Summer 8-11-2017

Examination and Comparison of the Performance of Common Non-Parametric and Robust Regression Models

Gregory F. Malek
Stephen F Austin State University, malekgf@jacks.sfasu.edu

Examination and Comparison of the Performance of Common Non-Parametric and Robust Regression Models

By

GREGORY FRANK MALEK, Bachelor of Science in Mathematics

Presented to the Faculty of the Graduate School of

Stephen F. Austin State University

In Partial Fulfillment

of the Requirements

For the Degree of

Master of Science

STEPHEN F. AUSTIN STATE UNIVERSITY

August 2017

Examination and Comparison of the Performance of Common Non-Parametric and Robust Regression Models

By

GREGORY FRANK MALEK, Bachelor of Science in Mathematics

APPROVED:

Dr. Robert K. Henderson, Thesis Director

Dr. Kent E. Riggs, Committee Member

Dr. Gregory K. Miller, Committee Member

Dr. Garland Simmons, Committee Member

Richard Berry, D.M.A.

Dean of the Graduate School

ABSTRACT

Examination and Comparison of the Performance of Common Non-Parametric and Robust Regression Models

By

Gregory Frank Malek

Stephen F. Austin State University, Masters in Statistics Program,

Nacogdoches, Texas, U.S.A.

g_m_2002@live.com

This work investigated common alternatives to the least-squares regression method in the presence of non-normally distributed errors. An initial literature review identified a variety of alternative methods, including Theil Regression, Wilcoxon Regression, Iteratively Re-Weighted Least Squares, Bounded-Influence Regression, and Bootstrapping methods. These methods were evaluated using a simple simulated example data set, as well as various real data sets, including math proficiency data, Belgian telephone call data, and faculty salaries at the University of South Florida.

In addition, simulations were conducted of common error scenarios to test and evaluate each method. These simulations involved simple regression models in which the error terms were contaminated normal distributions with different amounts and magnitudes of contamination. The models were evaluated based on confidence interval coverage of regression coefficients, as well as bias and confidence interval width.

Finally, results were summarized, conclusions drawn, and suggestions for future applications of the results have been provided.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

This work will investigate common alternatives to the least-squares regression method in the presence of non-normally distributed errors, including extreme errors that might occur at specifically identified locations in the predictor variable space. The work will seek to compare and contrast the various alternatives considered to the traditional least squares results with the ultimate objective of identifying specific circumstances where an alternative to least squares might provide a more useful model of the relationship between a response of interest and an appropriate predictor variable, or set of predictors.

The first section will include a literature review of several common methods, including Theil Regression, Wilcoxon Regression, Iteratively Re-Weighted Least Squares, Bounded-Influence Regression, and Bootstrapping. A simple one-predictor-variable regression example will be used to illustrate each method.

The second section will utilize the simple example used to describe the various approaches in the first section to compare the methods to the traditional least squares approach. This evaluation will include comparisons in the presence of single outliers in each of the response variable space, the predictor variable space, and outliers in both variable spaces.

The third section will apply the selected methods to various data sets, including math proficiency data, Belgian telephone call data, and faculty salaries at the University of South Florida. Each of these data sets have known outliers, which will provide for further comparison of the models across several real data sets.

The fourth section will attempt through limited simulation work to explore the behavior of each of these methods in fitting a simple regression line in the presence of potentially one or more outliers. Responses considered will be not only coverage of respective confidence intervals, but also any bias exhibited by the estimators, as well as the width of their associated confidence intervals.

Finally, the conclusions section will summarize the results of this work, and provide a review of the key findings suggested by the results. In addition, suggestions for future work will be offered.

Section 1: Alternative Methods of Fitting a Line

1.1: Traditional Least Squares

The traditional least squares method of fitting a line to ordered pairs of (x, y) results was first introduced by Sir Francis Galton in his studies of the relationships between generations (**#1**). In his work, he found that successive generations tended to regress towards the mean level for their respective generation. As a result, this process of fitting a model to describe the relationship between a response variable and a set of identifiable predictor variables acquired the rather unfortunate name of “regression”.

The Normal Simple Linear Regression (NSLR) model having only one predictor variable produces the simple linear regression equation written as:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ for } i = 1,2, \dots, n, \tag{1}$$

where y_i is the response value associated with the i^{th} paired observation,

β_0 is an intercept parameter,

β_1 is a slope parameter,

x_i is the value of the predictor variable associated with the i^{th} paired observation,

and

ε_i is a random variable error term associated with the i^{th} paired observation.

