.2 of Hettmansperger and McKean 1998). We state these results in the following theorem.

**Theorem 3.4.6.** Under regularity and design conditions

\[ \lim_{m \to \infty} P \left( \sup_{\|\Delta\| \leq c} \|S_m(\Delta) - (S_m(0) - \gamma \Delta)\| \geq \epsilon \right) = 0 \tag{3.38} \]
\[ \lim_{m \to \infty} P \left( \sup_{\|\Delta\| \leq c} |D_m(\Delta) - Q_m(\Delta)| \geq \epsilon \right) = 0 \tag{3.39} \]

Finally we may apply Jaeckel’s (1972) result to obtain the asymptotic distribution of \( \hat{\Delta} \). See Hjort and Pollard (1993) for a general result. We state this result in the next theorem.

**Theorem 3.4.7.** Under regularity and design conditions

\[ \hat{\Delta} - \tilde{\Delta} \xrightarrow{P} 0 \]

Which leads to the following theorem.

**Theorem 3.4.8.**

\[ \hat{\Delta} \xrightarrow{D} N_p \left( \Delta, \frac{1}{\gamma^2_f} \Sigma \right) \]
where

$$\Sigma = \lim_{m \to \infty} \sum_{k=1}^{m} C_k^T \Sigma \varphi C_k$$

### 3.5 Inference

In this section we briefly discuss inference based on the estimates discussed in this chapter. Since inference will undoubtably be on $\beta$ and not $\Delta$ we have

$$\hat{\beta} \sim N_p \left( \beta, \tau^2 (X^T X)^{-1} \left( \sum_{k=1}^{n} X_k^T \hat{\Sigma} \varphi X_k \right) (X^T X)^{-1} \right)$$

Let

$$A = \begin{bmatrix} a(R(\hat{e}_1))^T \\ \vdots \\ a(R(\hat{e}_m))^T \end{bmatrix}$$

be a matrix of scores and let $A_c = (I_m - \frac{1}{m} J_m) A$ be the centered $A$. The estimate of $\hat{\Sigma} \varphi$ we use is

$$\hat{\Sigma} \varphi = \frac{1}{m-p} A_c^T A_c.$$

As an example we consider the two way repeated measures problem of section 2.3. As with the GEE solution to the one way problem (section 2.6) these tests are invariant to intercept.

For convenience we repeat the model statement,

$$y = \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + e = W \mu + e.$$
Let
\[ E = \begin{bmatrix} 1 & 0_{2n-1}^T \\ 1_{2n-1} & I_{2n-1} \end{bmatrix} \]
so that
\[ W \mu = WEE^{-1} \mu = [1X] \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \]

Where 1 is an \( mn \times 1 \) vector of ones, \( \alpha \) is an intercept parameter, \( \beta \) is \((2n-1) \times 1\) vector of parameters, and \( X \) is the last \( 2n-1 \) columns of \( W \). Further,
\[ E^{-1} = \begin{bmatrix} 1 & 0_{2n-1}^T \\ -1_{2n-1} & I_{2n-1} \end{bmatrix} \]
so that
\[ \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = E^{-1} \mu = \begin{bmatrix} 1 & 0_{2n-1}^T \\ -1_{2n-1} & I_{2n-1} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{n2} \\ \mu_{21} \\ \vdots \\ \mu_{22} \end{bmatrix} = \begin{bmatrix} \mu_{11} \\ \mu_{12} - \mu_{11} \\ \vdots \\ \mu_{n2} - \mu_{11} \\ \mu_{21} - \mu_{11} \\ \vdots \\ \mu_{22} - \mu_{11} \end{bmatrix}. \]

We now write down the tests of parallelism and of main effects in terms of the regression estimate \( \hat{\beta} \).
First note that the test for parallelism is to test

\[ H_{01} : K_1 \mu = 0 \]

where

\[
K_1 = \begin{bmatrix}
1 & -1 & 0 & \cdots & 0 & -1 & 1 & 0 & \cdots & 0 \\
0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 1 & -1 & 0 & \cdots & 0 & -1 & 1
\end{bmatrix}
\]

is an \( n-1 \times 2n \) contrast matrix. That is we are testing if all the slopes of treatment one are the same as the slopes for treatment two. Given parallelism, a test for main effects is to test

\[ H_{02} : k_2^T \mu = 0 \]

where \( k_2^T = [1^T - 1^T] \). That is we are testing if the grand mean for treatment one is the same as the grand mean for treatment two.

