INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the original text directly from the copy submitted. Thus, some dissertation copies are in typewriter face, while others may be from a computer printer.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyrighted material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is available as one exposure on a standard 35 mm slide or as a 17" × 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. 35 mm slides or 6" × 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Algorithms and bounds for rank estimators for several samples

Zubovic, Yvonne Marie, Ph.D.
The Ohio State University, 1988

Copyright ©1988 by Zubovic, Yvonne Marie. All rights reserved.
ALGORITHMS AND BOUNDS FOR RANK ESTIMATORS
FOR SEVERAL SAMPLES

DISSERTATION

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

By

Yvonne M. Zubovic, B.S., M.S.

* * * * *

The Ohio State University
1988

Dissertation Committee:
Sue Leurgans
Jason Hsu
Douglas Wolfe

Approved by
Sue Leurgans
Adviser
Department of Biostatistics
Copyright by
Yvonne M. Zubovic
1988
ACKNOWLEDGMENTS

I would like to thank my adviser Sue Leurgans for her guidance and support throughout the research for this dissertation. I also express my appreciation to Jay Aubuchon, Jason Hsu, and Doug Wolfe for their helpful suggestions. Finally, I thank my family for their patience and confidence during this period of time, and John Osowski for his constant support and understanding.
VITA

March 15, 1958 ........................................... Born - Akron, Ohio

1981 ................................................................... B. S., University of Akron, Akron, Ohio

1981-1983 ....................................................... Teaching Assistant, University of Akron, Akron, Ohio

1984 ................................................................... M. S., University of Akron, Akron, Ohio

1983-1984 ....................................................... University Fellowship, The Ohio State University, Columbus, Ohio

1984-1987 ....................................................... Teaching Assistant, The Ohio State University, Columbus, Ohio

1987-1988 ....................................................... Presidential Fellowship, The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Nonparametric Survival Analysis
# TABLE OF CONTENTS

ACKNOWLEDGMENTS................................................................................................................................. ii
VITA.................................................................................................................................................................. iii
LIST OF TABLES.................................................................................................................................................... vi
LIST OF FIGURES................................................................................................................................................... vii

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td></td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II.</td>
<td></td>
</tr>
<tr>
<td>NOTATION AND PREVIOUS WORK</td>
<td>4</td>
</tr>
<tr>
<td>2.0 Introduction</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Model and Notation</td>
<td>4</td>
</tr>
<tr>
<td>2.2 Review of Nonparametric Procedures for Estimating the Regression Coefficient</td>
<td>6</td>
</tr>
<tr>
<td>2.3 Review of the Computation of the Hodges-Lehmann Estimate</td>
<td>12</td>
</tr>
<tr>
<td>2.4 Estimating the Regression Coefficient using Censored Data</td>
<td>13</td>
</tr>
<tr>
<td>III.</td>
<td></td>
</tr>
<tr>
<td>THREE-SAMPLE CASE</td>
<td>18</td>
</tr>
<tr>
<td>3.0 Introduction</td>
<td>18</td>
</tr>
<tr>
<td>3.1 Description of the Jaeckel Estimator for the Three-Sample Model</td>
<td>19</td>
</tr>
<tr>
<td>3.2 Characterization of the Jaeckel Estimator for the Three-Sample Case</td>
<td>30</td>
</tr>
<tr>
<td>3.3 Relationship between the Jaeckel Estimator and the Two-Sample Hodges-Lehmann Estimators</td>
<td>40</td>
</tr>
<tr>
<td>3.4 Ideas for an Algorithm</td>
<td>60</td>
</tr>
<tr>
<td>3.5 Example</td>
<td>63</td>
</tr>
<tr>
<td>IV.</td>
<td></td>
</tr>
<tr>
<td>K-SAMPLE CASE</td>
<td>70</td>
</tr>
<tr>
<td>4.0 Introduction</td>
<td>70</td>
</tr>
<tr>
<td>4.1 Formulation of the Jaeckel Equations for the k-Sample Model</td>
<td>70</td>
</tr>
<tr>
<td>4.2 Characterization of the Jaeckel Estimator</td>
<td>76</td>
</tr>
<tr>
<td>4.3 Relationship between the Jaeckel Estimator and the Hodges-Lehmann Two-Sample Estimators</td>
<td>84</td>
</tr>
<tr>
<td>4.4 Ideas for an Algorithm</td>
<td>97</td>
</tr>
<tr>
<td>4.5 Example</td>
<td>98</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Latency Times for Virgin Rats Injected with Plasma from Rats in Various Phases of the Estrous Cycle</td>
<td>64</td>
</tr>
<tr>
<td>2. Ordered Differences between Pairs of Samples for the Latency Data</td>
<td>65</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Graphical representation of the contours of the function $U_{1,3}(\beta)$ in the $(\beta_1, \beta_2)$ plane</td>
<td>36</td>
</tr>
<tr>
<td>2. Graphical representation of the contours of the function $U_{1,2}(\beta)$ in the $(\beta_1, \beta_2)$ plane</td>
<td>37</td>
</tr>
<tr>
<td>3. Graphical representation of the contours of the function $S_1^*(\beta)$ in the $(\beta_1, \beta_2)$ plane</td>
<td>38</td>
</tr>
<tr>
<td>4. Graphical representation of the contours of the function $S_2^*(\beta)$ in the $(\beta_1, \beta_2)$ plane</td>
<td>39</td>
</tr>
<tr>
<td>5. Graphical representation of the Jaeckel estimator as the intersection of the zero contours of the functions $S_1^<em>(\beta)$ and $S_2^</em>(\beta)$</td>
<td>41</td>
</tr>
<tr>
<td>6. Graphical illustration of the incompatibility of the two-sample Hodges-Lehmann estimators of the location shift between pairs of samples in configuration I</td>
<td>43</td>
</tr>
<tr>
<td>7. Graphical illustration of the incompatibility of the two-sample Hodges-Lehmann estimators of the location shift between pairs of samples in configuration II</td>
<td>46</td>
</tr>
<tr>
<td>8. Graphical representation of the regions $R_1,\ldots, R_7$ for the proof of Theorem 3.4</td>
<td>53</td>
</tr>
<tr>
<td>9. Graphical representation of the triangle containing the Jaeckel estimate for the virgin rat data</td>
<td>68</td>
</tr>
<tr>
<td>10. Graphical representation of the zero contours of the functions $S_1^<em>(\beta)$ and $S_2^</em>(\beta)$ for the virgin rat data</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Graphical representation of the contours of the function</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>11.</td>
<td>$S_1^*(\beta)$ when viewing $\mathbb{R}^3$ along the equiangular line</td>
</tr>
<tr>
<td>12.</td>
<td>$S_2^*(\beta)$ when viewing $\mathbb{R}^3$ along the equiangular line</td>
</tr>
<tr>
<td>13.</td>
<td>$S_3^*(\beta)$ when viewing $\mathbb{R}^3$ along the equiangular line</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Investigators in the medical field, the biological sciences, and engineering often conduct experiments in which the variable of interest is a response time. A response time is the time that elapses until a specified response is elicited from the subject or, in the case of inanimate subjects, until a specified event occurs. The investigators of such experiments are often interested in modelling the relationship of the survival time of a subject with other variables characterizing the subject. Let us consider the following example.

A cancer research team is interested in determining which of three treatments is the most effective in prolonging the life of patients with a certain type of leukemia. They conduct a clinical trial in which a patient, upon diagnosis of the disease, is randomly assigned to one of the three treatment arms. Due to funding and eagerness, the duration of the trial is limited to five years. The patients are observed until they die. If the patient dies during the study, the time from diagnosis to death is recorded. For those patients who survive until the end of the study, the length of time in the study is recorded. At the end of the five years, N patients have entered the trial. Of these N patients, $n_1$ have been administered treatment 1, $n_2$ have been administered treatment 2, and $n_3$ have been administered treatment 3. For the $n_i$ patients receiving treatment $i$, the time until death or the length of time in the study has been recorded, as well as whether the patient died or not.
This example has several characteristics common to many other experiments. First, the variable of interest is a response time. In this example the response is the death of the patient and the response time is the survival time after diagnosis. In other clinical trials the response may be the appearance of a tumor or the relapse of a disease with response time corresponding to the tumor-free time or the time in remission, respectively. In reliability studies the event is often the failure of the item on test and the response time is the failure time. The response time will be referred to as the survival time, the failure time, or the lifetime and the response will be called death or failure.

Another characteristic of this example is that a patient can be classified according to which treatment he received. Often the subjects can be classified as belonging to one of several groups. In a life-testing experiment the group classification may indicate which compound composes a motor part or it may indicate under which stress conditions a machine component was tested.

Finally, in the above example the survival time is not known for every patient at the end of the study. For many reasons, such as a limited period of study, economic restrictions, and patient withdrawal, the survival time of some individuals cannot be observed directly but is known to exceed another observed value. Such observations are called right-censored observations.

For the above example there are two questions which are of interest to the researchers:

(1.) Are the survival times different for the patients who receive different treatments? If so, which treatment groups differ in their survival times?

(2.) How much of a difference is there in the survival times of the different treatment groups?

One addresses the answer to the first question through hypothesis testing. Many statistics have been proposed to test for the equality of several distributions. These tests
differ in the assumptions that are required and the types of alternatives that can be detected with relatively high power. To determine which of the treatments differ, multiple comparison techniques have been used. These techniques differ in the manner in which they control the overall level of significance.

To answer the second question one estimates some measure of the differences in the survival times between the various groups. This involves modelling the distributions of the survival times as a function of the group classification and then estimating the parameters of the model. Many approaches have been proposed for this estimation problem, some of which will be described in more detail in Chapter II.

In this dissertation we consider one approach to the estimation problem raised in the second question and address some of the difficulties that are encountered in trying to compute the estimates. This estimation procedure, proposed by Jaeckel (1972), is discussed briefly in Chapter II for the general linear model. In Chapter III we consider the Jaeckel estimation procedure for the special case of three classification groups with complete data. Chapter III contains characterizations of the Jaeckel estimator and a discussion of the relationship between the Jaeckel estimator and an estimator derived by the Hodges-Lehmann two-sample method. In Chapter IV we extend the results of the three-sample case to the case of four samples, and discuss the extension of the results to the case of k samples. Chapter V contains a description of an algorithm used to compute the Jaeckel estimates in the case of four samples with complete data. In Chapter VI we discuss how ideas from this algorithm can be used in computing estimates for the case of two samples with censored data. Finally, in Chapter VII we conclude with a discussion of some further areas of research in which the results of this dissertation may be applied.
CHAPTER II
NOTATION AND PREVIOUS WORK

2.0 Introduction

In this chapter we discuss some of the previous work done on the estimation problem described in the first chapter. We begin in the first section with a definition of the model and its assumptions, and we define the notation that is to be used throughout the dissertation, unless otherwise noted. In the second section we review various nonparametric estimation procedures. We discuss estimation procedures for the two-sample problem, the k-sample problem, the simple linear regression problem, and the general linear model setting. The third section contains a review of the development of fast, exact algorithms for computing the estimates discussed in the second section. The final section contains a description of the extension of the methods in the second section to the case of censored data.

2.1 Model and Notation

As indicated in the previous section, the investigator in a clinical trial or reliability study is often interested in modelling the distribution of the survival times of the subjects under study. Let Y represent the survival time and let \( X \) represent the row vector of covariates. Consider the linear model

\[
g(Y_i) = \alpha + X_i \beta + e_i, \quad i = 1, \ldots, N
\]  

(2.1)
where $g$ is a known monotone transformation, $\alpha$ is an unknown location parameter, $\beta = (\beta_1, \ldots, \beta_p)^T$ is a vector of unknown regression coefficients, and $e_i$ is the error associated with the $i$-th observation. Assume the $e_i, i = 1, \ldots, N$, are independent and identically distributed with cumulative distribution function (c.d.f.) $F(\cdot)$. The distribution function is assumed to be continuous, but its shape is unknown.

Since survival times are usually positive, a transformation to the entire real line is often desired. One such transformation uses the logarithmic function, namely $g(y) = \log(y)$. This model, referred to as the accelerated failure time or the log-linear model, has been used in survival analysis by a variety of authors. Cox (1972) discussed some of the implications of this model and suggested some settings in which it may be applicable. Mann, Schafer, and Singpurwalla (1974) and Doksum (1975) described reliability studies in which this model is appropriate. In the analysis of survival data, Prentice (1978) and Kalbfleisch and Prentice (1980) proposed this model as an alternative to the proportional hazards model of Cox (1972). Under the assumptions of this model in the $k$-sample setting, the failure times of the $i$-th group have c.d.f. $G(\theta_i t)$ for $\theta_i > 0$. Thus the difference in the effect of treatment $i$ over treatment $j$ is to accelerate or decelerate the time to failure, depending on the value of $\theta_i/\theta_j$. Using the logarithmic transformation, the time scale change can be regarded as a location shift on the log-time axis. Assume from now on that $Y_i$ represents the transformed survival time so that the model is

$$Y_i = \alpha + X_i \beta + e_i, \quad i = 1, \ldots, N.$$  (2.2)

Recall that the purpose is to estimate the vector of unknown regression coefficients, $\beta$. If the shape of the distribution function $F$ is known, one can use parametric methods
such as maximum likelihood to obtain an estimator for $\beta$. The testing and estimation procedures for various distributions are summarized in the books by Kalbfleisch and Prentice (1980), Lawless (1982), and Cox and Oakes (1984).

In analyzing survival data investigators often encounter two problems which make the use of parametric procedures undesirable. First, the investigator may have difficulty in specifying $F$. He may be faced with a choice of two or more distributions which appear to fit the data equally well, although possibly poorly, but which yield quite different estimates of the regression coefficients. Secondly, the estimator may be sensitive to outliers or extreme observations, as in the case of the normal least squares estimators. Hence a procedure which is less sensitive to misspecification of $F$ and less sensitive to outliers is desired. In the next section we describe a variety of nonparametric procedures which were proposed to reduce the sensitivity of the estimators to outliers.

2.2 Review of Nonparametric Procedures for Estimating the Regression Coefficient

Hodges and Lehmann (1963) derived estimators of location and shift in the one-sample and two-sample settings, respectively, which are more robust than classical estimators based on the mean. Because their method is the basis of rank estimation, we examine it in detail in the context of the two-sample problem.

Let $Y_1,\ldots, Y_{n_1}$ and $Y_{n_1+1},\ldots, Y_{n_1+n_2}$ denote independent random samples from continuous distributions with c.d.f.'s $F(y-\alpha)$ and $F(y-\alpha-\beta)$, respectively. This corresponds to letting $p=1$ and $X_i=0$ if $i=1,\ldots, n_1$ and $X_i=1$ if $i=n_1+1,\ldots, n_1+n_2$ in model (2.2). To estimate the shift parameter $\beta$, find the value $b$ for which $Y_1,\ldots, Y_{n_1}$ and $Y_{n_1+1}-b,\ldots, Y_{n_1+n_2}-b$ are aligned, that is, they appear to come from the same distribution. The criterion for evaluating alignment involves a statistic $T(Y_1,\ldots, Y_{n_1}; Y_{n_1+1},\ldots, Y_{n_1+n_2})$ used to test whether the shift parameter is zero. Assume that the statistic satisfies the following two conditions:
(i.) \( T(Y_1, \ldots, Y_{n_1}; Y_{n_1+1-t}, \ldots, Y_{n_1+n_2-t}) \) is a nonincreasing function of \( t \) for fixed 
\( (Y_1, \ldots, Y_{n_1}, Y_{n_1+1}, \ldots, Y_{n_1+n_2}) \). \hspace{1cm} (2.3) 

and

(ii.) Under \( H_0: \beta = 0 \), the distribution of \( T(Y_1, \ldots, Y_{n_1}; Y_{n_1+1}, \ldots, Y_{n_1+n_2}) \) is symmetric about some value \( \xi \) for every continuous \( F(\cdot) \). \hspace{1cm} (2.4)

Choose as the estimator of \( \beta \) the value \( b_{HL} \) such that \( T(Y_1, \ldots, Y_{n_1}; Y_{n_1+1-b_{HL}}, \ldots, Y_{n_1+n_2-b_{HL}}) \) is as close as possible to \( \xi \), namely

\[
b_{HL} = \frac{b_{HL}^* + b_{HL}^{**}}{2} \hspace{1cm} (2.5)
\]

where

\[
b_{HL}^* = \sup \{ \beta: T(Y_1, \ldots, Y_{n_1}; Y_{n_1+1-\beta}, \ldots, Y_{n_1+n_2-\beta}) > \xi \} \hspace{1cm} (2.6)
\]

and

\[
b_{HL}^{**} = \inf \{ \beta: T(Y_1, \ldots, Y_{n_1}; Y_{n_1+1-\beta}, \ldots, Y_{n_1+n_2-\beta}) < \xi \} \hspace{1cm} (2.7)
\]

Hodges and Lehmann derived \( b_{HL} \) by inverting a two-sample rank statistic of the form

\[
T(Y_1, \ldots, Y_{n_1}; Y_{n_1+1}, \ldots, Y_{n_1+n_2}) = \sum_{j=n_1+1}^{n_1+n_2} E[\psi(V(R_j))] \hspace{1cm} (2.8)
\]

where \( V(1)<\ldots<V(n_1+n_2) \) denote the order statistics from a sample of size \( n_1+n_2 \) from a distribution \( \psi \), and \( R_j \) is the rank of \( Y_j \) among \( Y_1, \ldots, Y_{n_1+n_2} \). They used Wilcoxon scores in which \( \psi \) is the rectangular distribution on \((0,1)\) and normal scores in which \( \psi \) is the normal distribution. They showed that

(i.) the distribution of \( b_{HL} \) is (absolutely) continuous if \( F \) is (absolutely) continuous;
(ii.) $b_{HL}$ is translation invariant;
(iii.) if $F$ is symmetric, then the distribution of $b_{HL}$ is symmetric about $\beta$;
(iv.) $b_{HL}$ is either exactly or approximately median unbiased; and
(v.) under suitable conditions, if $n_1/(n_1+n_2) \to \lambda$ as $n_1+n_2 \to \infty$, then $\sqrt{n_1+n_2} (b_{HL}-\beta)$ has a limiting normal distribution.

The estimator $b_{HL}$ is less sensitive to extreme observations than the classical estimator, $\bar{Y}_2-\bar{Y}_1$, where $\bar{Y}_i$ denotes the mean of the observations in sample $i$.

As noted above, the two-sample setting can be regarded as a special case of the simple linear regression model (2.2) with $p=1$. Adichie (1967) extended the method of Hodges and Lehmann to obtain a class of point estimators of $\beta$ in the simple linear regression setting which are more robust than the least squares estimators. He assumed that the underlying distribution $F$ is absolutely continuous and symmetric with a density that is absolutely continuous and square integrable. He estimated $\beta$ by inverting the rank statistic

$$T(Y_1,\ldots, Y_N) = \frac{1}{N} \sum_{j=1}^{N} (X_j-\bar{X}_N) \psi\left(\frac{R_j}{N+1}\right)$$

(2.9)

where $\bar{X}_N = N^{-1} \sum_{j=1}^{N} X_j$, $R_j$ is the rank of $Y_j$ among $Y_1,\ldots, Y_N$, and $\psi$ is a score function on $(0,1)$ of the form

$$\psi(u) = -[g'(G^{-1}(u)) / g(G^{-1}(u))]$$

(2.10)

for any absolutely continuous symmetric c.d.f. $G$. He discussed the small sample and asymptotic properties of his estimator, $b_A$. In particular, he showed that $b_A$ is translation invariant, the distribution of $b_A$ is symmetric about $\beta$ if $F$ is symmetric, and $\sqrt{n_1+n_2} (b_A-\beta)$ is asymptotically normal.
Sen (1968) applied the method of Hodges and Lehmann to derive an estimator of $\beta$ for the same model as Adichie but under weaker restrictions on $F$. He inverted a statistic which is proportional to Kendall's correlation coefficient, namely

$$T(Y_1-X_1\beta,\ldots,Y_N-X_N\beta) = \left( \frac{1}{n} \left( \frac{N}{2} \right) \right)^{1/2} \sum_{1 \leq i < j \leq N} c(X_j-X_i) c((Y_j-X_j\beta)-(Y_i-X_i\beta))$$

(2.11)

where $n = \sum_{1 \leq i < j \leq N} c(X_j-X_i)$ and $c(u) = 1$, 0, or -1 if $u > 0$, = 0, or < 0, respectively. Sen proved the invariance, the unbiasedness, and the asymptotic normality of his estimator.

Lehmann (1963) considered the robust estimation of $\beta$ in the k-sample setting for $k > 2$. Let $b_{i,j} = \text{med}(Y_r - Y_s)$ denote the median of the $n_{nj}$ pairwise differences $Y_r - Y_s$ for $r \in S_i$ and $s \in S_j$, where $S_a$ denotes the set of indices corresponding to observations from sample $a$. As Hodges and Lehmann (1963) showed, $b_{i,j}$ is the estimator for $\beta_i - \beta_j$, the shift parameter between the distributions of samples $i$ and $j$, obtained by inverting a statistic with Wilcoxon scores. Unfortunately, while $\beta_i - \beta_j = (\beta_i - \beta_m) + (\beta_m - \beta_j)$, the estimate $b_{i,j}$ does not necessarily equal $b_{i,m} + b_{m,j}$. In general, the $b_{i,j}$ do not satisfy linear relations that are satisfied by the parameters that they estimate, a property that Lehmann called incompatibility. Lehmann avoided the problem of incompatibility by using adjusted estimates in place of the $b_{i,j}$. He adjusted the estimates of $\beta_i - \beta_j$ by minimizing the sum of squares

$$\sum_{i \neq j} \sum [b_{i,j} - (\beta_i - \beta_j)]^2$$

(2.12)

which yielded the estimators

$$Z_{ij} = b_{i,-} - b_j$$

(2.13)
where $b_i = k^{-1} \sum_{m=1}^{k} b_{i,m}$. While the adjusted estimators $Z_{ij}$ are compatible, their use has two disadvantages. First, although the $b_{ij}$ depend only on observations from samples $i$ and $j$, the estimator $Z_{ij}$ depends on observations from all $k$ samples. Secondly, the estimator $b_{ij}$ is consistent when $n_i$ and $n_j$ approach infinity but for $Z_{ij}$ to be consistent, $n_m$ must tend to infinity for all $m$.

Spjøtvoll (1968) modified Lehmann's estimation procedure to resolve the inconsistency problem. He proposed minimizing a weighted sum of squares using the asymptotic variance of $b_{ij}$ as the weight. He noted that the estimators derived from minimizing

$$Q_1 = \sum_{i \neq j} \left( \frac{N}{n_i} + \frac{N}{n_j} \right)^{-1} \left[ b_{ij} - (\beta_i - \beta_j) \right]^2$$

are difficult to compute, but that they have the same asymptotic properties as the estimators derived by minimizing

$$Q_2 = \sum_{i \neq j} \left( \frac{n_i n_j}{N^2} \right) \left[ b_{ij} - (\beta_i - \beta_j) \right]^2$$  \hspace{1cm} (2.15)

if $n_i$ approaches infinity in such a way that $n_i/N \to \rho_i$, $0 < \rho_i < 1$. The weighted adjusted estimators obtained by minimizing (2.15) are given by

$$W_{ij} = N^{-1} \sum_{m=1}^{k} n_m \left( b_{i,m} - b_{j,m} \right).$$  \hspace{1cm} (2.16)

The $W_{ij}$ are consistent if $n_i$ and $n_j$ only tend to infinity but they still depend on observations from samples other than $i$ and $j$.

The $k$-sample problem is the special case of the multiple regression model in which $X_j$ is an indicator of whether or not the observation is from the $j$-th sample. Jureckova
(1971) extended the method of Hodges and Lehmann to derive a class of estimators of the regression coefficients in the general linear model (2.2). She considered the linear rank statistics for \( j = 1, \ldots, p \)

\[
S_j(Y - X_\beta) = S_j(Y_1 - X_1 \beta, \ldots, Y_N - X_N \beta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_{ij} - \bar{X}_j) a(R(Y_i - X_i \beta)) \tag{2.17}
\]

where \( \bar{X}_j = N^{-1} \sum_{m=1}^{N} X_{mj} \), the quantity \( R(Y_i - X_i \beta) \) is the rank of \( Y_i - X_i \beta \) among \( Y_1 - X_1 \beta, \ldots, Y_N - X_N \beta \), and \( a(i), i = 1, \ldots, N \) are scores generated by a nonconstant, nondecreasing, square-integrable function \( \phi(u) \) on \( (0,1) \) by either

\[
a(i) = E\phi(U(i)) \tag{2.18}
\]

or

\[
a(i) = \phi\left(\frac{i}{N+1}\right) \tag{2.19}
\]

where \( U(i) \) is the \( i \)-th order statistic for a sample of size \( N \) from a uniform distribution on \( (0,1) \). Since under her assumptions the vector \( S(Y - X_\beta) = (S_1(Y - X_\beta), \ldots, S_p(Y - X_\beta)) \) is asymptotically normal with mean 0, the estimator of \( \beta \) is the vector \( \beta_{JJ} \) for which \( S(Y - X_\beta) \) is as close to 0 as possible. For \( \beta_{JJ} \) to be as close as possible to 0, Jureckova used the criterion that \( \beta_{JJ} \) minimizes the function

\[
\sum_{j=1}^{p} |S_j(Y - X_\beta)|, \text{ that is, } \beta_{JJ} \minimizes \text{the L}_1\text{-norm of the vector } S(Y - X_\beta). \tag{2.20}
\]

She proved the asymptotic normality of \( \beta_{JJ} \) under her assumptions. Her assumptions contain restrictions on \( X \) which may be difficult to check in practice. Heiler and Willers (1982) proved the asymptotic normality of Jureckova's estimator under less restrictive conditions on \( X \).

As discussed above, Jureckova (1971) estimated \( \beta \) in the multiple regression model by finding the value of \( b \) for which a statistic \( S(Y - Xb) \) for testing \( \beta = \beta_0 \) is as close as
possible to its mean under the null distribution. Jaeckel (1972) used a different criterion for estimating $\beta$, but showed that the resulting estimates are asymptotically equivalent to those proposed by Jureckova. He estimated $\beta$ by the value $b_L$ for which the residuals $Y_1 - X_1 b_L,..., Y_N - X_N b_L$ are the least dispersed, that is $b_L$ minimizes a measure of the dispersion of the residuals. Rather than using the variance of the residuals as a measure of dispersion, Jaeckel used a dispersion function based on the ranks of the residuals. His estimates are less sensitive to extreme observations than the classical least squares estimate. Jaeckel's class of estimators will be discussed in more detail in Chapter III.

2.3 Review of the Computation of the Hodges-Lehmann Estimate

The estimators described above are derived by inverting a rank statistic, that is finding the value $b$ for which

$$T(Y - Xb) = E_0[T(Y - Xb)],$$

(2.20)

where $E_0[\cdot]$ denotes expectation under the null hypothesis $H_0: \beta = 0$. The use of such estimators has been hindered by the difficulty in computing them. Several approaches are possible for obtaining the estimate. One can iteratively approach the solution to (2.20) by evaluating $T(\cdot)$ at a trial value, checking for whether the trial value is a solution and, if necessary, selecting a new trial value and repeating the previous steps. Such methods as the bisection algorithm, false position, or the Illinois variant of false position can be used to find an approximation of the solution to (2.20) in the simple linear regression model. Another approach is to use the nature of $T(\cdot)$ as a step function and apply a divide-and-conquer algorithm to find an exact solution to (2.20). In either case, computation of the estimate can be costly in terms of computer time and storage space if a naive approach is taken for evaluating $T(\cdot)$ at a trial value.
Several authors have addressed the problem of computing estimates from rank statistics in both the one-sample and the two-sample uncensored data case in an attempt to reduce the amount of computer time and storage required. McKean and Ryan (1977) used a modified false position algorithm to approximate the estimator. Others, including Johnson and Kashdan (1978), Johnson and Mizoguchi (1978), Johnson and Ryan (1978), and Monahan (1984) considered various modifications of a fast, exact, divide-and-conquer method for computation of estimates derived from the Wilcoxon statistic. Aubuchon (1984) considered the extension of these fast, exact algorithms to the case of the general linear rank statistics in the one-sample setting. These approximating and exact algorithms will be discussed in detail in Chapters V and VI.

2.4 Estimation of the Regression Coefficient using Censored Data

In the estimation of β in model (2.2), the work previously described assumed that Y, the true lifetime, can be observed for all of the individuals under study. For many reasons, such as a limited period of study, economic restrictions, and patient withdrawal, the survival time of some individuals in a clinical trial or reliability study cannot be observed directly but is known to exceed another observed value. Such observations are called right-censored observations.

Let Y₁,..., Yₙ represent the true lifetimes of the N individuals under study and assume model (2.2) holds. Let C₁,..., Cₙ represent the censoring times of the individuals under study. Let Yⱼ* = min(Yⱼ, Cⱼ) denote the observed time to failure for the individuals, and let δⱼ = 1 if Yⱼ = Yⱼ* and 0 otherwise for j = 1,..., N. Here δⱼ represents the death indicator of the individuals. For simplicity, Yⱼ* denotes a monotone transformation of the observed lifetimes. In this case the same transformation is applied to the true lifetimes and the censoring times. As an example consider the accelerated failure time model. In this model a logarithmic transformation is applied to all of the observed times
so that \( Y_i^* = \log Y_i \) if \( \delta_i = 1 \) and \( Y_i = \log C_i \) if \( \delta_i = 0 \).

Under the assumptions of the accelerated failure time model, Louis (1981) proposed an estimator of the time scale change for two independent random samples with right-censored observations. He provided a point estimate and a confidence interval for the scale change based on the logrank statistic. The logrank statistic has the form

\[
T = \int_{-\infty}^{\infty} \left\{ \frac{R_2(s)}{R_1(s) + R_2(s)} dD_1(s) - \frac{R_1(s)}{R_1(s) + R_2(s)} dD_2(s) \right\}
\]

where \( R_i(s) \) denotes the number of individuals at risk in sample \( i \) at time \( s \) for \( i = 1, 2 \), that is the number of individuals who have not died or been censored prior to time \( s \), and \( D_i(s) \) denotes the number of deaths in sample \( i \) at time \( s \) for \( i = 1, 2 \). His estimator in the log-time axis is the value \( b_L \) such that

\[
T(b_L) = 0
\]

where

\[
T(b) = \int_{-\infty}^{\infty} \left\{ \frac{R_2(s+b)}{R_1(s) + R_2(s+b)} dD_1(s) - \frac{R_1(s)}{R_1(s) + R_2(s+b)} dD_2(s+b) \right\}
\]

Louis further proved the consistency and asymptotic normality of his estimator.

Under the same assumptions of the accelerated failure time model as Louis, Padgett and Wei (1982) derived a different estimator of the time scale change for two independent random samples with right-censored data. Although they derived the estimator by minimizing the Cramer-von Mises distance between the estimated survival distributions of the two samples, the estimator can also be derived by inverting the test statistic proposed by Efron (1967). Their estimator can be written in closed-form.

Let \( Y_1^*, ..., Y_{n_1}^*, Y_{n_1+1}^*, ..., Y_{n_1+n_2}^* \) denote the logarithms of the observed times in samples 1 and 2, respectively. Let \( Y_{(1)}^* \leq ... \leq Y_{(n_1)}^* \) and \( Y_{(n_1+1)}^* \leq ... \leq Y_{(n_1+n_2)}^* \) be the corresponding order statistics. Let \( F_1(Y_{(i)}^*) \), \( i = 1, ..., n_1 \) denote the left-continuous product limit estimator (Kaplan and Meier, 1958) of the survival distribution for sample 1
at \( Y(i)^* \). Define \( a_i \) to be the jump in the product limit estimator at time \( Y(i)^* \), namely

\[
\hat{a}_i = \begin{cases} \\
\frac{\Delta}{\hat{F}_1(Y(i)^*)} \quad & \text{if } i = 1, \ldots, n_1 - 1 \\
\frac{\Delta}{\hat{F}_1(Y(i)^*)} \quad & \text{if } i = n_1.
\end{cases}
\tag{2.23}
\]

Similarly, let \( \hat{F}_2(Y(n_i+j)^*) \), \( j = 1, \ldots, n_2 \) be the product limit estimator of the survival distribution for sample 2 at the observed times and let \( b_j \) be the jump in the product limit estimator at time \( Y(n_i+j)^* \). The estimator of the shift parameter, denoted \( \hat{b}_{PW} \), proposed by Padgett and Wei is the median of the distribution with probability mass function

\[
p(v) = \begin{cases} \\
a_i b_j & \text{if } v = Y(n_i+j)^* - Y(i)^*; \text{ for } i = 1, \ldots, n_1 \text{ and } j = 1, \ldots, n_2 \\
0 & \text{otherwise.}
\end{cases}
\tag{2.24}
\]

They showed that their estimator is consistent under mild conditions.

Wei and Gail (1983) derived a class of estimators of the time scale change in the two-sample right-censored data case under the assumptions of the accelerated failure time model. They inverted rank statistics for testing whether two survival distributions are equal. They considered statistics in Gill’s (1980) \( K^+ \) class, statistics of the form

\[
T = \int_{-\infty}^{\infty} W_\Omega(s) \left\{ \frac{R_2(s)}{R_1(s) + R_2(s)} dD_1(s) - \frac{R_1(s)}{R_1(s) + R_2(s)} dD_2(s) \right\}
\tag{2.25}
\]

where \( R_i(s) \) and \( D_i(s) \) are as in (2.21), and \( W_\Omega(s) \) is a predictable function of \( \{R_i(r), D_i(r); i = 1, 2 \text{ and } r < s\} \) that equals zero whenever the product \( R_1(s)R_2(s) \) is zero. This class includes the logrank statistic as well as those statistics proposed by Gehan (1965), Peto and Peto (1972), Tarone and Ware (1977), Prentice (1978), and Harrington and Fleming (1982). The properties of these statistics are discussed in papers by Leurgans (1983), (1984) and Andersen et al. (1982), among others. Following Hodges and Lehmann (1963) Wei and Gail proved the regularity property, the invariance property,
and the asymptotic normality of their estimators.

Tsiatis (1986) derived estimates for $\beta$ in the multiple linear regression model with right-censored data. He considered the linear rank statistics, for $j=1,\ldots, p$:

$$S_j(Y^*-X\beta) = (S_1(Y^*-X\beta), \ldots, S_p(Y^*-X\beta))$$

where

$$\bar{X}_j(u, \beta) = \frac{\sum_{m=1}^{N} X_{mj} Z_m(u+X_m\beta)}{\sum_{m=1}^{N} Z_m(u+X_m\beta)},$$

$$N_i(u) = I(Y_i^* \leq u, d_i=1),$$

$$Z_m(u) = I(Y_m \geq u),$$

and $I(A)$ is an indicator of whether or not the logical expression $A$ is true. Assume $W_N(u)$ is a $F_N(u)$ measurable, left-continuous, nonnegative function of the observations, which converges in probability to a deterministic function; $W_N(u, \beta)$ is $W_N(u)$ evaluated at the residuals, $Y_i^* - X_i\beta$, $i=1,\ldots, N$. Since under Tsiatis' regularity conditions the vector $\sim S_j(Y^*-X\beta) = (S_1(Y^*-X\beta), \ldots, S_p(Y^*-X\beta))$ is asymptotically normal with mean $0$, the estimator of $\beta$ is the vector $b_T$ for which $\sim S_j(Y^*-X\beta)$ is as close to $0$ as possible. He derives $b_T$ by solving the estimating equations

$$S_j(Y^*-Xb_T) = 0 \text{ for } j=1,\ldots, p.$$
asymptotically normal.