The NLSR model often includes the assumption that the error terms are normally, independently, and identically distributed with zero mean and a constant variance σ^2 (i.e., $\varepsilon_i \sim NID(0, \sigma^2)$).

Once this choice of model is specified, the generally unknown intercept and slope parameters defining the model must be estimated from the available sample observations (i.e., the observed ordered pairs (x_i, y_i) for $i = 1,2, \dots, n$).

The most popular method for obtaining these estimates is the traditional least-squares approach.

This method involves finding those estimates that minimize the sum of squared residuals, where each residual is the difference between the observed value of the response variable and its fitted value on the estimated regression line. Mathematically, the least-squares approach attempts to find a pair of estimates for the parameters which will minimize the sum of the squared residuals:

$$\sum_{i=1}^{n} {r_i}^2 = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2 \tag{2}$$

Here b_0 and b_1 are the estimates for the intercept and slope parameters, β_0 and β_1, respectively.

In order to minimize this summation, the derivatives of the expression with respect to each parameter estimate are obtained, set equal to zero, and then solved for b_0 and b_1 as expressions of the observed data (i.e., in terms of the x_i and y_i). Such an approach leads to the following set of equations, often referred to as the “normal equations”:

$$\frac{d}{db_0} \sum_{i=1}^{n} (y_i - b_0 + b_1 x_i))^2 = -2 \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i)) = 0 \tag{3}$$

$$\frac{d}{db_1} \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2 = -2 \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i)) x_i = 0 \tag{4}$$

Solving (3) provides the following expression for b_0 as a function of b_1:

$$b_0 = \bar{Y} - b_1 \bar{X}. \tag{5}$$

where $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

This expression (5) implies that the traditional least squares fitted line will necessarily pass through the ordered pair $(\overline{X}, \overline{Y})$. Substituting this expression (5) for b_0 in equation (4) produces the following expression for b_1:

$$b_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \overline{Y})}{\sum_{i=1}^{n} x_i(x_i - \overline{X})}. \tag{6}$$

Often, equation (6) is given instead by the equation (7) below, which is an equivalent expression:

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{X})(x_i - \overline{X})}. \tag{7}$$

The simple linear regression model can also be written in matrix form, allowing the traditional least squares approach to be extended to include multiple predictor variables. Let $\vec{\boldsymbol{Y}}$ be an $nx1$ column vector containing the n observations of the response. Let **X** be an nxp matrix, where p-1 is the number of predictors, as follows:

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1,(p-2)} & X_{1,(p-1)} \\ 1 & X_{21} & & \vdots & \vdots \\ \vdots & \vdots & \ddots & & \vdots \\ 1 & & \cdots & & X_{n,(p-1)} \end{pmatrix}, \tag{8}$$

where $\boldsymbol{X}_{ij}$ is the level of the j^{th} predictor variable associated with the i^{th} observed pair, for j = 1, …, p-1 and i = 1, …, n.

Also, let $\vec{\boldsymbol{\beta}}$ be a px1 column vector containing the p regression parameters (for the p-1 predictors, and an additional parameter for the intercept as the first component of this vector):

$$\vec{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}. \tag{9}$$

Finally, let $\vec{\boldsymbol{\varepsilon}}$ be an nx1 column vector of NID(0,σ^2) random variables.

The model can now be written in matrix form as:

$$\vec{\boldsymbol{Y}} = \boldsymbol{X}\vec{\boldsymbol{\beta}} + \vec{\boldsymbol{\varepsilon}}, \tag{10}$$

Simultaneous least-squares estimates for the parameters can be found by using the matrix operation:

$$\vec{\boldsymbol{b}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}\vec{\boldsymbol{Y}}, \tag{11}$$

where $\vec{\boldsymbol{b}}$ is a px1 vector of the least-squares estimates for regression parameters $\vec{\boldsymbol{\beta}}$ as given in (9),

$\boldsymbol{X^T}$ = the pxn transpose of the nxp matrix $\boldsymbol{X}$ as given in (8), and

$(\boldsymbol{X^T X})^{-1}$ = the inverse of the positive definite symmetric matrix $\boldsymbol{X^T X}$.