Note that an equivalent test of parallelism is to use

\[ K_1 = [1_{n-1} - I_{2-1} - 1_{2-1} I_{2-1}] \]

so that we are testing

\[ H_{01} : K_1 \mu = [1_{n-1} - I_{n-1} - 1_{n-1} I_{n-1}] \begin{bmatrix} \alpha \\ \alpha 1 + \beta \end{bmatrix} = [-I_{n-1} - 1_{n-1} I_{n-1}] \beta = 0. \]

Define

\[ K = [-I_{2n-1} - 1_{2n-1} I_{2n-1}] . \]
Then an asymptotic $\alpha$-level test of parallelism based on $\beta$ is to reject $H_{01}$ if

$$
(K\hat{\beta})^T \left[ K(X^TX)^{-1} \left( \sum_{k=1}^{n} X_k^T \Sigma \phi X_k \right) (X^TX)^{-1} K^T \right]^{-1} (K\hat{\beta}) > F_{\alpha,n-1,mn-2n}.
$$

For a test of main effects based on $\hat{\beta}$

$$
H_{02} : k_2^T \mu = [1_n^T - 1_n^T] \begin{bmatrix} \alpha \\ \alpha 1_{2n-1} + \beta \end{bmatrix} = k_2^T \mu = [1_{n-1}^T - 1_n^T] \beta = 0
$$

Define $k = [1_{n-1}^T| - 1_n]^T$. Then an $\alpha$-level test of main effects based on $\beta$ is to reject $H_{02}$. If

$$
(k^T \hat{\beta})^T \left[ k^T (X^TX)^{-1} \left( \sum_{k=1}^{n} X_k^T \Sigma \phi X_k \right) (X^TX)^{-1} k \right]^{-1} (k^T \hat{\beta}) > F_{\alpha,1,mn-2n}.
$$

3.5.1 Arnold Transformation Correction

We now complete the inference for the rank-based Arnold transformation estimates under exchangeable errors. Having already transformed to uncorrelated we need to only correct for the case when uncorrelated does not imply independence. Since uncorrelated is a special case of compound symmetry, we may use the above methods for the contrast component described in (2.5.2). As an example we again consider the profile analysis of the two way repeated measures problem.

Recall that for the contrast piece we have the linear model,

$$
z_2 = \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} \gamma_1^* \\ \gamma_2^* \end{bmatrix} + \epsilon_2 = W \gamma^* + \epsilon_2.
$$

Note that here $W_1 = 1_{m_1} \otimes I_{n-1}$, $W_2 = 1_{m_2} \otimes I_{n-1}$, and $\gamma_1$ and $\gamma_2$ are $(n-1) \times 1$
vectors. Recall that the test of parallelism is

\[ H_{01} : \mathbf{K}_1 \gamma_1 = 0 \]

where

\[ \mathbf{K}_1 = [I_{n-1} | - I_{n-1}] \]

As above we would like to write the test on \( \beta \). That is this test is also invariant to intercept. So we have,

\[ z_2 = \mathbf{WEE}^{-1}\gamma^* + \epsilon_2 = [1 \mathbf{X}] \begin{bmatrix} \alpha \\ \beta^* \end{bmatrix} \]

where,

\[ \mathbf{E} = \begin{bmatrix} 1 & \mathbf{0}^T_{2n-3} \\ \mathbf{1}_{2n-3} & \mathbf{I}_{2n-3} \end{bmatrix} \]

So we may write the test of parallelism as

\[ H_{01} : \mathbf{K}_1 \gamma^* = [I_{n-1} | - I_{n-1}] \begin{bmatrix} \alpha^* \\ \alpha^*1 + \beta^* \end{bmatrix} = \begin{bmatrix} \mathbf{0}^T \\ \mathbf{I}_{n-2} \end{bmatrix} \begin{bmatrix} \beta^* \end{bmatrix} = 0. \]

