Spearman Rank Regression

Jason C. Parcon
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Statistics and Probability Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1268

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
SPEARMAN RANK REGRESSION

by

Jason C. Parcon

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

Western Michigan University
Kalamazoo, Michigan
August 2003
The main purpose of this dissertation is to obtain an estimate of the slope parameter in a regression model that is robust to outlying values in both the \( x \)- and \( Y \)-spaces. The least squares method, though known to be optimal for normal errors, can yield estimates with infinitely large MSE's if the error distribution is thick-tailed. Regular rank-based methods like the Wilcoxon method are known to be robust to outlying values in the \( Y \)-space, but it is still grossly affected by outlying values in \( x \)-space.

This dissertation derives an estimate of the slope from an estimating function that is essentially the Spearman correlation between the \( x \) values and the residuals. It is shown to be a special case of the \( GR \) estimates of Sievers (1983) in the context of simple linear regression. Thus, it is robust to
outlying values in the $x$- and $Y$-spaces. Three proposed schemes are presented for obtaining multiple regression estimates. Two of these schemes, namely, the residual method and the bias adjustment method, are shown to yield estimates that are consistent.
TABLE OF CONTENTS

LIST OF TABLES .................................................................................................. v
LIST OF FIGURES ................................................................................................. vii

CHAPTER

I. INTRODUCTION ............................................................................................... 1
  1.1 Background ......................................................................................... 1
  1.2 Statement of the Problem .............................................................. 7
  1.3 Motivation and Solution .............................................................. 8

II. ESTIMATION OF \( \beta \) IN SIMPLE LINEAR REGRESSION ....................... 12
  2.1 The Estimate .................................................................................. 12
  2.2 The GR Estimator ........................................................................ 17
  2.3 Equivariance Properties ............................................................ 20
  2.4 Asymptotic Properties ............................................................. 21
  2.5 Asymptotic Relative Efficiency .............................................. 28
  2.6 Estimation of the Intercept Parameter ....................................... 31

III. MULTIPLE REGRESSION ESTIMATES OF \( \beta \) THAT REDUCE TO SIMPLE LINEAR REGRESSION ........................................ 33

ii
### Table of Contents—continued

3.1 Introduction ................................................................................ 33
3.2 The Residual Method ................................................................. 40
3.3 The Bias Adjustment Method .................................................... 46
3.4 The Residual Method for Three Independent Variables .......... 50
3.5 The Bias Adjustment Method for Three Independent Variables .... 55

IV. MATRIX WEIGHTS ........................................................................ 57

V. DESIGN MATRICES ...................................................................... 60
  5.1 Tied Ranks ................................................................................... 60
  5.2 The Two-Sample Problem .......................................................... 61
  5.3 Matrix Weights Method for Design Matrices ..................... 62

VI. EXAMPLES AND SIMULATIONS ................................................... 66
  6.1 Examples for Simple Linear Regression ......................... 66
  6.2 Simulations for Simple Linear Regression ...................... 74
  6.3 Examples for Multiple Regression ......................................... 82
  6.4 Simulations for Multiple Regression ..................................... 93

VII. CONCLUSIONS ........................................................................... 100

iii
Table of Contents—continued

BIBLIOGRAPHY ........................................................................................................ 102
LIST OF TABLES

6.1.1 Pilot-Plant Data Set ................................................................. 68
6.1.2 Parameter Estimates for the Uncontaminated Pilot-Plant Data..... 69
6.1.3 Parameter Estimates for the Contaminated Pilot-Plant Data ...... 69
6.1.4 California Mission Data from the 1832 Census ....................... 72
6.1.5 Simple Linear Regression Fits (With the San Luis Rey Data Point) ................................................................. 73
6.2.1 Simulation Results for the First Set of Models (Errors are Normally Distributed) ......................................................... 80
6.2.2 Simulation Results for the Second Set of Models (Y outliers are present) ................................................................. 80
6.2.3 Simulation Results for the Third Set of Models (x outliers are present, $\xi_1 = 0.06$) ......................................................... 81
6.2.4 Simulation Results for the Third Set of Models (x outliers are present, $\xi_1 = 0.12$) ......................................................... 81
6.3.1 Estimates of Multiple Regression Parameters .......................... 83
6.3.2 Estimates of Multiple Regression Parameters (Contaminated Data) ........................................................................... 85
6.3.3 Estimates of Multiple Regression Parameters (Modified HBK Data) ........................................................................... 90
List of Tables—continued

6.4.1 Simulation Results for Multiple Regression: Case 1 ......................... 95
6.4.2 Simulation Results for Multiple Regression: Case 2 ......................... 96
6.4.3 Simulation Results for Multiple Regression: Case 3 ......................... 97
6.4.4 Simulation Results for Multiple Regression: Case 4 ......................... 98


LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1.1</td>
<td>Scatter Diagram of the Pilot-Plant Data</td>
<td>69</td>
</tr>
<tr>
<td>6.1.2</td>
<td>Scatter Diagram of the Pilot-Plant Data ($x_1 = 123 \rightarrow 1230$)</td>
<td>70</td>
</tr>
<tr>
<td>6.1.3</td>
<td>Scatter Diagram of the California Mission Data</td>
<td>73</td>
</tr>
<tr>
<td>6.1.4</td>
<td>Scatter Diagram of the California Mission Data with Least Squares, Wilcoxon, and Spearman Fits</td>
<td>74</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Scatter Diagram for Simulated Data</td>
<td>79</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Standardized Residual Plots versus Index of Observations</td>
<td>91</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Background

Regression analysis is an important statistical tool applied in most sciences. It explores relationships that are readily described by straight lines, or their generalization to many dimensions. In linear regression, the main problem is to find an estimate of the slope parameter. The most predominant regression method is the classical least squares (LS) procedure, which is based on the $L_2$ norm, i.e., an estimate $\hat{\beta}_{LS}$ is chosen such that $\sum_i e_i^2$ is minimized, where $e_i = Y_i - \beta x_i$ is the $i^{th}$ residual. Its popularity has been mainly attributed to its intuitive appeal and ease of computation. It is optimal in many senses when the errors in a regression model have a normal distribution or when linear estimates are required (Gauss-Markov Theorem). This technique however, makes several strong assumptions about the underlying data, and the data can fail to satisfy these assumptions in different ways. For example, it is well known that the LS estimates may have inefficiently large standard errors when the error distribution...
has thicker tails than normal. Thick-tailed distributions are identified by the presence of more extreme or outlying values than what one would expect from a normal distribution. The resulting least squares estimates have inflated MSEs because they tend to get strongly “pulled” by these extreme values. It is in this light that methods on robust regression come into play. Robust regression methods downweight these outlying values and give more importance to the main cluster of points in the data set. Hence, these methods will produce estimates that can be more efficient (in the sense of having smaller standard errors) when the normality assumption is violated.

Outlying observations can occur both on the Y- and x-direction. Robust methods like the regular rank-based R-estimates (Hettmansperger and McKean, 1977) downweight regression observations with outlying Y-values. However, observations with outlying x-values still exhibit large influence. Outliers in the x-direction that have the potential for strongly affecting the regression coefficients are called leverage points. We now propose an estimator that reduces the influence of outlying values in both x- and Y-direction.
The existence of $\hat{\beta}_{LS}$, its ease of derivation, and its efficiency for normal errors are all positive consequences of squaring the residuals, but there is a negative consequence: the observations with large residuals have a much larger contribution to $\sum_i e_i^2$, and hence more effect on the resulting estimate. The sensitivity of $LS$ estimates to the effect of outlying observations has led to the development of more robust fitting methods. As cited by Rousseeuw and Leroy (1987), the first attempt toward a more robust regression estimator came from Edgeworth, who proposed the least absolute values regression estimator (or $L_1$ estimator), which is determined by minimizing $\sum_{i=1}^n |e_i|$. Note that unlike the classical least squares theory, the residuals in this case are treated equally. Unfortunately, the breakdown point (the proportion of outliers that the estimator can handle) of the $L_1$ estimator is no better than 0%. The $L_1$ estimator may have displayed robustness against $Y$-outliers, but it is still vulnerable to $x$-outliers. In fact, Ronchetti (1985) has shown that if the leverage point is at a considerable distance from the bulk of the data, the $L_1$ line passes right through it. There is also a substantial loss of efficiency for this estimator when the errors are normally
Huber (1973) suggested a more general way of choosing an estimate of $\beta$ by minimizing $\sum_{i=1}^{n} \rho(e_i)$, where $\rho$ is a convex function. Solutions $\hat{\beta}_\psi$ are called classical M-estimators, where $\psi = \rho'$. The use of the so-called Huber’s function for $\psi$ yields statistically more efficient estimators than $L_1$ regression at a model with normal errors, but the breakdown point is still asymptotically zero. Even with the judicious choice of the $\psi$ function, $M$-estimators remained sensitive to outlying $x$-values.

Mallows (1975) and Schweppe (see Handschin, Schweppe, Kohlas, and Fiechter, 1975) expanded the domain of the $\psi$ function to include design points, as well as the residuals, to downplay the effect of otherwise influential data appoints. The resulting estimators are called GM-estimators or generalized $M$-estimators.

Rank-based estimates proposed by Jaeckel (1972) and Jureckova (1971), and then later developed by Hettmansperger and McKean (1977), achieved some robustness against outliers while allowing the user a choice of scores for efficiency.
considerations. The use of Wilcoxon scores, for example, achieves good efficiency for a normal error distribution. The Wilcoxon regression estimate has an asymptotic relative efficiency (ARE) of .955 relative to the LS estimate. In fact, as long as the error distribution is symmetric and absolutely continuous, efficiency is never less than .864 (Hodges and Lehmann, 1956).

The unified approach to the rank-based analysis of linear models, as developed by Hettmansperger and McKean (1977), starts with defining a dispersion function $D(\beta)$ on the residuals $e_i(\beta) = Y_i - \beta x_i$, and estimation is done by minimizing $D(\beta)$. Note that this approach closely parallels the least squares method. In spite of this high ARE, this approach remains robust only when the $x$'s are fixed. If the $x$'s are a random sample from some underlying distribution (such is the case for non-designed experiments), then the possibility of gross errors is introduced and the method loses its robustness. The Wilcoxon estimate has influence function that, although bounded in $Y$-space, is still unbounded in $x$-space. The asymptotic breakdown of the Wilcoxon estimate is still 0.
Sievers (1983) and Naranjo and Hettmansperger (1994) generalized the Wilcoxon estimates to a class of weighted estimates with weights based on the $x$'s, and this generalization resulted to estimates having bounded influence in both $Y$- and $x$-spaces. These are called the generalized rank ($GR$) estimates, essentially since if constant weights are used, then the Wilcoxon estimates will be obtained. The breakdown point of the $GR$ estimates is less than 33%, but this decreases as the number of independent variables increases (Naranjo and Hettmansperger, 1994).

This work does not generally follow this unified approach, but meets it halfway. We start with obtaining a solution to an estimating function that is fundamentally the Spearman's correlation of the $x_i$ and the corresponding residuals $e_i$. It will be shown that by using an appropriate weight, the estimator derived is a special case of the $GR$ estimator in Sievers (1983) and Naranjo and Hettmansperger (1994), in the context of simple linear regression. This result, however, will be of particular importance when we derive an estimate of $\beta$ in multiple regression.
1.2 Statement of the Problem

Let \( \{(x_i, y_i): i = 1, 2, \ldots, n\} \) be a sequence of random variables such that

\[
Y_i = \alpha + x_i't\beta + \epsilon_i, \quad (1.2.1)
\]

where

\( Y_i \in \mathbb{R} \) is the \( i^{th} \) value of the response variable \( Y \)

\( x_i \in \mathbb{R}^p \) is the \( i^{th} \) row of a known \( n \times p \) matrix of predictor variable \( X \) \((p \geq 1)\)

\( \beta \) is a \( p \)-vector of unknown slope parameters

\( \alpha \) is the unknown intercept parameter

\( \epsilon_i \sim F \) continuous and unknown, is the \( i^{th} \) error term.

The problem of estimating the value of \( \beta \) will be our main focus. The estimation of \( \alpha \) shall be dealt with at the second stage. Specifically, this dissertation proposes to do the following:

(a) investigate the properties of the simple Spearman regression estimate,

(b) develop multiple regression estimators that reduce to simple Spearman rank regression \((p = 1)\),

(c) derive the consistency properties of the proposed multiple regression
estimates,

(d) investigate the small-sample performance of proposed estimates by simulation methods, and

(e) compare of the performance of the proposed multiple regression method with two alternative methods for multiple Spearman rank regression.

1.3 Motivation and Solution

We now propose a solution that is motivated from the derivation of the least squares and Wilcoxon estimates of the slope. We shall first consider the estimation of the slope in the simple linear regression model, i.e., (1.2.1) with \( p = 1 \) and, without loss of generality, the \( x_i \) observations are centered. The least squares estimator \( \hat{\beta}_{LS} \) is such that the dispersion function \( D(\beta) = \sum e_i^2 \) is minimized. This accordingly states that \( \hat{\beta}_{LS} \) satisfies

\[
\sum_i x_i e_i = 0. \tag{1.3.1}
\]

We shall refer to equation (1.3.1) as the estimating equation. In lieu of (1.3.1), we can view \( \hat{\beta}_{LS} \) as an estimate of the slope parameter such that the (Pearson's) correlation between the \( x_i \) values and their corresponding \( e_i \) values is 0.
The equation in (1.3.1) should clearly show why the $LS$ estimator is vulnerable to outliers. The magnitude of an $x$ or $Y$ value is given full consideration in obtaining an estimate of $\beta$. We have noted this in Section 1.1 as a negative consequence of squaring the residuals.

The Wilcoxon regression estimator $\hat{\beta}_w$, on the other hand, is obtained by minimizing the dispersion function $D(\beta) = \sum_i \left( R(e_i) - \frac{n+1}{2} \right) e_i$, that is, $\hat{\beta}_w$ is the solution to

$$\sum_i \left( R(e_i) - \frac{n+1}{2} \right) x_i = 0 \quad (1.3.2)$$

where $R(e_i) = \text{rank of } e_i$. Similar to the observation made for (1.3.1), we can look at (1.3.2) as the correlation between the $R(e_i)$ and the $x_i$. Note that even if an outlying $Y$ value is present in the data, (1.3.2) only takes into account its rank through $R(e_i)$, giving an estimate $\hat{\beta}_w$ that is robust to $y$ outliers. However, this estimate is still vulnerable to $x$ outliers since (1.3.2) directly takes into account the magnitude of the $x_i$ in obtaining an estimate of $\beta$. As noted earlier, unless the $x$-values are fixed, this method loses its robustness.

The proposed estimator seeks to circumvent the problem posed by (1.3.1)
and (1.3.2). Instead of using the $x_i$ magnitudes directly, consider using their ranks instead, i.e., obtain an estimator $\hat{\beta}_s$ that satisfies

$$\sum_i \left( R(e_i) - \frac{n+1}{2} \right) \left( R(x_i) - \frac{n+1}{2} \right) = 0. \quad (1.3.3)$$

Hence, even if the $x$-values are not fixed, their magnitude is not directly considered, unlike in the two previous methods. Note that (1.3.3) states that the estimate to be obtained is such that the Spearman correlation between the $x$ values and the residuals is 0. We therefore refer to $\hat{\beta}_s$ as the Spearman estimator of $\beta$. This estimator should be robust to outlying values in the $x$ and $Y$ spaces.

The idea of using a correlation coefficient to obtain estimators for $\beta$ is not new. Sen (1968) obtained an estimator of $\beta$ based on Kendall's rank correlation tau. The point estimator is the median of the set of slopes $s_{ij} = (Y_j - Y_i)/(x_j - x_i)$ joining pairs of points with $x_j \neq x_i$, and is unbiased. This is the same estimator derived by Theil (1950).

Although it is not customary to associate the $s_{ij}$ with least squares estimation, it can be shown that $\hat{\beta}_{ls}$ is a weighted mean of the $s_{ij}$, with the
weights proportional to \((x_j - \bar{x})^2\). It is well known that median estimators are in general less affected by outliers than a mean estimator (or a weighted mean estimator with weights determined without regard to whether a particular observation is an outlier). The Theil (1950) – Sen (1968) estimator should then be more robust to outliers than the least squares estimator.

Since the proposed slope estimator of \(\beta\) is just a special case of the GR-estimator in the context of simple linear regression, the challenge is to develop a Spearman estimator for multiple regression. Unfortunately, there is no clear concept of multidimensional ranks, so the method cannot be readily extended to multiple regression.