In Chapter VI we consider the computation of estimates derived by inverting weighted logrank statistics in the two-sample setting.
CHAPTER III
THREE-SAMPLE CASE

3.0 Introduction

In the previous chapter we discussed several procedures for estimating the vector of regression coefficients in a linear model. In this chapter we examine more closely the estimator which Jaeckel (1972) proposed, for the vector of regression coefficients for a special case of the linear model, namely the three-sample model.

In the first section we describe the Jaeckel procedure in detail for the three-sample, uncensored model. In this description, we define the Jaeckel dispersion function and formulate the Jaeckel estimating equations. We conclude this section with a discussion of the properties of the functions involved in the Jaeckel estimating equations. We characterize the Jaeckel estimator in the second section. We prove a theorem characterizing the estimator in terms of the ordered differences of pairs of samples. We also provide a geometric characterization of the estimator in terms of the functions described in the first section. In the third section, we discuss the incompatibility of the Hodges-Lehmann two-sample estimators and describe the relationship between the Jaeckel estimator and the Hodges-Lehmann two-sample estimators. The fourth section contains a brief discussion of an algorithm for computing the Jaeckel estimate in the three-sample case, and the details of the algorithm are provided in Chapter V. In the last section we provide an example to illustrate the ideas discussed in the first four sections.
3.1 Description of the Jaeckel Estimator for the Three-Sample Model

Consider the uncensored three-sample model given by

\[ Y_i = \alpha + X_j \beta + e_i, \quad i=1, \ldots, N, \quad (3.1) \]

where, for \( j=1 \) and \( 2 \), \( X_{ij} = 1 \) if observation \( i \) is from sample \( j \) and = 0 otherwise.

Without loss of generality, assume that the observations for individuals in sample 1 occur first in \( Y \), followed by the observations from sample 2, and finally followed by the observations from sample 3. Let \( S_j \) denote the set of indices of the observations in sample \( j \) for \( j=1, 2, 3 \). Then \( S_j = \{ a_j + 1, a_j + 2, \ldots, a_j + n_j \} \), where \( a_1=0 \), \( a_2=n_1 \), and \( a_3=n_1 + n_2 \). Let \( N=n_1 + n_2 + n_3 \).

For the case of three samples the model in (3.1) can be written as

\[ Y_i = \begin{cases} 
\alpha + \beta_j + e_i & \text{if } i \in S_j, \quad j=1, 2 \\
\alpha + e_i & \text{if } i \in S_3.
\end{cases} \quad (3.2) \]

Assuming that the \( e_i \) are independent and identically distributed with unknown and continuous distribution function \( F \) for all \( i=1, \ldots, N \), then the \( Y_i \) are independent and identically distributed with distribution \( F(y - \alpha - \beta_j) \) for \( i \in S_j \), where \( \beta_3 \) is defined to be zero. Thus \( \alpha + \beta_j \) is the location parameter for the \( j \)-th sample. Consequently, \( \beta_1 \) is the shift between the distribution of the \( Y \)'s from the first sample and of the \( Y \)'s from the third sample. Similarly, \( \beta_2 \) is the shift between the distributions of samples 2 and 3; \( \beta_1 - \beta_2 \) is the shift between the distributions of samples 1 and 2.

We wish to estimate \( \beta=(\beta_1, \beta_2)^T \). An intuitive approach to this estimation problem
is to find the value of \( \beta \) which minimizes the dispersion of the residuals. Let
\[ Z_i = Y_i - X_i \beta \]
denote the i-th residual, where \( X_i \) denotes the i-th row of \( X \) corresponding to the covariate associated with individual i. We need a measure of the dispersion of \( Z_1, \ldots, Z_N \). The measure should be nonnegative, and it should be small when the \( Z_i \)'s are close to each other and large when the \( Z_i \)'s are dispersed. Let \( D(\widetilde{Z}) = D(Z_1, \ldots, Z_N) \) be a measure of the variability in \( Z_1, \ldots, Z_N \) satisfying the following properties:

(i.) \( D(-\widetilde{Z}) = D(\widetilde{Z}) \) \hspace{1cm} (3.3)

(ii.) \( D(\widetilde{Z} + c) = D(\widetilde{Z}) \) for any scalar c. \hspace{1cm} (3.4)

Then \( D(\cdot) \) is an even, translation-invariant measure of dispersion. By property (3.4),
\[ D(\widetilde{Y} - \alpha \widetilde{X} \beta) = D(\widetilde{Y} - \alpha \widetilde{X} \beta) \]
and consequently, \( D(\widetilde{Y} - \alpha \widetilde{X} \beta) \) is independent of \( \alpha \) for fixed \( \widetilde{Y} \) and \( \widetilde{X} \). Thus by minimizing \( D(\widetilde{Y} - \alpha \widetilde{X} \beta) \), we can estimate \( \beta \) but not \( \alpha \). Since the estimate of \( \beta \) is the value which minimizes \( D(\widetilde{Y} - \widetilde{X} \beta) \), different dispersion measures will generate different estimators for \( \beta \). If \( D(\widetilde{Z}) \) is the variance of \( \widetilde{Z} \), then minimizing \( D(\widetilde{Y} - \widetilde{X} \beta) \) yields the least squares estimator of \( \beta \). The least squares estimates are sensitive to outliers, and upon examining the corresponding dispersion measure, we see that it is a quadratic function of the residuals. Jaeckel (1972) proposed using a dispersion measure that is a linear function of the residuals to generate an estimate of \( \beta \) which is less sensitive to outliers. His dispersion measure can be obtained from the least squares dispersion measure by replacing the square of the residual by the product of the residual and a score based on the rank of the residual. Let us consider his dispersion function.

Let \( a_N(i), i = 1, \ldots, N \) be a set of constants called scores satisfying the following properties:
(i.) \( a_N(1) \leq a_N(2) \leq \ldots \leq a_N(N) \)  \hspace{1cm} (3.5)

(ii.) \( a_N(1) \neq a_N(N) \)  \hspace{1cm} (3.6)

(iii.) \( a_N(i) + a_N(N-i+1) = 0 \)  \hspace{1cm} (3.7)

(iv.) \( \sum_{i=1}^{N} a_N(i) = 0 \).  \hspace{1cm} (3.8)

By properties (3.5) through (3.8), the \( a_N(i) \) are a set of symmetric, nondecreasing scores that are not all equal and are centered about zero. Let \( R(Z_i) \) denote the rank of \( Z_i \) among \( Z_1, \ldots, Z_N \). Jaeckel's dispersion function is defined as

\[
D(Z) = \sum_{i=1}^{N} a_N(R(Z_i)) Z_i. \tag{3.9}
\]

Since \( Z_i = Y_i - X_i \beta \), the dispersion function as a function of \( \beta \) is as follows:

\[
D(Y - X\beta) = \sum_{i=1}^{N} a_N(R(Y_i - X_i \beta)) (Y_i - X_i \beta). \tag{3.10}
\]

Hence Jaeckel's dispersion measure is a linear combination of the ordered residuals.

Since \( R(Z_i + c) = R(Z_i) \) for any scalar \( c \) and since (3.8) holds, \( D(Z + c \tilde{1}) = D(Z) \), that is, \( D(\cdot) \) is translation-invariant. By (3.7), \( D(-Z) = D(Z) \) and hence \( D(\cdot) \) is even. Jaeckel showed that for fixed \( Y \) and \( X \), \( D(Y - X\beta) \) is a nonnegative, continuous, and convex function of \( \beta \). Therefore we find the value of \( \beta \) which minimizes \( D(Y - X\beta) \).

To minimize \( D(Y - X\beta) \), find the value of \( \beta \) for which the partial derivatives, if they exist, are equal to zero. The function \( D(Y - X\beta) \) is differentiable almost everywhere. The partial derivative with respect to \( \beta_1 \) is given by
\[
\frac{\delta}{\delta \beta_1} D(Y - X\beta) = - \sum_{i=1}^{N} a_N(R(Y_i - X_{i1}\beta_1 - X_{i2}\beta_2)) X_{i1}
\]

\[
= - \sum_{i=1}^{N} a_N(R(Y_i - X_{i1}\beta_1 - X_{i2}\beta_2)) (X_{i1} - \bar{X}_1),
\]

(3.11)

where (3.11) holds because \(a_N(\cdot)\) is a step function in \(\beta_1\), (3.12) holds by condition (3.8), and

\[\bar{X}_1 = \frac{1}{N} \sum_{j=1}^{N} X_{j1}.\]

Similarly, the partial derivative with respect to \(\beta_2\) is given by

\[
\frac{\delta}{\delta \beta_2} D(Y - X\beta) = - \sum_{i=1}^{N} a_N(R(Y_i - X_{i1}\beta_1 - X_{i2}\beta_2)) (X_{i2} - \bar{X}_2)
\]

(3.13)

Observe that the summations in (3.12) and (3.13) are just \(-\sqrt{N} S_j(Y - X\beta)\), \(j = 1, 2\) where \(S_j(\cdot)\) is the linear rank statistic defined in (2.17). Thus solving for the vector \(\hat{\beta}\) for which the partial derivatives equal zero is equivalent to finding the \(\hat{\beta}\) for which the vector function \(S(Y - X\hat{\beta}) = (S_1(Y - X\hat{\beta}), S_2(Y - X\hat{\beta}))\) is equal to the zero vector. For the rest of the dissertation, define \(S_j(Y - X\hat{\beta}) = -\delta/\delta \beta_j (D(Y - X\hat{\beta}))\), so that \(S_j\) is proportional to the linear rank statistic in (2.17). Let us look at \(S(Y - X\hat{\beta})\) more closely.

For the three-sample model, the formula for \(S_j(Y - X\hat{\beta})\) can be simplified because the matrix \(X\) of covariates has a special form. Recall that \(X_{ij}\) is an indicator for whether or not observation \(i\) was taken from sample \(j\). That is,

\[
X_{i1} = \begin{cases} 
1 & \text{if } i = 1, \ldots, n_1 \\
0 & \text{otherwise}
\end{cases}
\]

(3.14)

and

\[
X_{i2} = \begin{cases} 
1 & \text{if } i = n_1 + 1, \ldots, n_1 + n_2 \\
0 & \text{otherwise}
\end{cases}
\]

(3.15)
Hence for the three-sample model $\bar{X}_j = n_j/N$. So $S_1(\bar{Y} - \bar{X}_j\beta)$ can be written as

$$S_1(\bar{Y} - \bar{X}_j\beta) = \sum_{i=1}^{n_1} a_N(R(Y_i - X_i\beta)) \left(1 - \frac{n_1}{N}\right) + \sum_{i=n_1+1}^{N} a_N(R(Y_i - X_i\beta)) \left(-\frac{n_1}{N}\right)$$

(3.16)

$$= \sum_{i=1}^{n_1} a_N(R(Y_i - X_i\beta)),$$

(3.17)

where the second equality follows from (3.8). Similarly,

$$S_2(\bar{Y} - \bar{X}_j\beta) = \sum_{i=n_1+1}^{n_1+n_2} a_N(R(Y_i - X_i\beta)).$$

(3.18)

Thus $S_1(\cdot)$ and $S_2(\cdot)$ are just the sums of the scores corresponding to individuals in samples 1 and 2, respectively. The score for individual $i$ depends on the rank of the $i$-th residual among all $N$ residuals and hence evaluation of $S_1(\cdot)$ and $S_2(\cdot)$ requires a joint ranking procedure.

Many of the weighted logrank statistics used in an analysis with censored data are extensions of the Wilcoxon statistic, in which the scores are linear. We now examine the form of $S_1(\cdot)$ and $S_2(\cdot)$ for Wilcoxon scores. Let the $a_N(\cdot)$ denote the standardized Wilcoxon scores, namely,

$$a_N(i) = \frac{\sqrt{12}}{N+1} \left(i - \frac{N+1}{2}\right).$$

(3.19)

The scores $a_N(\cdot)$ in (3.19) satisfy properties (3.5) through (3.8). If these scores are substituted into (3.17) and (3.18), then $S_1(\cdot)$ and $S_2(\cdot)$ have the following form:
\[ S_j(Y - X\hat{\beta}) = \sum_{i \in S_j} \frac{\sqrt{12}}{N+1} \left( R(Y_i - X_i\hat{\beta}) - \frac{N+1}{2} \right) \]

\[ = \frac{\sqrt{12}}{N+1} \left( \sum_{i \in S_j} R(Y_i - X_i\hat{\beta}) - \frac{n_j(N+1)}{2} \right) \quad (3.20) \]

So for the Wilcoxon scores, \( S_j(\cdot) \) is proportional to the sum of the ranks in sample \( j \) minus a constant depending on the sample sizes. We next examine how to express \( S_j(Y - X\hat{\beta}) \) as a function in terms of rankings of pairs of samples.

Consider the sum of the ranks in the first sample. The term \( R(Y_i - X_i\hat{\beta}) \) for an observation in sample 1 is just the rank of \( Y_i - \beta_1 \) among all of the \( Y_j - X_j\hat{\beta}_1 - X_j\hat{\beta}_2 \) for \( j=1,...,N \). Assuming that the underlying distribution is continuous, then with probability one, the rank of \( Y_i - \beta_1 \) is just one plus the number of residuals smaller than \( Y_i - \beta_1 \). To find the number of residuals less than \( Y_i - \beta_1 \), we can consider each sample separately. The sum of ranks can be written in the following way:

\[ \sum_{i \in S_1} R(Y_i - X_i\hat{\beta}) = \sum_{i \in S_1} \left[ 1 + \#\{ m \in S_1 : Y_i - \beta_1 > Y_m - \beta_1 \} + \#\{ j \in S_2 : Y_i - \beta_1 > Y_j - \beta_2 \} + \#\{ k \in S_3 : Y_i - \beta_1 > Y_k \} \right]. \quad (3.21) \]

Letting

\[ \psi(u) = \begin{cases} 
1 & \text{if } u > 0 \\
0 & \text{otherwise} 
\end{cases} \]

then (3.21) is just
\[ \sum_{i \in S_1} R(Y_i - X_i \beta) = \]
\[ n_1 + \sum_{i \in S_1} \sum_{m \in S_1} \psi(Y_i - Y_m) + \sum_{i \in S_1} \sum_{j \in S_2} \psi(Y_i - Y_j - (\beta_1 \beta_2)) + \sum_{i \in S_1} \sum_{k \in S_3} \psi(Y_i - Y_k - \beta_1). \]  
\[ (3.23) \]

Let
\[ U_{1,2}(\beta) = \sum_{i \in S_1} \sum_{j \in S_2} \psi(Y_i - Y_j - (\beta_1 \beta_2)) \]  
\[ (3.24) \]
and
\[ U_{1,3}(\beta) = \sum_{i \in S_1} \sum_{k \in S_3} \psi(Y_i - Y_k - \beta_1). \]  
\[ (3.25) \]

Then the sum of the ranks in sample 1 can be expressed as
\[ \sum_{i \in S_1} R(Y_i - X_i \beta) = \frac{n_1(n_1+1)}{2} + U_{1,2}(\beta) + U_{1,3}(\beta). \]  
\[ (3.26) \]

Similarly, the sum of ranks in sample 2 can be expressed as
\[ \sum_{i \in S_2} R(Y_i - X_i \beta) = \frac{n_2(n_2+1)}{2} + n_1n_2 - U_{1,2}(\beta) + U_{2,3}(\beta) \]  
\[ (3.27) \]

where
\[ U_{2,3}(\beta) = \sum_{j \in S_2} \sum_{k \in S_3} \psi(Y_j - Y_k - \beta_2) \]  
\[ (3.28) \]
and
\[ \sum_{i \in S_1} \sum_{j \in S_2} \psi(Y_j - Y_i - (\beta_2 \beta_1)) = n_1n_2 - U_{1,2}(\beta) \]  
\[ (3.29) \]
with probability one assuming the underlying distribution $F$ is continuous. Hence for the
Wilcoxon scores, $S_1(\cdot)$ and $S_2(\cdot)$ have the following form:

$$S_1(Y - X\beta) = \sqrt{\frac{12}{N+1}} \left\{ \frac{n_1(n_1+1)}{2} + U_{1,2}(\beta) + U_{1,3}(\beta) - \frac{n_1(N+1)}{2} \right\}$$

$$= \frac{\sqrt{12}}{N+1} \left\{ U_{1,2}(\beta) - \frac{n_1n_2}{2} + U_{1,3}(\beta) - \frac{n_1n_3}{2} \right\}$$

and

$$S_2(Y - X\beta) = \sqrt{\frac{12}{N+1}} \left\{ \frac{n_2(n_2+1)}{2} + n_1n_2 - U_{1,2}(\beta) + U_{2,3}(\beta) - \frac{n_2(N+1)}{2} \right\}$$

$$= \frac{\sqrt{12}}{N+1} \left\{ \frac{n_1n_2}{2} - U_{1,2}(\beta) + U_{2,3}(\beta) - \frac{n_2n_3}{2} \right\}.$$  

(3.30)

In summary, finding the value of $\beta$ which minimizes (3.10) for the Wilcoxon scores
in (3.19) is equivalent to finding $\beta$ such that the vector $(S_1(Y - X\beta), S_2(Y - X\beta))$ equals
0 where $S_j(Y - X\beta) = \frac{\delta}{\delta \beta_j} D(Y - X\beta)$. By (3.30) and (3.31) this is equivalent to finding
the $\beta_1$ and $\beta_2$ that simultaneously solve the system of equations:

$$U_{1,2}(\beta) - \frac{n_1n_2}{2} + U_{1,3}(\beta) - \frac{n_1n_3}{2} = 0$$  

(3.32)

$$\frac{n_1n_2}{2} - U_{1,2}(\beta) + U_{2,3}(\beta) - \frac{n_2n_3}{2} = 0.$$  

(3.33)

For notational convenience, let the functions on the left-hand side of equations (3.32) and
(3.33) be denoted by

$$S_1^*(\beta) = S_1^*(Y - X\beta) = U_{1,2}(\beta) - \frac{n_1n_2}{2} + U_{1,3}(\beta) - \frac{n_1n_3}{2}$$

(3.34)
and

\[ S_2^*(\beta) = S_2^*(Y - X\beta) = \frac{n_1n_2}{2} - U_{1,2}(\beta) + U_{2,3}(\beta) - \frac{n_2n_3}{2} \]  \hspace{1cm} (3.35)

Let us examine some of the properties of \( S_1^*(\beta) \) and \( S_2^*(\beta) \). To do this we first examine the properties of the functions \( U_{1,2}(\beta) \), \( U_{1,3}(\beta) \), and \( U_{2,3}(\beta) \).

**Property 3.1:** The function \( U_{r,s}(\beta) \) is a nonincreasing right-continuous step function of \( \beta_r - \beta_s \), where \( \beta_3 \) is defined to be zero. Furthermore, \( U_{r,s}(\beta) \) jumps down at the pairwise differences \( Y_i - Y_j \), \( i \in S_r \) and \( j \in S_s \); \( U_{r,s}(\beta) \) only takes on integer values between 0 and \( n_rn_s \), inclusive.

Property 3.1 follows from the definition of \( U_{r,s}(\beta) \), since \( \psi(Y_i - Y_j - (\beta_r - \beta_s)) \) is a step function which jumps down when \( \beta_r - \beta_s = Y_i - Y_j \), for \( i \in S_r \), \( j \in S_s \). The function \( U_{r,s}(\beta) \) is a function of \( \beta \) through sample \( r \) and sample \( s \) alone.

**Property 3.2:** The function \( U_{r,s}(\emptyset) = \sum_{i \in S_r} \sum_{j \in S_s} \psi(Y_i - Y_j) \) is the Mann-Whitney form of the Mann-Whitney-Wilcoxon statistic for testing whether or not the location shift between two samples is identically zero.

Property 3.2 implies that \( U_{1,2}(\beta) \) is the Mann-Whitney statistic evaluated using the aligned data \( Y_1 - \beta_1, \ldots, Y_{n_1} - \beta_1, Y_{n_1+1} - \beta_2, \ldots, Y_{n_1+n_2} - \beta_2 \). A similar result is true for \( U_{1,3}(\beta) \) and \( U_{2,3}(\beta) \). Hence \( U_{r,s}(\beta) \) is a function based on a pairwise ranking procedure.

**Property 3.3:** Let \( D_{(1)}^{(r,s)} \leq D_{(2)}^{(r,s)} \leq \ldots \leq D_{(n_rn_s)}^{(r,s)} \) denote the ordered pairwise differences
between sample \( r \) and sample \( s \), \( Y_i - Y_j, i \in S_r, j \in S_s \). Let \( A \) be an integer between 0 and \( n_r n_s \). If \( U_{rs}(\beta) = A \), then

\[ \beta_r - \beta_s \in \left[ D^{(r,s)}(n_r n_s - A), D^{(r,s)}(n_r n_s - A + 1) \right] \]

(3.36)

with the convention that

\[ D^{(r,s)}(0) = -\infty \quad \text{and} \quad D^{(r,s)}(n_r n_s + 1) = \infty. \]

(3.37)

From (3.34), the function \( S_1^*(\beta) \) is a linear combination of the Mann-Whitney-Wilcoxon statistics that compare sample 1 with each of the other samples. As a consequence of Property 3.1, \( S_1^*(\beta) \) has the following property.

**Property 3.4:** If \( \beta_2 \) is held fixed, then \( S_1^*(\beta) \) is a nonincreasing step function of \( \beta_1 \) that jumps down at the pairwise differences between samples 1 and 3, \( Y_i - Y_k, i \in S_1, k \in S_3 \) and at the pairwise differences between sample 1 and the aligned sample 2, \( Y_i - Y_j + \beta_2, i \in S_1, j \in S_2 \). If \( \beta_1 \) is held fixed, then \( S_1^*(\beta) \) is a nondecreasing step function of \( \beta_2 \) that jumps up at \( \beta_1 - (Y_i - Y_j), i \in S_1, j \in S_2 \). If \( \beta_1 - \beta_2 \) is held fixed, then \( S_1^*(\beta) \) is a nonincreasing step function of \( \beta_1 \) that jumps down at the pairwise differences \( Y_i - Y_k, i \in S_1, k \in S_3 \). The function \( S_1^*(\beta) \) ranges in value from \(-n_1(n_2+n_3)/2\) to \(n_1(n_2+n_3)/2\).

From Property 3.4, \( S_1^*(\beta) \) satisfies the following property.

**Property 3.5:** The function \( S_1^*(\beta) \) is a step function which jumps whenever \( \beta_1 \) crosses one of the pairwise differences \( Y_i - Y_k, i \in S_1, k \in S_3 \) or \( \beta_1 - \beta_2 \) crosses one of the pairwise differences \( Y_i - Y_j, i \in S_1, j \in S_2 \).
Similarly, the function $S_2^*(\beta)$ is a linear combination of the Mann-Whitney-Wilcoxon statistics which compare sample 2 with each of the other samples and $S_2^*(\beta)$ has the following property.

**Property 3.6:** If $\beta_1$ is held fixed, then $S_2^*(\beta)$ is a nonincreasing step function of $\beta_2$ that jumps down at the pairwise differences between samples 2 and 3, $Y_j - Y_k, j \in S_2, k \in S_3$ and at the pairwise differences $\beta_1 - (Y_i - Y_j), i \in S_1, j \in S_2$. If $\beta_2$ is held fixed, then $S_2^*(\beta)$ is a nondecreasing step function of $\beta_1$ that jumps up at $Y_1 - Y_j + \beta_2, i \in S_1, j \in S_2$. If $\beta_1 - \beta_2$ is held fixed, then $S_2^*(\beta)$ is a nonincreasing step function of $\beta_2$ that jumps down at the pairwise differences $Y_j - Y_k, j \in S_2, k \in S_3$. The function $S_2^*(\beta)$ ranges in value from $-n_2(n_1+n_3)/2$ to $n_2(n_1+n_3)/2$.

From Property 3.6, $S_2^*(\beta)$ satisfies the following property:

**Property 3.7:** The function $S_2^*(\beta)$ is a step function which jumps whenever $\beta_2$ crosses one of the pairwise differences $Y_j - Y_k, j \in S_2, k \in S_3$ or $\beta_1 - \beta_2$ crosses one of the pairwise differences $Y_i - Y_j, i \in S_1, j \in S_2$.

We estimate $\beta$ by the $\beta_1$ and $\beta_2$ for which $S_1^*(\beta)$ and $S_2^*(\beta)$ are zero. From Properties 3.5 and 3.7, $S_j^*(\beta), j=1, 2$ is a step function and thus either there is a region on which $S_j^*(\beta)$ is equal to zero or there is a boundary on which $S_j^*(\beta)$ jumps across zero. The solution to the system of equations $S_j^*(Y-X\beta) = 0, j=1, 2$ is the set of $\beta$ values at which $S(Y-X\beta)$ equals or jumps across 0. In the following section we characterize the solution to these equations.
3.2 Characterization of the Jaeckel Estimator for the Three-Sample Case

In the previous section we defined the Jaeckel estimator as the solution to the two equations \( S_1^*(\beta) = 0 \) and \( S_2^*(\beta) = 0 \), where \( S_1^*(\beta) \) and \( S_2^*(\beta) \) are the step functions defined in (3.34) and (3.35), respectively. In this section we provide two characterizations of Jaeckel's estimator. The first characterization is in terms of the ordered differences between pairs of samples and the second is in terms of the geometry of the functions \( S_1^*(\beta) \) and \( S_2^*(\beta) \).

Recall from the discussion at the end of Section 3.1 that the Jaeckel estimator is the vector \( \beta \) such that either \( S_j^*(\beta) \) equals or jumps across zero, for \( j=1,2 \). The following theorem characterizes the solution to the Jaeckel estimating equations in terms of the ordered differences between pairs of samples. Before stating the theorem, we consider the criterion for determining when \( \beta \) is a solution of the Jaeckel estimating equations. For \( \beta \) to be a solution, one of the following conditions must hold:

(i.) \( S_1^*(\beta) = 0 \) and \( S_2^*(\beta) = 0 \),
(ii.) \( S_1^*(\beta) = 0 \) and \( S_2^*(\beta) \) jumps across 0,
(iii.) \( S_1^*(\beta) \) jumps across 0 and \( S_2^*(\beta) = 0 \), or
(iv.) \( S_1^*(\beta) \) and \( S_2^*(\beta) \) jump across 0. \hspace{1cm} (3.38)

Since from Property 3.5, \( S_1^*(\beta) \) jumps whenever \( \beta_1 \) crosses \( D_{(i)}^{(1,3)} \), \( i=1,...,n_1n_3 \) or whenever \( \beta_1 - \beta_2 \) crosses \( D_{(j)}^{(1,2)} \), \( j=1,...,n_1n_2 \), the vector \( \beta \) is a jump point of \( S_1^*(\beta) \) if \( S_1^*(\beta) < 0 \) and \( S_1^*((\beta_1 - \beta_2)^T) > 0 \). Similarly, from Property 3.7, \( S_2^*(\beta) \) jumps whenever \( \beta_2 \) crosses \( D_{(k)}^{(2,3)} \), \( k=1,...,n_2n_3 \) or whenever \( \beta_1 - \beta_2 \) crosses \( D_{(j)}^{(1,2)} \), \( j=1,...,n_1n_2 \). Thus \( \beta \) is a jump point of \( S_2^*(\beta) \) if \( S_2^*(\beta) < 0 \) and \( S_2^*((\beta_1 - \beta_2)^T) > 0 \). We now state the theorem.
Theorem 3.1: Let M be an integer between \( \max\{1, n_1(n_2-n_3)/2, n_2(n_1-n_3)/2\} \) and \( \min\{n_1n_2, n_1(n_2+n_3)/2, n_2(n_1+n_3)/2\} \). Let \( D^{(r,s)}_{(i)} \) denote the \( i \)-th ordered difference between sample \( r \) and sample \( s \), as defined in Property 3.3. Let \( v = \lfloor n_1(n_3-n_2)/2 \rfloor \) and \( \gamma = \lfloor (n_2(n_1+n_3)+1)/2 \rfloor \), where \( \lfloor \cdot \rfloor \) is the greatest integer less than or equal to \( \cdot \). Let \( I_1(M), I_2(M), I_3(M), \) and \( I_4(M) \) be the intervals defined by

\[
I_1(M) = \left[ D^{(1,3)}_{(v+M)}, D^{(1,3)}_{(v+M+1)} \right],
\]
\[
I_2(M) = \left[ D^{(2,3)}_{(\gamma-M)}, D^{(2,3)}_{(\gamma-M+1)} \right],
\]
\[
I_3(M) = \left[ D^{(1,2)}_{(n_1n_2-M)}, D^{(1,2)}_{(n_1n_2-M+1)} \right] \quad \text{and}
\]
\[
I_4(M) = \left[ D^{(1,3)}_{(v+M)} - D^{(2,3)}_{(\gamma-M+1)}, D^{(1,3)}_{(v+M+1)} - D^{(2,3)}_{(\gamma-M)} \right].
\] (3.39)

If the intervals \( I_3(M) \) and \( I_4(M) \) overlap, then a solution to \( S(Y-X(3)=0 \) occurs in the region \( I_1(M) \times I_2(M) \), where \( I_1(M) \times I_2(M) \) denotes the Cartesian product between \( I_1(M) \) and \( I_2(M) \).

Proof: Since \( M \) is an integer between 1 and \( n_1n_2 \), there exists, with probability one, a vector \( \underline{b} \) such that \( U_{1,2}(\underline{b}) = M \). By Property 3.3, the vector \( \underline{b} \) satisfies

\[
b_1-b_2 \in \left[ D^{(1,2)}_{(n_1n_2-M)}, D^{(1,2)}_{(n_1n_2-M+1)} \right] = I_3(M).
\] (3.40)

Substituting \( U_{1,2}(\underline{b}) = M \) into (3.34), \( S_1^*(\underline{b}) = M - n_1n_2/2 + U_{1,3}(\underline{b}) - n_1n_3/2 \). If
n_1(n_2+n_3) is even, then for S_1^*(b) to equal zero, the quantity U_{1,3}(b) must be equal to 
\[ n_1(n_2+n_3)/2 - M. \] If n_1(n_2+n_3) is odd, then S_1^*(b) jumps across zero if 
\[ U_{1,3}(b) < n_1(n_2+n_3)/2 - M \text{ and } U_{1,3}((b_1, \ b_2)^T) > n_1(n_2+n_3)/2 - M. \] Thus, for b to be a solution to the first estimating equation, 
\[ U_{1,3}(b) = [ n_1(n_2+n_3)/2 ] - M, \] and hence

\[ b_1 \in \left[ D \begin{pmatrix} 1,3 \\ n_1n_3 - [n_1(n_2+n_3)/2] + M \end{pmatrix}, D \begin{pmatrix} 1,3 \\ n_1n_3 - [n_1(n_2+n_3)/2] + M+1 \end{pmatrix} \right] = I_1(M). \quad (3.41) \]

Substituting U_{1,2}(b) = M into (3.35), S_2^*(b) = n_1n_2/2 - M + U_{2,3}(b) - n_2n_3/2. If 
n_2(n_1-n_3) is even, then for S_2^*(b) to equal zero, U_{2,3}(b) must be equal to 
\[ M - n_2(n_1-n_3)/2. \] If n_2(n_1-n_3) is odd, for S_2^*(b) to jump across zero, 
\[ U_{2,3}(b) < M - n_2(n_1-n_3)/2, \text{ and } U_{2,3}((b_1, \ b_2)^T) > M - n_2(n_1-n_3)/2. \] Thus, U_{2,3}(b) must 
equal M - [ (n_2(n_1-n_3)+1)/2 ]. This implies that for S_2^*(b) to equal 0 or jump across 0,

\[ b_2 \in \left[ D \begin{pmatrix} 2,3 \\ n_2n_3 - [(n_2(n_1-n_3)+1)/2] - M \end{pmatrix}, D \begin{pmatrix} 2,3 \\ n_2n_3 - [(n_2(n_1-n_3)+1)/2] - M+1 \end{pmatrix} \right] = I_2(M). \quad (3.42) \]

For b to be a solution to S(Y-X\beta) = 0, (3.41) and (3.42) must hold simultaneously. If 
these two conditions hold, then b_1-b_2 must satisfy

\[ b_1-b_2 \in \left[ D \begin{pmatrix} 1,3 \\ (\gamma+M) - D \begin{pmatrix} 2,3 \\ (\gamma-M+1) \end{pmatrix} \right] = I_4(M). \quad (3.43) \]

But by condition (3.40), b_1-b_2 \in I_3(M). Hence for a solution to occur, b_1-b_2 must 
satisfy condition (3.40) and (3.43) simultaneously, that is, I_3(M) and I_4(M) must 
overlap. A vector b in the region of overlap that satisfies conditions (3.41) and (3.42) 
also satisfies the estimating equations simultaneously and hence is a solution. But this 
region of overlap is just a subset of the region I_1(M) \times I_2(M). Hence if I_3(M) and I_4(M)
overlap, then a solution is contained in $I_1(M) \times I_2(M)$.

(Q.E.D.)

Theorem 3.1 describes the Jaeckel estimator of $\beta$ in terms of the ordered pairwise differences. Let us now consider a geometric description of the location of the Jaeckel estimator. First we describe the contours of $S_1^*(\beta)$ in the $(\beta_1, \beta_2)$ plane. Since $S_1^*(\beta)$ is a linear combination of $U_{1,2}(\beta)$ and $U_{1,3}(\beta)$, we consider the geometry of these two functions first. From Property 3.1, $U_{1,3}(\beta)$ is a function of $\beta_1$ alone which jumps down when $\beta_1$ increases past a pairwise difference, $D^{(1,3)}_{(i)}$, for $i=1,\ldots,n_1n_3$. If we plot the lines with equations $\beta_1=D^{(1,3)}_{(i)}$ for $i=1,\ldots,n_1n_3$, then the function $U_{1,3}(\beta)$ is constant in the region between two adjacent lines. A typical example is shown in Figure 1.

Similarly, $U_{1,2}(\beta)$ is a function of $\beta_1-\beta_2$ alone which jumps down when $\beta_1-\beta_2$ increases past a pairwise difference $D^{(1,2)}_{(j)}$, for $j=1,\ldots,n_1n_2$. If the lines with equations $\beta_1-\beta_2=D^{(1,2)}_{(j)}$ are plotted, then $U_{1,2}(\beta)$ is constant between two adjacent diagonal lines. Figure 2 shows the contours of $U_{1,2}(\beta)$ for a typical example.