Define

\[ \mathbf{K} = \begin{bmatrix} \mathbf{0}^T \\ \mathbf{I}_{n-2} \end{bmatrix} | - \mathbf{I}_{n-1} \]

Then an \( \alpha \)-level test of parallelism based on \( \beta^* \) is to reject \( H_{01} \) if

\[ \frac{(\mathbf{K} \hat{\beta}^*)^T \left[ \mathbf{K}(\mathbf{X}^T\mathbf{X})^{-1} \left( \sum_{k=1}^{n} \mathbf{X}_k \hat{\Sigma}_k \mathbf{X}_k \right) \mathbf{X}^T\mathbf{X}^{-1} \mathbf{K}^T \right]^{-1} (\mathbf{K} \hat{\beta}^*)}{\hat{\tau}_\varphi^2 (n-1)} > F_{\alpha,n-1,m(n-1)-2(n-1)}. \]
CHAPTER IV

NUMERIC EXAMPLES AND SIMULATION RESULTS

In this chapter we further explore some of the models discussed in chapter two. Our discussion includes both numeric examples as well as simulation results. All calculations were done using the software package R (R Development Core Team 2004).

4.1 One-Way Models

Recall the one-way repeated measures model discussed in section (??) where there is only one treatment level. So that our linear model may be written as,

\[ y_j = \mu + \epsilon_k = \theta_1 + \gamma + \epsilon_k \text{ for } k = 1, \ldots, m. \]

4.1.1 Example

As an example we simulated a data set under this one way setup. There are \( m = 15 \) subjects and \( n = 6 \) repeated measures on each subject. These data were generated from a normal distribution under compound symmetry with \( \sigma = 1 \) and \( \rho = 0.8 \), \( e_1, \ldots, e_{15} \sim N_6(0, A(0.8)) \). Further, the mean vector for each of the subjects is

\[ \mu = [3.0, 3.15, 3.3, 3.45, 3.6, 3.75]^T. \]
That is $\theta = 3$ and $\gamma = [0.0, 0.15, 0.3, 0.45, 0.6, 0.75]^T$. The data are given in the table (1) below. Note that the Arnold transformed parameters are

$$
\mu^* = [8.27, 0.12, -0.03, -0.18, -0.33, -0.48]^T
$$

so that $\theta^* = E[y_{k1}] = 8.27$ and $\gamma^* = E[y_{k2}] = [0.12, -0.03, -0.18, -0.33, -0.48]^T$.

The results for the test of $H_0 : \gamma = 0$ for each of the methods discussed in section (??) are given in table (2) below.

The estimates of $\mu$ are given in table (3).
### Table 2

One-way Example Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friedman</td>
<td>17.6635</td>
<td>0.0034</td>
</tr>
<tr>
<td>Multivariate</td>
<td>18.0860</td>
<td>0.0028</td>
</tr>
<tr>
<td>Rank-GEE</td>
<td>4.0288</td>
<td>0.0025</td>
</tr>
<tr>
<td>AT-rank</td>
<td>4.1411</td>
<td>0.0024</td>
</tr>
</tbody>
</table>

### Table 3

One-way Example Estimates

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariate</td>
<td>$\hat{\mu} = [3.14, 3.33, 3.42, 3.44, 3.70, 3.71]^T$</td>
</tr>
<tr>
<td>GEE</td>
<td>$\hat{\mu} = [3.22, 3.38, 3.49, 3.51, 3.76, 3.83]^T$</td>
</tr>
<tr>
<td>AT-rank</td>
<td>$\hat{\gamma}^* = [-0.02, -0.05, -0.15, -0.36, -0.48]^T$</td>
</tr>
</tbody>
</table>

#### 4.1.2 Simulation Results

A small simulation study was conducted. There were 2000 simulations. All statistical programs were written in R. We looked at three error distributions, the normal, the Cauchy, and the t with three degrees of freedom ($t_3$). We simulated 18 subjects and 6 repeated measures per subject ($m = 18, n = 6$). For simplicity we chose to look at the subjects having a mean vector of the form, $\mu = [0, 0, 0, 0, 0, \mu_6]^T$.