Instead of dealing with multidimensional ranks directly, the proposed methods bypass the problem and instead exploit the relationships between simple linear regression and multiple regression. These methods are described in Chapter 3. Another proposed method, described in Chapter 4, will assign simple linear regression weights to each independent variable to come up with a multiple regression estimate.
2.1 The Estimate

Consider the simple linear regression model, i.e., model (1.2.1) with \( p = 1 \).

In vector form, we have

\[
Y = \alpha 1 + \beta x + \varepsilon
\]  

(2.1.1)

where

\[
Y = [Y_1, Y_2, \ldots, Y_n]' \text{ is the dependent variable}
\]
\[
x = [x_1, x_2, \ldots, x_n]' \text{ is the predictor variable}
\]
\[
\alpha = \text{unknown intercept parameter}
\]
\[
\beta = \text{unknown slope parameter}
\]
\[
1 = \text{an } n \times 1 \text{ vector of 1's}
\]
\[
\varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n]' \sim F \text{ continuous and unknown, is the error term.}
\]

Let \( \hat{\beta}_S \) be an estimator of \( \beta \) such that

\[
S(\hat{\beta}_S) = 0
\]  

(2.1.2)

where
\[ S(\beta) = \sum_{i=1}^{n} \left[ R(e_i) - \frac{n+1}{2} \right] \left[ R(x_i) - \frac{n+1}{2} \right] \quad (2.1.3) \]

\[ e_i = Y_i - \beta x_i \text{ denotes the } i^{th} \text{ residual} \]

\[ R(\cdot) = \text{ the rank function.} \]

It is clear that \( \hat{\beta}_s \) is an estimator of \( \beta \) such that the Spearman correlation between the \( e_i \) and the \( x_i \) is 0. Thus, we shall refer to \( \hat{\beta}_s \) as the Spearman estimator of the slope.

We shall refer to \( S(\beta), (2.1.3), \) as the Spearman estimating function. It entails the use of the rank of \( Y_i \) (through the rank of the \( e_i \)) in determining the Spearman estimate of \( \beta \). However, unlike the Wilcoxon estimating function, \( S(\beta) \) uses the rank of the \( x_i \) rather than directly using its magnitude. Consequently, the resulting estimate \( \hat{\beta}_s \) is robust not just to \( Y \) outliers, but to \( x \) outliers as well. This will be shown formally in Theorem 2.2.1.

The next two theorems will give us the form of \( \hat{\beta}_s \). Let

\[ s_y = \frac{Y_j - Y_i}{x_j - x_i} \quad (2.1.4) \]

be the slope of the line formed by the points \((x_3, Y_3)\) and \((x, Y')\), \(x \neq x_3\), in a
regression scatter diagram. Note that for $i < j$, there are $\binom{n}{2}$ such slopes. The next theorem gives us the relationship between $S(\beta)$ and the $s_{ij}$.

**Theorem 2.1.1:** $S(\beta)$ is a non-increasing step function which steps down at each sample slope $s_{ij}$. The size of the step at $s_{ij}$ equals $R(x_j) - R(x_i)$.

**Proof:**

Assume that the $Y_i$ are fixed. Without loss of generality, let $x_i < x_j$ for all $i < j$. Then there exists a sufficiently large $\beta$ such that

$$Y_j - \beta x_j < Y_i - \beta x_i.$$  

Similarly, there exists a sufficiently small $\beta$ such that

$$Y_j - \beta x_j > Y_i - \beta x_i.$$  

Hence, the value of $S(\beta)$ changes where the ordering of the residuals changes, i.e., at the value of $\beta$ such that

$$Y_j - \beta x_j = Y_i - \beta x_i,$$

i.e., at $\beta = s_{ij}$. Thus, $S(\beta)$ is a step function whose jumps occur at $s_{ij}$.

Now, for $\beta = s_{ij}$, suppose the residuals $Y_j - \beta x_j$ and $Y_i - \beta x_i$ are the $(k+1)^{th}$ and the $k^{th}$ among the ordered residuals. Then the change in $S(\beta)$ at $s_{ij}$ when $\beta$
moves from just below \( s_i \) to just above it, is

\[
\left\{ \left( R_j - \frac{n+1}{2} \right) \left( k + 1 - \frac{n+1}{2} \right) + \left( R_i - \frac{n+1}{2} \right) \left( k - \frac{n+1}{2} \right) \right\} \\
- \left\{ \left( R_j - \frac{n+1}{2} \right) \left( k - \frac{n+1}{2} \right) + \left( R_i - \frac{n+1}{2} \right) \left( k + 1 - \frac{n+1}{2} \right) \right\}
\]

\[= R_j - R_i, \]

where \( R_j = \text{Rank}(x_j) \). This implies that \( R_j - R_i > 0 \), since by assumption, \( x_i < x_j \) for all \( i < j \). Hence, \( S(\beta) \) is a non-increasing function of \( \beta \). ■

We are now ready to obtain the computational form of \( \hat{\beta}_s \). Define \( S = S(\beta) \), and define the following distribution function \( G \):

\[
G(t) = 1 - \frac{S(t) - \min S}{\max S - \min S}.
\]

Note that since \( S \) is just the numerator of the Spearman rank order correlation coefficient, it should be clear that

\[
\min S = -\frac{n(n^2 - 1)}{12}
\]

and

\[
\max S = \frac{n(n^2 - 1)}{12}.
\]

**Theorem 2.1.2:** Let \( \hat{\beta}_s \) be a solution to \( S(\beta) = 0 \). Then
\[ \hat{\beta}_s = \text{median}_{R_i, R_j} \frac{Y_j - Y_i}{x_j - x_i} \]

for \( i < j \), i.e., \( \hat{\beta}_s \) is the weighted median of the \( \binom{n}{2} \) pairwise slopes \( \frac{Y_j - Y_i}{x_j - x_i} \), with weights proportional to \( R_j - R_i \).

Proof:

Let \( T \) be a random variable with probability distribution \( G \). Since

\[ S(\hat{\beta}_s) = 0, \]

\[ P(T < \hat{\beta}_s) = 1 - \frac{n(n^2 - 1)/12}{n(n^2 - 1)/6} = \frac{1}{2}. \]

Hence, \( \hat{\beta}_s \) is the median of the distribution function \( G \), and thus, a weighted median of the \( s_{ij} \). From Theorem 2.1.1, the probability assigned to each

\[ s_{ij} \text{ is } q_{ij} = \frac{R_j - R_i}{\sum_{i<j} (R_j - R_i)}. \]

Therefore, \( \hat{\beta}_s \) is the weighted median of the \( \binom{n}{2} \) pairwise slopes, with each slope being given the weight \( q_{ij} = \frac{R_j - R_i}{\sum_{i<j} (R_j - R_i)} \).
Scholz (1978) and Sievers (1978) investigated the properties of a similar estimator, but for general weights $q_y$. This includes the Theil (1950) – Sen (1968) estimator ($q_y \sim \text{constant}$) and the Wilcoxon estimator ($q_y \sim x_j - x_l$) of Jaeckel (1972). As noted in Chapter 1, however, Jaeckel considered the use of dispersion measures, and these dispersion measures are certain linear combinations of the ordered residuals. Jaeckel did not use the general idea of correlations to derive the slope estimator.

2.2 The $GR$ Estimator

This section presents the generalized-rank ($GR$) estimator of the slope parameter in a regression model. The $GR$-estimates were proposed by Sievers (1983). Many of their properties were further developed by Naranjo and Hetttsmanperger (1994). $GR$ estimators are characterized as $R$ estimates which have influence functions that are bounded in both factor and design spaces and have positive breakdown. Thus, these estimators achieve robustness against both $x$- and $Y$-outliers (in the sense of bounded influence). This family of estimators is an essential part of this paper since it will be shown later that, in the context of
simple linear regression, the Spearman estimator is a special case of the \textit{GR} estimator. It then goes without saying that the Spearman estimator is robust to \textit{x}- and \textit{Y}-outliers.

Denote the \textit{GR} estimator of \( \beta \) by \( \hat{\beta}_{GR} \). Note that \( \hat{\beta}_{GR} \) satisfies

\[
S_{GR}(\hat{\beta}_{GR}) = 0
\]

where \( S_{GR}(\beta) \) is given by

\[
S_{GR}(\beta) = \sum_{i<j} b_{ij}(x_j - x_i) \text{sgn}(e_j - e_i) \quad (2.2.1)
\]

where the \( b_{ij} \) are weights that are specified functions of the \( x \)'s, \( e_j \) is the \( j^{th} \) residual, and that \( \text{sgn}(\cdot) \) is a function such that

\[
\text{sgn}(a - b) = \begin{cases} 
1 & \text{if } a > b \\
0 & \text{if } a = b \\
-1 & \text{if } a < b
\end{cases}
\]

The weights \( b_{ij} \) are positive and symmetric, i.e., \( b_{ij} = b_{ji} \). If \( b_{ij} = 1 \), then (2.2.1) will yield the Wilcoxon estimator. Hence, we refer to \( \hat{\beta}_{GR} \) as the generalized-rank estimator.

The next theorem shows that, with the use of a proper weighting scheme, the Spearman estimator is a special case of the \textit{GR} estimator. It formally shows
that the Spearman estimator achieves robustness in both the \( x \) and \( Y \)-spaces.

**Theorem 2.2.1:** If

\[
    b_{ij} = \begin{cases} 
    \frac{1}{n} \cdot \left( \frac{R_j - R_i}{x_j - x_i} \right) & \text{if } x_j \neq x_i \\
    0 & \text{otherwise}
    \end{cases}
\]

for \( i < j \), then \( \hat{\beta}_s = \hat{\beta}_{GR} \).

**Proof:**

For \( b_{ij} = \frac{1}{n} \cdot \left( \frac{R_j - R_i}{x_j - x_i} \right) \), \( i < j \), we have

\[
    0 = S_{GR}(\beta) = \sum_{i<j} \sum b_{ij} (x_j - x_i) \text{sgn}(e_j - e_i) = \frac{1}{n} \sum_{i<j} \sum(R_j - R_i) \text{sgn}[(y_j - \beta x_j) - (y_i - \beta x_i)]
\]

This clearly implies that

\[
    \sum_{i<j} \sum(R_j - R_i) \text{sgn}(s_{ij} - \beta) = 0 \quad (2.2.2)
\]

where \( s_{ij} \) is the expression in (2.1.4). Hence, \( S_{GR}(\beta) \) is a step function that changes values at \( s_{ij} \) of magnitude proportional to \( R_j - R_i \). Note that the
expression in (2.2.2) is essentially 0 if "half" of the $s_i$ is less than $\beta$ and if "half" of the $s_i$ is greater than $\beta$. Therefore, the solution to (2.2.2) is

$$\hat{\beta}_{GR} = \text{median} \frac{Y_j - Y_i}{x_j - x_i}, x_i \neq x_j$$

$$= \hat{\beta}_g. \blacksquare$$

2.3 Equivariance Properties

In some of the proofs in succeeding discussions, it will be convenient to say "without loss of generality, assume that $\beta = 0$." It is regression equivariance that enables us to say this, since it makes the results hold for all values of $\beta$.

**Definition 2.3.1:** Let $v$ be a real number. A regression estimator is said to be regression equivariant if

$$T(x_i, y_i + vx_i) = T(x_i, y_i) + v. \quad (2.2.3)$$

Thus, when we state that "without loss of generality, assume that $\beta = 0$," we are assuming that the results are valid at any parameter value through the application of (2.2.3). The following theorem now shows that the Spearman rank regression estimator is regression equivariant.
Theorem 2.3.2: $\hat{\beta}_s$ is regression equivariant.

Proof:

Let $Y_i^* = Y_i + vx_i$, where $v \in \mathbb{R}$. Let $\hat{\beta}_s^*$ be the estimate calculated from $(x_i, Y_i^*)$. Using (2.1.3), we have

$$S(\beta^*) = \sum_i \left[ R(x_i) - \frac{n+1}{2} \right] \left[ R(Y_i^* - \beta^*x_i) - \frac{n+1}{2} \right]$$

$$= \sum_i \left[ R(x_i) - \frac{n+1}{2} \right] \left[ R(Y_i + vx_i - x_i \beta^*) - \frac{n+1}{2} \right]$$

$$= \sum_i \left[ R(x_i) - \frac{n+1}{2} \right] \left[ R(Y_i - x_i(\beta^* - v)) - \frac{n+1}{2} \right]$$

The solution to $S(\beta^*) = 0$ is $\hat{\beta}_s = \hat{\beta}_s^* - v$, which implies that $\hat{\beta}_s = \hat{\beta}_s^* + v$. ■

2.4 Asymptotic Properties

We shall now establish the asymptotic properties of $\hat{\beta}_s$ by making full use of Theorem 2.2.1. Recall that if $b_0 = \frac{1}{n} \left( \frac{R_j - R_i}{x_j - x_i} \right)$, then $\hat{\beta}_s = \hat{\beta}_{GR}$. It will be convenient to represent the weights as follows:

$$w_{ij} = \begin{cases} -\left( \frac{1}{n} \right) b_{ij} & i \neq j \\ \left( \frac{1}{n} \right) \sum_{k} b_{ik} & i = j \end{cases} \quad (2.4.1)$$
Let \( W \) be the \( n \times n \) matrix of elements \( w_{ij} \). Note that \( W \) is symmetric and that its rows sum to zero. We now have the following list of assumptions.

**A1:** Let \( f \) be the probability density function of the errors. Then \( f \) is absolutely continuous, and \( 0 < I(f) < \infty \), where
\[
I(f) = \int_0^1 \varphi_f(u) \, du
\]
denotes the Fisher information and
\[
\varphi_f(u) = \frac{f'(F^{-1}(u))}{f(F^{-1}(u))}.
\]

**A2:** Let \( H = x(x'x)^{-1}x' \), the so-called "hat matrix." The diagonal matrix of \( H \) are referred to as the leverage values, and we shall denote them as \( h_{ii} \).

We will assume that
\[
\lim_{n \to \infty} \max_{1 \leq i \leq n} h_{ii} = 0.
\]

A2 is known as Huber's condition. Huber (1981) showed that if the errors, \( \varepsilon_i \), are iid with finite variance, then the least squares estimates possess asymptotic normality if and only if Huber's condition is satisfied.

**A3:** \[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sigma_x^2, \text{ where } \sigma_x^2 > 0.
\]

**A4:** \[
\lim_{n \to \infty} \frac{1}{n} x'Wx = c, \text{ where } c > 0.
\]

**A5:** \[
\lim_{n \to \infty} \frac{1}{n} x'W^2x = v, \text{ where } v > 0.
\]
A6: \( Wx \) satisfies condition A2.

A7: There exists a number \( 0 < \rho < 1 \) such that

\[
\rho_n = \frac{\sum_i (x_i - \bar{x}) (R(x_i) - \frac{n+1}{2})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i \left( R(x_i) - \frac{n+1}{2} \right)^2}} \to \rho
\]

The following theorem lists some important relationships that will be used in succeeding discussions.

**Theorem 2.4.1:** Let \( W \) be the \( n \times n \) matrix with elements \( w_{ij} \), given by (2.4.1),

where \( b_{ij} = \frac{1}{n} \left( \frac{R_j - R_i}{x_j - x_i} \right) \).