1.1.1: Traditional Least Squares – Simple Example

As a means of illustrating the fitting of the model described by equation (1) above, the open source statistical package R (**#2**) was used to generate the 20 ordered pairs displayed in Table 1 and plotted in Figure 1.

The least-squares estimates are 2.5954 for the intercept parameter, and 1.1996 for the slope parameter, respectively. The actual relationship had an intercept of 2.6 and a slope of 1.2. Figure 2 overlays the fitted regression line in blue, as well as the true linear model equation in red. Note that in this example, least-squares approximates the true line so well that the lines are virtually indistinguishable.

Table 1: Simple Example Data

X	Y	Least-Squares Fitted Values	Y - e (E[Y])
1	5.016602408	3.795081331	3.8
2	2.784688156	4.994730153	5
3	6.31812063	6.194378974	6.2
4	8.410386613	7.394027796	7.4
5	9.714981727	8.593676618	8.6
6	7.850099922	9.79332544	9.8
7	11.84050087	10.99297426	11
8	11.22351123	12.19262308	12.2
9	14.02204069	13.39227191	13.4
10	14.31045089	14.59192073	14.6
11	15.3130134	15.79156955	15.8
12	17.05827744	16.99121837	17
13	19.53999569	18.19086719	18.2
14	19.19822543	19.39051602	19.4
15	21.30742007	20.59016484	20.6
16	21.03991676	21.78981366	21.8
17	22.59635183	22.98946248	23
18	24.06133987	24.1891113	24.2
19	26.80814811	25.38876012	25.4
20	25.42083104	26.58840895	26.6

** E[Y] is the expectation of the response, given by equation Y = 2.6 + 1.2*X.

Figure 1: Plot of Simple Example Data from Table 1

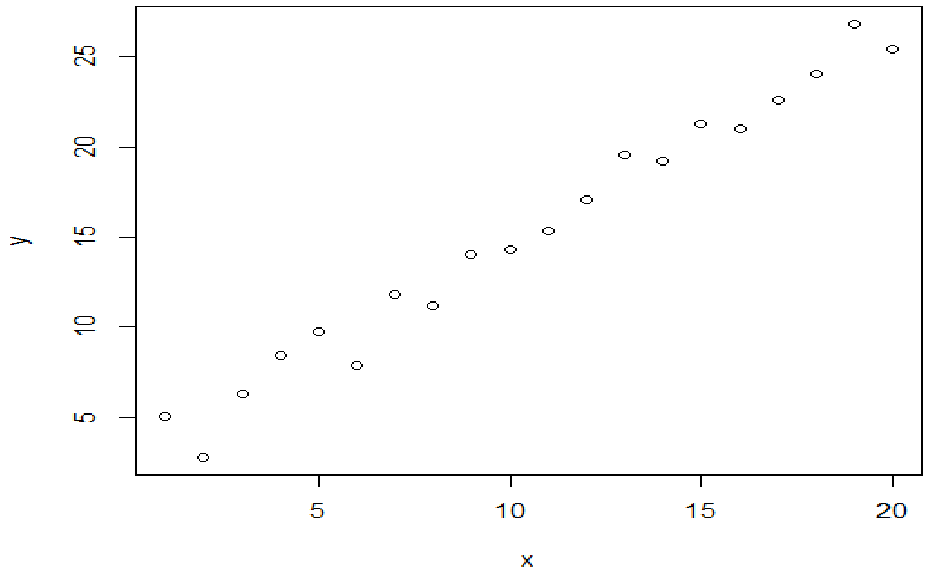
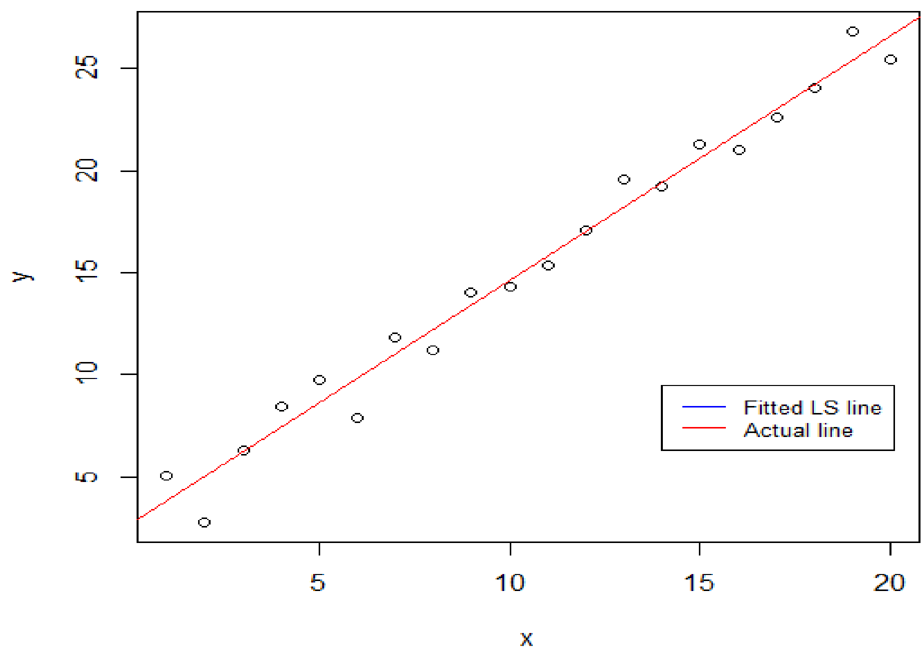