For the simulation results in the tables below (4-6), least squares (AT-LS) is the solution of Arnold. GEE and Arnold Transform rank (AT-R) use Wilcoxon scores. Multivariate (MV) uses sign rank Wilcoxon scores.

As expected least squares was best at the normal but very poor at other
### Table 4

Simulation Results for One-way Normal Data

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\mu_0$</th>
<th>Friedman</th>
<th>GEE</th>
<th>AT-LS</th>
<th>MV</th>
<th>AT-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.0445</td>
<td>0.0445</td>
<td>0.049</td>
<td>0.158</td>
<td>0.05</td>
</tr>
<tr>
<td>0</td>
<td>0.6</td>
<td>0.3105</td>
<td>0.343</td>
<td>0.395</td>
<td>0.4955</td>
<td>0.3565</td>
</tr>
<tr>
<td>0</td>
<td>0.8</td>
<td>0.5195</td>
<td>0.56</td>
<td>0.6255</td>
<td>0.6595</td>
<td>0.572</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.7295</td>
<td>0.789</td>
<td>0.8505</td>
<td>0.834</td>
<td>0.8</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>0.043</td>
<td>0.052</td>
<td>0.0545</td>
<td>0.155</td>
<td>0.051</td>
</tr>
<tr>
<td>0.25</td>
<td>0.6</td>
<td>0.378</td>
<td>0.423</td>
<td>0.4815</td>
<td>0.548</td>
<td>0.438</td>
</tr>
<tr>
<td>0.25</td>
<td>0.8</td>
<td>0.6145</td>
<td>0.6965</td>
<td>0.7465</td>
<td>0.7645</td>
<td>0.6875</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>0.8525</td>
<td>0.8985</td>
<td>0.946</td>
<td>0.912</td>
<td>0.908</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0.0555</td>
<td>0.0515</td>
<td>0.061</td>
<td>0.1595</td>
<td>0.0495</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6</td>
<td>0.5485</td>
<td>0.602</td>
<td>0.6735</td>
<td>0.6795</td>
<td>0.614</td>
</tr>
<tr>
<td>0.5</td>
<td>0.8</td>
<td>0.838</td>
<td>0.886</td>
<td>0.9295</td>
<td>0.8885</td>
<td>0.8885</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.9565</td>
<td>0.979</td>
<td>0.9915</td>
<td>0.9775</td>
<td>0.9825</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>0.058</td>
<td>0.0485</td>
<td>0.055</td>
<td>0.155</td>
<td>0.0475</td>
</tr>
<tr>
<td>0.75</td>
<td>0.6</td>
<td>0.878</td>
<td>0.9145</td>
<td>0.9515</td>
<td>0.898</td>
<td>0.9235</td>
</tr>
<tr>
<td>0.75</td>
<td>0.8</td>
<td>0.989</td>
<td>0.9955</td>
<td>0.9975</td>
<td>0.9895</td>
<td>0.9965</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>0.9995</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 5