Then the following relationships hold true:

\[
Wx = \frac{1}{n} \begin{bmatrix}
R(x_1) - \frac{n+1}{2} \\
R(x_2) - \frac{n+1}{2} \\
\vdots \\
R(x_n) - \frac{n+1}{2}
\end{bmatrix} \tag{2.4.2}
\]

\[
x'Wx = \frac{1}{n} \sum_i (x_i - \bar{x}) \left( R(x_i) - \frac{n+1}{2} \right) \tag{2.4.3}
\]

\[
x'W^2x = \frac{1}{n^2} \sum_i \left( R(x_i) - \frac{n+1}{2} \right)^2 \tag{2.4.4}
\]

\[= \frac{1}{n^2} \frac{n(n^2 - 1)}{12}. \]

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

Without loss of generality, assume that \( x_1 < x_2 < \ldots < x_n \), which implies

\[
R(x_i) = i. \quad \text{Thus, } b_{ij} = \frac{1}{n} \left( \frac{j - i}{x_j - x_i} \right). \quad \text{Now, the } i^{th} \text{ element of } Wx \text{ is given by}
\]

\[
\sum_j w_{ij} x_j. \quad \text{Note that,}
\]

\[
\sum_j w_{ij} x_j = w_n x_i + \sum_{j<i} w_{ij} x_j + \sum_{j>i} w_{ij} x_j = \frac{1}{n^2} \left[ x_i \sum_{k>i} \frac{k-i}{x_k - x_i} - \sum_{j=1}^{i-1} \frac{i-j}{x_i - x_j} - \sum_{j=i+1}^{n} \frac{j-i}{x_j - x_i} \right]
\]

\[
= \frac{1}{n^2} \left[ x_i \left( \frac{i-1}{x_i - x_1} + \frac{i-2}{x_i - x_2} + \cdots + \frac{1}{x_i - x_{i-1}} + \frac{1}{x_n - x_i} + \frac{2}{x_{i+2} - x_i} + \cdots + \frac{n-i}{x_i - x_1} \right) \right]
\]

\[
- \left( \frac{i-1}{x_i - x_{i-1}} \right) x_i - \left( \frac{1}{x_{i+1} - x_i} \right) x_{i+1} - \left( \frac{2}{x_{i+2} - x_i} \right) x_{i+2}
\]

\[
- \cdots - \left( \frac{n-i}{x_n - x_i} \right) x_n
\]

\[
= \frac{1}{n^2} \left[ \left( \frac{i-1}{x_i - x_1} \right) (x_i - x_1) + \left( \frac{i-2}{x_i - x_2} \right) (x_i - x_2) + \cdots + \left( \frac{1}{x_i - x_{i-1}} \right) (x_i - x_{i-1}) + \left( \frac{1}{x_{i+1} - x_i} \right) (x_i - x_{i+1}) + \right.
\]

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

\[
\sum_j w_j x_j = \frac{1}{n^2} \left[ \sum_{j=1}^{i-1} j - \sum_{j=1}^{n-i} j \right]
\]

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

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

\[
= \frac{1}{n} \left( i - \frac{n+1}{2} \right)
\]

\[
= \frac{1}{n} \left( R(x_i) - \frac{n+1}{2} \right).
\]

Hence,

\[
Wx = \frac{1}{n} \left[ \begin{array}{c} R(x_1) - \frac{n+1}{2} \\ R(x_2) - \frac{n+1}{2} \\ \vdots \\ R(x_n) - \frac{n+1}{2} \end{array} \right].
\]

We notice that \( W \) is a matrix that centers the ranks of the \( x \) values.

It should be fairly easy to obtain the values of \( x'Wx \) and \( x'W^2x \). Note
that \( x'Wx \) is just the inner product of two vectors: \( x \) and \( Wx \). Therefore,

\[
x'Wx = \frac{1}{n} \sum_i x_i \left( R(x_i) - \frac{n + 1}{2} \right)
\]

\[
= \frac{1}{n} \sum_i (x_i - \bar{x}) \left( R(x_i) - \frac{n + 1}{2} \right).
\]

Also, \( x'W^2x = (Wx)'Wx \). Thus,

\[
x'W^2x = \frac{1}{n^2} \sum_i \left( R(x_i) - \frac{n + 1}{2} \right)^2
\]

\[
= \frac{1}{n^2} \frac{n(n^2 - 1)}{12}.
\]

Let us now discuss some of the asymptotic properties of the \( GR \) estimator.

Denote the scale parameter \( \tau \) as

\[
\tau = \frac{1}{\sqrt{12} \int f^2(x) \, dx}.
\] (2.4.5)

The following lemma is stated with proof in Hettmansperger and McKean (1998).

It establishes the asymptotic normality of \( GR \) estimates.

**Lemma 2.4.2:** Under the assumptions A1, A4, A5, and A6, \( \hat{\beta}_{GR} \) has an approximate \( N(\beta, \tau^2 (x'Wx)^{-1} x'W'x (x'Wx)^{-1}) \) distribution.

Thus, taking in mind Theorem 2.2.1, we can make full use of Lemma 2.4.2 to ascertain the asymptotic normality of \( \hat{\beta}_s \). This is now shown in the next
Theorem 2.4.3: Consider the assumptions A1, A3, A4, A5, A6 and A7. Then
\[ \sqrt{n} (\hat{\beta}_s - \beta) \] has a limiting normal distribution with mean 0 and variance \[ \frac{\tau^2}{\sigma^2_s}, \]
where \( \tau \) is given in (2.4.5), \( \rho \) is given in A7, and \( \sigma^2_s \) is given in A3.

Proof:
We only need to establish the form of the variance, since Theorem 2.2.1 and Lemma 2.4.2 have already established the rest of the theorem.

Using the relationships in Theorem 2.4.1, we have
\[ \tau^2 (x'Wx)^{-1} x'W^2x (x'Wx)^{-1} = \tau^2 \frac{1}{n^2} \sum_i \left( R(x_i) - \frac{n+1}{2} \right)^2 \left[ \frac{1}{n} \sum_i (x_i - \bar{x})(R(x_i) - \frac{n+1}{2}) \right] \]

\[ \tau^2 (x'Wx)^{-1} x'W^2x (x'Wx)^{-1} = \tau^2 \frac{1}{n} \left[ \sum_i (x_i - \bar{x})(R(x_i) - \frac{n+1}{2}) \right] \frac{1}{n} \left[ \sum_i (x_i - \bar{x})^2 \right] \frac{1}{n} \sum_i (x_i - \bar{x})^2 \]

Now, from A3 and A7, we have
\[ \tau^2 (x'Wx)^{-1} x'W^2x (x'Wx)^{-1} = \frac{1}{n} \frac{\tau^2}{\rho^2 \sigma^2_s} \frac{1}{n} \sum_i (x_i - \bar{x})^2 \]

Therefore, the asymptotic variance of \( \sqrt{n} (\hat{\beta}_s - \beta) \) is \( \frac{\tau^2}{\rho^2 \sigma^2_s} \).
2.5 Asymptotic Relative Efficiency

The asymptotic variance of $\hat{\beta}_s$ derived in Theorem 2.4.3 provides a basis for efficiency comparisons of the Spearman estimator relative to the least squares and Wilcoxon estimators. Recall that $\tau = \frac{1}{\sqrt{12} \int f^2(x) dx}$ and assume that $\sigma^2 < \infty$ is the variance of the underlying error distribution. We now give a working definition of asymptotic relative efficiency.

**Definition 2.5.1:** If two estimates $\hat{\eta}_1$ and $\hat{\eta}_2$ have (asymptotic) variances $m_1$ and $m_2$, respectively, then the (asymptotic) efficiency of $\hat{\eta}_1$ with respect to $\hat{\eta}_2$ is

$$ARE(\hat{\eta}_1 \mid \hat{\eta}_2) = \frac{m_2}{m_1}.$$ 

Note that $ARE(Wil|LS) = \frac{\sigma^2}{\tau^2}$. It is the familiar efficiency of the Wilcoxon test to the Student's $t$-test. It is known for instance that $ARE(Wil|LS) \leq 0.955$ when the error distribution $F$ is normal, that $ARE(Wil|LS)$ can be arbitrarily large for error distributions with heavy tails, and that $ARE(Wil|LS) \geq 0.864$ for all continuous $F$, with the equality attained for a particular $F$ (Hodges...
and Lehmann, 1956). Because of this high efficiency, \( \hat{\beta}_{WL} \) merits serious consideration as an alternative to the classical least squares estimate \( \hat{\beta}_{LS} \).

**Theorem 2.5.2:** The asymptotic efficiency of the Spearman rank regression estimator with respect to the least squares and Wilcoxon estimators are

\[
ARE(\text{Spearman} \mid \text{LS}) = \frac{\sigma^2}{\tau^2 \rho^2}
\]

\[
ARE(\text{Spearman} \mid \text{Wil}) = \rho^2.
\]

**Proof:**

The asymptotic relative efficiencies are obtained by a straightforward application of Definition 2.5.1 and Theorem 2.4.3. Note that \( \text{Var}(\hat{\beta}_{LS}) \doteq \frac{1}{n} \frac{\sigma^2}{\sigma_x^2} \).

Therefore,

\[
ARE(\text{Spearman} \mid \text{LS}) = \frac{\frac{1}{n} \frac{\sigma^2}{\sigma_x^2}}{\frac{1}{n} \frac{\tau^2}{(\sigma_x^2 \rho^2)}}
= \frac{\sigma^2}{\tau^2 \rho^2}
\]

and
\[
\text{ARE(Spearman | Wil)} = \frac{\frac{1}{n} \frac{\tau^2}{\sigma_z^2}}{\frac{1}{n} \frac{\tau^2}{(\sigma_z^2 \rho^2)}}
\]

\[= \rho^2. \quad \blacksquare\]

Note that \(\rho_n\) in A7 is the correlation of the pairs \((x_i, R(x_i)), i = 1, 2, \ldots, n\). Thus, A7 is warranted since the \(x_i\) are monotonic with its rank. We can see that the asymptotic variance of \(\hat{\beta}_s\) is minimized if \(\rho_n \to 1\). Sen (1968) cites two cases where \(\rho_n = 1\) are of special interest. These two cases relate to how the independent variables are configured. In either of those cases, we can refer to the independent variables as being optimally designed if \(\rho_n = 1\). They are asymptotically optimally designed if \(\rho_n \to 1\) as \(n \to \infty\).

Theorem 2.5.2 states that the Spearman rank regression estimator is at best as efficient as the Wilcoxon estimator. For optimum or asymptotically optimum designs they are asymptotically equally efficient. This is not really unexpected, since the Wilcoxon and the least squares estimators utilize the exact values of \(x_1, x_2, \ldots, x_n\) in the estimating function, whereas the Spearman estimator utilizes only their ordering. Thus, there is always a loss of efficiency when we use
the Spearman estimator. But if the independent variable has clusters of outlying points then the downweighting that Spearman regression provides may be necessary.

2.6 Estimation of the Intercept Parameter

We now propose an estimate of the unknown intercept parameter $\alpha$.

Note that the estimating function used to obtain the Spearman estimate of the slope,

$$ S(\beta) = \sum_{i=1}^{n} \left[ R(Y_i - \beta x_i) - \frac{n+1}{2} \right] \left[ R(x_i) - \frac{n+1}{2} \right], $$

is translation invariant in the sense that it is not affected by the value of the unknown parameter $\alpha$. Thus, $\hat{\beta}_s$ provides no information relative to the value of $\alpha$.

The intercept parameter requires the specification of a location functional, $T(e_i)$. Take $T(e_i) = \text{median}(e_i)$. Since we assume, without loss of generality, that $T(e_i) = 0$, we can set $\alpha = \text{median}(Y_i - \beta x_i)$. This leads to estimating $\alpha$ using the median of the full-model Spearman residuals, i.e.,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[ \hat{\alpha}_s = \text{median}(Y_i - \hat{\beta}_s x_i). \] (2.6.1)

Note that the intercept for R-fits is estimated in the same manner. Furthermore, the said method for obtaining the estimate of the intercept is analogous to least squares, since the least squares estimate of the intercept is the mean of the residuals \( Y_i - \hat{\beta}_{LS} x_i \).
CHAPTER 3

MULTIPLE REGRESSION ESTIMATES OF $\beta$
THAT REDUCE TO SIMPLE LINEAR REGRESSION

3.1 Introduction

We now consider the problem of estimating the regression parameter in multiple regression. Sections 3.2 and 3.3 will consider the multiple regression model with $p = 2$ parameters. We shall postpone the discussion of multiple regression models with $p > 2$ in Section 3.5.

Suppose we fit the multiple regression model of the form

$$Y = \alpha 1 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

or equivalently,

$$Y = \alpha 1 + X\beta + \epsilon$$

where $X = [x_1, x_2]$ is an $n \times 2$ matrix of independent variables, $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ is a $2 \times 1$ vector of parameters, $\alpha$ is the intercept parameter, $Y$ is the response vector, and $\epsilon \sim F$ continuous and unknown, is the error term. We would like to obtain an estimate of $\beta$ that is robust to outliers in both factor and design spaces.
The examples given in Chapter 5 provide a simple illustration of how the presence of outliers can obscure the parameter estimates. Unfortunately, detecting outliers when there are two or more independent variables is not trivial, especially in the presence of several outliers. Classical identification methods do not always find these outliers since many of them are based on the sample mean and covariance matrix, which are themselves affected by outliers. A small cluster of outliers will attract the sample mean and will inflate the covariance matrix in its direction.

Residual plots, in particular, are useful in detecting potential outliers. In the case of least squares, outliers affect the fit enough for them to be masked. See Rousseeuw and van Zomeren (1990) for an example. The sensitivity of least squares to influential data points does not guarantee that outliers will have large residuals, and so the residual plots in this case will not be of any help in flagging the outlying values. The occurrence of the so-called masking effect takes place.

The same problem can be pointed out for the regular rank-based or $R$-estimates. As mentioned in several occasions in this paper, $R$-estimates are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
robust to $Y$-outliers but not to $x$-outliers. It will be illustrated in Chapter 5 that, in the presence of leverage values in the data, the $R$-fits do not guarantee that the corresponding residuals for these leverage values will be large so that they can be flagged as outliers. The masking effect problem persists in $R$-fits in the presence of $x$-outliers.

This chapter aims to develop the concept of Spearman rank regression for multiple regression models with $p = 2$ and $p = 3$ parameters. Recall that in simple linear regression, the estimate of the regression parameter was obtained as a solution to $S'(\beta) = 0$, where the estimating function $S(\beta)$ is given by (2.1.3). As noted earlier, $S(\beta)$ does not utilize the magnitude of the $x$ values directly. To obtain a parameter estimate that is robust to $x$ outliers, $S(\beta)$ utilizes only the ordering or the ranks of the $x$ values. Since ranks are inherently one-dimensional, the realization of this chapter's objective is not straightforward. One has to entertain the existence of multidimensional ranks.

We now propose multiple regression schemes that bypass the problem of multidimensional ranks and instead explore the known relationships between
simple linear regression and multiple regression. This approach will offer the appealing advantage of being able to use the theory developed in Chapter 2 and other GR theory for simple linear regression to develop the theory for multiple Spearman regression since, as shown in Theorem 2.2.1, the Spearman regression estimate is a special case of the GR estimate for simple linear regression.

As a first step in exploring the relationship between simple linear regression and multiple regression, consider the model (3.1.1). Let \( \hat{\beta}_2^{(1)} \) be a full-model estimate of \( \beta_2 \) such that \( \sqrt{n} \left( \beta_2 - \hat{\beta}_2^{(1)} \right) = O_p(1) \). Then

\[
Y - \hat{\beta}_2^{(1)} x_2 = \alpha_1 1 + \beta_1 x_1 + \left( \beta_2 - \hat{\beta}_2^{(1)} \right) x_2 + \varepsilon, \tag{3.1.2}
\]

where \( x_1 \) and \( x_2 \) are linearly independent and are assumed centered. In addition to the assumptions stated in Section 2.4, assume also that there exists positive definite matrices \( \Sigma, C_1, \) and \( V_1 \) such that

A8: \( \frac{1}{n} X'X \to \Sigma \)

A9: \( \frac{1}{n} X'Wx_1X \to C_1 \)

A10: \( \frac{1}{n} X'W^2x_1X \to V_1 \)
A11: \( x_1 = O_p(1) \)

where \( X = [x_1, x_2] \) and \( W_{x_1} \) is the weight matrices given in Section 2.4 computed from \( x_1 \).