The contours of $S_1^*(\beta)$ are easily determined from the contours of $U_{1,2}(\beta)$ and $U_{1,3}(\beta)$. For $S_1^*(\beta)$ to remain constant, if $U_{1,2}(\beta)$ decreases by one, then $U_{1,3}(\beta)$ must increase by one. Thus the contours of $S_1^*(\beta)$ are strings of parallelograms joined at the corners, and each parallelogram has boundaries defined by the four equations below:

$$
\begin{align*}
\beta_1&=D^{(1,3)}_{(i_1)}, & \beta_1&=D^{(1,3)}_{(i_1+1)}, & \beta_1-\beta_2&=D^{(1,2)}_{(j_1)}, & \beta_1-\beta_2&=D^{(1,2)}_{(j_1+1)},
\end{align*}
$$

for some $i_1$ and $j_1$. A typical contour of $S_1^*(\beta)$ is the shaded region shown in Figure 3.

Having described the contours of $S_1^*(\beta)$, we now consider the contours of $S_2^*(\beta)$. Following an argument similar to the one for $S_1^*(\beta)$, the contours of $S_2^*(\beta)$ are also
chains of parallelograms linked at the corners, with each parallelogram having boundaries defined by the equations below:

\[ \beta_2 = D_{(i_2)}^{(2,3)}, \quad \beta_2 = D_{(i_2+1)}^{(2,3)}, \quad \beta_1 - \beta_2 = D_{(i_2)}^{(1,2)}, \quad \text{and} \quad \beta_1 - \beta_2 = D_{(j_2+1)}^{(1,2)}, \]  

(3.45)

for some \( i_2 \) and \( j_2 \). The shaded region in Figure 4 is a contour of \( S_2^*(\beta) \) in a typical example.

In the equations (3.44) which define the boundaries for a parallelogram in the \( S_1^*(\beta) \) contour, there is a relationship between the indices \( i_1 \) and \( j_1 \). Similarly, in the equations (3.45) which define the boundaries for a parallelogram in the \( S_2^*(\beta) \) contour, there is a relationship between the indices \( i_2 \) and \( j_2 \). This relationship is described in the next theorem. This theorem is a technical result which will be useful later.

**Theorem 3.2:** Let \( M_1 \) be an integer between 0 and \( n_1(n_2+n_3) \), and let \( M_2 \) be an integer between 0 and \( n_2(n_1+n_3) \). The equations (3.44) which define the boundaries of the parallelograms in the contour on which

\[ S_1^*(\beta) = M_1 - \frac{n_1(n_2+n_3)}{2} \]  

(3.46)

satisfy the relationship

\[ i_1 + j_1 = n_1(n_2+n_3) - M_1, \]  

(3.47)

where \( i_1 = 0, 1, ..., n_1n_3 \) and \( j_1 = 0, 1, ..., n_1n_2 \). Similarly, the equations (3.45) which define the boundaries of the parallelograms in the contour on which
\[ S_2^*(\beta) = M_2 - \frac{n_2(n_1+n_3)}{2} \quad (3.48) \]

satisfy the relationship

\[ i_2 - j_2 = n_2n_3 - M_2 \quad (3.49) \]

where \( i_2 = 0, 1, \ldots, n_2n_3 \) and \( j_2 = 0, 1, \ldots, n_1n_2 \).

**Proof:** Since \( S_1^*(\beta) = M_1 - \frac{n_1(n_2+n_3)}{2} \) for \( \beta \) in the contour,

\[ U_{1,2}(\beta) + U_{1,3}(\beta) = M_1. \quad (3.50) \]

Since \( \beta \) is in the parallelogram, \( \beta_1 \in \left[ D \left( \frac{1,3}{i_1} \right), D \left( \frac{1,3}{i_1+1} \right) \right] \) and thus \( U_{1,3}(\beta) = n_1n_3 - i_1 \). Furthermore, for \( \beta \) in the parallelogram, \( \beta_1 - \beta_2 \in \left[ D \left( \frac{1,2}{i_1} \right), D \left( \frac{1,2}{i_1+1} \right) \right] \) and thus

\[ U_{1,2}(\beta) = n_1n_2 - j_1. \]

Substituting into (3.50) yields the relationship

\[ n_1n_2 - j_1 + n_1n_3 - i_1 = M_1, \]

which implies that \( n_1(n_2+n_3) - M_1 = i_1 + j_1 \).

The proof for \( i_2 \) and \( j_2 \) is similar to that of \( i_1 \) and \( j_1 \). Since \( S_2^*(\beta) = M_2 - \frac{n_2(n_1+n_3)}{2} \) for \( \beta \) in the contour,

\[ U_{2,3}(\beta) - U_{1,2}(\beta) = M_2 - n_1n_2. \quad (3.51) \]

Since \( \beta \) is in the parallelogram, \( \beta_2 \in \left[ D \left( \frac{2,3}{i_2} \right), D \left( \frac{2,3}{i_2+1} \right) \right] \) and \( U_{2,3}(\beta) = n_2n_3 - i_2 \). Also,

for \( \beta \) in the parallelogram, \( \beta_1 - \beta_2 \in \left[ D \left( \frac{1,2}{i_2} \right), D \left( \frac{1,2}{i_2+1} \right) \right] \) and thus \( U_{1,2}(\beta) = n_1n_2 - j_2 \).

Upon substitution for \( U_{2,3}(\beta) \) and \( U_{1,2}(\beta) \) in (3.51), \( n_2n_3 - i_2 + j_2 - n_1n_2 = M_2 - n_1n_2 \).
Figure 1. Graphical representation of the contours of the function $U_{1,3}(\beta)$ in the $(\beta_1, \beta_2)$ plane.
Figure 2. Graphical representation of the contours of the function $U_{1,2}(\beta)$ in the $(\beta_1, \beta_2)$ plane.
Figure 3. Graphical representation of the contours of the function $S_1^*(\beta)$

in the $(\beta_1, \beta_2)$ plane.
Figure 4. Graphical representation of the contours of the function $S_2^*(\beta)$ in the ($\beta_1, \beta_2$) plane.
Now that we have considered the contours for both $S_j^*(\beta)$ and $S_j^*(\beta)$, we examine the region in which the Jaeckel estimator is contained. Let the zero contour for $S_j^*(\beta)$, $j=1, 2$ be the region on which $S_j^*(\beta)$ equals zero. If $S_j^*(\beta)$ does not equal zero for any $\beta$, then the zero contour is defined to be the boundary at which $S_j^*(\beta)$ jumps from negative in one contour to positive in the adjacent contour. To find the region in which $\beta_1$ and $\beta_2$ satisfy the Jaeckel estimating equations, find the intersection of the zero contour for $S_1^*(\beta)$ and the zero contour for $S_2^*(\beta)$. In Figure 5 the zero contours for $S_1^*(\beta)$ and $S_2^*(\beta)$ are shown and the region in which the Jaeckel estimating equations are satisfied is marked.

3.3 Relationship between the Jaeckel Estimator and the Two-Sample Hodges-Lehmann Estimators

After having characterized the Jaeckel estimator in Section 3.2, we consider the relationship between the Jaeckel estimator and the Hodges-Lehmann two-sample estimators. This relationship will be useful in the computation of the Jaeckel estimate, described in Chapter V. Before considering this relationship, we discuss the incompatibility of the Hodges-Lehmann two-sample estimators.

Recall from Property 3.2 that $U_{r,s}(0)$ is the Mann-Whitney-Wilcoxon statistic used to test whether the location shift between the distribution of samples $r$ and $s$ is zero. As discussed in Chapter II, Hodges and Lehmann (1963) showed that the estimator of $\beta_r-\beta_s$ obtained by inverting $U_{r,s}(\beta)$ is

$$b_{r,s} = \text{med} \ (Y_i - Y_j), \ i \in S_r, j \in S_s$$

$$= \text{med} \ D_{(k)}^{(r,s)}, \ k=1, \ldots, n_r n_s. \quad (3.52)$$
Figure 5. Graphical representation of the Jaeckel estimator as the intersection of the zero contours of the functions $S_1^*(\beta)$ and $S_2^*(\beta)$. 
In general, $b_{1,2}, b_{1,3},$ and $b_{2,3}$ are incompatible, that is $b_{1,3} \neq b_{1,2} + b_{2,3}$. Let us consider the geometry of $b_{1,2}, b_{1,3},$ and $b_{2,3}$ in the $(\beta_1, \beta_2)$ plane. If $n_r n_s$ is odd, then

$$b_{r,s} = D^{(r,s)}_k$$
where $k = \frac{n_r n_s + 1}{2}$. \hspace{1cm} (3.53)

If $n_r n_s$ is even, then $b_{r,s}$ is any number in the interval

$$\left[ D^{(r,s)}_0, D^{(r,s)}_{(j+1)} \right]$$
where $j = \frac{n_r n_s}{2}$. \hspace{1cm} (3.54)

The usual convention is to take as the median the midpoint of the interval, namely

$$b_{r,s} = \frac{D^{(r,s)}_j + D^{(r,s)}_{(j+1)}}{2}$$
\hspace{1cm} (3.55)

where $j = \frac{n_r n_s}{2}$. For the purpose of this dissertation, however, $b_{r,s}$ shall be defined as any point in the interval (3.54).

Figure 6 illustrates the incompatibility of $b_{1,2}, b_{1,3},$ and $b_{2,3}$ for a typical example with $n_i$ odd, $i=1, 2, 3$. Here the equations

$$\beta_r - \beta_s = D^{(r,s)}_j$$, $j=1, \ldots, n_r n_s$ \hspace{1cm} (3.56)

are graphed; recall that $\beta_3$ is equal to zero, allowing the graphing of the equations in the $(\beta_1, \beta_2)$ plane. The lines corresponding to $\beta_1 = b_{1,3}, \beta_2 = b_{2,3},$ and $\beta_1 - \beta_2 = b_{1,2}$ are marked.
Figure 6. Graphical illustration of the incompatibility of the two-sample Hodges-Lehmann estimators of the location shift between pairs of samples in configuration I.
The points labelled A, B, and C in Figure 6 have the following coordinates:

\[
A = (b_{1,3}, b_{2,3}) \\
B = (b_{1,2} + b_{2,3}, b_{2,3}) \\
C = (b_{1,3}, b_{1,3} - b_{1,2}). \tag{3.57}
\]

Each point A, B, and C is an estimate of \((\beta_1, \beta_2)\) obtained from Hodges-Lehmann estimation using pairs of samples. For example, for point C, \(\beta_1\) is estimated by the Hodges-Lehmann estimator of shift from samples 1 and 3, and \(\beta_1 - \beta_2\) is estimated by Hodges-Lehmann estimation using samples 1 and 2. The estimate of \(\beta_2\) is then the difference between the estimate of \(\beta_1\) and the estimate of \(\beta_1 - \beta_2\). If \(b_{1,3} = b_{1,2} + b_{2,3}\) then points A, B, and C coincide. In general, pairwise Hodges and Lehmann estimation of \(\beta_1\) and \(\beta_2\) will lead to points A, B, and C with different coordinates as in Figure 6. This illustrates the incompatibility property of the \(b_{r,s}\)'s.

In Figure 7, the equations (3.56) are graphed for an example with \(n_j\) even, \(i = 1, 2, 3\). The estimates \(b_{1,2}, b_{1,3},\) and \(b_{2,3}\) are in the shaded regions marked in the figure. The points A, B, and C have the coordinates

\[
A = (b_{1,3}^*, b_{2,3}^*) \\
B = (b_{1,2}^* + b_{2,3}^*, b_{2,3}^*) \\
C = (b_{1,3}^*, b_{1,3}^* - b_{1,2}^*). \tag{3.58}
\]

where \(b_{r,s}^*\) denotes one of the endpoints of the interval \([D^{(r,s)}(m), D^{(r,s)}(m+1)}\) where \(m = n_in_j/2\), namely, \(b_{1,2}^* = D^{(1,2)}(i), b_{1,3}^* = D^{(1,3)}(i),\) and \(b_{2,3}^* = D^{(2,3)}(k)\). Here the points A, B, and C denote the most extreme
incompatibility that could arise from selecting a value for $b_{r,s}$ in the interval

$$\left[ D_{(m)}^{(r,s)}, D_{(m+1)}^{(r,s)} \right] m = n_1 n_2/2. $$

The cases when some $n_i$ are odd and some $n_i$ are even are some combination of Figures 6 and 7.

Figures 6 and 7 illustrate the two possible configurations for the points A, B, and C. In Figure 6 the point A lies above the line on which points B and C lie. In terms of the estimates, for this configuration $b_{1,3} < b_{1,2} + b_{2,3}$. This case will be referred to as configuration I. For the case in which $b_{1,3} > b_{1,2} + b_{2,3}$, as illustrated in Figure 7, the point A lies below the line connecting points B and C. This case will be referred to as configuration II.

Figure 7 also illustrates that the convention for the choice of median when $n_1 n_2$ is even does not create the incompatibility of the two-sample Hodges-Lehmann estimates of $\beta_1$ and $\beta_2$. In this figure, $R_A$ denotes the intersection of the median band for the differences between samples 1 and 3 and the median band for the differences between samples 2 and 3. Any point $(b_1, b_2)$ in $R_A$ satisfies the condition that $b_1$ is a median of $D_{(i)}^{(1,3)}$, $i=1,\ldots, n_1 n_3$ and $b_2$ is a median of $D_{(j)}^{(2,3)}$, $j=1,\ldots, n_2 n_3$. Similarly, for any point $(b_1, b_2)$ in $R_B$, $b_2$ is a median of $D_{(j)}^{(2,3)}$, $j=1,\ldots, n_2 n_3$ and $b_1-b_2$ is a median of $D_{(k)}^{(1,2)}$, $k=1,\ldots, n_1 n_2$. Finally, any point $(b_1, b_2)$ in $R_C$ satisfies the condition that $b_1$ is a median of $D_{(i)}^{(1,3)}$, $i=1,\ldots, n_1 n_3$ and $b_1-b_2$ is a median of $D_{(k)}^{(1,2)}$, $k=1,\ldots, n_1 n_2$. Since for this example, regions $R_A$, $R_B$, and $R_C$ do not intersect simultaneously, the two-sample Hodges-Lehmann estimates do not coincide. This example shows the general incompatibility of the two-sample Hodges-Lehmann estimators. For a given set of data, the regions $R_A$, $R_B$, and $R_C$ may intersect, and we will refer to the estimates in this case as compatible.
Figure 7. Graphical illustration of the incompatibility of the two-sample Hodges-Lehmann estimators of the location shift between pairs of samples in configuration II.
We now consider the relationship between the two-sample Hodges-Lehmann estimates of $\beta_1$ and $\beta_2$, and the Jaeckel estimates of these parameters. We first consider the special case in which the Hodges-Lehmann estimates are compatible and then consider the general case. Before stating the theorem for the special case, we recall from the discussion preceding Theorem 3.1, that for $\beta$ to be a solution to the first estimating equation, either $S_i^*(\beta) = 0$ or $S_i^*(\beta)$ jumps across 0. The vector $\beta$ is a point at which $S_i^*(\beta)$ jumps across 0 if $S_i^*(\beta) < 0$ and $S_i^*((\beta_1, \beta_2)^T) > 0$. A similar condition holds for $S_2^*(\beta)$. We now state the theorem for the relationship between the Hodges-Lehmann two-sample estimates and the Jaeckel estimates in the special case.

**Theorem 3.3:** Let $b=(b_1, b_2)^T$ satisfy the conditions:

(i.) $b_1$ is a median of $D^{(1,3)}_{(i)}$, $i=1, \ldots, n_1n_3$

(ii.) $b_2$ is a median of $D^{(2,3)}_{(j)}$, $j=1, \ldots, n_2n_3$

(iii.) $b_1-b_2$ is a median of $D^{(1,2)}_{(k)}$, $k=1, \ldots, n_1n_2$.  

(3.59)

Then $b$ satisfies the Jaeckel estimating equations.

**Proof:** The proof will be split into cases, depending on whether $n_i$ is even or odd for $i=1, 2, and 3$.

**Case 1:** Suppose $n_1$, $n_2$, and $n_3$ are even. The products $n_1n_2$, $n_1n_3$, and $n_2n_3$ are also even. From conditions (i.), (ii.), and (iii.),

$$U_{1,2}(\beta) = \frac{n_1n_2}{2}, \quad U_{1,3}(b) = \frac{n_1n_3}{2}, \quad \text{and} \quad U_{2,3}(b) = \frac{n_2n_3}{2}. \quad (3.60)$$

Evaluating $S_1^*(\cdot)$ and $S_2^*(\cdot)$ at $\beta$ gives
\[ S_1^*(b) = U_{1,2}(b) - \frac{n_1n_2 + n_1n_3}{2} = 0 \quad (3.61) \]

and

\[ S_2^*(b) = \frac{n_1n_2}{2} - U_{1,2}(b) + U_{2,3}(b) - \frac{n_2n_3}{2} = 0. \quad (3.62) \]

Hence \( b \) is a Jaeckel estimate.

**Case 2:** Suppose \( n_1, n_2, \) and \( n_3 \) are odd. The products \( n_1n_2, n_1n_3, \) and \( n_2n_3 \) are also odd. From conditions (i.), (ii.), and (iii.),

\[ U_{1,2}(b) = \frac{n_1n_2 - 1}{2}, \quad U_{1,3}(b) = \frac{n_1n_3 - 1}{2}, \quad \text{and} \quad U_{2,3}(b) = \frac{n_2n_3 - 1}{2}. \quad (3.63) \]

Evaluating \( S_1^*(\cdot) \) and \( S_2^*(\cdot) \) at \( b \) gives

\[ S_1^*(b) = U_{1,2}(b) - \frac{n_1n_2 + n_1n_3}{2} = -1 \quad (3.64) \]

and

\[ S_2^*(b) = \frac{n_1n_2}{2} - U_{1,2}(b) + U_{2,3}(b) - \frac{n_2n_3}{2} = 0. \quad (3.65) \]

Evaluating \( S_1^*(\cdot) \) at \( (b_1^-, b_2^T) \) gives

\[ S_1^*((b_1^-, b_2^T)) = U_{1,2}((b_1^-, b_2^T)) - \frac{n_1n_2}{2} + U_{1,3}((b_1^-, b_2^T)) - \frac{n_1n_3}{2} = 1 \quad (3.66) \]

since

\[ U_{1,2}((b_1^-, b_2^T)) = \frac{n_1n_2 + 1}{2}, \quad U_{1,3}((b_1^-, b_2^T)) = \frac{n_1n_3 + 1}{2}, \quad \text{and} \quad U_{2,3}((b_1^-, b_2^T)) = \frac{n_2n_3 - 1}{2}. \quad (3.67) \]

Thus \( S_2^*(\cdot) \) equals 0 at \( b \) and \( S_1^*(\cdot) \) jumps across 0 at \( b \). Hence \( b \) satisfies the Jaeckel
estimating equations.

Case 3: Suppose \( n_1 \) is even, and \( n_2 \) and \( n_3 \) are odd. The products \( n_1n_2 \) and \( n_1n_3 \) are even, and \( n_2n_3 \) is odd. From conditions (i.), (ii.), and (iii.),

\[
U_{1,2}(b) = \frac{n_1n_2}{2}, \quad U_{1,3}(b) = \frac{n_1n_3}{2}, \quad \text{and} \quad U_{2,3}(b) = \frac{n_2n_3 - 1}{2}.
\]  

(3.68)

Evaluating \( S_1^*(\cdot) \) and \( S_2^*(\cdot) \) at \( b \) gives

\[
S_1^*(b) = U_{1,2}(b) - \frac{n_1n_2}{2} + U_{1,3}(b) - \frac{n_1n_3}{2} = 0
\]  

(3.69)

and

\[
S_2^*(b) = \frac{n_1n_2}{2} - U_{1,2}(b) + U_{2,3}(b) - \frac{n_2n_3}{2} = -1/2.
\]  

(3.70)

Evaluating \( S_2^*(\cdot) \) at \((b_1, b_2^-)^T\) gives

\[
S_2^*((b_1, b_2^-)^T) = \frac{n_1n_2}{2} - U_{1,2}((b_1, b_2^-)^T) + U_{2,3}((b_1, b_2^-)^T) - \frac{n_2n_3}{2} \geq 1/2
\]  

(3.71)

since

\[
U_{1,2}((b_1, b_2^-)^T) \leq \frac{n_1n_2}{2}, \quad U_{1,3}((b_1-, b_2)^T) = \frac{n_1n_3}{2},
\]

and \( U_{2,3}((b_1-, b_2)^T) = \frac{n_2n_3 + 1}{2} \). \hspace{1cm} (3.72)

Thus \( S_1^*(\cdot) \) equals 0 at \( b \) and \( S_2^*(\cdot) \) jumps across 0 at \( b \). Hence \( b \) is a solution to the Jaeckel estimating equations.

The remaining cases can be argued as in case 3. Thus \( b \) satisfies the Jaeckel estimating equations.

(Q.E.D.)
The geometric interpretation of Theorem 3.3 is that if the points A, B, and C in Figure 6 coincide, then they are the Jaeckel estimate of \( \beta \). If, as in Figure 7, there is no unique median \( b_{r,s} \), then Theorem 3.3 states that any point which is in the intersection of the regions corresponding to the medians \( b_{1,2}, b_{1,3}, \) and \( b_{2,3} \) is also a Jaeckel estimate of \( \beta \). Thus, by Theorem 3.3 we have shown that if there are compatible Hodges-Lehmann two-sample estimates, then these estimates satisfy the Jaeckel estimating equations.

Since Theorem 3.3 shows the relationship between the two-sample Hodges-Lehmann estimator and the Jaeckel estimator in the case of compatibility, we now consider the case in which the two-sample estimates are not compatible. If the two-sample Hodges-Lehmann estimates are incompatible, then either configuration I or configuration II holds. Let A, B, and C be the points with coordinates

\[
A = (b_{1,3}, b_{2,3}) \\
B = (b_{1,2} + b_{2,3}, b_{2,3}) \\
C = (b_{1,3}, b_{1,3} - b_{1,2}),
\]

where \( b_{1,3}, b_{2,3}, \) and \( b_{1,2} \) are medians of the pairwise differences between samples 1 and 3, between samples 2 and 3, and between samples 1 and 2, respectively. The choice of medians depends on the configuration. Let

\[
i_1 = \lfloor \frac{n_1 n_3 + 1}{2} \rfloor, \quad j_1 = \lfloor \frac{n_2 n_3 + 1}{2} \rfloor, \quad \text{and} \quad k_1 = \lfloor \frac{n_1 n_2 + 1}{2} \rfloor
\]

and

\[
i_2 = \lfloor \frac{n_1 n_3 + 1}{2} \rfloor, \quad j_2 = \lfloor \frac{n_2 n_3 + 1}{2} \rfloor, \quad \text{and} \quad k_2 = \lfloor \frac{n_1 n_2 + 1}{2} \rfloor.
\]

If \( n_1 n_3 \) is odd then \( i_1 \) and \( i_2 \) are both equal to \( (n_1 n_3 + 1)/2 \). Similarly for the \( j \)'s and \( k \)'s. If
If \( D_{(1,3)} - D_{(2,3)} < D_{(1,2)} \), then configuration I holds. For this configuration, the medians are

\[
b_{1,3} = D_{(1,3)}
\]

\[
b_{2,3} = D_{(2,3)}
\]

and

\[
b_{1,2} = D_{(1,2)}
\]  \hspace{1cm} (3.76)

If \( D_{(1,3)} - D_{(2,3)} > D_{(1,2)} \), then configuration II holds. The medians for this configuration are

\[
b_{1,3} = D_{(1,3)}
\]

\[
b_{2,3} = D_{(2,3)}
\]

and

\[
b_{1,2} = D_{(1,2)}
\]  \hspace{1cm} (3.77)

Thus, if \( n_1 n_3 \) is odd, then \( b_{r,s} \) is \( D_{(r,s)} (n_1 n_3 + 1)/2 \), for both configurations. If \( n_1 n_3 \) is even, then \( b_{r,s} \) is \( D_{(r,s)} (n_1 n_3/2) \) or \( D_{(r,s)} (n_1 n_3/2 + 1) \) depending on the configuration. In the case of \( n_1 n_3 \) even, the median \( b_{r,s} \) is chosen so that the points A, B, and C are the most extreme, that
is, so that the triangle with vertices A, B, and C is the largest it can be. We now state the theorem for the relationship between the two-sample Hodges-Lehmann estimators and the Jaeckel estimator for the general case.

**Theorem 3.4:** Let

\[
A = (b_{1,3}, b_{2,3})
\]
\[
B = (b_{1,2} + b_{2,3}, b_{2,3})
\]
\[
C = (b_{1,3}, b_{1,3} - b_{1,2})
\]

where \( b_{1,2}, b_{1,3}, \) and \( b_{2,3} \) are as defined in the preceding paragraph. Then the triangle with vertices A, B, and C contains the Jaeckel estimate of \( \beta \).

**Proof:** The proof is by contradiction. Let \( \beta = (b_1, b_2) \) denote the Jaeckel estimate. The lines of the triangle with vertices A, B, and C partition the \( (\beta_1, \beta_2) \) plane into seven regions and their boundaries. The regions will be denoted in the following manner.

\[
R_1 = \{(\beta_1, \beta_2): \beta_1 < b_{1,3}, \beta_2 > b_{2,3}, \text{ and } \beta_1 - \beta_2 < b_{1,2}\},
\]
\[
R_2 = \{(\beta_1, \beta_2): \beta_1 > b_{1,3}, \beta_2 > b_{2,3}, \text{ and } \beta_1 - \beta_2 < b_{1,2}\},
\]
\[
R_3 = \{(\beta_1, \beta_2): \beta_1 > b_{1,3}, \beta_2 > b_{2,3}, \text{ and } \beta_1 - \beta_2 > b_{1,2}\},
\]
\[
R_4 = \{(\beta_1, \beta_2): \beta_1 > b_{1,3}, \beta_2 < b_{2,3}, \text{ and } \beta_1 - \beta_2 > b_{1,2}\},
\]
\[
R_5 = \{(\beta_1, \beta_2): \beta_1 < b_{1,3}, \beta_2 < b_{2,3}, \text{ and } \beta_1 - \beta_2 > b_{1,2}\},
\]
\[
R_6 = \{(\beta_1, \beta_2): \beta_1 < b_{1,3}, \beta_2 < b_{2,3}, \text{ and } \beta_1 - \beta_2 < b_{1,2}\}, \text{ and}
\]
\[
R_7 = \{(\beta_1, \beta_2): \beta \text{ lies on or within the triangle } ABC\}.
\]

Assume \( \beta \) is not contained in the triangle ABC. Then \( \beta \) must lie outside of the triangle, namely in the union of the regions \( R_1, R_2, ..., R_6 \) or on the boundaries between them.
Figure 8 shows regions R1, R2,..., R7 for the two configurations of the triangle.

We will show that for any vector $\beta$ in R1, R2,..., R6 or on the boundaries between them, at least one of the functions $S_j^*(\beta)$, for $j=1$ or 2, does not equal or jump across zero and hence $\beta$ cannot be a solution. Recall from the discussion preceding Theorem 3.3 that for $\beta$ to be a solution to the first Jaeckel equation, either $S_1^*(\beta) = 0$ or $S_1^*((\beta_1, \beta_2)^T) > 0$ and $S_1^*(\beta) < 0$. A similar condition is needed for $\beta$ to satisfy the second equation. In checking if a vector $\beta$ from outside of the triangle is a solution, we will consider the regions in six groups in (i.) through (vi.) below.

Figure 8. Graphical representation of the regions R1,..., R7 for the proof of Theorem 3.4.
(i.) Suppose \( \beta_1 < b_{1,3} \) and \( \beta_1 - \beta_2 < b_{1,2} \), that is \( \beta \) is in \( R_1 \), \( R_6 \), or on the boundary between \( R_1 \) and \( R_6 \), namely \( \{(\beta_1, \beta_2) : \beta_1 < b_{1,3}, \beta_2 = b_{2,3}, \text{ and } \beta_1 - \beta_2 < b_{1,2}\} \).

**Configuration I:** If triangle ABC is in configuration I, then

\[
U_{1,3}(\beta) = \begin{cases} 
\frac{n_1 n_3}{2} & \text{if } n_1 n_3 \text{ is even since } \beta_1 < D^{(1,3)}_{(0)}, i = \frac{n_1 n_3}{2} \\
\geq \frac{n_1 n_3 + 1}{2} & \text{if } n_1 n_3 \text{ is odd since } \beta_1 < D^{(1,3)}_{(0)}, i = \frac{n_1 n_3 + 1}{2}.
\end{cases}
\]

(3.80)

So, \( U_{1,3}(\beta) > \frac{n_1 n_3}{2} \). Also,

\[
U_{1,2}(\beta) = \begin{cases} 
\geq \frac{n_1 n_2}{2} & \text{if } n_1 n_2 \text{ is even since } \beta_1 - \beta_2 < D^{(1,2)}_{(0)}, j = \frac{n_1 n_2}{2} + 1 \\
\geq \frac{n_1 n_2 + 1}{2} & \text{if } n_1 n_2 \text{ is odd since } \beta_1 - \beta_2 < D^{(1,2)}_{(0)}, j = \frac{n_1 n_2 + 1}{2}.
\end{cases}
\]

(3.81)

So, \( U_{1,2}(\beta) \geq \frac{n_1 n_2}{2} \). Hence \( S_1^*(\beta) = U_{1,2}(\beta) - \frac{n_1 n_2}{2} + U_{1,3}(\beta) - \frac{n_1 n_3}{2} > 0 \).

**Configuration II:** If triangle ABC is in configuration II, then

\[
U_{1,3}(\beta) = \begin{cases} 
\geq \frac{n_1 n_3}{2} & \text{if } n_1 n_3 \text{ is even since } \beta_1 < D^{(1,3)}_{(0)}, i = \frac{n_1 n_3}{2} + 1 \\
\geq \frac{n_1 n_3 + 1}{2} & \text{if } n_1 n_3 \text{ is odd since } \beta_1 < D^{(1,3)}_{(0)}, i = \frac{n_1 n_3 + 1}{2}.
\end{cases}
\]

(3.82)
So, $U_{1,3}(\beta) \geq \frac{n_{1n_3}}{2}$. Also,

$$U_{1,2}(\beta) = \begin{cases} \frac{n_{1n_2}}{2} & \text{if } n_{1n_2} \text{ is even since } \beta_1 - \beta_2 < D_{(1,2)}^{(1,2)}, j = \frac{n_{1n_2}}{2} \\ \geq \frac{n_{1n_2}+1}{2} & \text{if } n_{1n_2} \text{ is odd since } \beta_1 - \beta_2 < D_{(1,2)}^{(1,2)}, j = \frac{n_{1n_2}+1}{2} \end{cases}$$

(3.83)

So, $U_{1,2}(\beta) > \frac{n_{1n_2}}{2}$. Hence $S_1^*(\beta) > 0$.

Thus in regions R1, R6, and the boundary between them, $S_1^*(\beta) > 0$. This $\beta$ is not the jump point at which $S_1^*(\beta)$ crosses 0 since $S_1^*((\beta_1+,\beta_2)^T) > 0$, following from the right-continuity and monotonicity of $S_1^*(\beta)$ and the definition of the regions. Hence $\beta$ is not the Jaeckel estimate.

(ii.) Suppose $\beta_1 > b_1,3$ and $\beta_1 - \beta_2 > b_{1,2}$, that is, $\beta$ is in R3, R4, or on the boundary between them, namely $((\beta_1, \beta_2): \beta_1 > b_1,3, \beta_2 > b_2,3, \text{ and } \beta_1 - \beta_2 = b_{1,2})$. The proof in this case is similar to the proof in (i.) with all inequalities reversed.

**Configuration I:** Here, $U_{1,3}(\beta) \leq n_{1n_3}/2$, since $\beta_1 > D_{(0)}^{(1,3)}$, where $i = \lfloor (n_{1n_3}+1)/2 \rfloor$. Also, $U_{1,2}(\beta) < n_{1n_2}/2$, since $\beta_1 - \beta_2 > D_{(0)}^{(1,2)}$, where $j = \lfloor (n_{1n_2}+1)/2 \rfloor$. Hence $S_1^*(\beta) < 0$.

**Configuration II:** For this configuration, the function $U_{1,3}(\beta) < n_{1n_3}/2$, since $\beta_1 > D_{(0)}^{(1,3)}$, where $i = \lfloor (n_{1n_3}/2) +1 \rfloor$. Also, $U_{1,2}(\beta) \leq n_{1n_2}/2$, since $\beta_1 - \beta_2 > D_{(0)}^{(1,2)}$, where $j = \lfloor (n_{1n_2}+1)/2 \rfloor$. Hence $S_1^*(\beta) < 0$. 


Thus in regions R3, R4, and the boundary between them, $S_1^* (\beta) < 0$. This $\beta$ is not the jump point at which $S_1^* (\cdot)$ crosses 0 since there exists $(\beta_1', \beta_2)$ such that $\beta_1' < \beta_1$, but $\beta_1' > b_{1,3}$ and $\beta_1' - \beta_2 > b_{1,2}$. By (ii.), $S_1^* ((\beta_1', \beta_2)^T) < 0$ which implies $S_1^* ((\beta_1 - \beta_2)^T) < 0$ so that $S_1^* (\cdot)$ does not cross 0 at $\beta$. Thus the Jaeckel estimate is not in regions R3 or R4 or on their boundary.

(iii.) Suppose $\beta_2 < b_{2,3}$ and $\beta_1 - \beta_2 > b_{1,2}$, that is $\beta$ is in R4, R5, or on the boundary between R4 and R5, defined as $(\beta_1, \beta_2): \beta_1 = b_{1,3}, \beta_2 < b_{2,3},$ and $\beta_1 - \beta_2 > b_{1,2}$.

**Configuration I:** If triangle ABC is in configuration I, then $U_{2,3} (\beta) \geq n_2 n_3 / 2$ since $\beta_2 < D (^{(2,3)}_0), i = \lfloor (n_2 n_3 / 2) + 1 \rfloor$. Also, $U_{1,2} (\beta) < \frac{n_1 n_2}{2}$ as in (ii.). Hence

$$S_2^* (\beta) = \frac{n_1 n_2}{2} - U_{1,2} (\beta) + U_{2,3} (\beta) - \frac{n_2 n_3}{2} > 0.$$ 

**Configuration II:** If triangle ABC is in configuration II, then $U_{2,3} (\beta) > n_2 n_3 / 2$ since $\beta_2 < D (^{(2,3)}_0), i = \lfloor (n_2 n_3 / 2) + 1 \rfloor$. Also, $U_{1,2} (\beta) \leq \frac{n_1 n_2}{2}$ as in (ii.). Hence $S_2^* (\beta) > 0$.

Thus in regions R4, R5, and the boundary between them, $S_2^* (\beta) > 0$. This $\beta$ is not the jump point at which $S_2^* (\cdot)$ crosses 0 since $S_2^* ((\beta_1, \beta_2)^T) > 0$, following from the right-continuity of $S_2^* (\beta)$, the monotonicity of $S_2^* (\beta)$, and the definition of the regions. Hence $\beta$ is not the Jaeckel estimate.