Simulation Results for One-way Cauchy Data

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\mu_6$</th>
<th>Friedman</th>
<th>GEE</th>
<th>AT-LS</th>
<th>MV</th>
<th>AT-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.0485</td>
<td>0.018</td>
<td>0.0225</td>
<td>0.031</td>
<td>0.022</td>
</tr>
<tr>
<td>0</td>
<td>0.6</td>
<td>0.1855</td>
<td>0.076</td>
<td>0.0305</td>
<td>0.104</td>
<td>0.069</td>
</tr>
<tr>
<td>0</td>
<td>0.8</td>
<td>0.2865</td>
<td>0.132</td>
<td>0.0375</td>
<td>0.179</td>
<td>0.129</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.417</td>
<td>0.215</td>
<td>0.063</td>
<td>0.2655</td>
<td>0.218</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>0.0415</td>
<td>0.0145</td>
<td>0.0165</td>
<td>0.03</td>
<td>0.0185</td>
</tr>
<tr>
<td>0.25</td>
<td>0.6</td>
<td>0.2325</td>
<td>0.101</td>
<td>0.0385</td>
<td>0.151</td>
<td>0.096</td>
</tr>
<tr>
<td>0.25</td>
<td>0.8</td>
<td>0.376</td>
<td>0.199</td>
<td>0.0545</td>
<td>0.2375</td>
<td>0.211</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>0.5115</td>
<td>0.3235</td>
<td>0.0845</td>
<td>0.347</td>
<td>0.32</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0.0525</td>
<td>0.019</td>
<td>0.0165</td>
<td>0.0425</td>
<td>0.02</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6</td>
<td>0.3345</td>
<td>0.16</td>
<td>0.05</td>
<td>0.2005</td>
<td>0.168</td>
</tr>
<tr>
<td>0.5</td>
<td>0.8</td>
<td>0.524</td>
<td>0.322</td>
<td>0.0825</td>
<td>0.3345</td>
<td>0.312</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.6665</td>
<td>0.476</td>
<td>0.1345</td>
<td>0.4535</td>
<td>0.473</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>0.049</td>
<td>0.0235</td>
<td>0.018</td>
<td>0.046</td>
<td>0.0215</td>
</tr>
<tr>
<td>0.75</td>
<td>0.0</td>
<td>0.532</td>
<td>0.3315</td>
<td>0.074</td>
<td>0.3545</td>
<td>0.345</td>
</tr>
<tr>
<td>0.75</td>
<td>0.6</td>
<td>0.5475</td>
<td>0.32</td>
<td>0.0755</td>
<td>0.3505</td>
<td>0.3415</td>
</tr>
<tr>
<td>0.75</td>
<td>0.8</td>
<td>0.738</td>
<td>0.558</td>
<td>0.151</td>
<td>0.5275</td>
<td>0.566</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>0.869</td>
<td>0.7585</td>
<td>0.2185</td>
<td>0.6545</td>
<td>0.7555</td>
</tr>
<tr>
<td>0.9</td>
<td>0</td>
<td>0.0475</td>
<td>0.0135</td>
<td>0.012</td>
<td>0.08</td>
<td>0.016</td>
</tr>
<tr>
<td>0.9</td>
<td>0.6</td>
<td>0.8505</td>
<td>0.7155</td>
<td>0.2105</td>
<td>0.6315</td>
<td>0.719</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8</td>
<td>0.94</td>
<td>0.883</td>
<td>0.301</td>
<td>0.783</td>
<td>0.881</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>0.9755</td>
<td>0.953</td>
<td>0.381</td>
<td>0.861</td>
<td>0.952</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 6