**Lemma 3.1.1:** *Suppose that conditions A1 – A10 are true for \( x_1 \). Let \( \hat{\beta}_1 \) be the slope of the simple regression of \( Y - \hat{\beta}_2^{(1)}x_2 \) against \( x_1 \). Then under model (3.1.1),

\[
\hat{\beta}_1 = \beta_1 + \frac{\sqrt{3\tau}}{n} \left( x_1'W_{x_1}x_1 \right)^{-1} S_1(\beta_1) + \left( \beta_2 - \hat{\beta}_2^{(1)} \right) \left( x_1'W_{x_1}x_1 \right)^{-1} x_1'W_{x_2}x_2 + o_p \left( \frac{1}{n^{\frac{1}{2}}} \right) \quad (3.1.4)
\]

i.e.,

\[
\sqrt{n} \left[ \hat{\beta}_1 - \left( \beta_1 + \left( \beta_2 - \hat{\beta}_2^{(1)} \right) \left( x_1'W_{x_1}x_1 \right)^{-1} x_1'W_{x_2}x_2 \right) \right] = \sqrt{3\tau} \left( \frac{1}{n} x_1'W_{x_1}x_1 \right)^{-1} \frac{1}{n^{\frac{1}{2}}} S_1(\beta_1) + o_p(1).
\]

**Proof:**

Let \( \eta = \begin{bmatrix} \beta_1 & \beta_2 - \hat{\beta}_2^{(1)} \end{bmatrix}' \) and \( \hat{\eta}' = \begin{bmatrix} \hat{\beta}_1 & 0 \end{bmatrix}' \). The following expansion of \( S(\eta) \) is stated in Naranjo and Hettmansperger (1994):

\[
n^{-\frac{1}{2}} \begin{bmatrix} S_1(\hat{\eta}) \\ S_2(\hat{\eta}) \end{bmatrix} = n^{-\frac{1}{2}} \begin{bmatrix} S_1(\eta) \\ S_2(\eta) \end{bmatrix} - \frac{1}{\sqrt{3\tau}} \begin{bmatrix} x_1'W_{x_1}x_1 & x_1'W_{x_1}x_2 \\ x_2'W_{x_1}x_1 & x_2'W_{x_1}x_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ 0 - (\beta_2 - \hat{\beta}_2^{(1)}) \end{bmatrix} + o_p(1).
\]

The desired result follows from the fact that \( S_1(\hat{\eta}) \neq 0 \). ■

The quantities \( S_1(\beta_1) \) and \( \tau \) are as given in (2.2.1) and (2.4.5),

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
respectively. The weights are, as in Chapter 2, defined as \( b_y = \frac{1}{n} \left( \frac{R_y - R_i}{x_{ij} - x_{ii}} \right) \),

where \( x_i \) is the independent variable, and \( R_j = \text{rank}(x_{ij}) \).

As pointed out earlier, we shall take advantage of the fact that the Spearman estimator in simple regression is just a special case of the GR estimator. Let us now state an important lemma on the estimating function \( S_i(\beta_i) \). This lemma will be crucial in achieving the above mentioned objective.

**Lemma 3.1.2:** Under the assumptions of Lemma 3.1.1,

\[
\frac{\sqrt{3\pi}}{n} \left( x_i'W_{x_i}x_i \right)^{-1} S_i(\beta_i) \rightarrow 0 \text{ as } n \rightarrow \infty.
\]

Proof:

\[
\frac{\sqrt{3\pi}}{n} \left( x_i'W_{x_i}x_i \right)^{-1} S_i(\beta_i) = \frac{\sqrt{3\pi}}{\sqrt{n}} \left( \frac{1}{n} x_i'W_{x_i}x_i \right)^{-1} \frac{1}{n^{1/2}} S_i(\beta_i)
\]

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

since, from Theorem 5.2.9 of Hettmansperger and McKean (1998), we have,

\[
\frac{1}{n^{1/2}} S_i(\beta_i) = \frac{1}{n^{1/2}} S_i(0) - \frac{\sqrt{n}}{\sqrt{3\pi}} c \beta_i + o_p(1)
\]

uniformly for all \( \beta \) such that \( \sqrt{n} \| \beta \| \leq d \), for any \( d > 0 \), where \( c \) is given in A4 of Section 2.4, and further, from Corollary 5.2.4 of Hettmansperger and McKean

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
(1998), \( \frac{1}{n^{3/2}} S(0) \) has an asymptotic \( N\left(0, \frac{1}{3} v\right) \) distribution assuming that \( \beta = 0 \), where \( v \) is given in A5 of Section 2.4. This completes the proof. ■

The following lemma is immediate from Lemma 3.1.1 and Lemma 3.1.2:

**Lemma 3.1.3:** Under the assumptions of Lemma 3.1.1, we have

\[
\hat{\beta}_1 = \beta_1 + (\hat{\beta}_2 - \hat{\beta}_2^{(i)}) (x_1'W_{x_1}x_1)^{-1} x_1'W_{x_1}x_2 + o_p\left(n^{-\frac{1}{2}}\right).
\]

The next three sections will discuss the proposed methods for obtaining the estimate of the multiple regression parameter. The first method, discussed in Section 3.2, shall be referred to as the *residual method*, so named to recognize the important role that the residuals will play to obtain the estimate. The second method is the *bias adjustment method*. This method will attempt to adjust the inherent bias present if we fit a simple linear regression model when the true model is a multiple regression model. This is explained in Section 3.3.

The two methods described above will all take advantage of the relationships between simple linear regression and multiple regression. We first discuss the residual method.
3.2 The Residual Method

Consider the regression model (3.1.1):

\[ Y = \alpha 1 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon. \]

Our objective is to find an estimate for \( \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \). Observe that the results stated in Lemma 3.1.3 hold for the contiguous model (3.1.2) but not necessarily for model (3.1.1), where \( \beta_2 \) is fixed.

We shall first illustrate the process by obtaining the residual method estimate of \( \beta_2 \). It will be noticed that the method is symmetric; similar steps can be employed to obtain the residual method estimate of \( \beta_1 \).

Let \( \hat{\beta}_2^{(i)} \) be an initial (full-model) estimate of \( \beta_2 \) such that \( \beta_2 - \hat{\beta}_2^{(i)} = O_p \left( \frac{1}{\sqrt{n}} \right) \). We can use, for example, the full-model least squares or the Wilcoxon to obtain \( \hat{\beta}_2^{(i)} \).

Secondly, obtain the following residuals with respect to \( x_2 \):

\[ e_2 = Y - \hat{\beta}_2^{(i)} x_2 \]

\[ = \alpha 1 + \beta_1 x_1 + (\beta_2 - \hat{\beta}_2^{(i)}) x_2 + \varepsilon \quad (3.2.1) \]
Note that (3.2.1) is just a contiguous regression model that takes the form of (3.1.2). Regress $e_2$ against $x_1$ using Spearman regression. Then from Lemma 3.1.3:

$$\hat{\beta}_1 = \beta_1 + (\beta_2 - \hat{\beta}_2^{(1)}) \left( x_1' W_{x_1} x_1 \right)^{-1} x_1' W_{x_1} x_2 + o_p \left( n^{-\frac{1}{2}} \right)$$

The residuals $e_2^* = e_2 - \hat{\beta}_1 x_1$ are now free of $\beta_1$. Now

$$e_2^* = \alpha 1 + \beta_1 x_1 + (\beta_2 - \hat{\beta}_2^{(1)}) x_2 + \varepsilon - \left( \beta_1 x_1 + (\beta_2 - \hat{\beta}_2^{(1)}) x_1 \left( x_1' W_{x_1} x_1 \right)^{-1} x_1' W_{x_1} x_2 \right) + o_p \left( n^{-\frac{1}{2}} \right)$$

$$= \alpha 1 + (\beta_2 - \hat{\beta}_2^{(1)}) z_2 + \varepsilon + o_p \left( n^{-\frac{1}{2}} \right), \quad (3.2.2)$$

where $z_2 = (I - K_{w_1}) x_2$ and $K_{w_1} = x_1 \left( x_1' W_{x_1} x_1 \right)^{-1} x_1' W_{x_1}$. We readily see that (3.2.2) is just a simple linear regression model with $e_2^* = e_2 - \hat{\beta}_1 x_1$ as the dependent variable and $z_2$ as the independent variable. If we regress $e_2^*$ versus $z_2$, the resulting slope estimate $\med \frac{e_{2j}^* - e_{2i}^*}{R(z_{2j}) - R(z_{2i})}$ is an estimate of $\beta_2 - \hat{\beta}_2^{(1)}$.

Finally, the Spearman estimate of $\beta_2$ is given by

$$\hat{\beta}_2^S = \left( \med \frac{e_{2j}^* - e_{2i}^*}{R(z_{2j}) - R(z_{2i})} \right) + \hat{\beta}_2^{(1)}.$$
We now show that $\hat{\beta}_2^*$ is a consistent estimate for $\beta_2$. We first state two important lemmas.

**Lemma 3.2.1:** Let $G_n(t)$ be the empirical cdf with mass points at $\frac{e_{2j}^* - e_{2i}^*}{z_{2j} - z_{2i}}$ and probabilities proportional to $R(z_{2j}) - R(z_{2i})$. Thus, $G_n^{-1}(\frac{1}{2}) = \text{median}_{R(z_{2j})-R(z_{2i})} \frac{e_{2j}^* - e_{2i}^*}{z_{2j} - z_{2i}}$.

If $e_2^* = \alpha 1 + \left( \beta_2 - \hat{\beta}_2^{(i)} \right) z_2 + \varepsilon$, then for every $\lambda > 0$,

$$P \left[ \left| G_n^{-1}(\frac{1}{2}) - \left( \beta_2 - \hat{\beta}_2^{(i)} \right) \right| > \lambda \right] \to 0$$

as $n \to \infty$.

**Proof:**

Note that $e_2^* = \alpha 1 + \left( \beta_2 - \hat{\beta}_2^{(i)} \right) z_2 + \varepsilon$ is just a simple linear regression model with $e_2^*$ as the dependent variable and $z_2$ as the independent variable. Hence, the desired result follows from the consistency of $GR$ simple linear regression estimates. ■

**Lemma 3.2.2:** Let $H_n(t)$ be the empirical cdf with mass points $\frac{(e_{2j}^* - e_{2i}^*) - o_p(n^{-1/2})}{z_{2j} - z_{2i}}$ and probabilities proportional to $R(z_{2j}) - R(z_{2i})$. For a fixed $t \in (0,1)$ and for
every \( \lambda > 0 \),

\[
P\left[ |G_n(t) - H_n(t)| > \lambda \right] \to 0.
\]

Proof:

\( G_n(t) \) has jumps proportional to \( R(z_{2j}) - R(z_{2i}) \) at mass points

\[
s_{ij} = \frac{e_{2j}^* - e_{2i}^*}{z_{2j} - z_{2i}}.
\]

\( H_n(t) \) also has the same weights as \( G_n(t) \) at mass points

\[
s_{ij}^* = \frac{(e_{2j}^* - e_{2i}^*) + o_p\left(n^{-\frac{1}{2}}\right)}{z_{2j} - z_{2i}}. \quad \text{Now, for all pairs } (i, j),
\]

\[
P\left[ |s_{ij} - s_{ij}^*| > \lambda_2 \right] = P\left[ o_p\left(n^{-\frac{1}{2}}\right) > \lambda_2 \right] \to 0
\]

since \( |G_n(t) - H_n(t)| > \lambda \) implies \( |s_{ij} - s_{ij}^*| > \lambda_2 \) for some \( \lambda_2 > 0 \). The result then follows. \( \blacksquare \)

Note that our estimate of \( \beta_2 - \hat{\beta}_2^{(1)} \) was derived under the assumption that

\[
e_2^* = \alpha 1 + (\beta_2 - \hat{\beta}_2^{(1)})z_2 + \varepsilon, \quad \text{not } e_2^* = \alpha 1 + (\beta_2 - \hat{\beta}_2^{(1)})z_2 + \varepsilon + o_p\left(n^{-\frac{1}{2}}\right). \quad \text{See (3.2.2).}
\]

The next lemma states that an estimate of \( \beta_2 - \hat{\beta}_2^{(1)} \) assuming (3.2.2) converges almost surely to an estimate assuming \( e_2^* = \alpha 1 + (\beta_2 - \hat{\beta}_2^{(1)})z_2 + \varepsilon. \)

**Lemma 3.2.2:** Let \( H_n(t) \) be the empirical cdf of \( \frac{(e_{2j}^* - e_{2i}^*)}{z_{2j} - z_{2i}} - o_p\left(n^{-\frac{1}{2}}\right) \). Then for
every $\lambda > 0$, 

$$P \left[ \left| G_n^{-1} \left( \frac{1}{2} \right) - H_n^{-1} \left( \frac{1}{2} \right) \right| > \lambda \right] \rightarrow 0$$

as $n \rightarrow \infty$.

Proof:

We have 

$$P \left[ \left| G_n^{-1} \left( \frac{1}{2} \right) - H_n^{-1} \left( \frac{1}{2} \right) \right| > \lambda \right] = P \left[ G_n^{-1} \left( \frac{1}{2} \right) > H_n^{-1} \left( \frac{1}{2} \right) + \lambda \right] + P \left[ H_n^{-1} \left( \frac{1}{2} \right) > G_n^{-1} \left( \frac{1}{2} \right) + \lambda \right].$$

Consider the first term on the right hand side of the equality above.

$G_n^{-1} \left( \frac{1}{2} \right) > H_n^{-1} \left( \frac{1}{2} \right) + \lambda$ implies that there exists a $t \left( t = \frac{G_n^{-1} \left( \frac{1}{2} \right) + H_n^{-1} \left( \frac{1}{2} \right)}{2} \right)$ such that

$$G_n(t) < \frac{1}{2} \text{ and } H_n(t) \geq \frac{1}{2}.$$

Thus

$$P \left[ G_n^{-1} \left( \frac{1}{2} \right) > H_n^{-1} \left( \frac{1}{2} \right) + \lambda \right] \leq P \left[ G_n(t) < \frac{1}{2} \text{ and } H_n(t) > \frac{1}{2} \right]$$

$$\leq P \left[ H_n(t) - G_n(t) > \lambda \right], \text{ for some } \lambda > 0$$

$$\rightarrow 0 \text{ as } n \rightarrow \infty$$

by Lemma 3.2.1. A similar argument applies for the second term.

We are now ready to prove the consistency of $\hat{\beta}_2^S$. The result is given in the next theorem.
Theorem 3.2.3: Under the assumptions of Lemma 3.2.1 and Lemma 3.2.1, we have

\[ \hat{\beta}_2^p \rightarrow \beta_2. \]

Proof:

For \( \lambda > 0 \),

\[
P \left[ \left| H_n^{-1} \left( \frac{1}{2} \right) - \left( \beta_2 - \hat{\beta}_2^{(1)} \right) \right| > \lambda \right]
\]

\[
\leq P \left[ \left| H_n^{-1} \left( \frac{1}{2} \right) - C_n^{-1} \left( \frac{1}{2} \right) \right| > \lambda \right] + P \left[ \left| H_n^{-1} \left( \frac{1}{2} \right) - \left( \beta_2 - \hat{\beta}_2^{(1)} \right) \right| > \lambda \right].
\]

Note that the first expression on the right hand side of the inequality goes

to 0 as \( n \rightarrow \infty \) by invoking Lemma 3.2.2; the second expression goes to 0 as

\( n \rightarrow \infty \) from Lemma 3.2.1. Hence,

\[
P \left[ \left| H_n^{-1} \left( \frac{1}{2} \right) - \left( \beta_2 - \hat{\beta}_2^{(1)} \right) \right| > \lambda \right] \rightarrow 0
\]

for \( \lambda > 0 \), i.e.,

\[
P \left[ \left| (H_n^{-1} \left( \frac{1}{2} \right) + \hat{\beta}_2^{(1)}) - \beta_2 \right| > \lambda \right] \rightarrow 0.
\]

The one-step residual method estimates obtained above are dependent on

the choice of the initial estimates of \( \beta_1 \) and \( \beta_2 \). Thus, if we use initial estimates
that are not robust to \(x\) outliers like the least squares and Wilcoxon estimates, the desired robustness might not be immediately achieved. We can proceed in an iterative manner by using the one-step estimates \(\hat{\beta}_1^s\) and \(\hat{\beta}_2^s\) as the initial estimates of \(\beta_1\) and \(\beta_2\), respectively, and repeating the entire process until convergence is reached. We refer to the resulting estimates as the fully-iterated residual method estimates.

### 3.3 The Bias Adjustment Method

The bias adjustment method is a procedure that is similar to the residual method in the sense that it makes use of the relationships between simple linear regression and multiple regression to obtain an estimate of \(\beta\). Lemma 3.1.3 has shown that if the true model is a multiple regression model but we fit separate simple linear regression models, then the resulting estimates are subject to bias.

This section attempts to obtain estimates that will correct for the bias.