(iv.) Suppose $\beta_2 > b_{2,3}$ and $\beta_1 - \beta_2 < b_{1,2}$, that is $\beta$ is in R1, R2, or on the boundary between them, namely $(\beta_1, \beta_2): \beta_1 = b_{1,3}, \beta_2 > b_{2,3},$ and $\beta_1 - \beta_2 < b_{1,2}$). This proof is similar to (iii.) with all inequalities reversed.

**Configuration I:** Here, $U_{2,3} (\beta) < n_2 n_3 / 2$, since $\beta_2 > D (^{(2,3)}_0)$, with $i = \lfloor (n_2 n_3 / 2) + 1 \rfloor$. 

Also, \( U_{1,2}(\beta) \geq \frac{n_1 n_2}{2} \) as in (i.). Hence \( S_2^*(\beta) < 0 \).

**Configuration II:** For this configuration, \( U_{2,3}(\beta) \leq n_2 n_3/2 \), since \( \beta_2 > D^{(2,3)}(i) \), where 
\[
i = \lfloor (n_2 n_3 + 1)/2 \rfloor.
\] Also, \( U_{1,2}(\beta) > \frac{n_1 n_2}{2} \) as in (i.). Therefore, \( S_2^*(\beta) < 0 \).

Thus in regions \( R_1 \), \( R_2 \), and the boundary between them, \( S_2^*(\beta) < 0 \). Following the argument for \( S_1^*(\cdot) \), \( \beta \) is not the jump point at which \( S_2^*(\cdot) \) crosses 0 since there exists \( (\beta_1, \beta_2') \) such that \( \beta_2' < \beta_2 \), but \( \beta_2' > b_{2,3} \) and \( \beta_1 - \beta_2' < b_{1,2} \). By (iv.), 
\[S_2^*((\beta_1, \beta_2')^T) < 0 \] and thus \( S_2^*)((\beta_1, \beta_2')^T) < 0 \). Hence \( S_2^*(\cdot) \) does not cross 0 at \( \beta \).

Thus the Jaeckel estimate is not in regions \( R_1 \) or \( R_2 \) or on their boundary.

The only remaining regions that we need to consider are the boundary between \( R_2 \) and \( R_3 \), and the boundary between \( R_5 \) and \( R_6 \).

(v.) If \( \beta \) is on the boundary between \( R_2 \) and \( R_3 \), then \( \beta_1 > b_{1,3}, \beta_2 > b_{2,3}, \) and 
\[\beta_1 - \beta_2 = b_{1,2} \).

**Configuration I:** For this configuration,

\[
U_{1,3}(\beta) = \begin{cases} 
\leq \frac{n_1 n_3}{2} & \text{if } n_1 n_3 \text{ is even} \\
\leq \frac{n_1 n_3 - 1}{2} & \text{if } n_1 n_3 \text{ is odd from (ii.).} 
\end{cases}
\]  
(3.84)

So, \( U_{1,3}(\beta) \leq \frac{n_1 n_3}{2} \). Also,

\[
U_{1,2}(\beta) = \begin{cases} 
\frac{n_1 n_2}{2} - 1 & \text{if } n_1 n_2 \text{ is even since } \beta_1 - \beta_2 = D^{(1,2)}(i), j = \frac{n_1 n_2}{2} + 1 \\
\frac{n_1 n_2 - 1}{2} & \text{if } n_1 n_2 \text{ is odd since } \beta_1 - \beta_2 = D^{(1,2)}(i), j = \frac{n_1 n_2 + 1}{2}.
\end{cases}
\]
So, $U_{1,2}(\beta) < \frac{n_1 n_2}{2}$. Hence $S_i^*(\beta) < 0$. However, $U_{1,3}((\beta_1 - \beta_2)^T) \leq \frac{n_1 n_3}{2}$ and $U_{1,2}((\beta_1 - \beta_2)^T) = \frac{n_1 n_2}{2}$ so that $S_i^*((\beta_1 - \beta_2)^T) \leq 0$. Therefore, this $\beta$ is not the jump point where $S_i^*(\cdot)$ crosses 0.

**Configuration II:** For this configuration,

$$U_{1,3}(\beta) = \begin{cases} \frac{n_1 n_3}{2} & \text{if } n_1 n_3 \text{ is even} \\ \frac{n_1 n_3 - 1}{2} & \text{if } n_1 n_3 \text{ is odd from (ii.)} \end{cases}$$

So, $U_{1,3}(\beta) < \frac{n_1 n_3}{2}$. Also,

$$U_{1,2}(\beta) = \begin{cases} \frac{n_1 n_2}{2} & \text{if } n_1 n_2 \text{ is even since } \beta_1 - \beta_2 = D^{(1,2)}_0, j = \frac{n_1 n_2}{2} \\ \frac{n_1 n_2 - 1}{2} & \text{if } n_1 n_2 \text{ is odd since } \beta_1 - \beta_2 = D^{(1,2)}_0, j = \frac{n_1 n_2 + 1}{2} \end{cases}$$

So, $U_{1,2}(\beta) \leq \frac{n_1 n_2}{2}$. Hence $S_i^*(\beta) < 0$. However, $U_{1,3}((\beta_1 - \beta_2)^T) < \frac{n_1 n_3}{2}$ and $U_{1,2}((\beta_1 - \beta_2)^T) = \frac{n_1 n_2}{2}$ so that $S_i^*((\beta_1 - \beta_2)^T) \leq 0$. Therefore, this $\beta$ is not the jump point at which $S_i^*(\cdot)$ crosses 0.

Thus on the boundary between $R_2$ and $R_3$, $S_i^*(\beta) < 0$ and $\beta$ is not the jump point at which $S_i^*(\cdot)$ crosses 0. Thus the Jaeckel estimate is not on this boundary.

(vi.) If $\beta$ is on the boundary between $R_5$ and $R_6$, then $\beta_1 < b_{1,3}$, $\beta_2 < b_{2,3}$, and $\beta_1 - \beta_2 = b_{1,2}$. 
Configuration I: For this configuration, \( U_{2,3}(\beta) \geq n_2 n_3 / 2 \) from (iii.). Also,
\[ U_1,2(\beta) < n_1 n_2 / 2, \text{ since } \beta_1 - \beta_2 = D \frac{(1,2)}{j}, j = \lfloor (n_1 n_2 / 2) + 1 \rfloor. \] Hence \( S_2^*(\beta) > 0. \)

Configuration II: For this configuration, \( U_{2,3}(\beta) > \frac{n_2 n_3}{2} \) from (iii.). Also,
\[ U_1,2(\beta) \leq n_1 n_2 / 2, \text{ since } \beta_1 - \beta_2 = D \frac{(1,2)}{j}, j = \lfloor (n_1 n_2 + 1) / 2 \rfloor. \] Hence \( S_2^*(\beta) > 0. \)

Thus on the boundary between \( R_5 \) and \( R_6 \), \( S_2^*(\beta) > 0 \). This \( \beta \) is not the jump point at which \( S_2^*(-) \) crosses 0 since \( S_2^*((\beta_1, \beta_2))^T > 0 \), following from the right-continuity of \( S_2^*(\beta) \), the monotonicity of \( S_2^*(\beta) \), and the definition of the regions. Thus the Jaeckel estimate is not on this boundary.

From (i.),..., (vi.), the Jaeckel estimate does not occur in regions \( R_1, R_2, ..., R_6 \) or their common boundaries, but it does exist (Jaeckel 1972). Hence the Jaeckel estimate must lie in region \( R_7 \).

(Q.E.D.)

One consequence of Theorem 3.4 is that the Jaeckel estimates of \( \beta_1 \) and \( \beta_2 \) are bounded above and below by linear combinations of the medians of the differences between pairs of the samples 1, 2, and 3. The bounds are presented in the next corollary.

**Corollary 3.1:** Let \( \sim \) satisfy the Jaeckel estimating equations. Let \( b_{1,2}, b_{1,3}, \) and \( b_{2,3} \) be defined as in Theorem 3.4. Then

\[
\min\{b_{1,3}, b_{1,2} + b_{2,3}\} \leq b_1 \leq \max\{b_{1,3}, b_{1,2} + b_{2,3}\} \tag{3.88}
\]

and

\[
\min\{b_{2,3}, b_{1,3} - b_{1,2}\} \leq b_2 \leq \max\{b_{2,3}, b_{1,3} - b_{1,2}\}. \tag{3.89}
\]
Proof: From Theorem 3.4, \((b_1, b_2)\) lies in the triangle ABC. Referring to Figure 8, \(b_1\) must lie between the first coordinates of points A and B, and \(b_2\) must lie between the second coordinates of points A and C. Results (3.88) and (3.89) follow immediately from the definition of points A, B, and C.

(Q.E.D.)

3.4 Ideas for an Algorithm

In the previous sections we characterized the Jaeckel estimator in terms of the ordered pairwise differences and in terms of the geometry of \(S_1^*(\beta)\) and \(S_2^*(\beta)\). We also established the relationship between the Jaeckel estimator of \(\beta\) and the Hodges-Lehmann two-sample estimators of the location shifts. In this section we discuss how these characterizations can be used in the computation of the Jaeckel estimate of \(\beta\).

Since we established bounds for the Jaeckel estimate of \(\beta_1\) and \(\beta_2\) which are functions of the two-sample Hodges-Lehmann estimates, we begin by computing the intervals whose elements are medians of the pairwise differences as defined in (3.52) and (3.54). These intervals may be degenerate, that is they may be one point, as in the case when the number of differences is odd, but we shall refer to them as intervals for simplicity. If the conditions of Theorem 3.3 are satisfied for some \(b_1\) and \(b_2\) in the median intervals for the differences between samples 1 and 3, and between samples 2 and 3, respectively, then \((b_1, b_2)^T\) satisfies the Jaeckel estimating equations. Otherwise, either configuration I or configuration II holds, as determined by (3.76) and (3.77). The medians \(b_{1,2}, b_{1,3}, \) and \(b_{2,3}\) are selected according to (3.76) or (3.77), depending on the configuration. Using \(b_{1,2}, b_{1,3}, \) and \(b_{2,3}\), the bounds on \(\beta_1\) and \(\beta_2\) are determined as in Corollary 3.1. Furthermore, by Theorem 3.4 the difference \(\beta_1 - \beta_2\) for the Jaeckel
estimator satisfies

$$\min\{b_{1,2}, b_{1,3} - b_{2,3}\} \leq \beta_1 - \beta_2 \leq \max\{b_{1,2}, b_{1,3} - b_{2,3}\}. \quad (3.90)$$

Thus we find the indices i, j, and k of the ordered pairwise differences between samples 1 and 2 which satisfy

$$D^{(1,2)}_{(i)} = b_{1,2}, \quad \text{and} \quad D^{(1,2)}_{(j)} \leq b_{1,3} - b_{2,3} \leq D^{(1,2)}_{(k)}. \quad (3.91)$$

Since (3.90) is true for the Jaeckel estimates, we need to check whether a solution to the Jaeckel estimating equations occurs in the bands corresponding to

$$M_L \leq U_{1,2}(\beta) \leq M_U, \quad (3.92)$$

where $M_U = n_1 n_2 - \min\{i, j\}$ and $M_L = n_1 n_2 - \max\{i, k\}$. From Theorem 3.1, we check for overlap of the intervals $I_3(M)$ and $I_4(M)$, where $M$ ranges from $M_L$ to $M_U$. Thus, at most $M_U - M_L + 1$ pairs of intervals must be checked for overlap before a solution to the estimating equations is found. Once overlap is found, the Jaeckel estimate can be any $\beta$ in the region of overlap.

In summary, the Jaeckel estimate for the three-sample problem can be computed by the following steps:

(1.) Compute the median intervals for $b_{1,2}$, $b_{1,3}$, and $b_{2,3}$.

(2.) Check whether the conditions of Theorem 3.3 are satisfied. If they are satisfied by $\tilde{\beta} = (b_1, b_2)^T$, then $\tilde{\beta}$ is a Jaeckel estimate.
(3.) Select the medians \( b_{1,2}, b_{1,3}, \) and \( b_{2,3} \) using (3.76) and (3.77).

(4.) Find \( i, j, \) and \( k \) satisfying (3.91) and compute \( M_U \) and \( M_L \).

(5.) Set \( M = M_L \).

(6.) Compute \( I_3(M) \) and \( I_4(M) \) from Theorem 3.1 and check for overlap. If the intervals overlap, a solution to the estimating equations occurs in the region of overlap.

(7.) If the intervals do not overlap, then set \( M = M + 1 \) and go back to (6.).

(3.93)

In step (6.) of (3.93), for a given integer \( M \) we find two intervals and check for a solution by checking for overlap of the intervals. This step can be interpreted in the following manner. In setting the value of \( M \) we are selecting a contour band of the function \( U_{1,2}(\beta) \), that is, we set \( U_{1,2}(\beta) = M \). We then find the zero contours of \( S_1^*(\beta) \) and \( S_2^*(\beta) \) under the restriction that \( U_{1,2}(\beta) = M \). If the zero contours of \( S_1^*(\beta) \) and \( S_2^*(\beta) \) overlap, then a solution occurs in the region of overlap. Otherwise we select a new value of \( M \) and repeat the process.

The computation of the \( i \)-th ordered pairwise difference between samples \( r \) and \( s \) by finding and sorting all pairwise differences can be costly, even for moderate sample sizes. The algorithm above relies on this computation and hence can also require much storage and computer time. A variety of authors have developed algorithms by which \( D^{(r,s)}_{(i,j)} \) can be computed without finding and sorting all of the differences. In Chapter V we describe these algorithms and also provide an algorithm for the computation of the Jaeckel estimate of \( \beta \).
3.5 Example

In this section we consider a specific set of data to illustrate the results of the previous sections. Terkel and Rosenblatt (1968) studied whether substances in the blood are responsible for the appearance of maternal behavior in rats. Retrieving is one of the behaviors generally observed in rats within 48 hours after parturition. Virgin rats can be induced to begin retrieving after continuous exposure to pups for about 5 days. Terkel and Rosenblatt investigated the effect of injecting maternal plasma into virgin rats on the time required for the onset of retrieving in the rats.

The experiment involved 32 virgin female rats which were 60 days old. The subjects were assigned to one of four groups:

Group MP: rats injected with plasma taken from a donor rat within 48 hours after parturition,

Group PP: rats in proestrus injected with plasma from rats that were also in proestrus,

Group DP: rats in diestrus injected with plasma from donors that were in diestrus, and

Group S: rats injected with a saline solution.

Each group consisted of 8 rats and the time until the onset of retrieving was observed for each rat. Hettmansperger (1984) reported latency times that might have arisen from such an experiment. These times are given in Table 1, where one unit of time equals the length of one observation session.

For this section we only consider three of the treatments, namely the groups receiving maternal plasma, diestrus plasma, and saline. In Chapter IV we consider all
four treatments. Let the rats injected with maternal plasma, with diestrus plasma, and with saline be denoted by samples 1, 2, and 3, respectively. The ordered pairwise differences between samples 1 and 2, between samples 1 and 3, and between samples 2 and 3 are given in Table 2.

Table 1.
Latency Times for Virgin Rats Injected with Plasma from Rats in Various Phases of the Estrous Cycle

<table>
<thead>
<tr>
<th>Maternal Plasma</th>
<th>Proestrus Plasma</th>
<th>Diestrus Plasma</th>
<th>Saline</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.1</td>
<td>0.4</td>
<td>0.9</td>
</tr>
<tr>
<td>0.7</td>
<td>1.6</td>
<td>1.9</td>
<td>2.1</td>
</tr>
<tr>
<td>1.0</td>
<td>3.7</td>
<td>2.4</td>
<td>3.0</td>
</tr>
<tr>
<td>1.2</td>
<td>4.3</td>
<td>2.8</td>
<td>4.7</td>
</tr>
<tr>
<td>1.7</td>
<td>4.7</td>
<td>3.9</td>
<td>6.4</td>
</tr>
<tr>
<td>2.3</td>
<td>5.6</td>
<td>5.4</td>
<td>6.6</td>
</tr>
<tr>
<td>2.4</td>
<td>6.6</td>
<td>11.4</td>
<td>8.5</td>
</tr>
<tr>
<td>3.1</td>
<td>8.8</td>
<td>20.4</td>
<td>10.0</td>
</tr>
</tbody>
</table>

The Hodges-Lehmann two-sample estimate of the shift between the distribution of the group receiving maternal plasma and the group receiving diestrus plasma, $b_{1,2}$, is in the interval $[-1.9, -1.8)$. Similarly, $b_{1,3} \in [-4.0, -3.7)$ and $b_{2,3} \in [-0.8, -0.6)$. Using (3.76) we find that
Table 2.
Ordered Differences between Pairs of Samples for the Latency Data

<table>
<thead>
<tr>
<th></th>
<th>MP - PP</th>
<th>MP - DP</th>
<th>MP - S</th>
<th>PP - DP</th>
<th>PP - S</th>
<th>DP - S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-8.3</td>
<td>-19.9</td>
<td>-9.5</td>
<td>-19.3</td>
<td>-8.9</td>
<td>-9.6</td>
</tr>
<tr>
<td>2</td>
<td>-8.1</td>
<td>-19.7</td>
<td>-9.3</td>
<td>-18.8</td>
<td>-8.4</td>
<td>-8.1</td>
</tr>
<tr>
<td>3</td>
<td>-7.8</td>
<td>-19.4</td>
<td>-9.0</td>
<td>-16.7</td>
<td>-7.4</td>
<td>-8.1</td>
</tr>
<tr>
<td>4</td>
<td>-7.6</td>
<td>-19.2</td>
<td>-8.8</td>
<td>-16.1</td>
<td>-6.9</td>
<td>-7.6</td>
</tr>
<tr>
<td>5</td>
<td>-7.1</td>
<td>-18.7</td>
<td>-8.3</td>
<td>-15.7</td>
<td>-6.3</td>
<td>-7.2</td>
</tr>
<tr>
<td>6</td>
<td>-6.5</td>
<td>-18.1</td>
<td>-8.0</td>
<td>-14.8</td>
<td>-5.7</td>
<td>-6.6</td>
</tr>
<tr>
<td>7</td>
<td>-6.4</td>
<td>-18.0</td>
<td>-7.8</td>
<td>-13.8</td>
<td>-5.5</td>
<td>-6.2</td>
</tr>
<tr>
<td>8</td>
<td>-6.1</td>
<td>-17.3</td>
<td>-7.7</td>
<td>-11.6</td>
<td>-5.3</td>
<td>-6.1</td>
</tr>
<tr>
<td>9</td>
<td>-5.9</td>
<td>-10.9</td>
<td>-7.6</td>
<td>-10.3</td>
<td>-5.3</td>
<td>-6.1</td>
</tr>
<tr>
<td>10</td>
<td>-5.7</td>
<td>-10.7</td>
<td>-7.5</td>
<td>-9.8</td>
<td>-5.0</td>
<td>-6.0</td>
</tr>
<tr>
<td>11</td>
<td>-5.6</td>
<td>-10.4</td>
<td>-7.3</td>
<td>-7.7</td>
<td>-4.8</td>
<td>-5.7</td>
</tr>
<tr>
<td>12</td>
<td>-5.4</td>
<td>-10.2</td>
<td>-6.9</td>
<td>-7.1</td>
<td>-4.8</td>
<td>-4.7</td>
</tr>
<tr>
<td>13</td>
<td>-5.1</td>
<td>-9.7</td>
<td>-6.8</td>
<td>-6.7</td>
<td>-4.4</td>
<td>-4.6</td>
</tr>
<tr>
<td>14</td>
<td>-4.9</td>
<td>-9.1</td>
<td>-6.2</td>
<td>-5.8</td>
<td>-4.2</td>
<td>-4.6</td>
</tr>
<tr>
<td>15</td>
<td>-4.9</td>
<td>-9.0</td>
<td>-6.1</td>
<td>-4.8</td>
<td>-3.8</td>
<td>-4.5</td>
</tr>
<tr>
<td>16</td>
<td>-4.6</td>
<td>-8.3</td>
<td>-6.1</td>
<td>-4.3</td>
<td>-3.6</td>
<td>-4.3</td>
</tr>
<tr>
<td>17</td>
<td>-4.4</td>
<td>-4.9</td>
<td>-5.9</td>
<td>-3.8</td>
<td>-3.4</td>
<td>-4.2</td>
</tr>
<tr>
<td>18</td>
<td>-4.3</td>
<td>-4.7</td>
<td>-5.9</td>
<td>-2.8</td>
<td>-3.1</td>
<td>-4.0</td>
</tr>
<tr>
<td>19</td>
<td>-4.2</td>
<td>-4.4</td>
<td>-5.7</td>
<td>-2.6</td>
<td>-2.9</td>
<td>-3.8</td>
</tr>
<tr>
<td>20</td>
<td>-4.2</td>
<td>-4.2</td>
<td>-5.6</td>
<td>-2.3</td>
<td>-2.9</td>
<td>-3.6</td>
</tr>
<tr>
<td>21</td>
<td>-4.0</td>
<td>-3.7</td>
<td>-5.4</td>
<td>-1.7</td>
<td>-2.7</td>
<td>-3.1</td>
</tr>
<tr>
<td>22</td>
<td>-3.9</td>
<td>-3.4</td>
<td>-5.4</td>
<td>-1.7</td>
<td>-2.3</td>
<td>-2.8</td>
</tr>
<tr>
<td>23</td>
<td>-3.8</td>
<td>-3.2</td>
<td>-5.4</td>
<td>-1.3</td>
<td>-2.1</td>
<td>-2.7</td>
</tr>
<tr>
<td>24</td>
<td>-3.7</td>
<td>-3.1</td>
<td>-5.2</td>
<td>-1.2</td>
<td>-1.9</td>
<td>-2.6</td>
</tr>
<tr>
<td>25</td>
<td>-3.6</td>
<td>-3.0</td>
<td>-4.9</td>
<td>-1.1</td>
<td>-1.9</td>
<td>-2.5</td>
</tr>
<tr>
<td>26</td>
<td>-3.5</td>
<td>-2.9</td>
<td>-4.7</td>
<td>-0.8</td>
<td>-1.9</td>
<td>-2.3</td>
</tr>
<tr>
<td>27</td>
<td>-3.5</td>
<td>-2.7</td>
<td>-4.3</td>
<td>-0.8</td>
<td>-1.7</td>
<td>-1.9</td>
</tr>
<tr>
<td>28</td>
<td>-3.3</td>
<td>-2.3</td>
<td>-4.2</td>
<td>-0.7</td>
<td>-1.4</td>
<td>-1.7</td>
</tr>
<tr>
<td>29</td>
<td>-3.3</td>
<td>-2.3</td>
<td>-4.2</td>
<td>-0.3</td>
<td>-1.2</td>
<td>-1.2</td>
</tr>
<tr>
<td>30</td>
<td>-3.2</td>
<td>-2.2</td>
<td>-4.1</td>
<td>-0.2</td>
<td>-1.0</td>
<td>-1.1</td>
</tr>
<tr>
<td>31</td>
<td>-3.2</td>
<td>-2.1</td>
<td>-4.0</td>
<td>0.2</td>
<td>-1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>32</td>
<td>-3.1</td>
<td>-1.9</td>
<td>-4.0</td>
<td>0.4</td>
<td>-1.0</td>
<td>-0.8</td>
</tr>
<tr>
<td>33</td>
<td>-3.0</td>
<td>-1.8</td>
<td>-3.7</td>
<td>0.7</td>
<td>-0.8</td>
<td>-0.6</td>
</tr>
<tr>
<td>34</td>
<td>-3.0</td>
<td>-1.7</td>
<td>-3.5</td>
<td>0.8</td>
<td>-0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>35</td>
<td>-2.7</td>
<td>-1.6</td>
<td>-3.5</td>
<td>0.9</td>
<td>-0.4</td>
<td>-0.2</td>
</tr>
<tr>
<td>36</td>
<td>-2.6</td>
<td>-1.6</td>
<td>-3.3</td>
<td>1.2</td>
<td>0.0</td>
<td>-0.2</td>
</tr>
<tr>
<td>37</td>
<td>-2.5</td>
<td>-1.5</td>
<td>-3.0</td>
<td>1.2</td>
<td>0.0</td>
<td>0.3</td>
</tr>
<tr>
<td>38</td>
<td>-2.5</td>
<td>-1.4</td>
<td>-2.5</td>
<td>1.3</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>39</td>
<td>-2.4</td>
<td>-1.4</td>
<td>-2.4</td>
<td>1.5</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>40</td>
<td>-2.3</td>
<td>-1.2</td>
<td>-2.3</td>
<td>1.7</td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td>41</td>
<td>-2.0</td>
<td>-1.2</td>
<td>-2.3</td>
<td>1.8</td>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>42</td>
<td>-2.0</td>
<td>-1.1</td>
<td>-2.0</td>
<td>1.9</td>
<td>0.7</td>
<td>1.4</td>
</tr>
</tbody>
</table>
Table 2. (Continued)

<table>
<thead>
<tr>
<th></th>
<th>MP - PP</th>
<th>MP - DP</th>
<th>MP - S</th>
<th>PP - DP</th>
<th>PP - S</th>
<th>DP - S</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>-1.9</td>
<td>-0.9</td>
<td>-1.8</td>
<td>1.9</td>
<td>0.9</td>
<td>1.5</td>
</tr>
<tr>
<td>44</td>
<td>-1.6</td>
<td>-0.8</td>
<td>-1.6</td>
<td>2.3</td>
<td>1.3</td>
<td>1.8</td>
</tr>
<tr>
<td>45</td>
<td>-1.4</td>
<td>-0.7</td>
<td>-1.6</td>
<td>2.4</td>
<td>1.6</td>
<td>1.9</td>
</tr>
<tr>
<td>46</td>
<td>-1.3</td>
<td>-0.7</td>
<td>-1.4</td>
<td>2.7</td>
<td>1.7</td>
<td>2.4</td>
</tr>
<tr>
<td>47</td>
<td>-1.2</td>
<td>-0.5</td>
<td>-1.3</td>
<td>2.8</td>
<td>1.9</td>
<td>2.9</td>
</tr>
<tr>
<td>48</td>
<td>-1.1</td>
<td>-0.4</td>
<td>-1.1</td>
<td>2.8</td>
<td>2.2</td>
<td>3.0</td>
</tr>
<tr>
<td>49</td>
<td>-0.9</td>
<td>-0.2</td>
<td>-0.9</td>
<td>3.2</td>
<td>2.2</td>
<td>3.3</td>
</tr>
<tr>
<td>50</td>
<td>-0.6</td>
<td>-0.1</td>
<td>-0.7</td>
<td>3.3</td>
<td>2.4</td>
<td>4.5</td>
</tr>
<tr>
<td>51</td>
<td>-0.6</td>
<td>0.0</td>
<td>-0.6</td>
<td>3.4</td>
<td>2.6</td>
<td>4.8</td>
</tr>
<tr>
<td>52</td>
<td>-0.6</td>
<td>0.1</td>
<td>-0.4</td>
<td>3.7</td>
<td>2.6</td>
<td>5.0</td>
</tr>
<tr>
<td>53</td>
<td>-0.4</td>
<td>0.3</td>
<td>-0.4</td>
<td>3.8</td>
<td>2.8</td>
<td>6.7</td>
</tr>
<tr>
<td>54</td>
<td>-0.4</td>
<td>0.3</td>
<td>-0.2</td>
<td>3.9</td>
<td>3.4</td>
<td>8.4</td>
</tr>
<tr>
<td>55</td>
<td>-0.1</td>
<td>0.4</td>
<td>0.1</td>
<td>4.2</td>
<td>3.5</td>
<td>9.3</td>
</tr>
<tr>
<td>56</td>
<td>0.1</td>
<td>0.5</td>
<td>0.1</td>
<td>4.3</td>
<td>3.6</td>
<td>10.4</td>
</tr>
<tr>
<td>57</td>
<td>0.1</td>
<td>0.6</td>
<td>0.2</td>
<td>4.7</td>
<td>3.8</td>
<td>10.5</td>
</tr>
<tr>
<td>58</td>
<td>0.6</td>
<td>0.7</td>
<td>0.3</td>
<td>4.9</td>
<td>4.1</td>
<td>11.9</td>
</tr>
<tr>
<td>59</td>
<td>0.7</td>
<td>0.8</td>
<td>0.3</td>
<td>5.2</td>
<td>4.5</td>
<td>13.8</td>
</tr>
<tr>
<td>60</td>
<td>0.8</td>
<td>1.2</td>
<td>0.8</td>
<td>6.0</td>
<td>4.7</td>
<td>14.0</td>
</tr>
<tr>
<td>61</td>
<td>1.2</td>
<td>1.3</td>
<td>1.0</td>
<td>6.2</td>
<td>5.7</td>
<td>15.7</td>
</tr>
<tr>
<td>62</td>
<td>1.3</td>
<td>1.9</td>
<td>1.4</td>
<td>6.4</td>
<td>5.8</td>
<td>17.4</td>
</tr>
<tr>
<td>63</td>
<td>1.5</td>
<td>2.0</td>
<td>1.5</td>
<td>6.9</td>
<td>6.7</td>
<td>18.3</td>
</tr>
<tr>
<td>64</td>
<td>2.0</td>
<td>2.7</td>
<td>2.2</td>
<td>8.4</td>
<td>7.9</td>
<td>19.5</td>
</tr>
</tbody>
</table>

\[
D^{(1,3)}_{(i_1)} - D^{(2,3)}_{(j_1)} = -3.7 - (-0.8) < -1.9 = D^{(1,2)}_{(k_1)},
\tag{3.94}
\]

so that configuration I holds. Thus the points A, B, and C from (3.73) have coordinates A = (-4.0, -0.6), B = (-2.4, -0.6), and C = (-4.0, -2.2). To illustrate the geometric characterization of the Jaeckel estimator of \( \hat{\beta} \), the triangle containing the Jaeckel estimate is shown in Figure 9. The zero contours for \( S_1^*(\hat{\beta}) \) and \( S_2^*(\hat{\beta}) \) are shown in Figure 10. From this figure, the solution to the Jaeckel estimating equations is the region \( R_J \) where

\[
R_J = \{ \hat{\beta} : -3.5 \leq \beta_1 \leq -3.3, -1.2 \leq \beta_2 \leq -1.1, \text{and} -2.3 \leq \beta_1 - \beta_2 \leq -2.2 \}.
\tag{3.95}
\]
Now we illustrate the use of the characterization of the Jaeckel estimator in terms of the ordered pairwise differences. From Corollary 3.1, if \( b \) satisfies the Jaeckel estimating equations, then 
\[-4.0 \leq b_1 \leq -2.4 \text{ and } -2.2 \leq b_2 \leq -0.6.\]
Furthermore, at point A,
\[\beta_1 - \beta_2 = -3.4 \text{ and at points B and C, } \beta_1 - \beta_2 = -1.8.\]
Hence we also have the bound 
\[-3.4 \leq b_1 - b_2 \leq -1.8.\]
From Table 2, 
\[-3.4 = D^{(1,2)}_{(22)} \text{ and } -1.8 = D^{(1,2)}_{(33)}.\]
Thus to find the Jaeckel estimate of \( \beta \) we check whether the intervals \( I_3(M) \) and \( I_4(M) \) from Theorem 3.1 overlap, where \( M \) is between 42 and 31. For this example, since \( \nu = 0 \) and \( \gamma = 64 \),

\[I_3(M) = \left[ D^{(1,2)}_{(64-M)}, D^{(1,2)}_{(55-M)} \right], \tag{3.96}\]
and

\[I_4(M) = \left[ D^{(1,3)}_{(M)}, D^{(2,3)}_{(65-M)}, D^{(1,3)}_{(M+1)}, D^{(2,3)}_{(64-M)} \right]. \tag{3.97}\]

From Table 2 if \( M = 31 \), then \( I_3(31) = [-1.8, -1.7) \) and \( I_4(31) = [-3.5, -3.4) \). Since \( I_3(31) \) and \( I_4(31) \) do not overlap there is no solution to the estimating equations in the band for which \( U_{1,2}(\beta) = 31 \). Similarly, for \( M = 32, 33, \) and \( 34 \) there is no overlap between \( I_3(M) \) and \( I_4(M) \). These intervals are given by:

\[I_3(32) = [-1.9, -1.8) \text{ and } I_4(32) = [-3.4, -2.9), \]
\[I_3(33) = [-2.1, -1.9) \text{ and } I_4(33) = [-2.9, -2.5), \text{ and} \]
\[I_3(34) = [-2.2, -2.1) \text{ and } I_4(34) = [-2.5, -2.4). \tag{3.98}\]

From (3.98), the endpoints of the interval \( I_3(M) \) are decreasing and the endpoints of \( I_4(M) \) are increasing as \( M \) increases. Thus the intervals \( I_3(M) \) and \( I_4(M) \) are getting closer to one another. If \( M = 35 \), then the intervals given by
overlap and hence there is a solution to the Jaeckel estimating equations in the region of overlap, namely $R_j$. If $M = 36$, then $I_3(36) = [-2.3, -2.3]$ and $I_4 = [-2.1, -1.3)$ do not overlap. We can stop iterating because the intervals $I_3(M)$ and $I_4(M)$ will begin moving further apart as $M$ increases, due to the monotonicities of the endpoints noted above.

Thus the Jaeckel estimating equations equations are solved by $(\beta_1, \beta_2)^T$ in $R_j$.

Figure 9. Graphical representation of the triangle containing the Jaeckel estimate for the virgin rat data.
Figure 10. Graphical representation of the zero contours of the functions $S_1^*(\beta)$ and $S_2^*(\beta)$ for the virgin rat data.
CHAPTER IV
K-SAMPLE CASE

4.0 Introduction

In Chapter III we examined the Jaeckel estimator for the three-sample model. In this chapter we extend the previous results to the case of k-samples. The first section contains a formulation of the Jaeckel estimating equations for the k-sample case. In the second section we provide a characterization of the solution to the Jaeckel estimating equations for the case of k-samples, based on the geometry of the functions involved in the equations. The third section contains a result which describes the relationship between the Jaeckel estimator and the Hodges-Lehmann two-sample estimators for the four-sample model. This result is important in the computation of the estimates, as described in Chapter V. The fourth section contains a brief discussion of an approach for computing the estimates in the four-sample case. In the final section is given an example for the four-sample case.

4.1 Formulation of the Jaeckel Equations for the k-Sample Model

Consider the uncensored k-sample model, k > 2, given by

\[ Y_i = \alpha + X_i \beta + e_i, \quad i=1,..., N, \tag{4.1} \]

where \( X_i \) denotes a row vector corresponding to the covariate associated with individual
i. For the \( k \)-sample model, \( X_{ij} = 1 \) if observation \( i \) is from sample \( j \) and = 0 otherwise for \( j=1, 2, ..., k-1 \). As in Chapter III, let \( S_j \) denote the set of indices of the observations in sample \( j \) for \( j=1, 2, ..., k \). Without loss of generality let \( S_j = \{ a_j + 1, a_j + 2, ..., a_j + n_j \} \) where \( a_1=0 \), and \( a_j= \sum_{i=1}^{j-1} n_i \), \( j=2, ..., k \). Let \( N=n_1 + n_2 + ... + n_k \).