Simulation Results for One-way $t_3$ Data

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\mu_5$</th>
<th>Friedman</th>
<th>GEE</th>
<th>AT-LS</th>
<th>MV</th>
<th>AT-GEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.0385</td>
<td>0.032</td>
<td>0.0365</td>
<td>0.1015</td>
<td>0.031</td>
</tr>
<tr>
<td>0</td>
<td>0.6</td>
<td>0.241</td>
<td>0.211</td>
<td>0.189</td>
<td>0.342</td>
<td>0.21</td>
</tr>
<tr>
<td>0</td>
<td>0.8</td>
<td>0.416</td>
<td>0.3675</td>
<td>0.2975</td>
<td>0.51</td>
<td>0.3705</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.6075</td>
<td>0.566</td>
<td>0.4585</td>
<td>0.676</td>
<td>0.5665</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>0.0535</td>
<td>0.0385</td>
<td>0.035</td>
<td>0.0975</td>
<td>0.038</td>
</tr>
<tr>
<td>0.25</td>
<td>0.6</td>
<td>0.2975</td>
<td>0.2525</td>
<td>0.224</td>
<td>0.3965</td>
<td>0.27</td>
</tr>
<tr>
<td>0.25</td>
<td>0.8</td>
<td>0.525</td>
<td>0.49</td>
<td>0.4025</td>
<td>0.596</td>
<td>0.4865</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>0.7375</td>
<td>0.708</td>
<td>0.571</td>
<td>0.766</td>
<td>0.716</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0.0495</td>
<td>0.0385</td>
<td>0.0425</td>
<td>0.1</td>
<td>0.0365</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6</td>
<td>0.451</td>
<td>0.402</td>
<td>0.323</td>
<td>0.519</td>
<td>0.4015</td>
</tr>
<tr>
<td>0.5</td>
<td>0.8</td>
<td>0.728</td>
<td>0.6955</td>
<td>0.566</td>
<td>0.7525</td>
<td>0.699</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.891</td>
<td>0.875</td>
<td>0.744</td>
<td>0.886</td>
<td>0.881</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>0.0445</td>
<td>0.0425</td>
<td>0.049</td>
<td>0.123</td>
<td>0.0455</td>
</tr>
<tr>
<td>0.75</td>
<td>0.6</td>
<td>0.767</td>
<td>0.7425</td>
<td>0.6075</td>
<td>0.7685</td>
<td>0.7555</td>
</tr>
<tr>
<td>0.75</td>
<td>0.8</td>
<td>0.941</td>
<td>0.937</td>
<td>0.8285</td>
<td>0.925</td>
<td>0.939</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>0.9915</td>
<td>0.99</td>
<td>0.931</td>
<td>0.9815</td>
<td>0.993</td>
</tr>
<tr>
<td>0.9</td>
<td>0</td>
<td>0.0435</td>
<td>0.033</td>
<td>0.041</td>
<td>0.1195</td>
<td>0.0385</td>
</tr>
<tr>
<td>0.9</td>
<td>0.6</td>
<td>0.9845</td>
<td>0.986</td>
<td>0.898</td>
<td>0.958</td>
<td>0.9865</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8</td>
<td>0.9995</td>
<td>1</td>
<td>0.976</td>
<td>0.994</td>
<td>0.9995</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.9915</td>
<td>0.9985</td>
<td>1</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
distributions. So we will only concentrate on the other methods. Friedman's performed well on the simulations, having the correct level and good power. However there is no estimate for Friedman's and therefore no stage two analyses. Multivariate was a liberal test for two out of the three distributions. The Arnold transform rank solution seemed to do at least as well as the GEE. It should be noted that other scores could be used to improve performance of the two new methods. For example at the Cauchy, sign scores \( \varphi(u) = \text{sgn}(2u - 1) \) could have been used.

4.2 Arnold's Transformation Analysis of Two-Way Models

Recall the two-way repeated measures model discussed in section (2.2) where each subject is assigned to one of two treatment levels. So that our linear model may be written as,

\[
\begin{align*}
\mathbf{y}_k &= \mathbf{\mu}_1 + \mathbf{e}_k = \mathbf{\theta}_1 \mathbf{1} + \mathbf{\gamma}_1 + \mathbf{e}_k \text{ for } k = 1, \ldots, m_1. \\
\mathbf{y}_k &= \mathbf{\mu}_2 + \mathbf{e}_k = \mathbf{\theta}_2 \mathbf{1} + \mathbf{\gamma}_2 + \mathbf{e}_k \text{ for } k = m_1 + 1, \ldots, m.
\end{align*}
\]

4.2.1 Examples

In this subsection we consider some examples for the two-way repeated measures problem. The first two are from simulated data and demonstrate the robustness properties of the Arnold transformation rank method compared to that of least squares.

Repeated measures data were generated from a normal distribution under compound symmetry with \( \sigma = 1, \rho = 0.75, m_1 = m_2 = 10, \)

\[ \mathbf{\mu}_1 = [1, 3, 4, 6]^T, \text{ and } \mathbf{\mu}_2 = [2, 4, 5, 7]^T. \]
So that we have a main effect problem. That is $H_{01} : \gamma_1 = \gamma_2$ is true however, $H_{02} : \theta_1 = \theta_2$ false.

The results for the simulated data are reported in table (7). Suppose that the observation vector for the first subject on treatment 2 were zeroed, $v_1 \leftarrow 0$. The results for this outlier data set are reported in table (8).

Table 7

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Main Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Stat</td>
<td>p-value</td>
</tr>
<tr>
<td>Least Square</td>
<td>0.5614 0.6422</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>0.4762 0.6998</td>
</tr>
</tbody>
</table>

Table 8

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Main Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Stat</td>
<td>p-value</td>
</tr>
<tr>
<td>Least Square</td>
<td>0.8470 0.4726</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>0.4928 0.6884</td>
</tr>
</tbody>
</table>

So that the rank method gives the correct analysis with and without the outlier however least squares only give the correct analysis on the original data.