To illustrate the method, suppose we regress \(e_1\) on \(x\) and \(e_2\) on \(x\) using Spearman regression. See (3.2.1) and (3.2.2). Using the arguments in Section 3.2, we can easily show that the resulting estimates \(\hat{\beta}_1\) and \(\hat{\beta}_2\) converge in
probability to $\beta_1 + (\beta_2 - \hat{\beta}_2^{(1)})(x_1'W_sx_1)^{-1}x_1'W_sx_2 + o_p\left(n^{-\frac{1}{2}}\right)$ and $\beta_2 + (\beta_1 - \hat{\beta}_1^{(1)})(x_2'W_sx_2)^{-1}x_2'W_sx_1 + o_p\left(n^{-\frac{1}{2}}\right)$, respectively, where $\hat{\beta}_1^{(1)}$ and $\hat{\beta}_2^{(1)}$ are initial estimates of $\beta_1$ and $\beta_2$ obtained from using, for example, least squares or Wilcoxon regression. Thus, $\hat{\beta}_1$ converges in probability to $\beta_1$ plus a bias term, and $\hat{\beta}_2$ converges in probability to $\beta_2$ plus a bias term.

Now, let $c_1 = (x_1'W_sx_1)^{-1}x_1'W_sx_2$ and $c_2 = (x_2'W_sx_2)^{-1}x_2'W_sx_1$. Note that

$$\hat{\beta}_1 = \beta_1 + c_1 (\beta_2 - \hat{\beta}_2^{(1)}) + o_p\left(n^{-\frac{1}{2}}\right)$$

$$\hat{\beta}_2 = \beta_2 + c_2 (\beta_1 - \hat{\beta}_1^{(1)}) + o_p\left(n^{-\frac{1}{2}}\right),$$

which are simply two equations in two unknowns. If we solve the above equations simultaneously, then we obtain

$$\beta_1 = \frac{\hat{\beta}_1 - c_1c_2 \hat{\beta}_2^{(1)} - c_1 (\hat{\beta}_2 - \hat{\beta}_2^{(1)})}{1 - c_1c_2}$$

$$\beta_2 = \frac{\hat{\beta}_2 - c_1c_2 \hat{\beta}_1^{(1)} - c_2 (\hat{\beta}_1 - \hat{\beta}_1^{(1)})}{1 - c_1c_2}.$$

We can therefore take

$$\hat{\beta}_1^g = \frac{\hat{\beta}_1 - c_1c_2 \hat{\beta}_1^{(1)} - c_1 (\hat{\beta}_2 - \hat{\beta}_2^{(1)})}{1 - c_1c_2}.$$
and

\[ \hat{\beta}_2^s = \frac{\hat{\beta}_2 - c_1c_2\hat{\beta}_2^{(i)} - c_2(\hat{\beta}_1 - \hat{\beta}_1^{(i)})}{1 - c_1c_2} \]

as the Spearman estimates of \( \beta_1 \) and \( \beta_2 \), respectively.

**Theorem 3.3.1:** The bias-adjustment method estimates \( \hat{\beta}_1^s \) and \( \hat{\beta}_2^s \) are consistent for \( \beta_1 \) and \( \beta_2 \), respectively.

**Proof:**

Note that from (3.3.1), \( \hat{\beta}_1 - c_1(\beta_2 - \hat{\beta}_2^{(i)}) \) converges to \( \beta_1 \) and \( \hat{\beta}_2 - c_2(\beta_1 - \hat{\beta}_1^{(i)}) \) converges to \( \beta_2 \). Hence, \( \hat{\beta}_1^s \) is converges to

\[ \frac{\beta_1 + c_1(\beta_2 - \hat{\beta}_2^{(i)}) - c_1c_2\hat{\beta}_1^{(i)} - c_2(\beta_2 + c_2(\beta_1 - \hat{\beta}_1^{(i)}) - \hat{\beta}_2^{(i)})}{1 - c_1c_2} = \beta_1 \]

and therefore, \( \hat{\beta}_1^s \) is consistent for \( \beta_1 \). We can apply a similar argument to prove the consistency of \( \hat{\beta}_2^s \) for \( \beta_2 \).

Similar to the residual method explained in the previous section, the bias-adjustment method requires the use of initial estimates which may be non-robust to the presence of \( x \) outliers. We refer to the estimates derived above as one-step bias-adjustment method estimates.
We can also use the one-step bias-adjustment estimates $\hat{\beta}_1^s$ and $\hat{\beta}_2^s$ as the new initial estimates of $\beta_1$ and $\beta_2$ and repeating the whole process until convergence is attained. This fully iterated procedure has the property of being independent of the choice of initial estimates.

It is recommended that one use this method with caution. Recall from Section 2.4 that

\[ x'Wx = \frac{1}{n} \sum_i (x_i - \bar{x}) \left( R(x_i) - \frac{n+1}{2} \right) \]  

and

\[ x'Wz = \frac{1}{n} \sum_i (z_i - \bar{z}) \left( R(x_i) - \frac{n+1}{2} \right). \]

Denote by $\hat{\beta}_{LS}(Y \mid x)$ the least squares estimate of the slope parameter in a simple linear regression model with $Y$ as the dependent variable and $x$ as the independent variable. It is easy to show that

\[ c_1 = \left( x_1'Wx_1 \right)^{-1} x_1'Wx_2 \]

\[ = \frac{\hat{\beta}_{LS}(x_2 \mid R(x_1))}{\hat{\beta}_{LS}(x_1 \mid R(x_1))} \]  

and

\[ c_2 = \left( x_2'Wx_2 \right)^{-1} x_2'Wx_1 \]

\[ = \frac{\hat{\beta}_{LS}(x_1 \mid R(x_2))}{\hat{\beta}_{LS}(x_2 \mid R(x_2))}. \]
We can readily see that the presence of leverage points can adversely affect the values of \( c_1 \) and \( c_2 \), and eventually the resulting Spearman estimates, owing to the fact that \( c_1 \) and \( c_2 \) are functions of least squares estimates. Thus, even if the Spearman residual method is computationally more rigorous than the Spearman bias adjustment method, the latter has the disadvantage of being less robust.

3.4 The Residual Method for Three Independent Variables

The purpose of this section is to derive the residual method estimates when there are three independent variables in a multiple regression model. We can actually generalize the method for any number of independent variables, but as we have outlined in Section 3.2, this method can be computationally rigorous.

Thus, we consider

\[
Y = \alpha 1 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.
\]

We first state the general version of Lemma 3.1.1 given in Naranjo and Hettmansperger (1994). Consider the model

\[
Y = \alpha 1 + X\beta + Z\gamma + \varepsilon,
\]  
(3.4.1)
where $Z$ is an $n \times q$ centered matrix of constants and $\gamma = \frac{\theta}{\sqrt{n}}$, for $\theta \neq 0$. Here, $X$ is an $n \times p$ matrix of constants. Since an intercept term is present, we can assume without loss of generality that $X$ is also a centered matrix.

**Lemma 3.4.1:** Suppose A1 to A10 are true for the augmented matrix $[X \, Z]$; then under (3.4.1),

$$
\hat{\beta}_{GR} = \beta + \frac{\sqrt{3}\tau}{n} (X'WX)^{-1} S(\beta) + (X'WX)^{-1} X'WZ \gamma + o_p(n^{-\frac{1}{2}})
$$

where $\tau = \frac{1}{\sqrt{12}} \int f^2(x)dx$.

As an initial step, obtain the least squares or Wilcoxon estimates of $\beta_1$, $\beta_2$, and $\beta_3$ and denote these as $\hat{\beta}_1^{(i)}$, $\hat{\beta}_2^{(i)}$, and $\hat{\beta}_3^{(i)}$, respectively. Then we have

$$
\beta_1 - \hat{\beta}_1^{(i)} = O\left(\frac{1}{\sqrt{n}}\right), \beta_2 - \hat{\beta}_2^{(i)} = O\left(\frac{1}{\sqrt{n}}\right), \text{ and } \beta_3 - \hat{\beta}_3^{(i)} = O\left(\frac{1}{\sqrt{n}}\right).
$$

Assume that A11 is true for each of the three independent variables in the model. Let $\beta'_1 = \beta_1 - \hat{\beta}_1^{(i)}$, $\beta'_2 = \beta_2 - \hat{\beta}_2^{(i)}$, and $\beta'_3 = \beta_3 - \hat{\beta}_3^{(i)}$ and consider the following sequences of models:

$$
e_1 = \alpha 1 + \beta'_1 x_1 + \beta'_2 x_2 + \beta'_3 x_3 + \epsilon
$$

$$
e_2 = \alpha 1 + \beta'_1 x_1 + \beta'_2 x_2 + \beta'_3 x_3 + \epsilon
$$
\[
e_i = \alpha 1 + \beta'_i x_i + \beta'_2 x_2 + \beta'_3 x_3 + \epsilon
\]

where \( e_1 = Y - \hat{\beta}'_2 x_2 - \hat{\beta}'_3 x_3 \), \( e_2 = Y - \hat{\beta}'_1 x_1 - \hat{\beta}'_3 x_3 \), and \( e_3 = Y - \hat{\beta}'_1 x_1 - \hat{\beta}'_2 x_2 \). From Lemma 3.4.1, the Spearman (or GR) estimates of \( \beta_1 \), \( \beta_2 \), and \( \beta_3 \) are given by

\[
\hat{\beta}_1 = \beta_1 + \beta'_2 (x'_i W_{x_1} x_i)^{-1} x'_i W_{x_1} x_2 + \beta'_3 (x'_i W_{x_1} x_i)^{-1} x'_i W_{x_1} x_3 + o_p \left( n^{-\frac{1}{2}} \right)
\]

\[
\hat{\beta}_2 = \beta_2 + \beta'_1 (x'_2 W_{x_2} x_2)^{-1} x'_2 W_{x_2} x_1 + \beta'_3 (x'_2 W_{x_2} x_2)^{-1} x'_2 W_{x_2} x_3 + o_p \left( n^{-\frac{1}{2}} \right) \quad (3.4.2)
\]

\[
\hat{\beta}_3 = \beta_3 + \beta'_1 (x'_3 W_{x_3} x_3)^{-1} x'_3 W_{x_3} x_1 + \beta'_2 (x'_3 W_{x_3} x_3)^{-1} x'_3 W_{x_3} x_2 + o_p \left( n^{-\frac{1}{2}} \right)
\]

Let \( e^*_k = e_k - \hat{\beta}_k x_k \), \( k = 1, 2, 3 \). Then we have

\[
e^*_1 = \alpha 1 + \beta^*_2 z_{12} + \beta^*_3 z_{23} + \epsilon \quad (3.4.3)
\]

\[
e^*_2 = \alpha 1 + \beta^*_1 z_{21} + \beta^*_3 z_{33} + \epsilon \quad (3.4.4)
\]

\[
e^*_3 = \alpha 1 + \beta^*_1 z_{31} + \beta^*_2 z_{32} + \epsilon \quad (3.4.5)
\]

where \( z_{kl} = (I - K_{W_{x_k}}) x_l \) and \( K_{W_{x_k}} = x_k (x'_k W_{x_k} x_k)^{-1} x'_k W_{x_k} \). From Lemma 3.1.1 (or Lemma 3.4.1), the following are Spearman estimates of \( \beta^*_1 \), \( \beta^*_2 \), and \( \beta^*_3 \):

\[
\hat{\beta}^*_1 = \beta^*_1 + \beta^*_2 (z_{21} W_{x_1} z_{x_1})^{-1} z'_{21} W_{x_{x_1}} z_{22} + o_p \left( \frac{1}{\sqrt{n}} \right), \quad \text{from (3.4.5)}
\]

\[
\hat{\beta}^*_2 = \beta^*_2 + \beta^*_3 (z_{12} W_{x_2} z_{z_2})^{-1} z'_{12} W_{x_{x_2}} z_{13} + o_p \left( \frac{1}{\sqrt{n}} \right), \quad \text{from (3.4.3)}
\]
\[
\hat{\beta}_3 = \beta_3 + \beta'_1(z_{23} W_{33} z_{23})^{-1} z_{23} W_{33} z_{21} + o_p\left(\frac{1}{\sqrt{n}}\right), \text{ from (3.4.4).}
\]

Now, let \( e_1^{**} = e_1^{*} - \hat{\beta}_2^* z_{12} \), \( e_2^{**} = e_2^{*} - \hat{\beta}_3^* z_{23} \), and \( e_3^{**} = e_3^{*} - \hat{\beta}_4^* z_{31} \). Then we have

\[
e_1^{**} = \alpha 1 + \beta_1^* z_1^* + \epsilon + o_p\left(\frac{1}{\sqrt{n}}\right)
\]

\[
e_2^{**} = \alpha 1 + \beta_2^* z_2^* + \epsilon + o_p\left(\frac{1}{\sqrt{n}}\right)
\]

\[
e_3^{**} = \alpha 1 + \beta_3^* z_3^* + \epsilon + o_p\left(\frac{1}{\sqrt{n}}\right)
\]

where \( z_1^* = (I - K W_{32})z_{13} \), \( z_2^* = (I - K W_{32})z_{21} \), and \( z_3^* = (I - K W_{33})z_{32} \). Note that the three equations above are just simple linear regression models. Hence, the following random variables converge in probability to \( \beta_1^* \), \( \beta_2^* \), and \( \beta_3^* \) are given by

\[
\hat{\beta}_1^* = \text{median}_{R(z_{32})} \frac{e^{**}_{2j} - e^{**}_{2i}}{z_{2j}^* - z_{2i}^*}
\]

\[
\hat{\beta}_2^* = \text{median}_{R(z_{33})} \frac{e^{**}_{3j} - e^{**}_{3i}}{z_{3j}^* - z_{3i}^*}
\]

and
\[ \hat{\beta}_3^* = \text{median}_{R(i)-R(n_i)} \frac{\tilde{e}_{ii}^{**} - e_{ii}^{**}}{z_{ij} - z_{ii}}. \]

See the arguments in Section 3.2. Since \( \beta_1^* = \beta_1 - \hat{\beta}_1^{(1)} \), \( \beta_2^* = \beta_2 - \hat{\beta}_2^{(1)} \), and \( \beta_3^* = \beta_3 - \hat{\beta}_3^{(1)} \), then the (one-step) residual method estimates of \( \beta_1 \), \( \beta_2 \), and \( \beta_3 \) are

\[ \hat{\beta}_1^g = \hat{\beta}_1^* + \hat{\beta}_1^{(1)} \]
\[ \hat{\beta}_2^g = \hat{\beta}_2^* + \hat{\beta}_2^{(1)} \]
and

\[ \hat{\beta}_3^g = \hat{\beta}_3^* + \hat{\beta}_3^{(1)} \]

respectively. The above estimates can be shown to be consistent. See Theorems 3.2.1 and 3.2.2.

Again, as what was stated in Section 3.2, we might not be able to immediately achieve the desired robustness if we use least squares or Wilcoxon regression to obtain the initial estimates since these estimates are not robust against \( x \) outliers. Hence, we can again use the estimates above as our initial estimates for \( \beta_1 \), \( \beta_2 \), and \( \beta_3 \) and then proceed in an iterative manner until convergence to obtain the final estimates.
3.5 The Bias Adjustment Method for Three Independent Variables

We shall now extend the bias-adjustment method to regression models with three independent variables. Note that the bias-adjustment method estimates can also be obtained for regression models with more than three independent variables but the computational requirements would increase drastically. Recall in (3.4.2) that, if the true model is a multiple regression model but we fit separate simple linear regression models, we obtain the following estimates for $\beta_1$, $\beta_2$, and $\beta_3$:

$$\hat{\beta}_1 = \beta_1 + \beta_2^* (x_1' W_{x_1} x_1)^{-1} x_1' W_{x_1} x_2 + \beta_3^* (x_1' W_{x_1} x_1)^{-1} x_1' W_{x_1} x_3 + o_p \left( \frac{1}{\sqrt{n}} \right)$$

$$\hat{\beta}_2 = \beta_2 + \beta_1^* (x_2' W_{x_2} x_2)^{-1} x_2' W_{x_2} x_1 + \beta_3^* (x_2' W_{x_2} x_2)^{-1} x_2' W_{x_2} x_3 + o_p \left( \frac{1}{\sqrt{n}} \right)$$

$$\hat{\beta}_3 = \beta_3 + \beta_1^* (x_3' W_{x_3} x_3)^{-1} x_3' W_{x_3} x_1 + \beta_2^* (x_3' W_{x_3} x_3)^{-1} x_3' W_{x_3} x_2 + o_p \left( \frac{1}{\sqrt{n}} \right).$$

Thus, $\hat{\beta}_k$ is converges in probability to $\beta_k$ plus a bias term, $k = 1, 2, 3$.