Extending the three-sample case to the case of \( k \) samples, the model in (4.1) can be written as

\[
Y_i = \begin{cases} 
\alpha + \beta_j + e_i & \text{if } i \in S_j, j=1, 2, ..., k-1 \\
\alpha + e_i & \text{if } i \in S_k.
\end{cases}
\]  

(4.2)

Assume that the \( e_i \) are independent and identically distributed with cumulative distribution function \( F \) for all \( i=1, ..., N \). The function \( F \) is unknown and it is assumed to be continuous.

We wish to estimate \( \beta=(\beta_1, \beta_2, ..., \beta_{k-1})^T \) by the value of \( \beta \) which minimizes the dispersion of the residuals proposed by Jaeckel (1972). Letting \( Y_i - \bar{X}_i \beta \) denote the \( i \)-th residual, Jaeckel's dispersion function is defined as

\[
D(Y-\bar{X} \beta) = \sum_{i=1}^{N} a_N(R(Y_i - X_i \beta)) (Y_i - X_i \beta) 
\]  

(4.3)

where \( a_N(i), i=1, ..., N \) is a set of scores satisfying the properties (3.5) through (3.8).

To minimize \( D(Y-\bar{X} \beta) \), find the value of \( \beta \) for which the partial derivatives equal zero. The function \( D(Y-\bar{X} \beta) \) is differentiable almost everywhere and its partial derivative with respect to \( \beta_j \) is given by:

\[
\frac{\delta}{\delta \beta_j} D(Y-\bar{X} \beta) = -\sum_{i=1}^{N} a_N(R(Y_i - X_i \beta)) (X_{ij} - \bar{X}_j),
\]  

(4.4)
where $\bar{X}_j = N^{-1} \sum_{m=1}^{N} X_{mj}$. As in Chapter III, let $S_j(\bar{X} - \bar{X}) = -8/8 \beta_j (D(\bar{X} - \bar{X}))$ for $j=1, \ldots, k-1$. Thus we solve for the $\beta$ value for which the vector function $S(\bar{X} - \bar{X}) = (S_1(\bar{X} - \bar{X}), S_2(\bar{X} - \bar{X}), \ldots, S_{k-1}(\bar{X} - \bar{X}))$ is equal to the zero vector. Let us look at $S(\bar{X} - \bar{X})$ more closely.

As in the three-sample case, the $X$ matrix for the k-sample model has a special form so that the expression of $S_j(\bar{X} - \bar{X})$ can be simplified. Recall that

$$X_{ij} = \begin{cases} 1 & \text{if } i \in S_j \\ 0 & \text{otherwise} \end{cases}$$

and $\bar{X}_j = n_j/N$. So $S_j(\bar{X} - \bar{X})$ can be written as

$$S_j(\bar{X} - \bar{X}) = \sum_{i \in S_j} a_N(R(Y_i - X_{ij}\beta)).$$

Hence $S_j(\cdot)$ is just the sum of the scores corresponding to individuals in sample $j$, where the score for individual $i$ depends on the rank of the $i$-th residual among all $N$ residuals. Thus, in general, the evaluation of $S_j(\cdot)$ requires a joint ranking procedure.

As in the three-sample case, we are interested in the form of $S_j(\cdot)$ for Wilcoxon scores. Let the $a_N(\cdot)$ denote the standardized Wilcoxon scores of (3.19). Substituting these scores into (4.6), $S_j(\cdot)$ has the following form for $j=1, \ldots, k-1$:

$$S_j(\bar{X} - \bar{X}) = \frac{\sqrt{12}}{N+1} \left( \sum_{i \in S_j} R(Y_i - X_{ij}\beta) \right) - \frac{n_j(N+1)}{2}$$

For the Wilcoxon scores, $S_j(\cdot)$ is proportional to the sum of the ranks in sample $j$ minus a constant depending on the sample sizes. Following the three-sample case, we can express $S_j(\bar{X} - \bar{X})$ in terms of rankings of pairs of samples.
As in the previous chapter, since the rank of \( Y_i - X_i \beta \) is just one plus the number of residuals smaller than \( Y_i - X_i \beta \), we can find the number of residuals less than \( Y_i - X_i \beta \) by considering each sample separately. Let

\[
\psi(u) = \begin{cases} 
1 & \text{if } u > 0 \\
0 & \text{otherwise}
\end{cases}
\]  
\( (4.8) \)

and

\[
U_{r,s}(\beta) = \sum_{i \in S_r} \sum_{j \in S_s} \psi(Y_i - Y_j - (\beta_r - \beta_s)), \quad (4.9)
\]

where \( \beta_k = 0 \). The function \( U_{r,s}(\beta) \) satisfies

\[
U_{s,r}(\beta) = n_{rs} - U_{r,s}(\beta), \quad (4.10)
\]

with probability one since the underlying distribution \( F \) is assumed to be continuous. The sum of ranks can be written in the following way:

\[
\sum_{i \in S} R(Y_i - X_i \beta) = \frac{n_i(n_i+1)}{2} + \sum_{m \neq j} U_{j,m}(\beta). \quad (4.11)
\]

Hence, for the Wilcoxon scores \( S_j(\cdot) \) has the following form:

\[
S_j(Y - X \beta) = \frac{\sqrt{12/N+1}}{N+1} \left\{ \frac{n_i(n_i+1)}{2} + \sum_{m \neq j} U_{j,m}(\beta) - \frac{n_i(N+1)}{2} \right\}. \quad (4.12)
\]

Using (4.10), \( S_j(Y - X \beta) \) can be written as
\[ S_j(Y - X\beta) = \frac{\sqrt{12}}{N+1} \left\{ \sum_{m=1}^{j-1} \left\{ \frac{ni_in_j}{2} - U_{m,j}(\beta) \right\} + \sum_{m=j+1}^{k} \left\{ U_{j,m}(\beta) - \frac{ni_in_j}{2} \right\} \right\}, \]

for \( j=1, 2, \ldots, k-1, \) \hspace{1cm} (4.13)

with the convention that a sum over the empty set is zero.

Consider the sum of the \( S_j(-) \) for a given value of \( \beta \). This sum is

\[ \sum_{j=1}^{k-1} S_j(b) = \frac{\sqrt{12}}{N+1} \left( \sum_{j=1}^{k-1} \sum_{i \in S_j} R(Y_i - X_i\beta) - \frac{nk(N+1)}{2} \right) \]

\[ = \frac{\sqrt{12}}{N+1} \left( \frac{N(N+1)}{2} - \sum_{i \in S_k} R(Y_i - X_i\beta) - \frac{(N-n_k)(N+1)}{2} \right) \]

\[ = \frac{\sqrt{12}}{N+1} \left( \frac{nk(N+1)}{2} - \sum_{i \in S_k} R(Y_i - X_i\beta) \right) \] \hspace{1cm} (4.14)

where the second equality holds since the ranks of all observations sum to \( N(N+1)/2 \).

Thus, the sum of the \( (N+1)(12)^{-1/2} S_j(\beta) \) for \( j=1, \ldots, k-1 \) is between \(-nk(N-n_k)/2 \) and \( nk(N-n_k)/2 \). Hence, there is the additional constraint that the sum of the \( S_j(\cdot) \) cannot be too large or too small, because of the nature of the \( S_j(\cdot) \) as functions of the ranks of the residuals.

In summary, finding the value of \( \beta \) which minimizes (4.3) for the Wilcoxon scores in (3.19) is equivalent to finding \( \beta \) such that the vector \( (S_1(\cdot) - \bar{X}\beta), S_2(\cdot) - \bar{X}\beta), \ldots, S_{k-1}(\cdot) - \bar{X}\beta) \) equals 0 under the constraint that the sum of the \( (N+1)(12)^{-1/2} S_j(\beta) \) for \( j=1, \ldots, k-1 \) is between \(-nk(N-n_k)/2 \) and \( nk(N-n_k)/2 \). By (4.13) this is equivalent to finding the \( \beta_1, \beta_2, \ldots, \beta_{k-1} \) that simultaneously solve the system of equations:
\[
\sum_{m=1}^{j-1} \left( \frac{n_i n_m}{2} - U_{m,j}(\beta) \right) + \sum_{m=j+1}^{k} \left( U_{j,m}(\beta) - \frac{n_i n_m}{2} \right) = 0, \quad j=1, 2, \ldots, k-1
\]  

(4.15)

under the constraint that

\[
\sum_{m=1}^{k-1} U_{j,m}(\beta) - \frac{n_i n_k}{2} = 0.
\]  

(4.16)

For notational convenience, let

\[
S_j^*(\beta) = S_j^*(Y - X\beta) = \sum_{m=1}^{j-1} \left( \frac{n_i n_m}{2} - U_{m,j}(\beta) \right) + \sum_{m=j+1}^{k} \left( U_{j,m}(\beta) - \frac{n_i n_m}{2} \right),
\]

for \(j=1, 2, \ldots, k-1\).  

(4.17)

It follows from Properties 3.1 and 3.2 for \(U_{r,s}(\beta)\) that the functions \(S_j^*(\beta)\) have the following property:

**Property 4.1:** The function \(S_j^*(\beta)\) is a right-continuous step function which jumps whenever \(\beta_i - \beta_j\) crosses one of the pairwise differences \(Y_i - Y_j, i \in S_r, j \in S_s\), for \(i < j\), or \(\beta_j - \beta_m\) crosses one of the pairwise differences \(Y_s - Y_t, s \in S_j, t \in S_m\), for \(m > j\) and \(j=1, \ldots, k-1\).

From Property 4.1, the function \(S_j^*(\beta)\) is a linear combination of the Mann-Whitney-Wilcoxon statistics that compare sample \(j\) with each of the other samples.

To estimate the vector \(\beta\), find \(\beta_1, \beta_2, \ldots, \beta_{k-1}\) such that \(S_j^*(\beta)\) equals zero for \(j=1, \ldots, k-1\). From Property 4.1, \(S_j^*(\beta)\) is a step function and thus either there is a region on which \(S_j^*(\beta)\) is equal to zero or there is a boundary at which \(S_j^*(\beta)\) jumps across
zero. The solution to the system of equations $S_j^*(\beta_{-X\beta}) = 0$, $j=1, 2, ..., k-1$ is the set of $\beta$ values at which $S(\beta_{-X\beta})$ equals or jumps across 0. In the section that follows we characterize the solution to the estimating equations.

4.2 Characterization of the Jaeckel Estimator

Previously we defined the Jaeckel estimator for the k-sample case as the solution to the system of equations (4.15) under constraint (4.16). In this section we describe a geometric characterization of this solution, in terms of the functions $S_j^*(\beta)$, $j=1, ..., k-1$. We also provide a characterization based on the pairwise ordered differences.

Let us now consider the geometry of the function $S_j^*(\beta)$ for $j=1, ..., k-1$. The contours of $S_j^*(\beta)$ are easily determined from the contours of $U_{ij}(\beta)$ for $i < j$ and $U_{jm}(\beta)$ for $j < m$. As an example, we consider the contours for the four-sample case since these contours can be pictured in $R^3$.

The function $U_{r,s}(\beta)$ is a function of $\beta_r - \beta_s$ alone that jumps down when $\beta_r - \beta_s$ increases past a pairwise difference $D_{r,s}^{(r,s)}$, $j=1, ..., n_r n_s$. If we plot the planes with equations $\beta_r - \beta_s = D_{r,s}^{(r,s)}$, for $j=1, ..., n_r n_s$, then $U_{r,s}(\beta)$ is constant between two adjacent planes. Hence, $S_j^*(\beta)$ changes value whenever $\beta_i - \beta_j$ crosses one of the planes with equation $\beta_i - \beta_j = Y_r - Y_s$, $i \in S_r, j \in S_s$, for $i < j$, or $\beta_j - \beta_m$ crosses one of the planes with equation $\beta_j - \beta_m = Y_s - Y_t$, $s \in S_j, t \in S_m$, for $m > j$. The planes $\beta_1 = D_{(1)}^{(1,4)}$, $\beta_2 = D_{(1)}^{(2,4)}$ and $\beta_3 = D_{(1)}^{(3,4)}$ are perpendicular to the $\beta_1$, $\beta_2$, and $\beta_3$ axes, respectively. The plane $\beta_1 - \beta_2 = D_{(1)}^{(1,2)}$ is perpendicular to the $(\beta_1, \beta_2)$ plane and parallel to the plane $\beta_1 = \beta_2$. Similar descriptions hold for the $\beta_1 - \beta_3 = D_{(1)}^{(1,3)}$ and $\beta_2 - \beta_3 = D_{(1)}^{(2,3)}$ planes. The function $S_j^*(\beta)$ is constant on polygonal solids that are joined at the corners. These solids are the regions bounded by the planes described above. Thus, the contour on
which $S_1^*(\beta)$ equals $M_1 - n_1(n_2+n_3+n_4)/2$, where $M_1$ is an integer between 0 and $n_1(n_2+n_3+n_4)/2$, is the intersection of the sets

$$\begin{align*}
D_{(r_1)}^{(1,4)} &\leq \beta_1 < D_{(r_1+1)}^{(1,4)}, \\
D_{(s_1)}^{(1,2)} &\leq \beta_1 - \beta_2 < D_{(s_1+1)}^{(1,2)}, \\
D_{(t_1)}^{(1,3)} &\leq \beta_1 - \beta_3 < D_{(t_1+1)}^{(1,3)},
\end{align*}$$

(4.18)

for $r_1$, $s_1$, and $t_1$ satisfying $r_1 + s_1 + t_1 = n_1(n_2+n_3+n_4) - M_1$. The intersection of the sets on which $\beta_1 = D_{(r_1)}^{(1,4)}$ and $\beta_1 = D_{(r_1+1)}^{(1,4)}$ planes serve as a "base" and "ceiling" for the cylinder. Thus, if we were looking at $\mathbb{R}^3$ along the equiangular line, the contour on which $S_1^*(\beta)$ equals $M_1 - n_1(n_2+n_3+n_4)/2$ for the values of $\beta_1$ in the interval $[D_{(r_1)}^{(1,4)}, D_{(r_1+1)}^{(1,4)}]$ is a string of parallelograms joined at their corners, each parallelogram corresponding to a different $(s_1, t_1)$ pair. The contour on which $S_1^*(\beta)$ equals $M_1 - n_1(n_2+n_3+n_4)/2$ for the values of $\beta_1$ in the intervals

$$\begin{align*}
\left[ D_{(r_1)}^{(1,4)}, D_{(r_1+1)}^{(1,4)} \right] \\
\left[ D_{(r_1+1)}^{(1,4)}, D_{(r_1+2)}^{(1,4)} \right] \\
\left[ D_{(r_1+2)}^{(1,4)}, D_{(r_1+3)}^{(1,4)} \right]
\end{align*}$$

(4.19)

are shown in Figure 11.

Similarly, the contour on which $S_2^*(\beta)$ is equal to $M_2 - n_2(n_1+n_3+n_4)/2$, where $M_2$ is an integer between 0 and $n_2(n_1+n_3+n_4)/2$, is the intersection of the sets
D^{(2,4)}_{(r_2)} \leq \beta_2 < D^{(2,4)}_{(r_2+1)}$

D^{(1,2)}_{(s_2)} \leq \beta_1 - \beta_2 < D^{(1,2)}_{(s_2+1)}$

D^{(2,3)}_{(t_2)} \leq \beta_2 - \beta_3 < D^{(2,3)}_{(t_2+1)}, \tag{4.20}

for r_2, s_2, and t_2 satisfying $r_2 + t_2 - s_2 = n_2(n_3+n_4) - M_2$. The contours for $S_2^*(\beta)$ are similar to those of $S_1^*(\beta)$. The intersection of the sets on which $D^{(1,2)}_{(s_2)} \leq \beta_1 - \beta_2 < D^{(1,2)}_{(s_2+1)}$ and $D^{(2,3)}_{(t_2)} \leq \beta_2 - \beta_3 < D^{(2,3)}_{(t_2+1)}$ for some $(s_2, t_2)$ is a four-sided cylinder, with opposite sides parallel, and which is parallel to the equiangular line. Only one pair of sides is common to the cylinder of an $S_1^*(\beta)$ contour, namely the $\beta_1 - \beta_2 = D^{(1,2)}_{(s_2)}$ and $\beta_1 - \beta_2 = D^{(1,2)}_{(s_2+1)}$ planes. The third set, namely the $\beta_2 = D^{(2,4)}_{(r_2)}$ and $\beta_2 = D^{(2,4)}_{(r_2+1)}$ planes, gives a bottom and top to the cylinder. Thus, if we were looking at $R^3$ along the equiangular line, the contour on which $S_2^*(\beta)$ equals $M_2 - n_2(n_1+n_3+n_4)/2$ for the values of $\beta_2$ in the interval $D^{(2,4)}_{(2,4)}, D^{(2,4)}_{(2,4)}$ is a chain of parallelograms linked at the corners, each parallelogram corresponding to a different $(s_2, t_2)$ pair. The contour on which $S_2^*(\beta)$ equals $M_2 - n_2(n_1+n_3+n_4)/2$ for the values of $\beta_2$ in the intervals

$$\begin{bmatrix} D^{(2,4)}_{(r_2)}, D^{(2,4)}_{(r_2+1)} \\ D^{(2,4)}_{(r_2+1)}, D^{(2,4)}_{(r_2+2)} \\ D^{(2,4)}_{(r_2+2)}, D^{(2,4)}_{(r_2+3)} \end{bmatrix} \tag{4.21}$$

are shown in Figure 12.

Finally, the contour on which $S_3^*(\beta)$ equals $M_3 - n_3(n_1+n_2+n_4)/2$, where $M_3$ is an integer between 0 and $n_3(n_1+n_2+n_4)/2$, is the intersection of the sets
Figure 11. Graphical representation of the contours of the function $S_1^*(\beta)$ when viewing $\mathbb{R}^3$ along the equiangular line.
Figure 12. Graphical representation of the contours of the function $S_2^*(\beta)$ when viewing $\mathbb{R}^3$ along the equiangular line.
\[ D^{(3,4)}_{(r_3)} \leq \beta_3 < D^{(3,4)}_{(r_3+1)} \]

\[ D^{(1,3)}_{(s_3)} \leq \beta_1 - \beta_3 < D^{(1,3)}_{(s_3+1)} \]

\[ D^{(2,3)}_{(t_3)} \leq \beta_2 - \beta_3 < D^{(2,3)}_{(t_3+1)} \]

(4.22)

for \( r_3, s_3, \) and \( t_3 \) satisfying \( r_3 - s_3 - t_3 = n_3n_4 - M_3 \). The contours for \( S_3^* (\beta) \) are similar to those of \( S_1^* (\beta) \) and \( S_2^* (\beta) \). The intersection of the sets on which \( D^{(1,3)}_{(s_3)} \leq \beta_1 - \beta_3 < D^{(1,3)}_{(s_3+1)} \) and \( D^{(2,3)}_{(t_3)} \leq \beta_2 - \beta_3 < D^{(2,3)}_{(t_3+1)} \) for some \( (s_3, t_3) \) is a four-sided cylinder, with opposite sides parallel. This cylinder is also parallel to the equiangular line. The third set, namely the \( \beta_3 = D^{(3,4)}_{(r_3)} \) and \( \beta_3 = D^{(3,4)}_{(r_3+1)} \) planes, gives a bottom and top to the cylinder. The contour on which \( S_3^* (\beta) \) equals \( M_3 - n_3(n_1+n_2+n_4)/2 \) for the values of \( \beta_3 \) in the intervals

\[
\begin{bmatrix}
D^{(3,4)}_{(r_3)} , D^{(3,4)}_{(r_3+1)} \\
D^{(3,4)}_{(r_3+1)} , D^{(3,4)}_{(r_3+2)} \\
D^{(3,4)}_{(r_3+2)} , D^{(3,4)}_{(r_3+3)}
\end{bmatrix}
\]

are shown in Figure 13.

Now that we have considered the case of \( k = 4 \), we look at the contours for \( S_j^* (\beta) \) for any \( k > 2 \). In this case \( S_j^* (\beta) \) changes value whenever \( \beta_i - \beta_j \) crosses one of the hyperplanes with equation \( \beta_i - \beta_j = Y_r - Y_s, i \in S_r, j \in S_s \), for \( i < j \), or \( \beta_j - \beta_m \) crosses one of the hyperplanes with equation \( \beta_j - \beta_m = Y_s - Y_t, s \in S_j, t \in S_m \), for \( m > j \). The function \( S_j^* (\beta) \) is constant on polygonal solids that are joined at the corners. These solids are the regions bounded by the hyperplanes with the following equations:

\[
\beta_j = D^{(i,k)}_{(r_{j,k})}, \quad \beta_j = D^{(i,k)}_{(r_{j,k}+1)};
\]
Figure 13. Graphical representation of the contours of the function $S_3^*(\beta)$ when viewing $R^3$ along the equiangular line.
\( \beta_i - \beta_j = D_{(i,j)}^{(i,j)}(r_{i,j}), \) and \( \beta_i - \beta_j = D_{(r_{i,j}+1)}^{(i,j)}, \) for \( i = 1, \ldots, j-1; \) and

\( \beta_j - \beta_m = D_{(j,m)}^{(j,m)}, \) and \( \beta_j - \beta_m = D_{(r_{j,m}+1)}^{(j,m)}, \) for \( m = j+1, \ldots, k-1 \) \quad (4.24)

for some \( r_{i,j}, \ldots, r_{j-1,j}, r_{j,j+1}, \ldots, r_{j,k}. \)

Now that we have considered the contours for \( S_j^*(\beta), j = 1, 2, \ldots, k-1, \) we can describe the region which contains the Jaeckel estimator. Let the zero contour for \( S_j^*(\beta) \) be the region on which \( S_j^*(\beta) \) equals zero, as defined in Chapter III. To find the region in which \( \beta_1, \beta_2, \ldots, \beta_{k-1} \) satisfy the Jaeckel estimating equations, find the intersection of the zero contours for \( S_j^*(\beta), j = 1, \ldots, k-1. \) Even in the case of four samples, the intersection of the zero contours for \( S_1^*(\beta), S_2^*(\beta), \) and \( S_3^*(\beta) \) is not easily described. We consider another approach to describing the intersection of the zero contours for the four-sample case.

The planes \( \beta_1 - \beta_2 = D_{(i,2)}^{(1,2)}(r_{i,2}) \) for \( i = 1, \ldots, n_1n_2, \beta_1 - \beta_3 = D_{(i,3)}^{(1,3)}(r_{i,3}) \) for \( j = 1, \ldots, n_1n_3, \) and \( \beta_2 - \beta_3 = D_{(m)}^{(2,3)}(r_{m}), \) for \( m = 1, \ldots, n_2n_3, \) partition \( \mathbb{R}^3 \) into cylinders which are parallel to the equiangular line. Each cylinder is bounded by at least three and at most six of the above planes, depending on the values of the ordered pairwise differences. Within a given cylinder, the function \( U_{1,2}(\beta) \) is constant, since \( \beta_1 - \beta_2 \) is between two consecutive ordered pairwise differences, \( D_{(i,2)}^{(1,2)}(r_{i,2}) \) and \( D_{(i+1,2)}^{(1,2)}, \) for some \( i. \) Similarly, the functions \( U_{1,3}(\beta) \) and \( U_{2,3}(\beta) \) are constant in the cylinder. The function \( U_{1,4}(\beta), \) however, is not constant throughout the cylinder. This function jumps each time \( \beta_1 \) crosses one of the differences \( D_{(j,4)}^{(1,4)}, j = 1, \ldots, n_1n_4. \) Consequently, in each cylinder, \( S_1^*(\beta) \) is constant between the planes \( \beta_1 = D_{(i,4)}^{(1,4)}(r_{i,4}) \) and \( \beta_1 = D_{(i+1,4)}^{(1,4)}, \) for each \( j, \) but \( S_1^*(\beta) \) jumps each time...
$\beta_1$ crosses one of these pairwise differences. In a given cylinder zero may or may not be in the range of $S_1^*(\beta)$.

Following a similar argument with the function $U_{2,4}(\beta)$, in a given cylinder $S_2^*(\beta)$ is constant between the planes $\beta_2 = D_{(i)}^{(2,4)}$ and $\beta_2 = D_{(i+1)}^{(2,4)}$, for each $i$, but $S_2^*(\beta)$ jumps each time $\beta_2$ crosses one of the pairwise differences between samples 2 and 4. Similarly, due to the nature of $U_{3,4}(\beta)$, the function $S_3^*(\beta)$ is constant between the planes $\beta_3 = D_{(m)}^{(3,4)}$ and $\beta_3 = D_{(m+1)}^{(3,4)}$, for each $m$, but $S_3^*(\beta)$ jumps each time $\beta_3$ crosses one of the $D_{(m)}^{(3,4)}$. In a given cylinder zero may or may not be in the range of the functions $S_2^*(\beta)$ and $S_3^*(\beta)$.

Thus, to determine whether or not there is a solution to the Jaeckel equations in a given cylinder, we find the subset of the cylinder for which $S_1^*(\beta)$ equals or jumps across 0, the subset of the cylinder on which $S_2^*(\beta)$ equals or jumps across 0, and the subset on which $S_3^*(\beta)$ equals or jumps across 0. These three subsets may or may not be empty, depending on whether or not zero is in the range of $S_j^*(\beta)$, $j=1, 2,$ and 3 for that cylinder. If the intersection of these three subsets is not empty, then a point in the intersection satisfies the Jaeckel estimating equations simultaneously, and hence is a solution. This characterizes the Jaeckel estimator, based on the zero contours for $S_j^*(\beta)$, $j=1, 2,$ and 3. In the next section we consider the relationship of the Jaeckel estimator and the Hodges-Lehmann two-sample estimators.

4.3 Relationship between the Jaeckel Estimator and the Hodges-Lehmann Two-Sample Estimators

In this section we examine the relationship between the Jaeckel estimator and the Hodges-Lehmann two-sample estimators. We first consider the case of four samples in detail and then consider the extension to the k-sample case.
Recall from Property 3.2 that $U_{r,s}(0)$ is the Mann-Whitney-Wilcoxon statistic used to test whether the location shift between the distribution of samples $r$ and $s$ is zero. Using Hodges-Lehmann estimation, we can estimate the shift between any pair of samples by inverting the Wilcoxon statistic. From this procedure the estimator is the median of the pairwise differences between the two samples, denoted $b_{r,s}$ as in (3.52). The convention described in Chapter III for defining the median when there is an even number of differences will be used here. In the three-sample case, the three Hodges-Lehmann estimates of shift are represented by three lines in the $(\beta_1, \beta_2)$ plane, namely $\beta_1 = b_{1,3}$, $\beta_2 = b_{2,3}$, and $\beta_1 - \beta_2 = b_{1,2}$. For the $k$-sample case, the Hodges-Lehmann estimates of shift are represented by $k(k-1)/2$ hyperplanes in $R^{k-1}$.

In general, the estimators $b_{r,s}$ are not compatible in that they do not satisfy certain linear relationships satisfied by the parameters they estimate. We examine the relationship of the Jaeckel estimator for $\underline{\beta}$ and the estimators for $\underline{\beta}$ obtained from the two-sample Hodges-Lehmann estimators of shift.

We first consider the case of four samples since in this case the estimator of $\underline{\beta}$ can be pictured in $R^3$. Later, we discuss the case for $k > 4$. The estimators $b_{1,2}$, $b_{1,3}$, and $b_{2,3}$ are of interest because they satisfy $U_{1,2}(\underline{\beta}) = n_1 n_2 / 2$, $U_{1,3}(\underline{\beta}) = n_1 n_3 / 2$, and $U_{2,3}(\underline{\beta}) = n_2 n_3 / 2$, respectively, and these $U_{r,s}(\underline{\beta})$ are terms of $S^*_j(\underline{\beta})$. In $R^3$, these estimates are represented by three planes defined by the equations $\beta_1 - \beta_2 = b_{1,2}$, $\beta_1 - \beta_3 = b_{1,3}$, and $\beta_2 - \beta_3 = b_{2,3}$. In considering the estimators for shift derived from samples 1, 2, and 3, there are only two possible configurations for the three planes when $b_{1,2} \neq b_{1,3} - b_{2,3}$, namely,

**Configuration I:** $b_{1,2} < b_{1,3} - b_{2,3}$

and

**Configuration II:** $b_{1,2} > b_{1,3} - b_{2,3}$.

(4.25)
Each of the three lines, defined by the intersection of two of the above planes, is parallel to the equiangular line. Thus, the three planes enclose a triangular cylinder which is parallel to the equiangular line. For each configuration, the three planes divide $\mathbb{R}^3$ into seven regions, denoted

$R_1 = \{(\beta_1, \beta_2): \beta_1 - \beta_3 < b_{1,3}, \beta_2 - \beta_3 > b_{2,3}, \text{ and } \beta_1 - \beta_2 < b_{1,2}\},$

$R_2 = \{(\beta_1, \beta_2): \beta_1 - \beta_3 > b_{1,3}, \beta_2 - \beta_3 > b_{2,3}, \text{ and } \beta_1 - \beta_2 < b_{1,2}\},$

$R_3 = \{(\beta_1, \beta_2): \beta_1 - \beta_3 > b_{1,3}, \beta_2 - \beta_3 > b_{2,3}, \text{ and } \beta_1 - \beta_2 > b_{1,2}\},$

$R_4 = \{(\beta_1, \beta_2): \beta_1 - \beta_3 > b_{1,3}, \beta_2 - \beta_3 < b_{2,3}, \text{ and } \beta_1 - \beta_2 > b_{1,2}\},$

$R_5 = \{(\beta_1, \beta_2): \beta_1 - \beta_3 < b_{1,3}, \beta_2 - \beta_3 < b_{2,3}, \text{ and } \beta_1 - \beta_2 > b_{1,2}\},$

$R_6 = \{(\beta_1, \beta_2): \beta_1 - \beta_3 < b_{1,3}, \beta_2 - \beta_3 < b_{2,3}, \text{ and } \beta_1 - \beta_2 < b_{1,2}\},$ and

$R_7 =$ the triangular cylinder bounded by the three planes. (4.26)

Given the values of the six Hodges-Lehmann estimates of the shifts between any pair of samples, we can determine whether configuration I or II holds. Furthermore, we can determine in which of the regions $R_1$ through $R_7$ the point $(b_{1,4}, b_{2,4}, b_{3,4})$ can be found. From these determinations we can restrict the possible regions in which the Jaeckel estimate can occur, and these restricted regions are specified by the planes defined by the two-sample estimates of shift. However, these restricted regions are not easy to visualize and are not easy to use. Thus, we want to find a bounded region which contains the Jaeckel estimate, a region that is more easily specified using only estimates of $\beta_1, \beta_2,$ and $\beta_3$ based on two-sample estimates of the shift between the distributions of the samples. First we consider the possible ways of estimating $\beta_1, \beta_2,$ and $\beta_3$ using only $b_{1,2}, b_{1,3}, b_{2,3}, b_{1,4}, b_{2,4},$ and $b_{3,4}$.

As indicated previously, the equations $\beta_1 - \beta_2 = b_{1,2}, \beta_1 - \beta_3 = b_{1,3}, \beta_2 - \beta_3 = b_{2,3},$ $\beta_1 = b_{1,4}, \beta_2 = b_{2,4},$ and $\beta_3 = b_{3,4}$ represent six planes in $\mathbb{R}^3$. The possible ways of estimating $\beta_1, \beta_2,$ and $\beta_3$ from these six estimates of shift are represented by the
intersection points of these six planes. For example, $\beta_1$ can be estimated directly by $b_{1,4}$.

The parameter $\beta_1$ can also be estimated by the intersection of the two planes $\beta_1 - \beta_2 = b_{1,2}$ and $\beta_2 = b_{2,4}$ to give the estimate $b_{1,2} + b_{2,4}$. Similarly, the intersection of the two planes $\beta_1 - \beta_3 = b_{1,3}$ and $\beta_3 = b_{3,4}$ gives the estimate $b_{1,3} + b_{3,4}$. Intersecting the three planes $\beta_1 - \beta_3 = b_{1,3}$, $\beta_2 - \beta_3 = b_{2,3}$, and $\beta_2 = b_{2,4}$ gives the estimate $b_{1,3} - b_{2,3} + b_{2,4}$, and intersecting the three planes $\beta_1 - \beta_2 = b_{1,2}$, $\beta_2 - \beta_3 = b_{2,3}$, and $\beta_3 = b_{3,4}$ gives the estimate $b_{1,2} + b_{2,3} + b_{3,4}$. Let $H_{L1}$ denote the set of possible estimates, namely

$$H_{L1} = \{ b_{1,4}, b_{1,2} + b_{2,4}, b_{1,3} + b_{3,4}, b_{1,3} - b_{2,3} + b_{2,4}, \text{ and } b_{1,2} + b_{2,3} + b_{3,4} \}.$$

Similarly, the intersection of the planes gives the set of possible estimates of $\beta_2$ as

$$H_{L2} = \{ b_{2,4}, b_{1,4} - b_{1,2}, b_{2,3} + b_{3,4}, b_{2,3} - b_{1,3} + b_{1,4}, \text{ and } b_{1,3} - b_{1,2} + b_{3,4} \}.$$

The intersection of planes gives as the possible estimates of $\beta_3$ the set

$$H_{L3} = \{ b_{3,4}, b_{1,4} - b_{1,3}, b_{2,4} - b_{2,3}, b_{1,4} - b_{1,2} - b_{2,3}, \text{ and } b_{1,2} + b_{2,4} - b_{1,3} \}.$$

In the following theorem we provide a description of the relationship between the two-sample Hodges-Lehmann estimates of $\beta_1$, $\beta_2$, and $\beta_3$, and the Jaeckel estimates of these parameters.

**Theorem 4.1:** Let $b_{\sim, j} = (b_1, b_2, b_3)^T$ satisfy the conditions:

$$S_j^*(\bar{b}) = 0 \text{ for } j = 1, 2, 3 \quad (4.27)$$

subject to the constraint

$$\sum_{j=1}^{3} S_j^*(b) \in \left[ -\frac{n_4(N - n_4)}{2}, \frac{n_4(N - n_4)}{2} \right]. \quad (4.28)$$

Then the point $b_{1,1}$ lies in the rectangular solid defined by
RS = \{ (\beta_1, \beta_2, \beta_3): \beta_i \in [\min H_{L_i}, \max H_{L_i}], i=1, 2, 3 \}. \quad (4.29)

**Proof:** The proof can be outlined as follows. For $\beta = (\beta_1, \beta_2, \beta_3)$ such that $\beta_1 < \min H_{L_1}$ we show that either $S_1^*(\beta)$, $S_2^*(\beta)$, $S_3^*(\beta)$, or $\{S_1^*(\beta)+S_2^*(\beta)+S_3^*(\beta)\}$ is too large and hence the estimating equations cannot be solved simultaneously. Similarly, for $\beta_1 > \max H_{L_1}$, we show that either $S_1^*(\beta)$, $S_2^*(\beta)$, $S_3^*(\beta)$, or $\{S_1^*(\beta)+S_2^*(\beta)+S_3^*(\beta)\}$ is too small and the estimating equations cannot be solved simultaneously. Thus for a solution to occur, it must be that

$$\min H_{L_1} \leq \beta_1 \leq \max H_{L_1}. \quad (4.30)$$

Similarly, we show that for a solution to the system we need

$$\min H_{L_2} \leq \beta_2 \leq \max H_{L_2} \quad (4.31)$$

and

$$\min H_{L_3} \leq \beta_3 \leq \max H_{L_3}. \quad (4.32)$$

Thus, for a solution to the system of equations to occur, we show that (4.30), (4.31), and (4.32) must hold simultaneously.