Next suppose we have the following interaction set-up $m_1 = m_2 = 10, \rho = 0.75, \sigma = 1$,

$$\mu_1 = [1/3, 0, 1/3, 0]^T, \text{ and } \mu_2 = [0, 1/3, 0, 1/3]^T.$$  

That is $H_{01} : \gamma_1 = \gamma_2$ is false.
Table 9

Interaction Example - Without Outlier

<table>
<thead>
<tr>
<th></th>
<th>Interaction</th>
<th></th>
<th>Main Effects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test Stat</td>
<td>p-value</td>
<td>Test Stat</td>
<td>p-value</td>
</tr>
<tr>
<td>Least Square</td>
<td>3.4755</td>
<td>0.0202</td>
<td>0.5200</td>
<td>0.6094</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>3.4905</td>
<td>0.0218</td>
<td>0.8993</td>
<td>0.1902</td>
</tr>
</tbody>
</table>

Table 10

Interaction Example - With Outlier

<table>
<thead>
<tr>
<th></th>
<th>Interaction</th>
<th></th>
<th>Main Effects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test Stat</td>
<td>p-value</td>
<td>Test Stat</td>
<td>p-value</td>
</tr>
<tr>
<td>Least Square</td>
<td>1.8261</td>
<td>0.1498</td>
<td>-0.0298</td>
<td>0.3050</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>4.0962</td>
<td>0.0096</td>
<td>0.3518</td>
<td>0.3646</td>
</tr>
</tbody>
</table>

The results for the simulated data set are displayed in table (9). Now suppose, if the response vector for first subject from each treatment group are interchanged, \( v_1 \leftrightarrow u_1 \). The results for this outlier data set are in table (10).

4.2.2 Simulation Results

A small simulation study was conducted to evaluated the small sample performance. Two situations were considered. The case where there is an interaction effect and the case there there is no interaction, only a main effect. That is we
are considering, 

$$\mu_1 = \theta_1 \mathbf{1} \text{ and } \mu_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \gamma_{24} \end{bmatrix}$$

for various $\theta_1$ and $\gamma_{24}$. Also, $m_1 = m_2 = 10 \sigma^2 = 1$ and $\rho = 0.75$.

The resulting power curves are in figures (1 - 2). Where red represents least squares and black represents Wilcoxon.

The results are similar to the results for the one sample problem in the previous section. That is AT-R is slightly less powerful than the least squares methods at the normal, however more powerful than least squares at heavier tailed distributions then the normal.

4.3 Rank Methods Profile Analysis

4.3.1 Example

As a final example we perform all methods on a simulated data set similar to those discussed in the previous section. Repeated measures data were generated from a normal distribution under compound symmetry with $\sigma = 1, \rho = 0.75, m_1 = 42, m_2 = 47$,

$$\mu_1 = [11, 13, 14, 16]^T \text{ and } \mu_2 = [12, 14, 15, 17]^T.$$ 

All methods give the correct answer.

Suppose that the observation vector for the first subject on both treatments is zeroed, $\mathbf{u}_1 \leftarrow \mathbf{0}$ and $\mathbf{v}_1 \leftarrow \mathbf{0}$. The results become Now all methods except LS
Table 11

Main Effect Example without Outlier

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Main Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td>AT-R</td>
<td>0.5762</td>
</tr>
<tr>
<td>LS</td>
<td>0.7515</td>
</tr>
<tr>
<td>AT-RC</td>
<td>0.5827</td>
</tr>
<tr>
<td>Jaeckel</td>
<td>0.6547</td>
</tr>
</tbody>
</table>

Table 12

Main Effects Example with Outlier

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Main Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td>AT-R</td>
<td>0.4286</td>
</tr>
<tr>
<td>LS</td>
<td>0.8638</td>
</tr>
<tr>
<td>AT-RC</td>
<td>0.4189</td>
</tr>
<tr>
<td>Jaeckel</td>
<td>0.6339</td>
</tr>
</tbody>
</table>
give the correct answer.