Figure 2: Least-Squares Fit for Simple Example,
including Actual Relationship



1.2: Theil Regression

While the traditional least squares estimators have some very nice properties when the errors are normally and independently distributed with zero mean and constant variance, these estimators can be less than optimal in situations where such conditions on the error distributions fail to hold. As a result, alternative approaches to obtaining intercept and slope parameter estimates for a simple regression model have been suggested in the literature.

The Theil-Sen estimators are one such pair of estimators (**#3**). The Theil-Sen estimator for the SLR slope is given by:

$$b_1 = median(m_{ij}), \tag{12}$$

where

$$m_{ij} = \frac{y_j - y_i}{x_j - x_i}, \text{ for } 1 \le i < j \le n. \tag{13}$$

This estimator is simply the median of all pairwise slopes among the n observations. The intercept estimator is given by:

$$b_0 = median(y_k - b_1 x_k), \text{ for } k = 1,2, \dots n. \tag{14}$$

1.2.1: Theil-Sen - Simple Example.

Using the data displayed in Table 1, the Theil-Sen slope estimator determined from equations (12) and (13) is $b_1 = 1.1735$. The corresponding intercept estimator from equation (14) is $b_0 = 2.8677$.

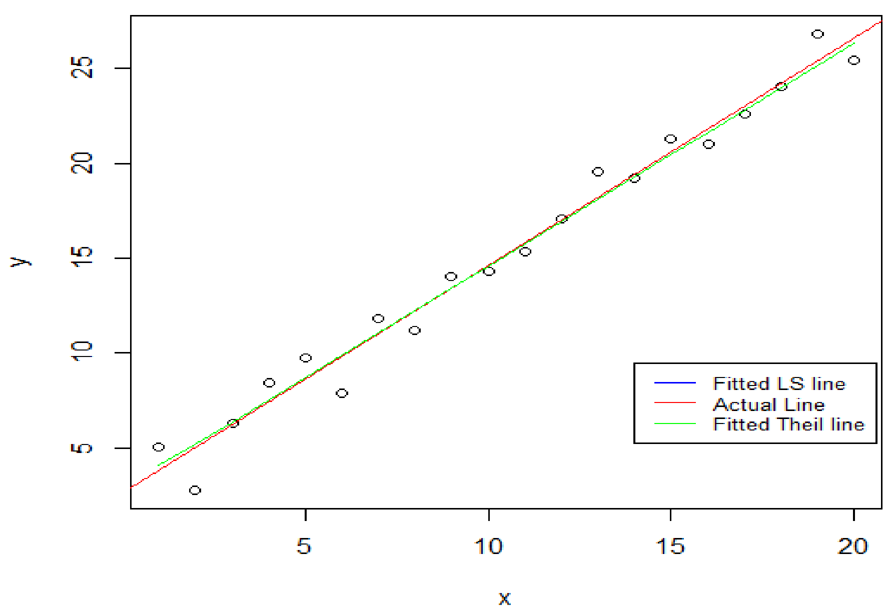
Table 2 below summarizes the comparison between the Theil and Least-Squares estimators for the data in Table 1. While the Theil-Sen intercept estimator is ~10% larger than the true intercept, the slope estimate is within 3% of its true value.