Suppose that $\beta$ satisfies $\beta_1 < \min H_{L_1}$. Then $\beta_1 < \min H_{L_1} \leq b_{1,4}$ which implies that $U_{1,4}(\beta) > n_{14}/2$. One of two relationships can hold for $\beta_1 - \beta_2$, either $\beta_1 - \beta_2 < b_{1,2}$ or $\beta_1 - \beta_2 > b_{1,2}$. Consider these two cases separately.

**Case 1:** If $\beta_1 - \beta_2 < b_{1,2}$ then $U_{1,2}(\beta) > n_{12}/2$. One of two relationships can hold for $\beta_1 - \beta_3$, either $\beta_1 - \beta_3 < b_{1,3}$ or $\beta_1 - \beta_3 > b_{1,3}$. 
If $\beta_1 - \beta_3 < b_{1,3}$, then $U_{1,3}(\beta) > n_1 n_3 / 2$ and hence $S^*_1(\beta) > 0$. Thus $\beta$ is not a solution to the estimating equations.

If $\beta_1 - \beta_3 > b_{1,3}$, then $U_{1,3}(\beta) < n_1 n_3 / 2$ and thus $S^*_1(\beta)$ may equal or cross zero. Since $\beta_1 - \beta_3 > b_{1,3}$ and $\beta_1 < \min \{HL_1 \leq b_{1,3} - b_{3,4}\}$, $\beta$ must satisfy $\beta_3 < b_{3,4}$ which implies that $U_{3,4}(\beta) > n_3 n_4 / 2$. For $\beta$ to be a solution, $S^*_2(\beta)$ and $S^*_3(\beta)$ must equal zero and these two functions depend on $\beta_2 - \beta_3$, so we consider the value of $\beta_2 - \beta_3$. One of two relationships can hold for $\beta_2 - \beta_3$, either $\beta_2 - \beta_3 < b_{2,3}$ or $\beta_2 - \beta_3 > b_{2,3}$.

If $\beta_2 - \beta_3 > b_{2,3}$ then $U_{2,3}(\beta) < n_2 n_3 / 2$ and hence $S^*_3(\beta) > 0$ so that $\beta$ is not a solution to the estimating equations.

If $\beta_2 - \beta_3 < b_{2,3}$ then $U_{2,3}(\beta) > n_2 n_3 / 2$. For this case, $\beta_3 < \beta_1 - b_{1,3} < \min \{HL_1 - b_{1,3} \leq b_{1,3} - b_{2,3} + b_{2,4} - b_{1,3}\}$ so that $\beta_2 < \beta_3 + b_{2,3} < b_{2,4}$. Thus, $\beta_1 < b_{1,4}$, $\beta_2 < b_{2,4}$, and $\beta_3 < b_{3,4}$ which implies that $U_{1,4}(\beta) > n_1 n_4 / 2$, $U_{2,4}(\beta) > n_2 n_4 / 2$, and $U_{3,4}(\beta) > n_3 n_4 / 2$. Thus the constraint (4.28) is not satisfied. Hence $\beta$ is not a solution in this case.

Thus, if $\beta_1 - \beta_2 < b_{1,2}$, then $\beta$ does not satisfy the estimating equations and hence is not a solution.

**Case 2:** If $\beta_1 - \beta_2 > b_{1,2}$, then $U_{1,2}(\beta) < n_1 n_2 / 2$ and hence $S^*_1(\beta)$ may equal or cross zero. Since $\beta_1 - \beta_2 > b_{1,2}$ and $\beta_1 < \min \{HL_1 \leq b_{1,2} - b_{2,4}\}$, the vector $\beta$ must satisfy $\beta_2 < b_{2,4}$ which implies that $U_{2,4}(\beta) > n_2 n_4 / 2$. For $\beta$ to be a solution, $S^*_2(\beta)$ and $S^*_3(\beta)$ must equal zero and these two functions depend on $\beta_2 - \beta_3$, so we consider the value of $\beta_2 - \beta_3$. One of two relationships can hold for $\beta_2 - \beta_3$, either $\beta_2 - \beta_3 < b_{2,3}$ or $\beta_2 - \beta_3 > b_{2,3}$.

If $\beta_2 - \beta_3 < b_{2,3}$, then $U_{2,3}(\beta) > n_2 n_3 / 2$. This implies that $S^*_2(\beta) > 0$ and so $\beta$ can not be a solution to the estimating equations.
If $\beta_2 - \beta_3 > b_{2,3}$ then $U_{2,3}(\beta) < n_{2n3} / 2$ and thus $S_2^*(\beta)$ may equal or cross zero. For this case, $\beta_3 < \beta_2 - b_{2,3} < \beta_1 - b_{1,2} - b_{2,3} < \min HL_1 - b_{1,2} - b_{2,3} \leq b_{1,2} + b_{2,3} + b_{3,4} - b_{1,2} - b_{2,3}$ so that $\beta_3 < b_{3,4}$. We consider $\beta_1 - \beta_3$ to check whether $S_3^*(\beta)$ is too large or small. One of two relationships can hold for $\beta_1 - \beta_3$, either $\beta_1 - \beta_3 < b_{1,3}$ or $\beta_1 - \beta_3 > b_{1,3}$.

If $\beta_1 - \beta_3 > b_{1,3}$, then $U_{1,3}(\beta) < n_{1n3} / 2$. Since $\beta_1 - \beta_3 > b_{1,3}$ and $\beta_1 < \min HL_1 \leq b_{1,3} - b_{3,4}$, $\beta$ must satisfy $\beta_3 < b_{3,4}$ which implies that $U_{3,4}(\beta) > n_{3n4} / 2$. Thus, $S_3^*(\beta) > 0$ and $\beta$ is not a solution in this case.

If $\beta_1 - \beta_3 < b_{1,3}$, then $U_{1,3}(\beta) > n_{1n3} / 2$. Here, $\beta_1 < b_{1,4}, \beta_2 < b_{2,4},$ and $\beta_3 < b_{3,4}$ which implies that $U_{1,4}(\beta) > n_{1n4} / 2, U_{2,4}(\beta) > n_{2n4} / 2,$ and $U_{3,4}(\beta) > n_{3n4} / 2$. Thus the constraint (4.28) is not satisfied. Hence $\beta$ is not a solution in this case.

Thus, if $\beta_1 - \beta_2 > b_{1,2}$, then $\beta$ does not satisfy the estimating equations and hence is not a solution.

From Case 1 and Case 2, if $\beta_1 < \min HL_1$, then $\beta$ does not satisfy the estimating equations and hence is not a solution.

Suppose now that $\beta$ satisfies $\beta_1 > \max HL_1$. It follows from Case 1 and Case 2 above with all inequalities reversed, that either $S_1^*(\beta), S_2^*(\beta),$ or $S_3^*(\beta)$ is too small or that the constraint is not satisfied. Thus $\beta$ does not satisfy the estimating equations and hence is not a solution.

Since if $\beta_1 < \min HL_1$ or $\beta_1 > \max HL_1$, then $\beta$ is not a solution to the estimating equations, $\beta$ must satisfy (4.30).

A similar argument can be used to show that if $\beta_2 < \min HL_2$ or $\beta_2 > \max HL_2$, then $\beta$ is not a solution to the estimating equations. Thus $\beta$ must satisfy (4.31). Finally, a
similar argument can be used to show that if $\beta_3 < \min HL_3$ or $\beta_3 > \max HL_3$, then $\hat{\beta}$ is not a solution to the estimating equations. Thus $\hat{\beta}$ must satisfy (4.32). The result follows from these arguments.

(Q.E.D.)

After having proved Theorem 4.1, we prove a technical lemma which establishes that the union of the previously mentioned restricted regions contains the Jaeckel estimate. The union of these regions is contained in RS, and thus provides sharper bounds for the estimator than those set forth in Theorem 4.1. However, the bounds in Theorem 4.1 are much easier to describe and prove useful later in computing the estimate.

As indicated earlier, the estimates $b_{1,2}$, $b_{1,3}$, and $b_{2,3}$ satisfy $U_{1,2}(\hat{\beta}) = n_1 n_2 / 2$, $U_{1,3}(\hat{\beta}) = n_1 n_3 / 2$, and $U_{2,3}(\hat{\beta}) = n_2 n_3 / 2$, respectively. However, to completely determine the value of $S_j^*(\hat{\beta})$, $j=1, 2, 3$ at a given $\hat{\beta}$, we need information about the ranks of the observations in sample 4 compared to each of the aligned samples 1, 2, and 3. That is, we need the values of $U_{1,4}(\hat{\beta})$, $U_{2,4}(\hat{\beta})$, and $U_{3,4}(\hat{\beta})$ at the given value of $\hat{\beta}$.

Thus, we also look at the estimators $b_{1,4}$, $b_{2,4}$, and $b_{3,4}$ and consider their relationship to a vector $\hat{\beta}$ in regions $R_1$, $R_2$, ..., $R_7$. By examining the value of $S_1^*(\hat{\beta})$, $S_2^*(\hat{\beta})$, and $S_3^*(\hat{\beta})$ for $\hat{\beta}$ in regions $R_1$, $R_2$, ..., $R_7$ we can restrict the area in which a solution to the Jaeckel estimating equations is possible by ruling out regions of $R^3$ in which $S_j^*(\hat{\beta})$ is too large or too small. We now consider the possible regions in which a solution may occur.

Define the restricted regions as follows:

$$RR_1 = R_1 \cap \{ \hat{\beta}: \beta_1 > b_{1,4}, \beta_2 < b_{2,4} \}$$

$$RR_2 = R_2 \cap \{ \hat{\beta}: \beta_2 < b_{2,4}, \beta_3 > b_{3,4} \}$$

$$RR_3 = R_3 \cap \{ \hat{\beta}: \beta_1 < b_{1,4}, \beta_3 > b_{3,4} \}$$
We now show that a solution may occur only in the restricted regions.

**Lemma 4.1:** If \( \beta \) satisfies the Jaeckel estimating equations, then \( \beta \) is contained in the union of the restricted regions, \( RR1 \cup RR2 \cup ... \cup RR7 \), where the restricted regions are defined as in (4.33).

**Proof:** Assume \( \beta \) satisfies the Jaeckel equations and \( \beta \) is not contained in the union of the regions \( RR1, RR2, ..., RR7 \). Then \( \beta \) must lie outside of the union of the restricted regions. We will show for any \( \beta \) outside of \( RR1 \cup RR2 \cup ... \cup RR7 \), that at least one of the functions \( S_j^*(P) \) for \( j=1, 2, \) or \( 3 \) is either too large or too small in value to satisfy the estimating equations, or that the constraint is not satisfied. Hence \( \beta \) outside the restricted region cannot be a solution. Since the Jaeckel estimate exists (Jaeckel 1972), it must be contained in the union of the restricted regions, \( RR1, RR2, ..., RR7 \).

For a vector \( \beta \) in region \( R1 \), \( U_{1,2}(\beta) > n_1n_2/2 \) and \( U_{1,3}(\beta) > n_1n_3/2 \) so \( S_1^*(\beta) \) is too large if \( \beta_1 < b_{1,4} \). Also, \( U_{2,3}(\beta) < n_2n_3/2 \) so \( S_2^*(\beta) \) is too small if \( \beta_2 > b_{2,4} \). Hence \( \beta \) in \( R1 \) can only be a solution to the estimating equations if \( \beta_1 > b_{1,4} \) and \( \beta_2 < b_{2,4} \). Thus a solution may occur in the restricted region \( RR1 = R1 \cap \{ \beta: \beta_1 > b_{1,4} \} \cap \{ \beta: \beta_2 < b_{2,4} \} \).

Similarly, for \( \beta \) in region \( R2 \), \( S_3^*(\beta) \) is too large if \( \beta_3 < b_{3,4} \) and \( S_2^*(\beta) \) is too small...
if $\beta_2 > b_{2,4}$. Thus a solution may also occur in $\text{RR2} = \text{R2} \cap \{ \beta : \beta_2 < b_{2,4} \} \cap \{ \beta : \beta_3 > b_{3,4} \}$.

If $\beta$ is in region R3, then $S_3^*(\beta)$ is too large if $\beta_3 < b_{3,4}$ and $S_1^*(\beta)$ is too small if $\beta_1 > b_{1,4}$. Hence a solution may be contained in $\text{RR3} = \text{R3} \cap \{ \beta : \beta_1 < b_{1,4} \} \cap \{ \beta : \beta_3 > b_{3,4} \}$.

For a $\beta$ in region R4, the function $S_2^*(\beta)$ is too large if $\beta_2 < b_{2,4}$ and $S_1^*(\beta)$ is too small if $\beta_1 > b_{1,4}$. Thus $\text{RR4} = \text{R4} \cap \{ \beta : \beta_1 < b_{1,4} \} \cap \{ \beta : \beta_2 > b_{2,4} \}$ may contain a solution.

Similarly, for a $\beta$ in region R5, $S_2^*(\beta)$ is too large if $\beta_2 < b_{2,4}$ and $S_3^*(\beta)$ is too small if $\beta_3 > b_{3,4}$. So a solution may occur in $\text{RR5} = \text{R5} \cap \{ \beta : \beta_2 > b_{2,4} \} \cap \{ \beta : \beta_3 < b_{3,4} \}$.

If $\beta$ is in region R6, then $S_1^*(\beta)$ is too large if $\beta_1 < b_{1,4}$ and $S_3^*(\beta)$ is too small if $\beta_3 > b_{3,4}$. Thus $\text{RR6} = \text{R6} \cap \{ \beta : \beta_1 > b_{1,4} \} \cap \{ \beta : \beta_3 < b_{3,4} \}$ may contain a solution.

Finally if $\beta$ is in region R7, then the quantity $N(N+1)/2 - \{ S_1^*(\beta) + S_2^*(\beta) + S_3^*(\beta) \}$ is too large if $\beta_1 < b_{1,4}$, $\beta_2 < b_{2,4}$, and $\beta_3 < b_{3,4}$. Similarly, it is too small if $\beta_1 > b_{1,4}$, $\beta_2 > b_{2,4}$, and $\beta_3 > b_{3,4}$. Thus a solution may be contained in $\text{RR7} = \text{R7} \cap \{ \beta : \beta_1 < b_{1,4}, \beta_2 < b_{2,4}, \text{and } \beta_3 < b_{3,4} \} \cup \{ \beta : \beta_1 > b_{1,4}, \beta_2 > b_{2,4}, \text{and } \beta_3 > b_{3,4} \}$.

(Q.E.D.)

Thus, only the restricted regions RR1, RR2, ..., RR7 may contain a solution to the Jaeckel estimating equations. Each restricted region is defined by the relationship of $\beta$ to the six planes $\beta_1 - \beta_2 = b_{1,2}$, $\beta_1 - \beta_3 = b_{1,3}$, $\beta_2 - \beta_3 = b_{2,3}$, $\beta_1 = b_{1,4}$, $\beta_2 = b_{2,4}$, and $\beta_3 = b_{3,4}$. Many of the restricted regions will be empty, depending on the relationship of the point $(b_{1,4}, b_{2,4}, b_{3,4})$ to the triangular cylinder defined by the other three planes.

Thus, once the values of the six Hodges-Lehmann estimates are known for a given data
set, many of the restricted regions can be determined to be empty due to the values of $b_{r,s}$.

To illustrate this determination of restricted regions as empty, we consider the following example.

Suppose that $(b_{1,4}, b_{2,4}, b_{3,4}) \in R_1$. Then the following relationships hold between the estimates:

$$
\begin{align*}
&b_{1,4} - b_{2,4} < b_{1,2}, \\
&b_{1,4} - b_{3,4} < b_{1,3}, \text{ and} \\
&b_{2,4} - b_{3,4} > b_{2,3}.
\end{align*}
$$

(4.34)

Thus, if $\beta_1 < b_{1,4}$ and $\beta_2 > b_{2,4}$ then $\beta_1 - \beta_2 < b_{1,4} - b_{2,4} < b_{1,2}$. This implies that RR4 is empty and RR7 is further restricted. Similarly, if $\beta_1 < b_{1,4}$ and $\beta_3 > b_{3,4}$ then $\beta_1 - \beta_3 < b_{1,4} - b_{3,4} < b_{1,3}$, which implies that RR3 is empty and that RR2 is further restricted. Furthermore, if $\beta_2 > b_{2,4}$ and $\beta_3 < b_{3,4}$ then $\beta_2 - \beta_3 > b_{2,4} - b_{3,4} > b_{2,3}$. This implies that RR5 is empty and that RR6 is further restricted.

In summary, if $(b_{1,4}, b_{2,4}, b_{3,4}) \in R_1$, then the only regions which may contain a solution to the estimating equations are

$$
\begin{align*}
&\text{RR1}, \; \text{RR2} \cap \{\beta: \beta_1 > b_{1,4}\}, \; \text{RR6} \cap \{\beta: \beta_2 < b_{2,4}\}, \text{ and} \\
&\text{RR7} \cap \{\beta: \beta_1 < b_{1,4}, \beta_2 > b_{2,4}\}^c.
\end{align*}
$$

(4.35)

A similar restriction of regions can be described for $(b_{1,4}, b_{2,4}, b_{3,4})$ in any of the restricted regions RR1 through RR7.

After having discussed the relationship of the Jaeckel estimator for $\beta$ and the two-sample Hodges-Lehmann estimators of $\beta$ for the four-sample case, we now consider the
Following the restricted regions approach presented for the four-sample case, we can describe the regions which may contain a solution to the Jaeckel estimating equations for the k-sample case. These regions, however, are even more difficult to work with than those of the four-sample case. Thus we want to find a bounded region easily specified from the two-sample Hodges-Lehmann estimators, that contains the Jaeckel estimator. Before describing this region, we examine in more detail the estimating equations.

As evident from the extension of the three-sample case to the four-sample case, there is a relationship between the estimating equations for the (k-1)-sample case and the k-sample case. To indicate the number of samples on which the estimating equations depend, we adopt for the remainder of this chapter the notation

\[ S_{j,k}^*(\beta) = \sum_{m=1}^{j-1} \left\{ \frac{n_{jm}m}{2} - U_{m,j}(\beta) \right\} + \sum_{m=j+1}^{k} \left\{ U_{j,m}(\beta) - \frac{n_{jm}m}{2} \right\}, \]

for \( j = 1, 2, ..., k-1. \) (4.36)

Since \( U_{r,s}(\beta) \) is a function of \( \beta \) through \( \beta_r - \beta_s \) alone, no confusion arises in writing \( U_{r,s}((\beta_1, \beta_2, ..., \beta_{k-2})^T) \) in the (k-1)-sample case as \( U_{r,s}((\beta_1, \beta_2, ..., \beta_{k-2}, \beta_{k-1})^T) \) as long as \( r \) and \( s \) are less than \( k-1. \) The following relationship holds for the estimating equations:

\[ S_{j,k}^*(\beta) = S_{j,(k-1)}^*(\beta) + \left\{ U_{j,k}(\beta) - \frac{n_j k}{2} \right\}, \quad 1 \leq j < k-1. \] (4.37)

Based on this relationship between the equations for the case of \( k \) samples and the equations for the case of \( (k-1) \) samples, we conjecture the following extension of
Theorem 4.1 to the general case of k samples.

Conjecture 4.1 Let $\underline{b}_{1:k}=(b_1, b_2, \ldots, b_{k-1})^T$ satisfy the conditions:

$$S_j^*(\underline{b}) = 0 \text{ for } j=1, 2, \ldots, k-1 \quad (4.38)$$

subject to the constraint

$$\sum_{j=1}^{k-1} S_j^*(\underline{b}) \in \left[ -\frac{n_k(N-n_k)}{2}, \frac{n_k(N-n_k)}{2} \right]. \quad (4.39)$$

Then the point $\underline{b}_{1:k}$ lies in the rectangular solid defined by

$$RS = \{ \underline{b} : \beta_j \in [\min \text{ HL}_j, \max \text{ HL}_j], j=1, 2, \ldots, k-1 \}, \quad (4.40)$$

where HL$_j$ denotes the set of Hodges-Lehmann two-sample estimators for $\beta_j$.

While we have not found any examples which disprove this conjecture, we have not yet shown that the conjecture is true. From (4.37) it appears that a proof by induction on $k$ is possible, but part of the inductive step has not been completed. Given that the conjecture holds for four samples, let us consider how to prove the conjecture for $k=5$.

From (4.37),

$$S_{j,5}^*(\underline{b}) = S_{j,4}^*(\underline{b}) + \left\{ U_{j,5}(\underline{b}) - \frac{n_{15} n_5}{2} \right\}, \quad j = 1, 2, 3. \quad (4.41)$$

If $\beta_1 < \min \text{ HL}_1$, then from the proof of Theorem 4.1 either $S_{j,4}^*(\underline{b}) > 0$ for at least one $j$. 
= 1, 2, 3 or \( \sum_{j=1}^{3} S_j^*(b) > 0 \), depending on the relationships of \( \beta_i - \beta_m \) to \( b_{i,m} \), for each \( i \) and \( m \). Consider the case for which \( S_{2,4}^*(\mathbf{p}) > 0 \) for some \( \mathbf{p} \) such that \( \beta_1 < \min H_{L_1} \). If \( U_{2,5}(\mathbf{p}) > 0 \), then \( S_{2,5}^*(\mathbf{p}) > 0 \) and \( \mathbf{p} \) will not be a solution to the Jaeckel equations.

However, if \( U_{2,5}(\mathbf{p}) < 0 \), then \( S_{2,5}^*(\mathbf{p}) \) might equal or jump across zero. Thus for this set of inequalities, namely \( \beta_2 > b_{2,5} \) and the inequalities for which \( S_{2,4}^*(\mathbf{p}) > 0 \) is true, the result does not immediately follow. If it can be shown that \( \mathbf{p} \) is not a solution to the estimating equations for this set of inequalities, then the argument should extend to cover the remaining cases and the inductive step of the proof of the conjecture should follow.

### 4.4 Ideas for an Algorithm

In Sections 4.1 through 4.3 we formulated the Jaeckel estimating equations, characterized the Jaeckel estimator geometrically, and established a relationship between the Jaeckel estimator and the Hodges-Lehmann two-sample estimators. In this section we describe the use of these results in the computation of the Jaeckel estimate in the four-sample case.

In Section 3.4 the approach for computing the Jaeckel estimate in the three-sample case can be summarized in the following manner. First compute the Hodges-Lehmann two-sample estimates of \( \beta_1, \beta_2 \), and \( \beta_1 - \beta_2 \). If these are not a solution to the Jaeckel estimating equations, then they determine the bounds on the Jaeckel estimates of \( \beta_1 \) and \( \beta_2 \). From these bounds we obtain a range of integer values which correspond to the values of the contours of the function \( U_{1,2}(\mathbf{p}) \) in which a solution to the estimating equations may occur. We iteratively set \( U_{1,2}(\mathbf{p}) \) equal to one of these contour values. Then we find the intervals for \( \beta_1 \) and \( \beta_2 \) in this contour where \( S_1^*(\mathbf{p}) \) and \( S_2^*(\mathbf{p}) \) are equal to or jump across zero. We check whether these intervals overlap and when we find a region of overlap, a solution to the estimating equations occurs in this region of
This approach can be extended to the case of four samples. The two-sample Hodges-Lehmann estimates of shift are computed and applying Theorem 4.1 we obtain bounds on the Jaeckel estimates of $\beta_1$, $\beta_2$, and $\beta_3$ from these two-sample estimates. These bounds determine a range of values for the contours of the functions $U_{1,2}(\beta)$ and $U_{1,3}(\beta)$ where a solution to the estimating equations can occur. We restrict $\beta$ to the cylinder in which the pair $(U_{1,2}(\beta), U_{1,3}(\beta))$ equals a specified pair $(A, B)$ falling within the range of contour values described above. We can determine the regions within this cylinder in which the functions $S_1^*(\beta)$, $S_2^*(\beta)$, and $S_3^*(\beta)$ equal or jump across zero. If these regions overlap, a solution to the estimating equations occurs in the region of overlap. Otherwise, we select a new pair $(A, B)$ and repeat the process until we find a solution. Because we have a finite number of pairs $(A, B)$ determined by the bounds, a finite number of iterations is required. In Chapter V we provide the details of the algorithm.

4.5 Example

In this section we consider the example described in Section 3.5 in which maternal behavior was induced in virgin rats by injecting them with plasma from donor rats at various phases of the estrous cycle. In Section 3.5 we considered only the three groups injected with maternal plasma, diestrus plasma, and saline. In this section we consider all four treatments.

Let the groups of rats injected with maternal plasma, proestrus plasma, diestrus plasma, and saline be denoted by samples 1, 2, 3, and 4, respectively. The data is found in Table 1 and the ordered differences between pairs of samples are given in Table 2. From Table 2, the medians of the pairwise differences between each pair of samples are
in the intervals:

\[ b_{1,2} \in [-3.1, -3.0), \quad b_{1,3} \in [-1.9, -1.8), \quad b_{1,4} \in [-4.0, -3.7), \]

\[ b_{2,3} \in [0.4, 0.7), \quad b_{2,4} \in [-1.0, -0.8), \quad b_{3,4} \in [-0.8, -0.6). \]  \hspace{1cm} (4.42)

From Theorem 4.1 the bounds on the Jaeckel estimates are given by

\[ \beta_j \in [\min HL_j, \max HL_j], \text{ for } j=1, 2, 3. \]

For this example,

\[ \min HL_1 = \min \{ -4.0, -3.1 + (-1.0), -1.9 + (-0.8), -1.9 - (-0.7) + (-1.0), 
-3.1 + 0.4 + (-0.8) \} = -4.1, \]

\[ \max HL_1 = \max \{ -3.7, -3.0 + (-0.8), -1.8 + (-0.6), -1.8 - (-0.4) + (-0.8), 
-3.0 + 0.7 + (-0.6) \} = -2.4, \]

\[ \min HL_2 = \min \{ -1.0, -4.0 - (-3.0), 0.4 + (-0.8), 0.4 - (-1.8) + (-4.0), 
-1.9 - (-3.0) + (-0.8) \} = -1.8, \]

\[ \max HL_2 = \max \{ -0.8, -3.7 - (-3.1), 0.7 + (-0.6), 0.7 - (-1.9) + (-3.7), 
-1.8 - (-3.1) + (-0.6) \} = 0.7, \]

\[ \min HL_3 = \min \{ -0.8, -4.0 - (-1.8), -1.0 - (-0.7), -4.0 - (-3.0) - (0.7), 
-1.0 + (-3.1) - (-1.8) \} = -2.3, \]

\[ \max HL_3 = \max \{ -0.6, -3.7 - (-1.9), -0.8 - 0.4, -3.7 - (-3.1) - 0.4, 
-0.8 + (-3.0) - (-1.9) \} = -0.6. \]  \hspace{1cm} (4.43)

We find the range of \( A \) and \( B \) that are described in the previous section. The bounds on

\[ \beta_1 - \beta_2 \text{ and } \beta_1 - \beta_3 \] are given by the following
\[
\text{min } \text{HL}_1 - \text{max } \text{HL}_2 = -4.8 \leq \beta_1 - \beta_2 \leq -0.6 = \text{max } \text{HL}_1 - \text{min } \text{HL}_2,
\]

and

\[
\text{min } \text{HL}_1 - \text{max } \text{HL}_3 = -3.5 \leq \beta_1 - \beta_3 \leq -0.1 = \text{max } \text{HL}_1 - \text{min } \text{HL}_3.
\]

Since \(-4.8 \in [-4.9, -4.6) = [D^{(1,2)}_{(15)}, D^{(1,2)}_{(16)}]\) and \(-0.6 = D^{(1,2)}_{(50)} = D^{(1,2)}_{(51)} = D^{(1,2)}_{(52)}\), the integer \(A\) ranges in value from 14 to 49. Similarly, since \(-3.5 \in [-3.7, -3.4) = [D^{(1,3)}_{(21)}, D^{(1,3)}_{(22)}]\) and \(-0.1 = D^{(1,2)}_{(50)}\), the integer \(B\) ranges in value from 14 to 43. As discussed in the previous section, we select a pair \((A, B)\) falling within the range of contour values described above, we restrict \(\beta\) to the cylinder in which the pair \((U_{1,2}(\beta), U_{1,3}(\beta))\) equals \((A, B)\), and we determine the regions within this cylinder in which the functions \(S_1^*(\beta), S_2^*(\beta), \) and \(S_3^*(\beta)\) equal or jump across zero. If these regions overlap, a solution to the estimating equations occurs in the region of overlap. The details for determining these regions are given in Chapter V.
5.0 Introduction

As discussed in Chapter II, several authors have addressed the problem of computing estimates from rank statistics in the one-sample and two-sample problems with uncensored data. If done naively, the computation of these estimates requires a large amount of the computer time and storage. For example, the Hodges-Lehmann two-sample estimator of shift is the median of the pairwise differences. One could compute the estimate in a straightforward way by computing the $n_1n_2$ differences, sorting the differences, and finding the median. This approach would require an array of $n_1n_2$ elements to store the differences and $n_1n_2$ steps to compute the differences. Even for moderate sample sizes, this approach can be costly. A fast, exact algorithm, called the divide-and-conquer algorithm, has been developed to reduce the amount of computer time and storage required to compute the rank estimates in the one-sample and two-sample problem. This algorithm is of use in computing the Jaeckel estimates of the regression coefficient in the $k$-sample case with Wilcoxon scores. In this chapter we first discuss the divide-and-conquer algorithm for the two-sample case. In the second section we provide an algorithm to be used in the $k$-sample case.

5.1 Divide-and-Conquer Algorithm for the Two-Sample Hodges-Lehmann Estimator

We consider a fast, exact method to find the solution to the equation $T(t) = c$, where
T(\cdot) is a step function and \(c\) is a constant. This divide-and-conquer method is an iterative procedure in which a trial value is selected, \(T(\cdot)\) is evaluated at the trial value, the stopping criterion is checked, and the process is repeated if necessary. At each iteration, the trial value is selected from a set containing only a finite number of possible solutions to the equation, the true solution being among these. We provide an outline for this algorithm in the two-sample problem, along with the details for its use.

Suppose \(Y(1) \leq \ldots \leq Y(n_1)\) and \(Y(n_1+1) \leq \ldots \leq Y(n_1+n_2)\). Let \(T(\cdot)\) be a nonincreasing, right-continuous step function which jumps at the pairwise differences, \(Y(j)-Y(n_1+i)\) for \(i=1,\ldots, n_2\) and \(j=1,\ldots, n_1\). The function \(T(\cdot)\) can be written as

\[
T(t) = \sum_{i=1}^{n_2} \sum_{j=1}^{n_1} W(i, j) \psi(Y(j) - Y(n_1+i) - t),
\]

where \(\psi(u) = 1\) if \(u > 0\) and 0 otherwise, and \(W(i, j), i=1,\ldots, n_2\) and \(j=1,\ldots, n_1\) are weights which determine the step size. Since \(T(\cdot)\) is a step function with a finite number of jumps, for a given \(c\) such that \(\min T(t) < c < \max T(t)\) for \(-\infty < t < \infty\), either \(T(\cdot)\) equals \(c\) on an interval or \(T(\cdot)\) steps across \(c\) at some jump point, \(t^*\). Note that if \(T(\cdot)\) steps across \(c\) at some \(t^*\) then \(t^*\) is a pairwise difference, \(Y(j)-Y(n_1+i)\) for some \((i, j)\). Similarly, if \(T(t)=c\) on some interval \(t_L \leq t < t_R\), then \(t_L\) and \(t_R\) are pairwise differences of samples 1 and 2. Thus the set of all pairwise differences, \(S_0 = \{Y(j)-Y(n_1+i): i=1,\ldots, n_2\) and \(j=1,\ldots, n_1\}\) contains either the value \(t^*\) or the endpoints \(t_L\) and \(t_R\). The divide-and-conquer algorithm to be described finds \(t^*\) or \(t_L\) and \(t_R\) by a fast, exact method. Although we have defined \(T(\cdot)\) to be right-continuous, the algorithm could just as easily be described for a left-continuous step function. The use of the algorithm in the case of a left-continuous step function is illustrated in Chapter VI.

The basic algorithm returns \(\hat{t}\), where \(\hat{t}\) is either \(t^*\) or one of the endpoints, \(t_L\) or \(t_R\).
A second routine can be used to find the other endpoint, but for now we only consider the
details of the basic routine. The algorithm begins with the set $S_0$ and $p=0$. We visualize
the elements of $S_0$ as the elements of an $n_2 \times n_1$ rectangular array in which the $(i, j)$
element, denoted $A(i, j)$, is the difference $Y(j) - Y(n_1+i)$. At step $p$, a trial value $t_p \in S_p$ is
selected and the set $S_0$ is partitioned into three subsets. Let $L_p$ be the subset of $S_0$ with
elements less than $t_p$. Similarly, $E_p$ and $G_p$ are the subsets of $S_0$ with elements equal to
$t_p$ and greater than $t_p$, respectively. So $T(t_p) = |G_p|$ and $T(t_p^-) = |G_p| + |E_p|$, where
$|L_p|$, $|G_p|$, and $|E_p|$ denote the total weights corresponding to elements in the sets $L_p$,
$G_p$, and $E_p$, respectively. The sets $L_p$, $E_p$, and $G_p$ are never formed. Instead, the
boundaries of the sets in the conceptual array are found and from these boundaries $T(t_p)$
and $T(t_p^-)$ are computed. On the basis of these values, we can determine whether or not
$t_p$ is a jump point or an endpoint. Otherwise, $\hat{t}$ is in the set $S_p \cap L_p$ or $S_p \cap G_p$. The
algorithm can be outlined as follows.

1. Sort $Y_j$, $i=j, \ldots, n_1$ and denote by $Y(1) \leq \ldots \leq Y(n_1)$.
2. Sort $Y_{n_1+i}$, $i=1, \ldots, n_2$ and denote by $Y(n_1+1) \leq \ldots \leq Y(n_1+n_2)$.
3. Let $S_0 = \{ Y(j) - Y(n_1+i) ; i=1, \ldots, n_2 \text{ and } j=1, \ldots, n_1 \}$.
4. Set $p=0$.
5. Do until $\hat{t}$ is found:
   5.1. Choose $t_p \in S_p$.
   5.2. Find $l_1(i)$ and $r_1(i)$, $i=1, \ldots, n_2$.
   5.3. Compute $T(t_p)$ and $T(t_p^-)$.
   5.4. Check if $\hat{t} = t_p$.
      If $T(t_p^-) > c$ and $T(t_p) < c$ then $\hat{t} = t_p$ is a jump point.
      If $T(t_p^-) > c$ and $T(t_p) = c$ then $\hat{t} = t_p$ is the left endpoint.
If $T(t_p) = c$ and $T(t_p) < c$ then $t = t_p$ is the right endpoint.

If $T(t_p) < c$ then $S_{p+1} = S_p \cap L_p$.

If $T(t_p) > c$ then $S_{p+1} = S_p \cap G_p$.