4.3.2 Simulation Results

In this subsection we briefly examine some simulation results for the all of
the new rank methods. We also include the least squares analysis as a reference
point.

\[
\begin{bmatrix}
0 \\
0.05 \\
0.10 \\
0.15
\end{bmatrix}
\]

\[\mu_1 = \theta_1 1 \text{ and } \mu_2 = \gamma_2\]

The data below were generated under compound symmetry with the fol-
lowing: \(m_1 = m_2 = 21, \sigma^2 = 1, \text{ and } \rho = 0.75\) Below, red represents least squares,
black represents AT-R, blue represents AT-RC, and green represents Jaeckel.

All the rank based methods appear to be quite similar. All performing
much better at the heavier tailed distributions as well as the contaminated nor-
mals.
Figure 1

Power Curve for Interaction, $H_{01} : \gamma_1 = \gamma_2$
Figure 2

Power Curves for Main Effects, $H_{01}: \theta_1 = \theta_2$

77
Figure 3

Power Curves for Interaction, $H_{01} : \gamma_1 = \gamma_2$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 4

Power Curves for Main Effects, $H_{01} : \theta_1 = \theta_2$
Figure 5

Contaminated Normal Power Curves for $\sigma = 3$
Figure 6

Contaminated Normal Power Curves for $\sigma = 5$
Figure 7

Contaminated Normal Power Curves for $\sigma = 10$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CHAPTER V

CONCLUSIONS

5.1 Summary

We have developed three rank-based solutions to the general repeated measures problem under exchangeable within subject errors. That is the Arnold transformation (AT) approach, the generalized estimating approach (GEE), and the Jaeckel approach.

Arnold's transformation allows us to transform observations having a compound symmetry variance covariance structure to observations which are uncorrelated (working independence) for the general repeated measures problem. If we are willing to make the assumption that for the underlying error distribution having uncorrelated errors implies that the errors are independent, we may make use of the usual univariate analysis. That is we may use the regression analysis discussed in chapter one. Arnold's transformation also greatly simplifies the analysis of several data problems. In particular the location problems described in chapter two. That is the test of main effect in the one-way case results in a univariate one-sample test. Similarly the two sample profile analysis test of main effects becomes a univariate two sample shift problem.

The theory developed in chapter three gives the asymptotic distribution of the estimate which minimizes Jaeckel's dispersion function for general rank scores. We have developed a general contiguity result. That is our contiguity result does not depend on the underlying variance covariance structure. We have also developed a linearity result under the exchangeable errors assumption. This
theory also extends the AT theory in that we may weaken the relatively strong assumption that uncorrelated implies independence.

Finally Arnold's transformation completes a robust rank-based GEE solution to the general repeated measures problem under exchangeable errors. That is AT gives a robust estimate of compound symmetry variance covariance.

In chapter four we demonstrate some of the robustness properties of our new rank-based methods. As with the usual univariate rank estimates our estimates seem to be robust to aberrant observations. Further at distributions with heavier tails then the normal the estimates all seem to out perform the least squares equivalents. That is tests based on these estimates are more powerful at the heavier tailed distributions.

5.2 Future Work

The two most obvious extensions are to eliminate the exchangeable error assumption and extend to unbalanced data cases. We are using balance to refer to an equal number of repeated measures per subject. It does not seem possible to extend Arnold's transformation in either way. Thompson's theory is general in that it does not depend on the variance covariance structure, however it is written in terms of balanced data. Hence we may need to extend this theory in that way.

In developing the linearity result we assumed exchangeable errors and made use of the fact that the errors then have equal marginals. As such we may be able to easily extend these results to error distributions having such a property. An example of such an error distribution is one which has an AR(1) variance
covariance structure, that is

\[
V = \begin{bmatrix}
1 & \rho & \cdots & \rho^{n-1} \\
\rho & 1 & \cdots & \rho^{n-2} \\
\vdots & \ddots & \ddots & \vdots \\
\rho^{n-1} & \rho^{n-2} & \cdots & 1
\end{bmatrix}
\]

We have set up the rank GEE estimates. However much still needs to be done to complete the asymptotic theory. We also need other initial estimates for \( V \).
REFERENCES


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


R Development Core Team (2004), *R: A Language and Environment for Statistical Computing*.


Thompson, G. L (1990), Asymptotic Distribution of Rank Statistics under Dependencies with Multivariate Application *Journal of Multivariate Analysis*, 33, 183-211.