Since the equations above are simply three equations in three unknowns, we can solve the equations simultaneously to obtain the Spearman estimates of $\beta_1$, $\beta_2$, and $\beta_3$. Although these estimates are consistent (see Theorem 3.3.1), we again
remark that these estimates are less robust than the Spearman residual method estimates, since the estimates would be functions of constants that are not robust to $x_1$, $x_2$, and $x_3$. 
CHAPTER 4

MATRIX WEIGHTS

In Chapter 3, we proposed methods of estimating the multiple regression parameter $\beta$ by exploiting the relationships between multiple regression and simple linear regression. This is in lieu of developing a concept for multidimensional rank, an intuitive step in obtaining a Spearman estimate of $\beta$.

We discuss in this chapter another scheme of obtaining a Spearman estimate of $\beta$. This method entails assigning separate weights to the individual components of each independent variable in the multiple regression model. The weights are separate in the sense that they are computed independently for each predictor variable. Recall that the simple linear regression estimate for the slope parameter $\beta$ is the point $\hat{\beta}_s$ in the parameter space that satisfies $S(\hat{\beta}_s) = 0$, where the estimating function $S(\beta)$ is given by

$$S(\beta) = \sum_{i<j} b_{ij}(x_j - x_i) \text{sgn}(e_j - e_i).$$
Here, \( b_0 = \frac{1}{n} \left( \frac{R(x_j) - R(x_i)}{x_j - x_i} \right) \) and the \( e_j = Y_j - \beta x_j \) are the full-model residuals.

Note that the weight \( b_0 \) is a function of the independent variable \( x \).

Consider now the multiple regression model with \( p \) parameters, i.e.,

\[
Y = \alpha 1 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon.
\]

and suppose we modify the above estimating function to obtain the following gradient functions:

\[
S_1(\beta) = \sum_{i<j} b_{ij}^{(1)}(x_{ij} - x_{ii}) \text{sgn}(e_j - e_i)
\]
\[
S_2(\beta) = \sum_{i<j} b_{ij}^{(2)}(x_{ij} - x_{ii}) \text{sgn}(e_j - e_i)
\]
\[
\vdots
\]
\[
S_p(\beta) = \sum_{i<j} b_{ij}^{(p)}(x_{ij} - x_{ii}) \text{sgn}(e_j - e_i)
\]

where \( e_j = Y_j - \beta_1 x_{1j} - \beta_2 x_{2j} - \ldots - \beta_p x_{pj} \) and \( b_{ij}^{(k)} = \frac{1}{n} \left( \frac{R(x_{kj}) - R(x_{ki})}{x_{kj} - x_{ki}} \right) \) are weights based on \( x_k, k = 1, 2, \ldots, p \). Hence, a separate weight is assigned to each component of \( X = [x_1 \ x_2 \ \ldots \ \ x_p] \). Therefore, we can define a point \( \hat{\beta}_s = \begin{bmatrix} \hat{\beta}_1^s & \hat{\beta}_2^s & \ldots & \hat{\beta}_p^s \end{bmatrix} \) in the parameter space that satisfies \( S(\hat{\beta}_s) = 0 \), where \( S(\beta) \) is given by
\[ S(\beta) = \begin{bmatrix} S_1(\beta) & S_2(\beta) & \cdots & S_p(\beta) \end{bmatrix} \]

\[ = \sum_{i<j} \begin{bmatrix} b_0^{(1)} \\ b_0^{(2)} \\ \vdots \\ b_0^{(p)} \end{bmatrix} (x_j - x_i) \text{sgn}(e_j - e_i) \]

and \( x_j = [x_{1j}, x_{2j}, \ldots, x_{pj}] \). The matrix weights estimate of \( \beta \) is therefore given by \( \hat{\beta}_S \).
5.1 Tied Ranks

Consider the simple linear regression model in Chapter 2 where several of the values of the independent variable \( x \) are not distinct. Without loss of generality, assume that \( x_1 \leq x_2 \leq \ldots \leq x_n \); they are already assumed to be not all equal. Denote by

\[
m = \sum_{i<j} \text{sgn}(x_j - x_i)
\]

(5.1.1)

the number of positive differences \( x_j - x_i, 1 \leq i < j \leq n \). Recall the estimating function in (2.2.1) which we state here for convenience:

\[
S(\beta) = \sum_{i<j} b_{ij}(x_j - x_i)\text{sgn}(e_j - e_i).
\]

If we denote the weights \( b_{ij} \) as

\[
b_{ij} = \begin{cases} 
\frac{1}{n} & \text{if } x_j \neq x_i, \\
0 & \text{otherwise}
\end{cases}
\]

then similar to the arguments used in Chapter 2, we obtain \( \hat{\beta}_s \) as the weighted
median of the m pairwise slopes \( s_{ij} = \frac{Y_j - Y_i}{x_j - x_i} \), with weights proportional to \( R(x_j) - R(x_i) \), i.e.,

\[
\hat{\beta}_s = \text{median} \frac{Y_j - Y_i}{x_j - x_i}.
\]

Since there are ties among the x observations, the rank that we assign to each of the observations in a tied x group is the arithmetic average of the integer ranks that are associated with the group.

5.2 The Two-Sample Problem

In this section, we discuss a specific application of Section 5.1. Let \( x = \begin{bmatrix} x_1 & x_2 & \cdots & x_{n_1} & x_{n_1+1} & x_{n_1+2} & \cdots & x_{n_1+n_2} \end{bmatrix}' \), where \( n = n_1 + n_2 \), \( n_i < n \), \( x_i = 0 \) if \( i = 1, \ldots, n_1 \), and \( x_i = 1 \) if \( i = n_1 + 1, \ldots, n \). Then we have a regression problem that is essentially the two-sample regression problem. The model is

\[ Y_i = \alpha + \beta x_i + \Box, \quad i = 1, \ldots, n \]

where the \( \Box \) are iid \( F \), continuous.

It should be easy to see that \( m = n_1n_2 \). See (5.1.1). Following the assignment of ranks in the previous section, the rank assigned to each
observation in the first tied \( x \) group \((x_1, x_2, ..., x_{n_1})\) is given by \( \frac{n_1 + 1}{2} \); the rank assigned to each observation in the second tied \( x \) group \((x_{n_1+1}, x_{n_1+2}, ..., x_{n_1+n_2})\) is given by \( n_1 + \frac{n_2 + 1}{2} \). Thus, \( R(x_j) - R(x_i) \) is just a constant for \( i = 1, ..., n_1 \) and \( j = n_1 + 1, ..., n \). Therefore, from the previous section, the Spearman estimator of \( \beta \) is

\[
\hat{\beta}_\varphi = \text{median} \left( Y_j - Y_i \right),
\]

for \( i = 1, ..., n_1 \) and \( j = n_1 + 1, ..., n \). Note that the above expression is simply the Hodges-Lehmann estimate of the shift for the two-sample problem (Hodges and Lehmann, 1963).

5.3 Matrix Weights Method for Design Matrices

Let us now consider implementing the matrix weights method discussed in the previous chapter when the independent variable is a design matrix. Suppose we have \( X = \begin{bmatrix} 1_n & x_1 & x_2 \end{bmatrix} \) where \( x_1 = \begin{bmatrix} 0_{n_1} & 1_{n_2}' & 0_{n_3}' \end{bmatrix} \), \( x_2 = \begin{bmatrix} 0_{n_1}' & 0_{n_2}' & 1_{n_3}' \end{bmatrix} \), \( 1_n \)
is a vector of 1’s with \( n \) elements, and \( n_1 + n_2 + n_3 = n \). Let \( \beta = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 \end{bmatrix}' \).

The regression model for comparison of group means is given by
\[ Y = X\beta + \varepsilon \]

where \( \varepsilon \sim F, \) continuous and unknown. Note that the full-model residuals \( e_i = \)

\[ Y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}, \]

can be expressed as

\[
e_i = \begin{cases} 
Y_i - \beta_0 & \text{if } i = 1, \ldots, n_1 \\
Y_i - \beta_0 - \beta_1 & \text{if } i = n_1 + 1, \ldots, n_1 + n_2, \\
Y_i - \beta_0 - \beta_2 & \text{if } i = n_1 + n_2 + 1, \ldots, n 
\end{cases}
\]

so that

\[
e_j - e_i = Y_j - Y_i \quad \text{if } 1 \leq i < j \leq n_1; \\
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad n_1 + 1 \leq i < j \leq n_1 + n_2; \\
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad n_1 + n_2 + 1 \leq i < j \leq n
\]

\[
= (Y_j - Y_i) - \beta_1 \quad \text{if } i = 1, \ldots, n_1; j = n_1 + 1, \ldots, n_1 + n_2 \\
= (Y_j - Y_i) - \beta_2 \quad \text{if } i = 1, \ldots, n_2; j = n_1 + n_2 + 1, \ldots, n \\
= (Y_j - Y_i) - (\beta_2 - \beta_1) \quad \text{if } i = n_1 + 1, \ldots, n_1 + n_2; j = n_1 + n_2 + 1, \ldots, n.
\]

Again, since there are ties among the \( x \) observations, the rank that we assign to each of the observations in a tied \( x \) group is the mean of the integer ranks that are associated with the group. Hence, for the given design matrix above, we have
\[ R(x_h) - R(x_u) = 0 \quad \text{if} \quad 1 \leq i < j \leq n_i; \]
\[ n_i + 1 \leq i < j \leq n_i + n_2; \]
\[ n_i + n_2 + 1 \leq i < j \leq n; \]
\[ i = 1, \ldots, n_i; \quad j = n_i + n_2 + 1, \ldots, n \]
\[ = \frac{n}{2} \quad \text{if} \quad i = 1, \ldots, n_i; \quad j = n_i + 1, \ldots, n_1 + n_2 \]
\[ = -\frac{n}{2} \quad \text{if} \quad i = n_i + 1, \ldots, n_1 + n_2; \quad j = n_1 + n_2 + 1, \ldots, n \]

Using the results in Chapter 4, we obtain 
\[ S(p) = \begin{bmatrix} S_1(p) & S_2(p) \end{bmatrix} \]
where

\[ S_1(p) = \frac{1}{2} \left\{ \sum_{(i,j) \in A} \text{sgn} \left( Y_j - Y_i \right) - \text{sgn} \left( Y_j - Y_i - (\beta_2 - \beta_1) \right) \right\} \]

and

\[ S_2(p) = \frac{1}{2} \left\{ \sum_{(i,j) \in B} \text{sgn} \left( Y_j - Y_i \right) - \text{sgn} \left( Y_j - Y_i - (\beta_2 - \beta_1) \right) \right\}. \]
Here, \( A = \{1, \ldots, n_1\} \times \{n_1+1, \ldots, n_1+n_2\} \), \( B = \{n_1+1, \ldots, n_1+n_2\} \times \{n_1+n_2+1, \ldots, n\} \), and \( C = \{1, \ldots, n_1\} \times \{n_1+n_2+1, \ldots, n\} \). Hence, the Spearman estimate of \( \beta \) is \( \hat{\beta}_s \), where \( \hat{\beta}_s \) satisfies \( S(\hat{\beta}_s) = 0 \).
6.1 Examples for Simple Linear Regression

This section will present two examples implementing simple Spearman rank regression. The first example compares the performance of Spearman rank regression with the least squares method and the Wilcoxon using a data set that is well-behaved, but we change one of the $x$-values to study the effect of leverage values on the estimate. The second example aims to achieve the purpose of the first example, but it uses a data set with actual influential observations.

Example 6.1.1: Pilot Plant Data

Table 6.1.1 contains the Pilot-Plant data from Rousseeuw and Leroy (1987, p. 22), originally cited by Daniel and Wood (1971). The dependent variable ($Y$) used is the acid content determined by titration, and the predictor variable ($x$) is the organic acid content determined by extraction and weighing.
We now fit a simple linear regression model for the data and estimate the slope parameter using least squares, Wilcoxon, and Spearman rank regression. The scatter diagram given in Figure 6.1.1 suggests a very strong linear relationship between \( x \) and \( Y \). Thus, we should expect only marginal differences among the three estimates obtained. The results are given in Table 6.1.2.

Now, suppose that one of the \( x \) values has been erroneously recorded. Suppose the first observation \( x_1 \) was recorded as 1230, rather than 123. We now have a data point that behaves very differently from the main cluster of points. Again, we fit a simple linear regression model and estimate the regression parameters.

Figure 6.1.2 depicts the least squares and Wilcoxon lines as almost coinciding and were pulled towards the direction of the outlying value. However, the Spearman method gave more importance to the main cluster of points and downweighted the outlying value. Thus, the estimate was virtually unchanged from the previous estimate where the uncontaminated data set was used.
Table 6.1.1  Pilot-Plant Data Set

<table>
<thead>
<tr>
<th>Observation</th>
<th>Extraction (x)</th>
<th>Titration (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>123</td>
<td>76</td>
</tr>
<tr>
<td>2</td>
<td>109</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>104</td>
<td>71</td>
</tr>
<tr>
<td>5</td>
<td>57</td>
<td>55</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td>48</td>
</tr>
<tr>
<td>7</td>
<td>44</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>66</td>
</tr>
<tr>
<td>9</td>
<td>16</td>
<td>41</td>
</tr>
<tr>
<td>10</td>
<td>28</td>
<td>43</td>
</tr>
<tr>
<td>11</td>
<td>138</td>
<td>82</td>
</tr>
<tr>
<td>12</td>
<td>105</td>
<td>68</td>
</tr>
<tr>
<td>13</td>
<td>159</td>
<td>88</td>
</tr>
<tr>
<td>14</td>
<td>75</td>
<td>58</td>
</tr>
<tr>
<td>15</td>
<td>88</td>
<td>64</td>
</tr>
<tr>
<td>16</td>
<td>164</td>
<td>88</td>
</tr>
<tr>
<td>17</td>
<td>169</td>
<td>89</td>
</tr>
<tr>
<td>18</td>
<td>167</td>
<td>88</td>
</tr>
<tr>
<td>19</td>
<td>149</td>
<td>84</td>
</tr>
<tr>
<td>20</td>
<td>167</td>
<td>88</td>
</tr>
</tbody>
</table>
Figure 6.1.1  Scatter Diagram of the Pilot-Plant Data

Table 6.1.2  Parameter Estimates for the Uncontaminated Pilot-Plant Data

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Intercept</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Squares</td>
<td>35.4583</td>
<td>0.32161</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>35.3548</td>
<td>0.32258</td>
</tr>
<tr>
<td>Spearman</td>
<td>35.6271</td>
<td>0.32042</td>
</tr>
</tbody>
</table>

Table 6.1.3  Parameter Estimates for the Contaminated Pilot-Plant Data

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Intercept</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Squares</td>
<td>65.5815</td>
<td>0.01916</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>65.1571</td>
<td>0.01798</td>
</tr>
<tr>
<td>Spearman</td>
<td>35.9473</td>
<td>0.31408</td>
</tr>
</tbody>
</table>
Example 6.1.1 is a useful illustration of how an outlying $x$ value can obscure the results for the least squares method and even for the Wilcoxon. As mentioned before, the Wilcoxon is robust to $Y$ outliers but it loses its robustness in the presence of $x$ outliers.

The next example is uses a data set with actual influential observations. It was taken from Rousseeuw and Leroy (1987, p. 27)
Example 6.1.2: Table 6.1.4 gives the 1832 census data on livestock population $(x)$ and crop production $(y)$ of 21 California missions as cited by Davis (2002). The corresponding scatter diagram is in Figure 6.1.3.

The San Luis Rey mission is of special interest in Figure 6.1.3., which clearly indicates that the data from the said mission is a leverage point. We shall now fit a simple linear regression model for the above data using the least squares method, the Wilcoxon, and Spearman rank regression.