5.5. $p = p + 1$.

6. End do. (5.2)

To determine the complexity of the algorithm, we consider either the maximum number of steps required or the expected number of steps required. The number of steps, denoted $u(n_1, n_2)$, in either the worst-case or on average is usually described using the notation $O(v(n_1, n_2))$, where $u(n_1, n_2)$ is $O(v(n_1, n_2))$ if $|u(n_1, n_2)/v(n_1, n_2)|$ remains bounded as $n_1$ and $n_2$ approach infinity. To use this algorithm to find a solution to the equation $T(t) = c$, we need to consider the details of how to compute $T(t)$ in $O(n_1 + n_2)$ steps, how to select a trial value, and when to stop iterating. To compute $T(t)$, one visualizes the $n_2 \times n_1$ rectangular array of elements $A(i, j)$. The elements in the array correspond to the elements of the set $S_0$, the set of all pairwise differences. Since the $Y_j$'s and the $Y_{n_1+i}$'s are sorted in ascending order, the $A(i, j)$ increase as $j$ increases, for each $i=1,\ldots, n_2$ and the $A(i, j)$ decrease as $i$ increases, for each $j=1,\ldots, n_1$. Thus the smallest element lies in the bottom left corner, namely $A(n_2, 1)$, and the largest element lies in the top right corner of the array, namely $A(1, n_1)$. For a given trial value $t$, the partitioning of $S_0$ into subsets of differences less than $t$, differences greater than $t$, and differences equal to $t$, denoted $L$, $G$, and $E$, respectively, can be accomplished in $O(n_1 + n_2)$ steps by making use of the patterns just noted. Given a trial value of $t$, for each row $i$ define pointers $l_1(i)$ and $r_1(i)$ such that
If \( Y(j) - Y(n1+i) > t \) for all \( i = 1, \ldots, n2 \), then \( l1(i) = 0 \). Similarly, if \( Y(j) - Y(n1+i) < t \) for all \( i = 1, \ldots, n2 \), then \( r1(i) = n2 + 1 \). Because the elements in any row are monotone, the pointers \( l1(\cdot) \) and \( r1(\cdot) \) indicate which elements of that row belong to the sets \( L, G, \) and \( E \). Since for fixed \( j \) the pairwise differences decrease as \( i \) increases, once the pointer \( l1(i) \) for the \( i \)-th row is found, the elements \( Y(j) - Y(n1+i+1) \) for \( j \leq l1(i) \) must also be less than \( t \). Thus the pointer for the \((i+1)\)-st row is known to be greater than or equal to the pointer for the \( i \)-th row, that is \( l1(i+1) \geq l1(i) \). Similarly, once the pointer \( r1(i+1) \) for the \((i+1)\)-st row is found, the elements \( Y(j) - Y(n1+i) \) must also be greater than \( t \) for \( j \geq r1(i+1) \). Thus the pointer for the \( i \)-th row is known to be less than or equal to the pointer for the \((i+1)\)-st row, that is \( r1(i+1) \geq r1(i) \). To find the pointers, begin in the upper left corner and move to the right and downward. Using this approach, at most \( n1+n2 \) differences need to be computed, and the pointers can be found in \( O(n1+n2) \) steps.

To compute \( T(t) \) and \( T(t^-) \) we can make use of the pointers for each row. Using these pointers, the number of differences in row \( i \) for which \( Y(j) - Y(n1+j) > t \) is \( n1 - r1(i)+1 \) and the number of differences for which \( Y(i) - Y(n1+j) \geq t \) is \( n1 - l1(i) \). If the weights \( W(i, j) \) are all equal to 1, as in the Wilcoxon statistic, then

\[
T(t) = n2(n1+1) - \sum_{i=1}^{n2} r1(i) \quad \text{and} \quad T(t^-) = n1n2 - \sum_{i=1}^{n2} l1(i).
\]

(5.4)

If the weight function is not constant, it must be possible to compute \( W(i, j) \) in \( O(n1+n2) \) steps to maintain \( O(n1+n2) \) steps for the calculation of \( T(t) \). Aubuchon (1984) discussed the computation of estimates for general linear signed-rank statistics in the one-sample...
setting. These ideas can be extended to the two-sample setting. If the weights are such that the weight contributed by row i to $T(\cdot)$ depends only on the pointers for row i, then the computation can be done in the required number of steps.

Next one must consider how to select a trial value $t_p \in S_p$. Several methods have been suggested. The trial value at stage $p$ can be chosen to be the weighted median of the row medians of the elements in $S_p$, where the weights correspond to the number of elements of $S_p$ in that row, as used by Johnson and Mizoguchi (1978). The trial value could be a randomly chosen element in $S_p$, with each element having an equal chance of being selected. Or the trial value could be a random row median, with the probabilities being proportional to the number of elements in the row which are in $S_p$. These last two methods are suggested by Monahan (1984).

Finally, we consider when to stop iterating. Johnson and Mizoguchi (1978) continue to iterate until either a solution is found or the number of elements in $S_p$ is less than $n_1 + n_2$, and the solution can be found by sorting the elements. Using their selection of trial value and their stopping rule, the complexity in the worst-case is $O((n_1+n_2) \log(n_1+n_2))$. Monahan (1984) stops iterating when a solution is found, the number of elements in $S_p$ is less than $n_1 + n_2$, or when the desired solution is the largest or smallest element of $S_p$. Since finding the largest or smallest element of $S_p$ can be done in $O((n_1+n_2))$ time, the total complexity of $O((n_1+n_2) \log(n_1+n_2))$ is maintained for Monahan's routine.

In summary, using this type of routine to find $\hat{t}$ requires an initial $O(n_1 \log n_1)$ steps to sort $Y_1, \ldots, Y_{n_1}$, and another $O(n_2 \log n_2)$ steps for the sorting of $Y_{n_1+1}, \ldots, Y_{n_1+n_2}$. Then, at each iteration $O(n_1+n_2)$ steps would be required for the evaluation of $T(t)$ from the pointers $l_1(\cdot)$ and $r_1(\cdot)$. The selection of the trial value requires $O(n_1+n_2)$ steps for the methods described above. The routine requires a storage vector of size $n_1$ for the $Y_j$'s, a
vector of size \( n_2 \) for the \( Y_{n_1+1} \)'s, two vectors of size \( n_2 \) for the pointers \( l_1(\cdot) \) and \( r_1(\cdot) \), and possibly a weight vector of size \( n_2 \).

In the next section we show how the partitioning ideas used in the divide-and-conquer algorithms can be used in computing the Jaeckel estimate of the vector of regression coefficients in the \( k \)-sample case.

5.2 Details of \( K \)-Sample Algorithm for the Jaeckel Estimator

Brief discussions of the approach for computing the Jaeckel estimate of \( \hat{\beta} \) were given for the three-sample and four-sample problems in Sections 3.4 and 4.4, respectively. In this section we provide the details of an algorithm for the computation of the Jaeckel estimate. Since the geometry for the case of \( k \) samples is difficult to visualize, the algorithm will be described for the four-sample case. We then discuss the extensions to the case of \( k \) samples. The algorithm uses the partitioning ideas of the divide-and-conquer routines described in the previous section.

Consider the four-sample model. The Jaeckel estimate is the solution to the system of equations \( S_j(\hat{\beta}) = 0 \), where \( S_j(\hat{\beta}) \), \( j = 1, 2, 3 \) is given in (4.17). In the discussion at the end of Section 4.2, we described the contours of \( S_j(\hat{\beta}) \) in terms of the contours of the functions \( U_{r,s}(\hat{\beta}) \), \( 1 \leq r < s \leq 4 \). Let \( A \) be an integer between 0 and \( n_1n_2 \) and let \( B \) be an integer between 0 and \( n_1n_3 \). Recall that the intersection of the contours on which \( U_{1,2}(\hat{\beta}) \) and \( U_{1,3}(\hat{\beta}) \) equal \( A \) and \( B \), respectively, is a four-sided cylinder that is parallel to the equiangular line. The trace of this cylinder in the \( \beta_3 = 0 \) plane is a parallelogram. Since \( U_{1,2}(\hat{\beta}) = A \) in this cylinder, \( \beta_1 - \beta_2 \) satisfies

\[
\beta_1 - \beta_2 \in \left[ D_{(n_1n_2-A)}^{(1,2)}, D_{(n_1n_2-A+1)}^{(1,2)} \right]
\]

(5.5)
Similarly, $\beta_1 - \beta_3$ satisfies

$$\beta_1 - \beta_3 \in \left[ D_{(n_1n_3-B)}^{(1,3)}, D_{(n_1n_3-B+1)}^{(1,3)} \right]$$

(5.6)

From conditions (5.5) and (5.6), it follows that $\beta_2 - \beta_3$ must satisfy

$$\beta_2 - \beta_3 \in \left[ D_{(n_1n_3-B)}^{(1,3)} - D_{(n_1n_2-A+1)}^{(1,2)}, D_{(n_1n_3-B+1)}^{(1,3)} - D_{(n_1n_2-A)}^{(1,2)} \right]$$

(5.7)

Thus in the cylinder the range of values of the function $U_{2,3}(\beta)$ depends on the quantities $D_{(n_1n_3-B)}^{(1,3)} - D_{(n_1n_2-A+1)}^{(1,2)}$ and $D_{(n_1n_3-B+1)}^{(1,3)} - D_{(n_1n_2-A)}^{(1,2)}$. Let $C_L(A, B)$ be the value of the function $U_{2,3}(\beta)$ evaluated at a vector $\beta$ such that $\beta_2 - \beta_3 = D_{(n_1n_3-B)}^{(1,3)} - D_{(n_1n_2-A+1)}^{(1,2)}$. Let $C_R(A, B)$ be the value of $U_{2,3}(\beta)$ evaluated at a vector $\beta$ for which $\beta_2 - \beta_3 = D_{(n_1n_3-B+1)}^{(1,3)} - D_{(n_1n_2-A)}^{(1,2)}$. Observe that $C_R(A, B) \leq C_L(A, B)$, implied by the monotonicity of $U_{2,3}(\beta)$.

The sides of the cylinder are defined by the equations $\beta_1 - \beta_2 = D_{(n_1n_3-B)}^{(1,3)} - D_{(n_1n_2-A+1)}^{(1,2)}$, $\beta_1 - \beta_3 = D_{(n_1n_3-B+1)}^{(1,3)} - D_{(n_1n_2-A)}^{(1,2)}$, and $\beta_1 - \beta_3 = D_{(n_1n_3-B)}^{(1,3)} - D_{(n_1n_2-A+1)}^{(1,2)}$. The trace of this cylinder in the $\beta_3 = 0$ plane is a parallelogram with vertices

$$L_1 = (D_{(n_1n_3-B)}^{(1,3)}, D_{(n_1n_3-B)}^{(1,3)} - D_{(n_1n_2-A+1)}^{(1,2)}, L_2 = (D_{(n_1n_3-B)}^{(1,3)}, D_{(n_1n_3-B)}^{(1,3)} - D_{(n_1n_2-A)}^{(1,2)},$$

$$R_1 = (D_{(n_1n_3-B+1)}^{(1,3)}, D_{(n_1n_3-B+1)}^{(1,3)} - D_{(n_1n_2-A+1)}^{(1,2)}, R_2 = (D_{(n_1n_3-B+1)}^{(1,3)}, D_{(n_1n_3-B+1)}^{(1,3)} - D_{(n_1n_2-A)}^{(1,2)}).$$

(5.8)
Let us consider the value of the functions $S_j^*(\beta)$, $j=1, 2, 3$ at points in this parallelogram. From Property 4.1, the minimum value of the function $S_1^*(\beta)$ in the parallelogram is at the vertex $R_1$ and the maximum value in the parallelogram of $S_1^*(\beta)$ is at the vertex $L_2$. Similarly, the minimum and maximum values of the function $S_2^*(\beta)$ in the parallelogram are at the vertices $R_2$ and $L_1$, respectively. The minimum and maximum values of the function $S_3^*(\beta)$ in the parallelogram are at the vertices $L_1$ and $R_2$, respectively. Thus, we can determine whether a solution occurs in the parallelogram by checking the values of the functions at the vertices. For example, if $S_1^*(R_1) < 0$ and $S_1^*(L_2) > 0$, then $S_1^*(\beta)$ equals or jumps across zero in the parallelogram. If both $S_1^*(R_1)$ and $S_1^*(L_2)$ have the same sign, then there is no solution to the first estimating equation in the parallelogram.

Given any value of $\beta_3$, we can check whether a solution to the estimating equations occurs in the parallelogram by checking the values of the functions $S_j^*(\beta)$ at the vertices. From the preceding paragraph, we see that given a cylinder and a value of $\beta_3$, we can determine if a solution to the Jaeckel equations occurs. Now we consider how to find the regions within a given cylinder where $S_1^*(\beta)$, $S_2^*(\beta)$, and $S_3^*(\beta)$ equal or cross zero. If these regions overlap, then a solution occurs in the region of overlap. Consider the value of the functions $U_{1,2}(\cdot)$, $U_{1,3}(\cdot)$, and $U_{2,3}(\cdot)$ at the vector $\beta + t_1$, where $t$ is a constant. Since $(\beta_1+t) - (\beta_2+t) = \beta_1 - \beta_2$, the function $U_{1,2}(\beta + t_1)$ equals $U_{1,2}(\beta)$. Following a similar argument, $U_{1,3}(\beta + t_1) = U_{1,3}(\beta)$ and $U_{2,3}(\beta + t_1) = U_{2,3}(\beta)$. The functions $U_{1,4}(\cdot)$, $U_{2,4}(\cdot)$, and $U_{3,4}(\cdot)$, however, do not remain constant along the equiangular line. Given we are restricted to the cylinder in which $U_{1,2}(\beta) = A$ and $U_{1,3}(\beta) = B$, the value of $t_1$ for which $S_1^*(\beta + t_1)$ equals or jumps across zero can be found in the following manner. Since
\[ 0 = S_1^*(\beta + t_1 \perp) = U_{1,2}(\beta) + U_{1,3}(\beta) + U_{1,4}(\beta + t_1 \perp) - \frac{n_1(n_2+n_3+n_4)}{2} \]

\[ = A + B + U_{1,4}(\beta + t_1 \perp) - \frac{n_1(n_2+n_3+n_4)}{2}, \]  

(5.9)

we must have

\[ U_{1,4}(\beta + t_1 \perp) = \frac{n_1(n_2+n_3+n_4)}{2} - A - B. \]  

(5.10)

Thus,

\[ \beta_1 + t_1 \in \left[ D^{(1,4)}_{(n_1(n_4-n_2-n_3)/2 +A+B)} - D^{(1,4)}_{(n_1(n_4-n_2-n_3)/2 +A+B+1)} \right] \]  

(5.11)

which implies that

\[ t_1 \in \left[ D^{(1,4)}_{(n_1(n_4-n_2-n_3)/2 +A+B)} - \beta_1, D^{(1,4)}_{(n_1(n_4-n_2-n_3)/2 +A+B+1)} - \beta_1 \right] \]

(5.12)

Thus, at a given vector \( \beta \) we must add \( t_1 \) from (5.12) to each of \( \beta_1, \beta_2, \) and \( \beta_3 \) for \( S_1^*(\cdot) \) to equal or cross zero.

Following a similar argument, given we are at the vector \( \beta \) in the cylinder, we must add \( t_2 \) to each vector element for \( S_2^*(\cdot) \) to equal or cross zero, where \( t_2 \) is given by

\[ t_2 \in \left[ D^{(2,4)}_{(n_2(n_4+n_1-n_3)/2 -A+C(A,B))} - D^{(2,4)}_{(n_2(n_4+n_1-n_3)/2 -A+C(A,B)+1)} - \beta_2 \right] \]

(5.13)

and \( C(A, B) = U_{2,3}(\beta) \) for the particular value of \( \beta \) in this cylinder. Similarly, we must
add \( t_3 \) to each element for \( S_3^*(\cdot) \) to equal or cross zero, where

\[
t_3 \in \left[ D_{(n_4+2n_1+n_2)/2 - B-C(A,B))} - \beta_3, D_{(n_4+2n_1+n_2)/2 - B-C(A,B)+1} - \beta_3 \right]
\]

(5.14)

For a solution to the Jaeckel estimating equations to occur in the given cylinder, \( S_j^*(\cdot) \), \( j=1, 2, 3 \) must have zero contours that all intersect, that is the intervals for \( t_1, t_2, \) and \( t_3 \) must overlap.

In summary, the Jaeckel estimate for the four-sample problem can be computed by the following steps:

(1.) Compute the median intervals for \( b_{1,2}, b_{1,3}, \) and \( b_{2,3} \).

(2.) Select the bounds on \( \beta_j, j=1, 2, 3 \) using Theorem 4.1.

(3.) Find the range of values for A and B, determined from the bounds on \( \beta_1 - \beta_2 \) and \( \beta_1 - \beta_3 \).

(4.) Initialize (A, B).

(5.) Compute \( t_1, t_2, \) and \( t_3 \) for the points \( R_1, R_2, L_1, \) and \( L_2, \) and check for overlap. If the intervals overlap, a solution to the estimating equations occurs in the region of overlap.

(6.) If no overlap, then select a new (A, B) and go back to (5.).

(5.15)

To use the above algorithm we need to consider how to find the range of values for A and B, how to determine whether the intervals for \( t_1, t_2, \) and \( t_3 \) overlap, and how to select a new trial pair (A, B) if there is no overlap between the intervals. First we consider how to find the range of values for A and B. Theorem 4.1 provides bounds for
\( \beta_1, \beta_2, \) and \( \beta_3. \) Since \( \min \text{HL}_j \leq \beta_j \leq \max \text{HL}_j \), for \( j = 1, 2, \) and \( 3, \) \( \min \text{HL}_1 - \max \text{HL}_j \leq \beta_1 \) \( - \beta_j \leq \max \text{HL}_4 - \min \text{HL}_j. \) Thus, the bounds on \( \beta_1 - \beta_2 \) are \( [\min \text{HL}_1 - \max \text{HL}_2, \max \text{HL}_1 - \min \text{HL}_2] \) and the bounds on \( \beta_1 - \beta_3 \) are \( [\min \text{HL}_1 - \max \text{HL}_3, \max \text{HL}_1 - \min \text{HL}_3]. \) To find the range for A we find the indices \( i_A, j_A, k_A, \) and \( m_A \) of the ordered pairwise differences between samples 1 and 2 which satisfy

\[
D_{(i_A)}^{(1,2)} \leq \min \text{HL}_1 - \max \text{HL}_2 \leq D_{(j_A)}^{(1,2)},
\]

and

\[
D_{(k_A)}^{(1,2)} \leq \max \text{HL}_1 - \min \text{HL}_2 \leq D_{(m_A)}^{(1,2)}.\]

(5.16)

Following a similar argument, to find the range for B we find the indices \( i_B, j_B, k_B, \) and \( m_B \) of the ordered pairwise differences between samples 1 and 3 which satisfy

\[
D_{(i_B)}^{(1,3)} \leq \min \text{HL}_1 - \max \text{HL}_3 \leq D_{(j_B)}^{(1,3)},
\]

and

\[
D_{(k_B)}^{(1,3)} \leq \max \text{HL}_1 - \min \text{HL}_3 \leq D_{(m_B)}^{(1,3)}.\]

(5.17)

Thus we need to check whether a solution to the Jaeckel estimating equations occurs in the cylinders for which

\[
A_L \leq U_{1,2}(\bar{\beta}) \leq A_U,
\]

(5.18)

and

\[
B_L \leq U_{1,3}(\bar{\beta}) \leq B_U,
\]

(5.19)

where \( A_U = n_1 n_2 - \min\{i_A, k_A\}, A_L = n_1 n_2 - \max\{j_A, m_A\}, B_U = n_1 n_3 - \min\{i_B, k_B\} \)
and $B_L = n_1 n_3 - \max \{ j_B, m_B \}$. This gives us the range of values for $A$ and $B$ for the algorithm.

Next we consider how to determine whether the intervals for $t_1$, $t_2$, and $t_3$ overlap. Let $LE_j$ and $RE_j$ denote the left and right endpoints, respectively, of the interval for $t_j$. If $\max \{ LE_1, LE_2, LE_3 \} \leq \min \{ RE_1, RE_2, RE_3 \}$, then the intervals for $t_1$, $t_2$, and $t_3$ overlap simultaneously. Otherwise, the intervals do not overlap simultaneously. If the intervals do not overlap simultaneously, then we must select new trial values $A$ and $B$ and repeat the process. The choice of new trial values depends on the configuration of the intervals.

We want to move in a direction which will bring closer together the intervals that are the furthest apart. We define the distance, denoted $DI(I_1(A, B), I_2(A, B), I_3(A, B))$, between three intervals, $I_j(A, B)$, $j=1, 2, 3$ in the following manner:

$$DI(I_1(A, B), I_2(A, B), I_3(A, B)) = \max \{ 0, \max \{ LE_1, LE_2, LE_3 \} - \min \{ RE_1, RE_2, RE_3 \} \}. \quad (5.20)$$

Now we describe how to select the new values of $A$ and $B$.

If $\max \{ LE_1, LE_2, LE_3 \} = LE_1$ and $\min \{ RE_1, RE_2, RE_3 \} = RE_2$, then decrease $A$.

If $\max \{ LE_1, LE_2, LE_3 \} = LE_1$ and $\min \{ RE_1, RE_2, RE_3 \} = RE_3$, then decrease $B$.

If $\max \{ LE_1, LE_2, LE_3 \} = LE_2$ and $\min \{ RE_1, RE_2, RE_3 \} = RE_1$, then increase $A$.

If $\max \{ LE_1, LE_2, LE_3 \} = LE_2$ and $\min \{ RE_1, RE_2, RE_3 \} = RE_3$, then increase $A$ or decrease $B$.

If $\max \{ LE_1, LE_2, LE_3 \} = LE_3$ and $\min \{ RE_1, RE_2, RE_3 \} = RE_1$, then increase $B$.

If $\max \{ LE_1, LE_2, LE_3 \} = LE_3$ and $\min \{ RE_1, RE_2, RE_3 \} = RE_2$, then decrease $A$ or increase $B$. \quad (5.21)
For the algorithm (5.15), divide-and-conquer routines can be used in step (1.) to find the median intervals for \(b_{1,2}, b_{1,3}, \text{ and } b_{2,3}.\) The intervals for \(t_1, t_2, \text{ and } t_3\) defined in (5.12), (5.13), and (5.14), respectively, can also be computed using divide-and-conquer routines since for each case, the endpoints of the intervals are linear combinations of the ordered pairwise differences. Hence step (5.) can be computed using ideas from the divide-and-conquer algorithms. By selecting the new values of \(A\) and \(B\) as in (5.21), the intervals for \(t_1, t_2, \text{ and } t_3\) move closer together at each iteration, following from the monotonicities of the endpoints of the intervals. We continue iterating until a solution is found.

The algorithm can be extended to the case of \(k\) samples. In the case of four samples we set \(U_{1,2}(\beta)\) and \(U_{1,3}(\beta)\) equal to \(A\) and \(B,\) as determined by the bounds on the Jaeckel estimates. Then we computed the intervals for \(t_1, t_2, \text{ and } t_3\) at the extreme points of the region determined by \(A\) and \(B,\) and checked for overlap. In the \(k\)-sample case, we set \(U_{1,j}(\beta)\) equal to \(A_j,\) for \(j=2,\ldots, k-1.\) The intervals for \(t_1,\ldots, t_{k-1}\) can be defined in the same manner as in the four-sample case and the intervals can be computed at the extreme points of the region defined by the \(A_j.\) The selection of the new \(A_j\) can be determined from the monotonicities of the \(S_j^*(\beta)\) as in the four-sample case. The extension to the case of \(k\)-samples, however, awaits the proof of Conjecture 4.1 since the range of values for the \(A_j\) depends on the bounds on the Jaeckel estimates.
6.0 Introduction

Many of the ideas motivating the divide-and-conquer algorithms discussed in the previous chapter are of use in computing the estimate of the location shift from a weighted logrank statistic. In the first section notation for the two-sample problem with censored data is defined and the weighted logrank statistic is described. An extension of Hodges-Lehmann estimation to censored data is examined. Some of the problems that arise in trying to compute estimators of a location shift based on the weighted logrank statistic are discussed. Algorithms for computing these estimators, which make use of the partitioning ideas used in the complete data case, are described in detail in the second section. Applications of the algorithms for computing estimates derived from three different statistics are given in the third section.

6.1 Two-Sample Censored-Data Problem

Consider the two-sample problem with right-censored data. For convenience in this chapter we change notation slightly from the previous chapters. Let \( Y_1^*, \ldots, Y_{n1}^* \) and \( Y_{n1+1}^*, \ldots, Y_{n1+n2}^* \) be independent random samples from continuous distributions with distribution functions \( F_1(y) \) and \( F_2(y) \), respectively. Here \( Y_1^*, \ldots, Y_{n1}^* \) and \( Y_{n1+1}^*, \ldots, Y_{n1+n2}^* \) represent the true lifetimes of the \( n_1+n_2 \) individuals under study. Let \( C_1^*, \ldots, C_{n1}^* \) and \( C_{n1+1}^*, \ldots, C_{n1+n2}^* \) represent the censoring times of the individuals in
sample 1 and sample 2, respectively. Let \( Y_i = \min(Y_i^*, C_i) \) denote the observed time to failure for the \( i \)-th individual and let \( \delta_i = 1 \) if \( Y_i^* = Y_i \) and 0 otherwise for \( i = 1, ..., n_1 + n_2 \).

Here \( \delta_i \) represents the death indicator of the \( i \)-th individual. For simplicity, \( Y_i \) has been used to denote the observed lifetimes of the subjects in the actual time scale. In general, \( Y_i \) may denote a monotone transformation of the observed lifetimes. In this case the same transformation is applied to the true lifetimes and the censoring times. As an example, consider the accelerated failure time model. In this model a logarithmic transformation is often applied to all of the observed times so that \( Y_i = \log Y_i^* \) if \( \delta_i = 1 \) and \( Y_i = \log C_i \) if \( \delta_i = 0 \).

As discussed in Chapter II, many of the statistics that have been used to test whether \( F_1 = F_2 \), based on the type of data described above, have the form

\[
T = \int_{-\infty}^{\infty} W_0(s) \left\{ \frac{R_2(s)}{R_1(s) + R_2(s)} dD_1(s) - \frac{R_1(s)}{R_1(s) + R_2(s)} dD_2(s) \right\}
\]

(6.1)

where \( R_i(s) \) denotes the number of individuals at risk in sample \( i \) at time \( s \) for \( i = 1, 2 \); \( D_i(s) \) denotes the number of deaths in sample \( i \) at time \( s \) for \( i = 1, 2 \); and \( W_0(s) \) is a predictable, nonnegative function of \( \{R_i(r), D_i(r); i = 1, 2 \text{ and } r < s\} \) that equals zero whenever the product \( R_1(s)R_2(s) \) is zero. If \( W_0(s) = 1 \) then (6.1) is the logrank statistic. The function \( W_0(s) \) can be thought of as a weight function and for this reason a statistic \( T \) defined by (6.1) is often called a weighted logrank statistic.

Suppose that the distributions of the true lifetimes for samples 1 and 2 satisfy the location family condition

\[
F_1(y) = F(y - \beta) \text{ and } F_2(y) = F(y)
\]

(6.2)
where \( F(\cdot) \) is a continuous, unknown distribution function. The parameter \( \beta \) is the location shift between the distribution of the two samples. Under condition (6.2), a test of whether \( F_1 = F_2 \) is equivalent to a test of whether or not there is a location shift, that is, a test of whether or not \( \beta = 0 \).

Suppose the true parameter, \( \beta \), is different from zero and interest lies in estimating its value. If the data were uncensored, then the Hodges-Lehmann procedure estimates \( \beta \) by the value for which a test statistic, when evaluated using the aligned data, is as close as possible to the center of its null distribution. That is, the estimate \( \hat{\beta} \) is the solution to

\[
T(b) = \xi, \tag{6.3}
\]

where \( T(b) \) is a test statistic, evaluated using \( Y_{1},..., Y_{n_1} \) and \( Y_{n_1+1-b},..., Y_{n_1+n_2-b} \), which satisfies (2.3) and (2.4), and \( \xi = E_{0}[T] \), the center of the null distribution of \( T \). Since \( T() \) is a step function, it may equal \( \xi \) or it may jump across \( \xi \). Therefore, when we speak of solving equation (6.3) we refer to finding the point \( b \) such that \( T(Y_{1},..., Y_{n_1}; Y_{n_1+1-b},..., Y_{n_1+n_2-b}) \) is as close as possible to \( \xi \), namely a point \( b \) in the interval where \( T(Y_{1},..., Y_{n_1}; Y_{n_1+1-b},..., Y_{n_1+n_2-b}) \) equals \( \xi \) or the point at which \( T(Y_{1},..., Y_{n_1}; Y_{n_1+1-b},..., Y_{n_1+n_2-b}) \) jumps across \( \xi \). To extend this estimation procedure to censored data we align the data for sample 2 and proceed as described above. This alignment procedure involves shifting the censoring times as well as the survival times. Wei and Gail (1983) suggested this extension of the Hodges-Lehmann estimation procedure for censored data since the weighted logrank statistics, in the large sample case, have a null distribution which is approximately normal with mean 0. Thus condition (2.4) is approximately satisfied. Furthermore, the procedure is only used with statistics which satisfy the monotonicity condition (2.3). The estimate of \( \beta \), then, is the value \( \hat{\beta} \)
satisfying equation (6.3) where

\[
T(b) = \int_{-\infty}^{\infty} W(s,b) \left\{ \frac{R_2(s+b)}{R_1(s) + R_2(s+b)} \, dD_1(s) - \frac{R_1(s)}{R_1(s) + R_2(s+b)} \, dD_2(s+b) \right\}
\]

(6.4)

and \(W(s,b)\) is \(W_0(s)\) evaluated using the original data for sample 1 and the aligned data for sample 2.

Several problems arise in trying to evaluate the various estimators which are derived by inverting a statistic of the form (6.4). One can iteratively approach the solution to (6.3) by evaluating \(T(\cdot)\) at a trial value, checking for whether the trial value is a solution and, if necessary, selecting a new trial value and repeating the previous steps. Such methods as the bisection algorithm, false position, or the Illinois variant of false position can be used to find an approximation of the solution to (6.3). Another approach is to make use of the nature of \(T(b)\) as a step function and apply a divide-and-conquer algorithm to find an exact solution to (6.3). In either case, computation of the estimate can be costly in terms of computer time and storage space if a naive approach is taken for evaluating \(T(\cdot)\) at a trial value. We look at this evaluation to determine what these problems are and then we describe a method which reduces the computer time and memory required for the evaluation.

For simplicity, first consider the evaluation of the weighted logrank statistic with a weight function of unity. Let \(T(t)\) be the logrank statistic evaluated at a given shift value of \(t\) for sample 2, written

\[
T(t) = \int_{-\infty}^{\infty} \left\{ \frac{R_2(s+t)}{R_1(s) + R_2(s+t)} \, dD_1(s) - \frac{R_1(s)}{R_1(s) + R_2(s+t)} \, dD_2(s+t) \right\}.
\]

(6.5)
Observe that $T(t) = I_1(t) - I_2(t)$ where

$$I_1(t) = \int_{-\infty}^{\infty} \frac{R_2(s+t)}{R_1(s) + R_2(s+t)} \, dD_1(s)$$  \hspace{1cm} (6.6)

and

$$I_2(t) = \int_{-\infty}^{\infty} \frac{R_1(s)}{R_1(s) + R_2(s+t)} \, dD_2(s+t)$$  \hspace{1cm} (6.7)

$$= \int_{-\infty}^{\infty} \frac{R_1(s-t)}{R_1(s-t) + R_2(s)} \, dD_2(s).$$  \hspace{1cm} (6.8)

Note that $dD_1(s)$ is nonzero only at the death times in sample 1 and similarly $dD_2(s)$ is nonzero only at the death times in sample 2. Rewriting the integral $I_1(t)$ gives

$$I_1(t) = \sum_{i=1}^{n_1} \frac{R_2(Y_i + t)}{R_1(Y_i) + R_2(Y_i + t)} \, \delta_i$$  \hspace{1cm} (6.9)

and rewriting the integral $I_2(t)$ gives

$$I_2(t) = \sum_{j=1}^{n_2} \frac{R_1(Y_{n_1+j} - t)}{R_1(Y_{n_1+j} - t) + R_2(Y_{n_1+j})} \, \delta_{n_1+j}.$$  \hspace{1cm} (6.10)

As seen from the form of the integrals in (6.9) and (6.10), the method used to compute $I_1(t)$ can also be used to calculate $I_2(t)$ if the roles of sample 1 and sample 2 are reversed, and if $-t$ is used in place of $t$. Thus, we focus our attention on the computation of $I_1(t)$.

Consider the quantities $R_1(Y_i)$ and $R_2(Y_{i} + t)$ needed for the computation of $I_1(t)$, where $Y_i$ is an observed time from sample 1. Let $\psi_R(c)$ denote the right-continuous version of $\psi$, that is $\psi_R(c) = 1$ if $c \geq 0$ and $0$ if $c < 0$. The numbers at risk at time $Y_i$, namely $R_1(Y_i)$ and $R_2(Y_{i} + t)$, can be written in the following form:
The quantity \( R_1(Y_i) \) can be computed easily once the order statistics for sample 1 are found, since \( R_1(Y_i) \) is the number of observations in sample 1 that are greater than or equal to \( Y_i \). Similarly, for each \( i = 1, \ldots, n_1 \), the quantity \( R_2(Y_{i+t}) \) is the number of observations in the aligned sample, \( Y_{n_1+1-t}, \ldots, Y_{n_1+n_2-t} \), that are greater than or equal to \( Y_i \). This is equivalent to the number of pairwise differences, \( Y_{n_1+j} - Y_i, j = 1, \ldots, n_2 \) that are greater than or equal to \( t \). A straightforward approach for computing \( R_2(Y_{i+t}) \) for \( i = 1, \ldots, n_1 \) might be the following. First sort the observations in the second sample. Next, for each \( Y_i \), \( i = 1, \ldots, n_1 \) compute the \( n_2 \) differences, \( Y_{n_1+j} - Y_i, j = 1, \ldots, n_2 \). Then for each \( Y_i \) find \( R_2(Y_{i+t}), i = 1, \ldots, n_1 \), the number of differences, \( Y_{n_1+j} - Y_i \), which are greater than or equal to \( t \). Now \( I_1(t) \) can be computed from the \( R_1(Y_i) \) and \( R_2(Y_{i+t}) \).

Using this type of routine to evaluate \( I_1(t) \) would require an initial \( O(n_1 \log n_1) \) steps for the sorting of sample 1, \( O(n_2 \log n_2) \) steps for the sorting of sample 2, and \( O(n_1 n_2) \) for the computation of the pairwise differences. These differences must be stored, requiring storage of \( n_1 n_2 \) elements. For each trial value \( t \), \( O(n_1 n_2) \) steps are needed for the computation of \( R_2(Y_{i+t}), i = 1, \ldots, n_1 \), and hence for the computation of \( I_1(t) \) from the quantities \( R_1(Y_i) \) and \( R_2(Y_{i+t}) \). Similarly, the evaluation of \( I_2(t) \) would require \( O(n_1 n_2) \) steps. Although sorting the two samples and finding the differences \( Y_{n_1+j} - Y_i, j = 1, \ldots, n_2 \), for each \( i = 1, \ldots, n_1 \) need only be done once rather than for each iteration of the algorithm, a computation of the number at risk in one sample at the death times of the other sample, requiring \( O(n_1 n_2) \) steps, must be done at each iteration. Even for moderate sample sizes, this approach can be quite costly. A less costly means of finding the
number of differences $Y_{n_1+j} - Y_i \geq t$, for each $i=1,\ldots, n_1$ and $j=1,\ldots, n_2$ is desired.

Suppose now that $T(\cdot)$ has a weight function other than $W(s,t)=1$. For example, let $W(s,t)=\prod_{i} F(s,t)$ where $F(s,t)$ denotes the product-limit estimator of the survival distribution for the pooled sample, $Y_1,\ldots, Y_{n_1}, Y_{n_1+1-t},\ldots, Y_{n_1+n_2-t}$, evaluated at time $s$. To compute the value of the statistic $T(\cdot)$ at a trial value $t$, one must find $F(s,t)$ at each death time in sample 1 and at each death time in the aligned sample 2. If a naive approach is taken the aligned data $Y_1,\ldots, Y_{n_1}, Y_{n_1+1-t},\ldots, Y_{n_1+n_2-t}$ must be sorted and then an iterative procedure used to compute $F(s,t)$ from these sorted values, the corresponding numbers of deaths and the corresponding numbers of censorings at each time. This method would require $O((n_1+n_2) \log (n_1+n_2))$ steps for the sorting and $O(n_1+n_2)$ steps for the computation of $F(s,t)$ at each observation. Note that, although for different trial values the $Y_i$'s and the $Y_{n_1+j-t}$'s maintain the same ordering among themselves, the $Y_i$'s and the $Y_{n_1+j-t}$'s may exchange ranks in the pooled sample. Thus, this sorting and computation would have to be done for every new trial value $t$.

In summary, the problem of evaluating $T(\cdot)$ at a trial value, $t$, reduces to (i.) finding the number of differences for which $Y_{n_1+j} - Y_i \geq t$ for each $i=1,\ldots, n_1$ and (ii.) computing a weight function of the $Y_i, i=1,\ldots, n_1$ and $Y_{n_1+j-t}, j=1,\ldots, n_2$. A method which can accomplish (i.) and (ii.) in $O(n_1+n_2)$ steps rather than the $O(n_1n_2)$ steps of the straightforward approach would greatly reduce the cost of evaluating $T(t)$. We discuss such a method in the sections that follow.

6.2 Details of Algorithm

As indicated in section 6.1, several approaches are possible for obtaining an estimate of the location shift between two samples by solving equation (6.3). If a closed-form solution is not known, then a root-finding technique can be used to solve
$T(b) - E_0[T] = 0$. Since for the weighted logrank statistic $T(\cdot)$ is a step function, an algorithm that does not require derivatives is needed. Some examples of root-finding techniques that find an approximation to the solution include the bisection algorithm, the false position algorithm, and a modification of false position. In these algorithms, successively smaller intervals that bracket a solution to the equation are found until a trial value selected from the interval is deemed close enough to the true solution. We provide an outline of such an algorithm and the details for its use.

Another approach to solving equation (6.3) is to use a fast, exact method such as one of the divide-and-conquer algorithms described in Chapter V. By the nature of equation (6.3) the solution is contained in or obtained from the set of all possible jump points of $T(b)$. The divide-and-conquer algorithms proceed by successively selecting a subset from the set of all possible jump points such that the subset contains the solution. An outline of one such algorithm and the details for its use are provided in the third example of section 6.3.

First we consider an algorithm that approximates a solution to equation (6.3) by bracketing it in successively smaller intervals until a trial value in the interval is close enough to the solution. At each iteration a trial value is selected from the current interval. The statistic $T(\cdot)$ is evaluated at this trial value and the stopping criterion is checked. If the trial value is not a solution, a new interval is created from the trial value and the endpoints of the current interval such that the new interval contains the solution. Note that at each iteration of the algorithm the statistic must be evaluated at a new trial value. As noted in section 6.1, the cost of this computation can be reduced by making use of some simple ideas of partitioning of sets. We give an outline of an algorithm which makes use of these ideas in computing the estimate of the shift and then give the details for using this algorithm.
We want to approximate the solution to $T(t) = c$ where $c = E_0[T(t)]$. The approximating algorithm can be outlined as follows.

1. Sort $Y_i$, $i = 1, ..., n_1$ and denote by $Y(1) \leq ... \leq Y(n_1)$.
2. Sort $Y_{n_1+j}$, $j = 1, ..., n_2$ and denote by $Y(n_1+1) \leq ... \leq Y(n_1+n_2)$.
3. Find corresponding $\delta(i)$ and $\delta(n_1+j)$.
4. Compute $R_1(Y(i))$ and $R_2(Y(n_1+j))$.
5. Set initial conditions.
   
   \[ a_0 = Y(n_1+1) - Y(n_1). \]
   \[ b_0 = Y(n_1+n_2) - Y(1). \]
   
   Compute $Q(a_0)$ and $Q(b_0)$ as in step 8.
6. Set $p = 0$.
7. Select trial value, $t_p$.
8. Compute $Q(t_p)$.
   
   Find $l_1(i)$ and $l_2(j)$.
   
   Compute $I_1(t_p)$ and $I_2(t_p)$.
   
   $T(t_p) = I_1(t_p) - I_2(t_p)$.
   \[ Q(t_p) = T(t_p) - c. \]
9. Check stopping criterion.
10. Update current conditions.
11. $p = p + 1$.
12. Go to step 7. (6.13)

To use this algorithm to approximate a solution to equation (6.3) we need to consider the details of how to compute $T(t)$ in $O(n_1+n_2)$ steps, how to select a trial value, and
when to stop iterating. To compute $T(t)$ we visualize an $n_1 \times n_2$ rectangular array of elements, where the element in the $i$-th row and the $j$-th column is $Y_{(n_1+j)} - Y(i)$. The elements in the array correspond to the elements of the set $S_0$, the set of all pairwise differences. Similar to the method described in Chapter V, for a given trial value $t$, the partitioning of $S_0$ into subsets of differences less than $t$ and differences greater than or equal to $t$, denoted $L$ and $GE$, respectively, can be accomplished in $O(n_1+n_2)$ steps by making use of row pointers. For each row a pointer, $l_1(\cdot)$, is found such that

\[ Y_{(n_1+j)} - Y(i) \begin{cases} < t & \text{for } 1 \leq j \leq l_1(i) \\ \geq t & \text{otherwise.} \end{cases} \quad (6.14) \]

As in the previous chapter, if $Y_{(n_1+j)} - Y(i) \geq t$ for all $j=1,..., n_2$, then $l_1(i)=0$. Because the elements in any row are nondecreasing, the pointer $l_1(\cdot)$ indicates which elements of that row belong to the sets $L$ and $GE$. Using the relationship $l_1(i+1) \geq l_1(i)$, the partitioning of $S_0$ into $L$ and $GE$ can be accomplished in $O(n_1+n_2)$ steps. The number at risk in the aligned sample 2 at each death time in sample 1 is the number of elements in row $i$ which belong to the set $GE$. Thus, using these pointers, $R_2(Y(i)+t) = n_2-l_1(i)$ for $i=1,..., n_1$. If a weight function besides $W(s,t)=1$ is used, one needs to be able to compute $W(Y(i), t)$ in $O(n_1+n_2)$ steps to maintain $O(n_1+n_2)$ steps for the calculation of $I_1(t)$. If $W(Y(i), t)$ is a function of the numbers at risk in the two samples at time $Y(i)$ alone, then this computation is straightforward. Similarly, if the weights are such that a weight vector of constants can be constructed and the pointers used to identify an element of the weight vector as the weight contributed to $I_1(t)$ by that row, then this computation can be accomplished. To compute $I_2(t)$ exchange the roles of samples 1 and 2 and replace
t by -t. If we use pointers, l_2(j), for each j=1,..., n_2 to indicate which differences, Y(i) - Y(n_1+j), are less than -t, then R_1(Y(n_1+j)-t) is just n_1-l_2(j). Again, if the weight at time Y(n_1+j) can be computed in O(n_1+n_2) steps, then I_2(t) can be computed in O(n_1+n_2) steps.

The selection of a trial value differs for the various algorithms. In the bisection algorithm the new trial value is the midpoint of the current interval:

\[ t_k = \frac{a_k + b_k}{2}. \]  

(6.15)

For the false position algorithm, the trial value is the x-intercept of the line joining the points \( (a_k, Q(a_k)) \) and \( (b_k, Q(b_k)) \), namely

\[ t_k = a_k + \frac{(b_k - a_k)Q(a_k)}{Q(a_k) - Q(b_k)}. \]  

(6.16)

For both the bisection algorithm and false position, the new interval is determined by the sign of the functional value of \( t_k \).

If \( Q(a_k)Q(t_k)<0 \) then \( a_{k+1}=a_k, b_{k+1}=t_k, Q(a_{k+1})=Q(a_k), Q(b_{k+1})=Q(t_k). \)  

(6.17)

If \( Q(a_k)Q(t_k)>0 \) then \( a_{k+1}=t_k, b_{k+1}=b_k, Q(a_{k+1})=Q(t_k), Q(b_{k+1})=Q(b_k). \)

The Illinois variant of false position uses the same trial value selection as in (6.16) but updates the current values differently.

If \( Q(a_k)Q(t_k)<0 \) then \( a_{k+1}=a_k, b_{k+1}=t_k, Q(a_{k+1})=\frac{Q(a_k)}{2}, Q(b_{k+1})=Q(t_k). \)  

(6.18)

If \( Q(a_k)Q(t_k)>0 \) then \( a_{k+1}=t_k, b_{k+1}=b_k, Q(a_{k+1})=Q(t_k), Q(b_{k+1})=\frac{Q(b_k)}{2}. \)
The stopping criterion for any of the algorithms can be based on $|t_k - a_k|$ or on $|Q(t_k)|$.

The divide-and-conquer algorithms differ from the approximating algorithms in the selection of the trial value $t_p$ and the stopping criterion. The same method for evaluating $T(-)$ at the trial value can be used in both the approximating and the exact algorithms. Since we have already described the divide-and-conquer algorithms in detail for the uncensored data case, we do not repeat the description here. We illustrate the use of such an algorithm in the censored data case for a specific example in section 6.3.3.

6.3 Specific Statistics

We now provide some examples of how the methods described in section 6.2 can be used in the censored data case. First we consider the logrank statistic and indicate how the above methods can be used in the evaluation of the statistic at a trial value. The second example involves a more complicated weight function than in the logrank statistic. This estimator is derived from inverting the statistic of Peto and Peto (1972). We describe how to compute the weights in a fashion which maintains the $O(n_1 + n_2)$ complexity for the evaluation of $T(\cdot)$. The final example is a special case in which a closed-form solution to equation (6.3) exists. It is the estimator of time scale change suggested by Padgett and Wei (1982) and is based on the statistic proposed by Efron (1967). The divide-and-conquer scheme is described for this example.

6.3.1 Logrank Estimator

Louis (1981) proposes an estimator of the time scale change for two independent random samples with right-censored observations by inverting the logrank statistic. The logrank statistic evaluated at a trial value $t$ has the form of (6.5). The distribution of $T$
under the null hypothesis of $\beta = 0$ may be approximated by a symmetric, mean zero distribution. Hence to estimate $\beta$ we must find the value $t$ for which $T(t)$ is as close to zero as possible. As in section 6.1 $T(t)$ can be written as $T(t) = I_1(t) - I_2(t)$ where

$$I_1(t) = \sum_{i=1}^{n_1} \frac{R_2(Y_i + t)}{R_1(Y_i) + R_2(Y_i + t)} \delta_i$$

and

$$I_2(t) = \sum_{j=1}^{n_2} \frac{R_1(Y_{n1+j} - t)}{R_1(Y_{n1+j} - t) + R_2(Y_{n1+j})} \delta_{n1+j}.$$

Noting that $R_2(Y_i + t) = \#\{j=1, \ldots, n_2: Y_{ni+j} - Y_i \geq t\}$ we see that $I_1(t)$ is a nonincreasing step function which jumps down at the differences $Y_{ni+j} - Y_i$, for $i=1, \ldots, n_1$ and $j=1, \ldots, n_2$. Similarly $-I_2(t)$ is a nonincreasing step function which jumps down at the same pairwise differences. Thus $T(t)$ is a nonincreasing step function which jumps at the pairwise differences.

The estimator of the location shift is the value of $t$ for which $T(t)$ steps across zero, that is $T(t) > 0$ and $T(t+) < 0$, or the midpoint of the interval of values for which $T(t) = 0$. To compute the estimate one could use a root-finding technique such as the bisection algorithm or the false position algorithm. These algorithms give an approximation to the solution. As discussed in section 6.2, at each iteration of the algorithm $T(\cdot)$ must be evaluated at a new trial value. This involves computing a function of the numbers at risk in the original sample and the aligned sample for each death time. Referring to the approximation algorithm of section 6.2 we focus attention on step 8 for the logrank statistic, namely evaluating $T(\cdot)$ at a trial value. Recall, this step of the algorithm is the
8. Compute $T(t_k)$.

Find $l_1(i)$ and $l_2(j)$.

Compute $I_1(t_k)$ and $I_2(t_k)$.

$$T(t_k) = I_1(t_k) - I_2(t_k). \quad (6.21)$$

To carry out this step of the algorithm we begin by visualizing an $n_1 \times n_2$ rectangular array in which the element in row $i$ and column $j$ is $Y(n_1+i)-Y(i)$. Using this array we set up a vector containing the pointers for each row, $l_1(i)$, $i=1,..., n_1$, as defined in (6.14).

Once the pointer for a row, $l_1(i)$, has been found, the contribution to $I_1(t)$ by the term corresponding to the time $Y(i)$ is computed as

$$\frac{R_2(Y(i) + t)}{R_1(Y(i)) + R_2(Y(i) + t)} \delta(i) = \frac{n_2 - l_1(i)}{R_1(Y(i)) + n_2 - l_1(i)} \delta(i) \quad (6.22)$$

where $R_1(Y(i))$ and $\delta(i)$ are in look-up tables from steps 3 and 4 of the approximating algorithm (6.13). Thus, the quantity

$$I_1(t) = \sum_{i=1}^{n_1} \frac{R_2(Y(i) + t)}{R_1(Y(i)) + R_2(Y(i) + t)} \delta(i) = \sum_{i=1}^{n_1} \frac{n_2 - l_1(i)}{R_1(Y(i)) + n_2 - l_1(i)} \delta(i) \quad (6.23)$$

can be computed in $O(n_1+n_2)$ steps.

Exchanging the roles of the two samples and replacing $t$ with $-t$, the same steps as were used above can be repeated to find the pointers, $l_2(j)$, such that
\[ Y(i) - Y(n_{i+j}) < -t \text{ for } 1 \leq i \leq l_2(j) \]
\[ \geq -t \text{ otherwise.} \]

This computation again requires \( O(n_1 + n_2) \) steps. From the pointers, \( l_2(t) \) can be computed using the relation

\[ l_2(t) = \sum_{j=1}^{n_2} \frac{n_1 - l_2(j)}{R_2(Y(n_{i+j})) + n_1 - l_2(j)} \delta(n_{i+j}). \]  

(6.25)

Now the logrank statistic evaluated at the trial value is just \( T(t) = I_1(t) - I_2(t) \). This evaluation has been accomplished in \( O(n_1 + n_2) \) steps.

### 6.3.2 Peto and Peto Estimator

In the previous example, the estimator was derived from the logrank statistic for which \( W(s, t) = 1 \). Suppose the estimator to be computed is derived from a weighted logrank statistic with a weight different from unity. As an example consider the estimator derived from inverting the statistic of Peto and Peto (1972), namely,

\[ T^\wedge(t) = \int_{-\infty}^{\infty} \frac{F(s, t)}{R_1(s) + R_2(s+t)} dD_1(s) - \frac{R_1(s)}{R_1(s) + R_2(s+t)} dD_2(s+t) \]

(6.26)

where \( \wedge F(s, t) \) denotes the product-limit estimator of the survival distribution for the pooled sample \( Y_1, \ldots, Y_{n_1}, Y_{n_1+1} - t, \ldots, Y_{n_1+n_2} - t \), evaluated at time \( s \). As for the logrank statistic, the null distribution of \( T \) can be approximated by a symmetric distribution with mean zero. Thus we find \( t \) such that \( T(t) \) is as close to zero as possible.

From section 6.3.1 we see that, using a similar algorithm as for the logrank statistic, \( T(t) \) can be evaluated in \( O(n_1 + n_2) \) steps if the weight function \( \wedge F(s, t) \) can be computed in
O(n1+n2) steps. This cannot be done if we must sort the Y1's and Yn1+j - t's to calculate the weight function for every value of t. A method which computes the weight function using the quantities already calculated is desired. We now discuss such a method.

Recall that in the step of the algorithm for evaluating T() at a trial value t, we must find l1(i) and l2(j), compute I1(t) and I2(t), and compute T(t)=I1(t)-I2(t). For the statistic T(t) from (6.26) the integral I1(t) has the form:

\[
I_1(t) = \sum_{i=1}^{n_1} \frac{R_2(Y(i) + t)}{R_1(Y(i)) + R_2(Y(i) + t)} \delta(i). 
\] (6.27)

As shown previously, R1(Y(i)) and \(\delta(i)\) can be obtained from look-up tables, and R2(Y(i) + t) can be computed from the pointer l1(i). We concentrate now on the evaluation of the weight function.

If there were no censoring, the product-limit estimate for the pooled sample Y1,...,Yn1,Yn1+1 - t,...,Yn1+n2 - t at a death time Y(i) would be

\[
\frac{R_1(Y(i)) + R_2(Y(i) + t) - \delta(i)}{n_1 + n_2}. 
\] (6.28)

Censored observations occurring prior to Y(i) inflate this value. We investigate how to compute this inflation factor.

Let m1(i) be the accumulated value of the inflation factor for the pooled product-limit estimate at time Y(i) due to censored observations in sample 1. We define m1(i) for i=1,..., n1 by
\[ m_1(0) = 1 \]

\[ m_1(i) = \begin{cases} 
  m_1(i-1) & \text{if } Y(i) \text{ is a death} \\
  m_1(i-1) \left(1 + \frac{1}{R_1(Y(i)) + R_2(Y(i) + t)} \right) & \text{otherwise.}
\end{cases} \quad (6.29) \]

The quantity \( R_1(Y(i)) \) can be obtained from a look-up table and \( R_2(Y(i) + t) = n_2 - l_1(i) \), so that the \( m_1(i) \) can be computed in the required \( O(n_1 + n_2) \) steps. Similarly, let \( m_2(j) \) be the accumulated value of the inflation factor for the pooled product-limit estimate at time \( Y(n_1+j) - t \) due to censored observations in sample 2. Define \( m_2(j) \) for \( j=1, \ldots, n_2 \) by

\[ m_2(0) = 1 \]

\[ m_2(j) = \begin{cases} 
  m_2(j-1) & \text{if } Y(n_1+j) \text{ is a death} \\
  m_2(j-1) \left(1 + \frac{1}{R_1(Y(n_1+j) - t) + R_2(Y(n_1+j))} \right) & \text{otherwise.}
\end{cases} \quad (6.30) \]

To get the actual inflation factor at time \( Y(i) \), we must take into account the \( Y_i \)'s and \( Y_{n_1+j} - t \)'s that were censored prior to time \( Y(i) \). This is just \( m_1(i) \times m_2(l_1(i)) \) since \( l_1(i) \) is the subscript of the largest \( Y(n_1+j) \) for which \( Y(n_1+j) - t < Y(i) \). Thus the product-limit estimate at a death time \( Y(i) \) is just

\[ \hat{F}(Y(i), t) = m_1(i) \ m_2(l_1(i)) \left( \frac{R_1(Y(i)) + R_2(Y(i) + t) - \delta(i)}{n_1 + n_2} \right) \quad (6.31) \]
Thus using $R_1(Y_{(i)}), R_2(Y_{(i)}+t)$, and $l_1(i)$ we can compute $\hat{F}(Y_{(i)}, t)$ in $O(n_1 + n_2)$ steps.

Now that the weight has been computed, $I_1(t)$ can be computed. A similar method can be used to compute $I_2(t)$ and subsequently, $T(t)$.

### 6.3.3 Padgett and Wei Estimator

Padgett and Wei (1982) derive an estimator of the time scale change for two independent random samples with right-censored data under the assumptions of the accelerated failure time model. They derive the estimator by minimizing the Cramer-von Mises distance between the estimated survival distributions of the two samples, but the estimator can also be derived by inverting the test statistic proposed by Efron (1967).

Their estimator has a closed-form and is shown to be consistent under some mild conditions.

Let $\hat{F}_1(Y_{(i)}), i=1,..., n_1$ denote the left-continuous product limit estimator of the survival distribution for sample 1 at $Y_{(i)}$. Define $a_i$ to be the jump in the product limit estimator at time $Y_{(i)}$, namely

$$a_i = \begin{cases} \hat{F}_1(Y_{(i)}) - \hat{F}_1(Y_{(i+1)}) & \text{if } i = 1, ..., n_1 - 1 \\ \hat{F}_1(Y_{(i)}) & \text{if } i = n_1. \end{cases}$$

(6.32)

Similarly, let $\hat{F}_2(Y_{(n_1+j)}), j=1,..., n_2$ be the product limit estimator of the survival distribution for sample 2 at the observed times and let $b_j$ be the jump in the product limit estimator at time $Y_{(n_1+j)}$. The estimator of the shift parameter, denoted $\hat{\beta}$, proposed by Padgett and Wei is the median of the distribution with probability mass function
\[ p(v) = \begin{cases} a_i b_j & \text{if } v = Y(n_{1+j}) - Y(i); \text{ for } i = 1, \ldots, n_1 \text{ and } j = 1, \ldots, n_2 \\ 0 & \text{otherwise.} \end{cases} \quad (6.33) \]

To find \( \hat{\beta} \) we consider the function

\[ T(t) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} a_i b_j \psi(Y(n_{1+j}) - Y(i) - t). \quad (6.34) \]

Note that \( T(\cdot) \) is a left-continuous nonincreasing step function which jumps at the pairwise differences, \( Y(n_{1+j}) - Y(i) \), for \( i = 1, \ldots, n_1 \) and \( j = 1, \ldots, n_2 \). For a given \( c \) such that \( 0 < c < \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} a_i b_j = 1 \), either \( T(\cdot) \) equals \( c \) on an interval, say \((t_L, t_R] \), or \( T(\cdot) \) steps across \( c \) at some jump point, \( t^* \). Finding \( \hat{\beta} \) is equivalent to finding the value (or the endpoints of the interval) where \( T(\cdot) \) jumps across (or equals) \( c = 1/2 \).

For this example we describe a divide-and-conquer algorithm which finds \( t^* \) or \( t_L \) and \( t_R \) by a fast, exact method. The basic algorithm returns \( \hat{t} \), where \( \hat{t} \) is either \( t^* \) or one of the endpoints, \( t_L \) or \( t_R \). Noting that \( t^* \) or \( t_L \) and \( t_R \) are pairwise differences, the set of all pairwise differences, \( S_0 \), contains \( \hat{t} \). At each iteration \( p \) of the algorithm, a trial value \( t_p \) is chosen from \( S_p \), a subset of \( S_0 \) known to contain \( \hat{t} \). The set \( S_0 \) is partitioned into three sets corresponding to those differences which are less than, equal to, and greater than \( t_p \). The three sets, denoted \( L_p, E_p, \) and \( G_p \), are never formed. Instead, the differences are calculated as needed so that the sum of the weights of the elements in the sets, denoted \( |L_p|, |E_p|, \) and \( |G_p| \) can be found in \( O(n_1 + n_2) \) steps. This calculation is done by working on a conceptual rectangular array in which the element in row \( i \) and column \( j \) is the difference \( Y(n_{1+j}) - Y(i) \). Since \( S_p \) is known to contain \( \hat{t} \), either \( L_p \cap S_p, E_p \cap S_p, \) or \( G_p \cap S_p \) must contain \( \hat{t} \). If \( |G_p| > c \) then \( \hat{t} \in S_p \cap G_p \). If \( |G_p| + |E_p| < c \) then \( \hat{t} \in S_p \cap \)
Otherwise \( t = t_p \) and we need to determine whether \( \hat{t} \) is a jump point or the endpoint of an interval. We can make the determination of whether \( t_p = \hat{t} \) or another iteration is required by checking the values of \( T(t_p) = |G_p| + |E_p| \) and \( T(t_p +) = |G_p| \).

The basic algorithm can be outlined as follows.

1. Sort \( Y_i \), \( i = 1, ..., n_1 \) and denote by \( Y(1) \leq ... \leq Y(n_1) \).
2. Sort \( Y_{n_1+j} \), \( j = 1, ..., n_2 \) and denote by \( Y(n_1+1) \leq ... \leq Y(n_1+n_2) \).
3. Compute \( a_i \) and \( b_j \).
4. Compute \( B_j = \sum_{m=1}^{i} b_m \).
5. Let \( S_0 = \{ Y(n_1+j) - Y(i) ; i = 1, ..., n_1 \text{ and } j = 1, ..., n_2 \} \) conceptually.
6. Set \( p = 0 \).
7. Do until \( \hat{t} \) is found:
   
   7.1. Choose \( t_p \in S_p \).
   
   7.2. Find \( L_p \), \( E_p \), and \( G_p \).
   
   7.3. Compute \( T(t_p) = |G_p| + |E_p| \) and \( T(t_p +) = |G_p| \).
   
   7.4. Check if \( \hat{t} = t_p \).
      
      If \( T(t_p) > c \) and \( T(t_p +) < c \) then \( \hat{t} = t_p \) is a jump point.
      
      If \( T(t_p) > c \) and \( T(t_p +) = c \) then \( \hat{t} = t_p \) is the left endpoint.
      
      If \( T(t_p) = c \) and \( T(t_p +) < c \) then \( \hat{t} = t_p \) is the right endpoint.
      
      If \( T(t_p) < c \) then \( S_{p+1} = S_p \cap L_p \).
      
      If \( T(t_p +) > c \) then \( S_{p+1} = S_p \cap G_p \).
   
   7.5. \( p = p + 1 \).
8. End do. (6.35)

To compute \( |G_p| \) and \( |G_p| + |E_p| \) we use the row pointers described in section 5.1.
For each row $i$ define pointers $l_1(i)$ and $r_1(i)$ such that

$$
Y(n_{1+j} - Y(i) < t \quad \text{for} \quad 1 \leq j \leq l_1(i) 
$$

$$
> t \quad \text{for} \quad r_1(i) \leq j \leq n_2
$$

$$
= t \quad \text{otherwise.} \quad (6.36)
$$

The contribution to $|G_p|$ from row $i$ is

$$
\frac{n_2}{\sum_{j=r_1(i)+1}^{r_1(i)}} a_i b_j = a_i \sum_{j=r_1(i)+1}^{r_1(i)} b_j 
$$

$$
= a_i \left( \sum_{j=1}^{r_1(i)-1} b_j - \sum_{j=1}^{r_1(i)-1} b_j \right) 
$$

$$
= a_i \left( 1 - B_{r_1(i)-1} \right) \quad (6.37)
$$

where $B_j = \sum_{m=1}^{j} b_m$. The last equality of (6.37) holds because by definition the jumps of the product-limit estimator sum to one. Hence,

$$
T(t_{p+}) = \sum_{i=1}^{n_1} a_i \left( 1 - B_{r_1(i)-1} \right) 
$$

$$
= 1 - \sum_{i=1}^{n_1} a_i B_{r_1(i)-1}. \quad (6.38)
$$

Similarly, the total weight of $|G_p| + |E_p|$ is

$$
T(t_p) = \sum_{i=1}^{n_1} \sum_{j=l_1(i)+1}^{r_1(i)} a_i b_j 
$$

$$
= 1 - \sum_{i=1}^{n_1} a_i B_{l_1(i)}. \quad (6.39)
$$
Next we consider how to select a trial value \( t_p \in S_p \) and we consider the stopping criterion. As discussed in the previous chapter, several methods have been suggested in the complete data case. The trial value at stage \( p \) can be chosen to be the weighted median of the row medians of the elements in \( S_p \), where the weights correspond to the number of elements of \( S_p \) in that row, as used by Johnson and Mizoguchi (1978). Monahan (1984) suggested that the trial value can be a randomly chosen element in \( S_p \), with each element having an equal chance of being selected, or that the trial value can be a random row median, with the probabilities being proportional to the number of elements in the row which are in \( S_p \). Any of these methods can be applied to the censored data case with some modification. The statistic \( T(t) \) from (6.4) jumps only at the differences for which either \( Y(i) \) or \( Y(n1+j) \) is a death. Thus differences for which both \( Y(i) \) and \( Y(n1+j) \) are censored should not be included as possible trial values. We modify the method for selecting a trial value so that only elements in \( S_p \) for which either \( Y(i) \) or \( Y(n1+j) \) is a death are included to ensure that the trial value is a possible jump point. The stopping criteria that are discussed in Chapter V are also applicable in the censored data case.

As illustrated in the examples of sections 6.3.1, 6.3.2, and 6.3.3, the partitioning ideas of the divide-and-conquer algorithms are useful in computing estimators derived from the weighted logrank statistics in the two-sample censored data case, even when approximating algorithms such as false position are used.
CHAPTER VII
CONCLUDING REMARKS

In the first chapter we described a typical experiment in which the investigators are interested in determining which of several treatments is most effective in treating a disease. In addition to answering the question of which treatment is most effective, the investigators want to quantify the difference between the treatment groups. In this dissertation we focused on this estimation problem. Because parametric procedures can be sensitive to outliers and to misspecification of the underlying distribution, we considered a nonparametric estimation procedure proposed by Jaeckel (1972). We examined this estimation procedure in detail for the special case of the general linear model in which the covariate is a vector of indicators of group membership. For the three-sample and the four-sample models we characterized the Jaeckel estimator for the Wilcoxon scores geometrically and in terms of the ordered differences between pairs of samples. We also established a relationship between the Jaeckel estimator of the vector of regression coefficients and the Hodges-Lehmann estimators of the location shift between two samples. Some of these results were extended to the k-sample problem. Using these results we described an algorithm for computing the Jaeckel estimate of the vector of regression coefficients. In addition, we discussed the application of ideas from divide-and-conquer algorithms to the computation of the location shift in the two-sample censored data problem.

In obtaining the above results many questions have been raised. The characterization
of the Jaeckel estimator for the k-sample problem is for the estimator derived from a dispersion function using Wilcoxon scores. This characterization exploits the relationship between the counting form and the ranking form of the linear rank statistic which is to be inverted. It appears to be more straightforward to extend the results to estimators derived from placement statistics than to estimators derived from general linear rank statistics. However, there may be other rank statistics which behave like the Wilcoxon statistic, so that similar arguments can be applied. If this were the case, it would be interesting to establish conditions for determining which dispersion functions yield estimators that are related to the corresponding two-sample Hodges-Lehmann estimators.

As discussed in Chapter II, some authors have proposed estimators for the location shift between the distributions of two samples for censored data. The question of estimation in the k-sample censored data model has received less attention. Tsiatis (1986) suggested an approach for estimation in the general linear model, fashioned after the approach of Jureckova (1972). It is interesting to consider how to extend the Jaeckel procedure to censored data. There remains the question of how to define the dispersion measure for the residuals when some of the observations are censored. Secondly, once a dispersion measure is defined for censored data, there is the problem of determining what conditions on the censoring distribution are necessary for the asymptotic results to hold. The extension of the Jaeckel procedure to censored data would raise the question of whether the estimators from this approach are asymptotically equivalent to estimators derived from an extension of Jureckova's procedure to censored data. It appears that the algorithm described in Chapter V may be of use in the computation of rank estimates in the k-sample censored data case. We might extend the Jaeckel approach by defining a score which depends on whether or not the residual corresponds to a censored observation. Then the geometric interpretation of the functions involved in the estimating
equations is similar to the uncensored data case with the exception that the jumps of the step functions depend on whether or not \( Y(i) \) and \( Y(j) \) of the pairwise difference \( Y(i) - Y(j) \) are censored. Since the geometric interpretation is similar, the algorithms may extend.

There are other questions concerning estimation with censored data that need to be addressed. In the extension of the results for the Jaeckel estimation procedure to more general scores in the dispersion measure, we raised the idea of deriving estimators by inverting placement statistics. It would be interesting to consider placement statistics for censored data and the estimators derived from them. A comparison between these estimators derived from placement statistics and from rank statistics for censored data might show whether one method of estimation is preferable over another. A similar comparison could be made between estimators for censored derived from other methods, for example M-estimators. Padgett and Wei (1982) derived an estimator of shift by minimizing the Cramer-von Mises distance between the estimated survival distributions of the two samples. It would be interesting to generalize this approach to other measures of distance between the two estimated survival distributions and compare the resulting estimators, both to each other and to the estimators derived using the Hodges-Lehmann procedure. Such comparisons would be useful for understanding the best approach for robust estimation in the censored data case.

Another generalization of the results of this dissertation to consider is the extension to the case of a continuous covariate. A direct extension of the results is difficult to imagine, even for the simple linear regression model, but it may be possible to attack the problem using an approach similar to the one described in this dissertation. An understanding of the nature of the problem for estimation in the k-sample model may give insight into the use of linear rank statistics for more general models.

There is still much work that can be done regarding the algorithm. A refinement of
the algorithm should be made before it is used in practice. This algorithm gives an exact solution to the Jaeckel estimating equations but it would be interesting to compare it to approximating algorithms, such as a steepest descent routine, to get a better idea of the time requirements, especially in the case of moderate sample sizes. Once the algorithm has been refined, it will be of great use in computing point estimates. It may also have implications in the areas of multiple comparisons and the computation of confidence intervals with exact coverage. It can also be used for bootstrapping to get estimates of the variances and for computing confidence intervals. Efron (1979) proposed the bootstrap method for estimating the bias and variance of a statistic. The method approximates the sampling distribution of a specified random variable $R(X, F)$, where $X_1, \ldots, X_n$ are independent and identically distributed random variables with distribution function $F$. To approximate the sampling distribution of $R(X, F)$, a bootstrap sample $X_1^*, \ldots, X_n^*$ is drawn from the sample probability distribution, $\hat{F}$, constructed by putting mass $1/n$ at each observation, and the bootstrap distribution of $R(X^*, \hat{F})$ is calculated. Often the bootstrap distribution of $R(X^*, \hat{F})$ cannot be calculated directly and is approximated using Monte Carlo techniques. To approximate the variance of the statistic using bootstrapping in these instances, it is desirable to have a computing routine which is inexpensive. Thus our algorithm may be useful in computing estimates of the variance of the Jaeckel estimator of $\beta$ using the bootstrap method.
LIST OF REFERENCES