The results in Table 6.1.5 indicate a disagreement in the results. This is expected, since, as illustrated in the previous example, the presence of a leverage point can greatly obscure the estimation process. Let us evaluate the estimates presented in Table 6.1.5 using the Figure 6.1.4. We notice right away that the Spearman estimates fit the bulk of the data nicely. However, the Least Squares and Wilcoxon fits were pulled towards the direction of the San Luis Rey data point.
<table>
<thead>
<tr>
<th>Mission</th>
<th>Livestock Population</th>
<th>Crop Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Diego de Alcala</td>
<td>18200</td>
<td>158675</td>
</tr>
<tr>
<td>San Carlos Borromeo</td>
<td>5818</td>
<td>103847</td>
</tr>
<tr>
<td>San Antonio de Padua</td>
<td>17491</td>
<td>84933</td>
</tr>
<tr>
<td>San Gabriel Arcangel</td>
<td>26342</td>
<td>233695</td>
</tr>
<tr>
<td>San Luis Obispo</td>
<td>8822</td>
<td>128751</td>
</tr>
<tr>
<td>San Francisco de Asis</td>
<td>9518</td>
<td>67117</td>
</tr>
<tr>
<td>San Juan de Capistrano</td>
<td>16270</td>
<td>83923</td>
</tr>
<tr>
<td>Santa Clara de Asis</td>
<td>20320</td>
<td>98356</td>
</tr>
<tr>
<td>San Buenaventura</td>
<td>7616</td>
<td>135303</td>
</tr>
<tr>
<td>Santa Barbara</td>
<td>5707</td>
<td>151143</td>
</tr>
<tr>
<td>La Purisima Concepcion</td>
<td>13985</td>
<td>191014</td>
</tr>
<tr>
<td>Santa Cruz</td>
<td>9236</td>
<td>58072</td>
</tr>
<tr>
<td>Nuestra Senora de la Soledad</td>
<td>12508</td>
<td>64808</td>
</tr>
<tr>
<td>San Jose</td>
<td>24180</td>
<td>222809</td>
</tr>
<tr>
<td>San Juan Bautista</td>
<td>12333</td>
<td>69577</td>
</tr>
<tr>
<td>San Miguel Arcangel</td>
<td>12970</td>
<td>105049</td>
</tr>
<tr>
<td>San Fernando Rey</td>
<td>9060</td>
<td>95172</td>
</tr>
<tr>
<td>San Luis Rey</td>
<td>57330</td>
<td>92656</td>
</tr>
<tr>
<td>Santa Ines</td>
<td>9860</td>
<td>179925</td>
</tr>
<tr>
<td>San Rafael Arcangel</td>
<td>5492</td>
<td>74609</td>
</tr>
<tr>
<td>San Francisco de Solano</td>
<td>5063</td>
<td>10991</td>
</tr>
</tbody>
</table>
Figure 6.1.3  Scatter Diagram of the California Mission Data

Table 6.1.5  Simple Linear Regression Fits (With the San Luis Rey Data Point)

<table>
<thead>
<tr>
<th>Method</th>
<th>Intercept</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Squares</td>
<td>99034.4</td>
<td>1.07329</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>89709.8</td>
<td>0.602897</td>
</tr>
<tr>
<td>Spearman</td>
<td>62748.5</td>
<td>3.261409</td>
</tr>
</tbody>
</table>
6.2 Simulations for Simple Linear Regression

The following are intended to provide empirical verification on the estimates of efficiency of the Spearman estimates with respect to the least squares and Wilcoxon estimates. We illustrate the computation of the efficiency using the contaminated normal distribution. The probability density function of the contaminated normal distribution consists of mixing the standard normal pdf
with a normal pdf having mean zero and variance $\delta^2 > 1$. For $0 < \xi < 1$, the pdf can be written as $f_\xi(x) = (1 - \xi)\phi(x) + \xi\delta^{-2}\phi(\delta^{-1}x)$ with $\sigma_i^2 = 1 + \xi(\delta^2 - 1)$. This distribution has heavier tails than the standard normal distribution and can be used to model data contamination. We can think of $\xi$ as the fraction of the data contaminated.

The models simulated are from

$$Y = 2 + 5x + \varepsilon$$

where $x \sim (1 - \xi_1) N(0,1) + \xi_1 N(0, 100)$ and $\varepsilon \sim (1 - \xi_2) N(0,1) + \xi_2 N(0, 100)$. The sample size used is 50. For each choice of $\xi_1$ and $\xi_2$ we considered 300 samples and computed the MSEs of the resulting estimates for $\hat{\beta}_s$.

The first set of models simulated is the case where $\xi_1 = 0, 0.06, 0.12$ and $\xi_2 = 0$. For $\xi_1 = 0$, we would like to examine the behavior of the three slope estimators under normal errors and a well-behaved data set. For $\xi_1 = 0.06, 0.12$, we would like to examine the behavior of the three slope estimates for models with $x$ values that have high variability.

In the second situation we generate outliers in the $Y$ direction. For this
purpose, we contaminate the distribution of the errors. We have $\xi_1 = 0$ and $\xi_2 = 0.06, 0.12$.

Finally, we generate a model whose $x$ values have high variability and contaminated errors. Thus, $\xi_1 = 0.06, 0.12$ and $\xi_2 = 0.06, 0.12$.

We note at this point that after the $x$ values are generated, they are held fixed throughout the simulations.

Scatter plots for the above models for selected values of $\xi_1$ and $\xi_2$ are given in Figure 6.2.1 (Panels (a) to (d)). The results of the simulations are shown in Tables 6.2.1 to Table 6.2.3.

The results obtained in Table 6.2.1 are not totally unexpected. It is known that under normal errors, the least squares estimate is the most efficient estimator. The simulations have shown that the Spearman estimator has also high efficiency (about 89% with respect to least squares) if there is no contamination in the design, but it loses its efficiency if the errors are normal and there are moderate amounts of contamination in the design. Also, as expected, the Wilcoxon estimator displayed efficiency estimates that are close to its
theoretical efficiency relative to least squares (approximately 0.954).

It is also worth taking note that the theoretical values of the efficiency of the Spearman estimates relative to the Wilcoxon have been verified by the simulations in Table 6.2.1.

Table 6.2.2 depicted another expected result, i.e., that least squares have inflated MSEs in the presence of $Y$ outliers. The Wilcoxon estimates, as expected, exhibited very high efficiency compared to least squares. However, the simulations showed that the Spearman estimates are even more efficient than the Wilcoxon estimates if there are moderate to high amounts of contamination in the error. The Spearman estimate should therefore be given consideration as an alternative to least squares and Wilcoxon estimators.

Simulation results with $x$ outlying values are given in Tables 6.2.3 and 6.2.4. They indicated that the Spearman estimates are more efficient than least squares, although the Wilcoxon estimates are still more efficient. We have to note, however, that all leverage values considered in this simulation are good leverage values. If there is a presence of bad leverage values, it already goes
without saying that the Spearman estimates will be the most efficient. Examples 6.1.1 and 6.1.2 depicted the sensitivity of the least squares and Wilcoxon estimates to bad leverage points. The least squares and Wilcoxon estimates were pulled toward the direction of the bad leverage points, but the Spearman estimates remained resistant. In these circumstances, the least squares and Wilcoxon estimates will have inflated MSEs and this would render them inefficient.
Figure 6.2.1. Scatter Diagrams for Simulated Data
Table 6.2.1  Simulation Results for the First Set of Models
(Errors are Normally Distributed)

<table>
<thead>
<tr>
<th>$\xi$</th>
<th>0</th>
<th>0.06</th>
<th>0.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE (LS)</td>
<td>0.01944</td>
<td>0.00877</td>
<td>0.00264</td>
</tr>
<tr>
<td>MSE (Wilcoxon)</td>
<td>0.02074</td>
<td>0.00918</td>
<td>0.00271</td>
</tr>
<tr>
<td>MSE (Spearman)</td>
<td>0.02182</td>
<td>0.01553</td>
<td>0.00470</td>
</tr>
<tr>
<td>$\text{eff}(\text{Spearman} \mid \text{LS})$</td>
<td>0.89113</td>
<td>0.56470</td>
<td>0.56265</td>
</tr>
<tr>
<td>$\text{eff}(\text{Spearman} \mid \text{Wilcoxon})$</td>
<td>0.95039</td>
<td>0.59090</td>
<td>0.57587</td>
</tr>
<tr>
<td>$\text{eff}(\text{Wilcoxon} \mid \text{LS})$</td>
<td>0.93764</td>
<td>0.95566</td>
<td>0.97704</td>
</tr>
<tr>
<td>$\rho^2$ (theoretical $\text{eff}(\text{Sp} \mid \text{Wil})$)</td>
<td>0.97106</td>
<td>0.69230</td>
<td>0.51195</td>
</tr>
</tbody>
</table>

Table 6.2.2  Simulation Results for the Second Set of Models
(Y outliers are present)

<table>
<thead>
<tr>
<th>$\xi$</th>
<th>0.06</th>
<th>0.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE (LS)</td>
<td>0.32565</td>
<td>0.46658</td>
</tr>
<tr>
<td>MSE (Wilcoxon)</td>
<td>0.03499</td>
<td>0.04511</td>
</tr>
<tr>
<td>MSE (Spearman)</td>
<td>0.03090</td>
<td>0.04293</td>
</tr>
<tr>
<td>$\text{eff}(\text{Spearman} \mid \text{LS})$</td>
<td>10.53953</td>
<td>10.86931</td>
</tr>
<tr>
<td>$\text{eff}(\text{Spearman} \mid \text{Wilcoxon})$</td>
<td>1.13257</td>
<td>1.05096</td>
</tr>
<tr>
<td>$\text{eff}(\text{Wilcoxon} \mid \text{LS})$</td>
<td>9.30586</td>
<td>10.34230</td>
</tr>
</tbody>
</table>
Table 6.2.3  Simulation Results for the Third Set of Models
(x outliers are present, $\xi_1 = 0.06$)

<table>
<thead>
<tr>
<th></th>
<th>$\xi_2$</th>
<th>0.06</th>
<th>0.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE (LS)</td>
<td>0.03322</td>
<td>0.05416</td>
<td></td>
</tr>
<tr>
<td>MSE (Wilcoxon)</td>
<td>0.00896</td>
<td>0.01447</td>
<td></td>
</tr>
<tr>
<td>MSE (Spearman)</td>
<td>0.01507</td>
<td>0.02059</td>
<td></td>
</tr>
<tr>
<td>$\text{eff}(\text{Spearman} \mid \text{LS})$</td>
<td>2.20391</td>
<td>2.63055</td>
<td></td>
</tr>
<tr>
<td>$\text{eff}(\text{Spearman} \mid \text{Wilcoxon})$</td>
<td>0.59415</td>
<td>0.70267</td>
<td></td>
</tr>
<tr>
<td>$\text{eff}(\text{Wilcoxon} \mid \text{LS})$</td>
<td>3.70933</td>
<td>3.74367</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2.4  Simulation Results for the Third Set of Models
(x outliers are present, $\xi_1 = 0.12$)

<table>
<thead>
<tr>
<th></th>
<th>$\xi_2$</th>
<th>0.06</th>
<th>0.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE (LS)</td>
<td>0.02186</td>
<td>0.00774</td>
<td></td>
</tr>
<tr>
<td>MSE (Wilcoxon)</td>
<td>0.00543</td>
<td>0.00268</td>
<td></td>
</tr>
<tr>
<td>MSE (Spearman)</td>
<td>0.01021</td>
<td>0.00505</td>
<td></td>
</tr>
<tr>
<td>$\text{eff}(\text{Spearman} \mid \text{LS})$</td>
<td>2.13983</td>
<td>1.53155</td>
<td></td>
</tr>
<tr>
<td>$\text{eff}(\text{Spearman} \mid \text{Wilcoxon})$</td>
<td>0.53117</td>
<td>0.53048</td>
<td></td>
</tr>
<tr>
<td>$\text{eff}(\text{Wilcoxon} \mid \text{LS})$</td>
<td>4.02855</td>
<td>2.88709</td>
<td></td>
</tr>
</tbody>
</table>
6.3 Examples for Multiple Regression

We now give examples to illustrate Spearman regression in multiple regression. All the data used in this section are simulated data, since these data possess known properties that are otherwise unknown if real data are used. Knowledge of these properties would enable us to evaluate the performance of the parameters in an objective manner.

Example 6.3.1: A well-behaved data set

For this discussion, we use an artificial data set simulated from the model

$$Y = 2 + 5x_1 + 1000x_2 + \varepsilon$$

(6.3.1)

where $\varepsilon \sim N(0,1)$. The values of $x_1$ and $x_2$ were generated from the $N(0, 36)$ and $N(0, 16)$ distributions, respectively. The sample size used is 100. The objective of this example is to evaluate the performance of the proposed estimators against the least squares and Wilcoxon estimators using a well-behaved data set. The results are given in Table 6.3.1.

The results above are not unexpected. If the data set does not contain any outlying values, we expect the estimators to be in total agreement, as
implied by the results in Chapter 3. Note that the values of the estimates coincide with the true values of the regression parameters. Hence, under this scenario, the question will boil down on the efficiency of the estimates. Estimates of efficiency are considered in Section 6.4.

Table 6.3.1  Estimates of Multiple Regression Regression Parameters

<table>
<thead>
<tr>
<th>Method</th>
<th>Intercept</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Squares</td>
<td>1.946</td>
<td>5.0043</td>
<td>1000.0020</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>1.956</td>
<td>5.0069</td>
<td>999.9962</td>
</tr>
<tr>
<td>Spearman (Residual Method, One-Step Estimates)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td>1.956</td>
<td>5.0064</td>
<td>999.9995</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td>1.957</td>
<td>5.0067</td>
<td>999.9993</td>
</tr>
<tr>
<td>Spearman (Bias-Adjustment Method, One-Step Estimates)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td>1.956</td>
<td>5.0059</td>
<td>1000.0000</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td>1.958</td>
<td>5.0068</td>
<td>999.9994</td>
</tr>
<tr>
<td>Residual Method Initial Values</td>
<td>1.957</td>
<td>5.0067</td>
<td>999.9994</td>
</tr>
<tr>
<td>Spearman (Matrix Weights)</td>
<td>1.957</td>
<td>5.0067</td>
<td>999.9994</td>
</tr>
</tbody>
</table>
Example 6.3.2: Example 6.3.1 continued

This example will continue to use data simulated from model (6.3.1). However, we modify the data so that the first 15 of them are leverage points, i.e., outlying in \( x \) space. Of the 15 leverage points, the first 10 of them do not follow the model, hence we refer to them as *bad leverage points*; the rest of the leverage points follow the model, and hence are referred to as *good leverage points*.

Table 6.3.2 gives us the estimates of the regression parameters using the different regression schemes. As expected, the least squares and Wilcoxon regression methods yielded estimates that are extremely far from the true values. The Spearman fully-iterated residual method and the matrix weight method proved to be robust against the presence of the bad leverage points, as they yielded estimates that are close to the actual values of the parameters. The Spearman one-step residual method estimates also yielded values close to the actual parameter values, but evidently these estimates needed the finetuning that the fully-iterated estimates provide. As mentioned in Section 3.2, it is possible that the desired robustness will not be immediately achieved by the use of the
one-step residual method since the initial estimates are themselves possibly non-
robust.

Table 6.3.2  Estimates of Multiple Regression Regression Parameters
(Contaminated Data)

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Squares</td>
<td>19569.21</td>
<td>-36.76</td>
<td>75.19</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>18799.75</td>
<td>-35.49</td>
<td>75.87</td>
</tr>
<tr>
<td>Spearman (Residual Method, One-Step Estimates)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td>550.223</td>
<td>3.714</td>
<td>971.715</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td>549.048</td>
<td>3.771</td>
<td>971.736</td>
</tr>
<tr>
<td>Spearman (Bias-Adjustment Method, One-Step Estimates)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td>-27120.6</td>
<td>-9380.9</td>
<td>7634.796</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td>-27141.3</td>
<td>-9382.9</td>
<td>7637.092</td>
</tr>
<tr>
<td>Residual Method Initial Values</td>
<td>3.647</td>
<td>4.912</td>
<td>999.9683</td>
</tr>
<tr>
<td>Spearman (Residual Method, Fully-Iterated Estimates)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td>3.640</td>
<td>4.911</td>
<td>999.9689</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td>3.641</td>
<td>4.911</td>
<td>999.9689</td>
</tr>
<tr>
<td>Spearman (Matrix Weights)</td>
<td>3.584</td>
<td>4.911</td>
<td>999.9712</td>
</tr>
</tbody>
</table>

We can also observe from the results the sensitivity of the Spearman bias-
adjustment method to the initial estimates of the parameters. When the initial
estimates were computed from least squares and the Wilcoxon, the one-step bias adjustment method yielded results that are no better than the initial estimates. When the initial values used are close to the actual parameter values, the method yielded satisfactory estimates.

**Example 6.3.3: Modified Hawkins, Bradu, and Kass Data**

It is easy to evaluate whether an estimate is “good” or not in the previous example since we know the true parameter values. If, as real life situations dictate, we do not know the actual parameter values, a good way of evaluating the goodness of the estimates is by the use of (standardized) residual plots. If the parameter estimates are robust to the (bad) leverage values, then the corresponding residuals of these points should be relatively much larger in absolute value than the residuals of the rest of the points.

For this discussion, we use the artificial data set proposed by Hawkins, Bradu, and Kass (1984). The original data set involves three independent variables, but for this example we only use the first two independent variables. There are a total of 75 observations where the first 14 of them are leverage...
points. The rest of the 61 points follow a linear model. Of the 14 leverage points, the first 10 of them do not follow the model, hence we refer to them as bad leverage points; the rest of the leverage points follow the model, and hence are referred to as good leverage points.

Since the third independent variable will be not be used for this example, an appropriate adjustment to the dependent variable was made to preserve the characteristics of the points in the data set. The least squares fit was obtained for the "clean" portion of the data set, i.e., observations 15 to 75.

The adjustment made to the dependent variable \( Y \) is given by

\[
\text{adjusted } Y = Y - \hat{\beta}_{3,LS} x_3
\]

where \( \hat{\beta}_{3,LS} \) is the least squares slope estimate corresponding to the third independent variable \( x_3 \).

Table 6.3.3 gives us the estimates of the regression parameters obtained using different schemes. Except for the Spearman residual one-step and fully-iterated methods and the matrix method, there is generally a disagreement of the results. Since the previous example has illustrated that the Spearman fully-
iterated residual method and the matrix weight method essentially produce robust estimates, let us compare the performance of these two methods with the least squares and Wilcoxon methods. Since the actual parameter values of the regression coefficients for the (modified) Hawkins, Bradu, and Kass data are not available, we shall utilize standardized residual plots to evaluate the performance of the estimates.

Figure 6.3.1 gives us the plots of the standardized residuals obtained from least squares, Wilcoxon, Spearman fully-iterated residual method (least squares initial values), and matrix weights method. The residuals corresponding to the bad leverage points are marked by the hollow circles and the residuals for the good leverage points by the hollow squares.

We see right away that the least squares and Wilcoxon methods did not flag the bad leverage points as potential outliers, but they flagged the good leverage points as good outliers. This is a prime example of the so called masking effect. The bad leverage points have exerted enough influence to drag the regression equation close to their response and mask the fact that they might
otherwise be outliers.

The Spearman methods, on the other hand, were able to correctly flag the bad leverage points (the benchmark used is two standard deviations from the 0 line). Since the bad leverage points were downweighted, more importance were given to the bulk of the data points that follow the linear model. Their corresponding residuals are relatively smaller in absolute value than the residuals corresponding to the bad leverage points.
### Table 6.3.3

Estimates of Multiple Regression Parameters
(Modified HBK Data)

<table>
<thead>
<tr>
<th>Method</th>
<th>Intercept</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Squares</td>
<td>-1.1330</td>
<td>0.7786</td>
<td>0.1230</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>-1.3424</td>
<td>0.4191</td>
<td>0.4607</td>
</tr>
<tr>
<td>Spearman (Residual Method, One-Step Estimates)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td>-0.5764</td>
<td>0.2417</td>
<td>0.1867</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td>-0.5778</td>
<td>0.2405</td>
<td>0.1888</td>
</tr>
<tr>
<td>Spearman (Bias-Adjustment Method, One-Step Estimates)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td>1.1104</td>
<td>-1.6240</td>
<td>1.1634</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td>-1.5088</td>
<td>1.4178</td>
<td>-0.1839</td>
</tr>
<tr>
<td>Residual Method Initial Values</td>
<td>-0.5827</td>
<td>0.2420</td>
<td>0.1889</td>
</tr>
<tr>
<td>Spearman (Residual Method, Fully-Iterated Estimates)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td>-0.5816</td>
<td>0.2415</td>
<td>0.1891</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td>-0.5816</td>
<td>0.2414</td>
<td>0.1892</td>
</tr>
<tr>
<td>Spearman (Matrix Weights)</td>
<td>-0.5815</td>
<td>0.2416</td>
<td>0.1890</td>
</tr>
</tbody>
</table>

* Method did not converge
(a) Least Squares

(b) Wilcoxon
Figure 6.3.1 Standardized Residual Plots versus Index of Observations

(c) Spearman Fully-Iterated Residual Method (Least Squares Initial Values)

(d) Matrix Weights Method
6.4 Simulations for Multiple Regression

This section aims to discuss some of the small sample properties of different multiple regression estimates. In particular, we would like to estimate the efficiency of the different Spearman estimates with respect to the least squares and Wilcoxon estimates.

The model simulated is

\[ Y = 2 + 5x_1 + 10x_2 + \varepsilon \]

where \( x_1 \) and \( x_2 \) come from some distribution (they are held fixed throughout the simulations). Further, we consider two cases for the distribution of the error. The first is \( N(0, 1) \) distribution, and the second case is a contaminated normal error distribution. The sample size used is 50. For each choice of distributions, 300 samples were obtained, and the mean, the MSEs, and the efficiency of the estimates with respect to least squares and Wilcoxon were computed. The next four tables give us the results for the following cases:

Case 1: \( x_1 \sim N(0, 1), x_2 \sim t_{25}, \varepsilon \sim N(0, 1) \)
Case 2: \( x_1 \sim N(0, 1), x_2 \sim t_{(20)}, \epsilon \sim 0.96\cdot N(0, 1) + 0.04\cdot N(0, 100) \)

Case 3: \( x_1 \sim 0.96\cdot N(0, 1) + 0.04\cdot N(0, 100), x_2 \sim t_{(3)}, \epsilon \sim N(0, 1) \)

Case 4: \( x_1 \sim 0.96\cdot N(0, 1) + 0.04\cdot N(0, 100), x_2 \sim t_{(3)}, \epsilon \sim 0.96\cdot N(0, 1) + 0.04\cdot N(0, 100) \)
Table 6.4.1  Simulation Results for Multiple Regression: Case 1

<table>
<thead>
<tr>
<th></th>
<th>Results for $\hat{b}_1$</th>
<th>Results for $\hat{b}_2$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Efficiency with respect to</td>
<td>Efficiency with respect to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least Squares</td>
<td>5.0004</td>
<td>0.00071</td>
<td>1.00000</td>
<td>999.9994</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>5.0006</td>
<td>0.00073</td>
<td>0.97260</td>
<td>1.00000</td>
</tr>
<tr>
<td>Spearman Residual Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.0008</td>
<td>0.00078</td>
<td>0.90970</td>
<td>0.93533</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.0008</td>
<td>0.00079</td>
<td>0.90049</td>
<td>0.92586</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.0008</td>
<td>0.00079</td>
<td>0.90510</td>
<td>0.93060</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.0008</td>
<td>0.00079</td>
<td>0.90027</td>
<td>0.92563</td>
</tr>
<tr>
<td>Spearman Bias-Adjustment Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.0009</td>
<td>0.00079</td>
<td>0.90100</td>
<td>0.92638</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.0008</td>
<td>0.00079</td>
<td>0.89908</td>
<td>0.92441</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.0007</td>
<td>0.00078</td>
<td>0.90651</td>
<td>0.93205</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.0007</td>
<td>0.00079</td>
<td>0.89835</td>
<td>0.92365</td>
</tr>
<tr>
<td>Residual Method Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.0008</td>
<td>0.00079</td>
<td>0.89992</td>
<td>0.92527</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.0008</td>
<td>0.00079</td>
<td>0.89994</td>
<td>0.92529</td>
</tr>
<tr>
<td>Matrix Weights Method</td>
<td>5.0012</td>
<td>0.00070</td>
<td>1.00875</td>
<td>1.03716</td>
</tr>
<tr>
<td>Method</td>
<td>Mean</td>
<td>MSE</td>
<td>Efficiency with respect to</td>
<td>Mean</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------</td>
<td>------</td>
<td>-----------------------------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Least Squares</td>
<td>Wilcoxon</td>
</tr>
<tr>
<td>Least Squares</td>
<td>5.00021</td>
<td>0.00325</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>5.00079</td>
<td>0.00084</td>
<td>3.86845</td>
<td>1.00000</td>
</tr>
<tr>
<td>Spearman Residual Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.00102</td>
<td>0.00089</td>
<td>3.65079</td>
<td>0.94347</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.00109</td>
<td>0.00087</td>
<td>3.71296</td>
<td>0.95980</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.00103</td>
<td>0.00087</td>
<td>3.74279</td>
<td>0.96752</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.00110</td>
<td>0.00087</td>
<td>3.71459</td>
<td>0.96023</td>
</tr>
<tr>
<td>Spearman Bias-Adjustment Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.00106</td>
<td>0.00092</td>
<td>3.52049</td>
<td>0.91005</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.00109</td>
<td>0.00087</td>
<td>3.71778</td>
<td>0.96105</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.00114</td>
<td>0.00087</td>
<td>3.73216</td>
<td>0.96477</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.00109</td>
<td>0.00087</td>
<td>3.72311</td>
<td>0.96243</td>
</tr>
<tr>
<td>Residual Method Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.00110</td>
<td>0.00087</td>
<td>3.72302</td>
<td>0.96240</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.00111</td>
<td>0.00087</td>
<td>3.71697</td>
<td>0.96084</td>
</tr>
<tr>
<td>Matrix Weights Method</td>
<td>5.00107</td>
<td>0.00078</td>
<td>4.14289</td>
<td>1.07094</td>
</tr>
</tbody>
</table>
Table 6.4.3  Simulation Results for Multiple Regression: Case 3

<table>
<thead>
<tr>
<th></th>
<th>Results for Beta-1-hat</th>
<th></th>
<th>Results for Beta-2-hat</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Efficiency with respect to</td>
<td>Efficiency with respect to</td>
<td>Efficiency with respect to</td>
<td>Efficiency with respect to</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>MSE</td>
<td>Least Squares</td>
<td>Wilcoxon</td>
</tr>
<tr>
<td>Least Squares</td>
<td>5.00091</td>
<td>0.00276</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>5.00046</td>
<td>0.00289</td>
<td>0.95286</td>
<td>1.00000</td>
</tr>
<tr>
<td>Spearman Residual Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.00153</td>
<td>0.00535</td>
<td>0.51507</td>
<td>0.81012</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.00151</td>
<td>0.00568</td>
<td>0.48530</td>
<td>0.78811</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.00148</td>
<td>0.00534</td>
<td>0.51583</td>
<td>0.80680</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.00152</td>
<td>0.00568</td>
<td>0.48515</td>
<td>0.78827</td>
</tr>
<tr>
<td>Spearman Bias-Adjustment Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.00164</td>
<td>0.00569</td>
<td>0.48468</td>
<td>0.77609</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.00147</td>
<td>0.00569</td>
<td>0.48491</td>
<td>0.78922</td>
</tr>
<tr>
<td>Wilcoxon Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.00148</td>
<td>0.00564</td>
<td>0.48915</td>
<td>0.77934</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.00148</td>
<td>0.00569</td>
<td>0.48464</td>
<td>0.78832</td>
</tr>
<tr>
<td>Residual Method Initial Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>5.00151</td>
<td>0.00568</td>
<td>0.48507</td>
<td>0.78831</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>5.00148</td>
<td>0.00569</td>
<td>0.48465</td>
<td>0.78900</td>
</tr>
<tr>
<td>Matrix Weights Method</td>
<td>5.00122</td>
<td>0.00372</td>
<td>0.74088</td>
<td>0.94935</td>
</tr>
</tbody>
</table>
Table 6.4.4 Simulation Results for Multiple Regression: Case 4

<table>
<thead>
<tr>
<th>Method</th>
<th>Results for Beta-1-hat</th>
<th></th>
<th></th>
<th>Results for Beta-2-hat</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>MSE</td>
<td>Efficiency with respect to</td>
<td>Mean</td>
<td>MSE</td>
<td>Efficiency with respect to</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Least Squares</td>
<td>Wilcoxon</td>
<td></td>
<td>Least Squares</td>
</tr>
<tr>
<td>Least Squares</td>
<td>4.99493</td>
<td>0.01136</td>
<td>1.00000</td>
<td></td>
<td>1000.025</td>
<td>0.04603</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>4.99359</td>
<td>0.00774</td>
<td>1.46873</td>
<td>1.00000</td>
<td></td>
<td>1000.004</td>
</tr>
<tr>
<td>Spearman Residual Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LS Initial Values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>4.99969</td>
<td>0.01081</td>
<td>1.05095</td>
<td>0.71555</td>
<td></td>
<td>1000.001</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>4.99977</td>
<td>0.01104</td>
<td>1.02925</td>
<td>0.70077</td>
<td></td>
<td>1000.001</td>
</tr>
<tr>
<td><strong>Wilcoxon Initial Values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>4.99979</td>
<td>0.01079</td>
<td>1.05354</td>
<td>0.71731</td>
<td></td>
<td>1000.001</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>4.99977</td>
<td>0.01103</td>
<td>1.03003</td>
<td>0.70131</td>
<td></td>
<td>1000.001</td>
</tr>
<tr>
<td>Spearman Bias-Adjustment Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LS Initial Values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>4.99772</td>
<td>0.01107</td>
<td>1.02688</td>
<td>0.69916</td>
<td></td>
<td>1000.002</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>4.99979</td>
<td>0.01107</td>
<td>1.02700</td>
<td>0.69924</td>
<td></td>
<td>1000.001</td>
</tr>
<tr>
<td><strong>Wilcoxon Initial Values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>4.99945</td>
<td>0.01092</td>
<td>1.04061</td>
<td>0.70851</td>
<td></td>
<td>1000.002</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>4.99982</td>
<td>0.01105</td>
<td>1.02851</td>
<td>0.70027</td>
<td></td>
<td>1000.001</td>
</tr>
<tr>
<td><strong>Residual Method Initial Values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-step</td>
<td>4.99973</td>
<td>0.01105</td>
<td>1.02875</td>
<td>0.70044</td>
<td></td>
<td>1000.001</td>
</tr>
<tr>
<td>Fully-Iterated</td>
<td>4.99980</td>
<td>0.01104</td>
<td>1.02930</td>
<td>0.70081</td>
<td></td>
<td>1000.001</td>
</tr>
<tr>
<td>Matrix Weights Method</td>
<td>4.99656</td>
<td>0.01042</td>
<td>1.09042</td>
<td>0.74243</td>
<td></td>
<td>1000.004</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The results in Tables 6.4.1 and 6.4.3 are already expected. Since the errors in cases 1 and 3 are normally distributed, we expect the least squares estimates to be the most efficient. We can observe from Table 6.4.1 that if the $x$ values are "well-behaved," then for normal errors the Spearman estimates obtained from the residual method are highly efficient with respect to the least squares and Wilcoxon estimates. However, if the errors are normal and if the $x$ values have high variability, we can observe from Table 6.4.3 that the residual method estimates are not as efficient.

The results for cases 2 and 4 are given in Tables 6.4.2 and 6.4.4. We immediately see that if we have contaminated errors, the Spearman estimates are more efficient than the least squares estimates, whether the $x$ values are "well-behaved" (Case 2), or with high variability (Case 4).

We note that the Spearman estimates produced by the matrix weights method are the most efficient among the Spearman estimates. A particularly interesting result is in Case 1, where the matrix weights estimates are almost as efficient as the least squares estimates.
CHAPTER 7

CONCLUSIONS

This dissertation has shown that the Spearman estimator is a viable alternative to the least squares and the Wilcoxon estimators in estimating regression parameters. In particular, it was able to demonstrate its robustness against $x$ outliers, both in the simple and multiple regression settings. It has also shown its potential for being infinitely more efficient than the least squares estimator when the error distribution is not normal.

Since the derivation of the Spearman estimator in simple linear regression involves ranking the $x$ values, the extension of the method to multiple linear regression is not straightforward. But since the Spearman simple linear regression estimate is just a special case of the generalized-rank ($GR$) estimates, the problem was overcome by exploiting the relationships between multiple regression and simple linear regression.

Three methods for obtaining multiple regression estimators were
developed, i.e., the residual method, the bias reduction method, and the matrix weights method. The residual method and the bias adjustment method were shown to provide consistent estimates. Although the discussion focused on illustrating the methods for models with two and three independent variables, the said methods could be used for models with any number of independent variables. Finally, the simulations have shown that the matrix weights estimates are the most efficient among the Spearman estimates.

Future studies are going to be devoted on, first, developing the asymptotic normality of the residual method estimates. Secondly, we need to derive the asymptotic properties of the matrix weights estimates; the simulations have already provided empirical verification that the matrix weights estimates are consistent.
BIBLIOGRAPHY