Table 2: Comparison of Least-Squares and Theil Estimators for Simple Example

Parameter	**Actual**	**Least-Squares Estimator**	**Theil-Sen Estimator**
Intercept	2.6	2.5954	2.8677
Slope	1.2	1.1996	1.1735

Figure 3 below displays the lines:

Figure 3: Theil fit vs LS fit for simple example, including actual fit.



1.3: Wilcoxon Regression

Another regression method that can be used when error terms from the model described in equation (1) fail to meet the requirements of being NID(0, σ^2) is the Wilcoxon regression method. With $\vec{\boldsymbol{Y}}$ the nx1 vector of observed response variable results, and $\overrightarrow{\boldsymbol{x_c}}$ the nx1 vector of corresponding predictor variable values centered about their mean, the Wilcoxon method produces a slope estimator by minimizing the Wilcoxon norm, as follows **(#4)**:

$$\hat{b}_1 = \mathrm{Min}_\beta\{||\vec{\boldsymbol{Y}} - \beta\overrightarrow{\boldsymbol{x_c}}||_W\}, \tag{15}$$

where

$$||\vec{v}||_W = \sum_{i=1}^{n} a_\phi(\mathrm{rank}(\vec{\boldsymbol{v}}_i))\vec{\boldsymbol{v}}_i, \quad \vec{\boldsymbol{v}} \in R^n, \tag{16}$$

$$a_\phi(i) = \phi(\tfrac{i}{n+1}), \tag{17}$$

$$\phi(u) = \sqrt{12}*(u - ½), \tag{18}$$

and rank($\vec{\boldsymbol{v}}_i$) ranks the entries of $\vec{\boldsymbol{v}}$ from least (rank = 1) to greatest (rank = n).

The estimate of the model intercept term is obtained as

$$\hat{b}_0 = median\left[\left(\vec{\boldsymbol{Y}} - \hat{b}_1\vec{\boldsymbol{x}}\right)\right]. \tag{19}$$

To minimize this function, the Wilcoxon norm was evaluated across 100,000 “guesses” of the slope estimator, where the guesses for the estimator were of equal distance apart, and spanned the range of all pairwise slopes from the original data. The guess which minimized the Wilcoxon norm over these 100,000 iterates was taken to be the slope estimator.

1.3.1: Wilcoxon Regression – Simple Example.

Table 4 displays the Wilcoxon slope and intercept estimates obtained as described above for the simple example data from Table 1. These estimates were obtained using an initial guess for the slope parameter of 3.9904, which is the maximum of the pairwise slopes. This resulted in the v_i values and ranks as shown in Table 3, which produced a score function (15) value of 306.1359. These results produced an intercept estimate (19) of -27.0875.

Table 3: Iteration of the First Guess for Wilcoxon Slope Estimate

V_i	Ranks	$a_\phi(\text{rank}(V_i))$
1.0262	1	-1.56709
-5.1961	2	-1.40214
-5.653	3	-1.23718
-7.5512	4	-1.07222
-10.237	5	-0.90726
-16.0923	6.5	-0.65983

-16.0923	6.5	-0.65983
-20.6996	8	-0.41239
-21.8915	9	-0.24744
-25.5935	10	-0.08248
-28.5819	11	0.082479
-30.8265	12	0.247436
-32.3352	13	0.412393
-36.667	14	0.57735
-38.5485	15	0.742307
-42.8064	16	0.907265
-45.2404	17	1.072222
-47.7658	18	1.237179
-49.0094	19	1.402136
-54.3871	20	1.567094

Then, the Wilcoxon objective function given by (15) was evaluated at 99,999 more potential estimates of the slope parameter, each of which decreased in an equally partitioned sequence ending near the minimum of the pairwise slopes. Therefore, the i^{th} guess g_i of the slope parameter was given as follows:

$$g_i = \max(pairwise\ slopes) - \frac{i-1}{100{,}000} * (\max(\text{pairwise slopes}) - \min(\text{pairwise slopes})),$$

$$\text{for i} = 1,2,\ldots,100{,}000.$$

Figure 4 below shows a histogram that plots the Wilcoxon score function evaluated at each guess of the slope parameter. A blue, vertical line is drawn at the estimate value that minimizes the Wilcoxon score function, and is therefore chosen as the Wilcoxon slope estimator.

Figure 4: Plot of Score Function Evaluated at Respective Slope Estimates

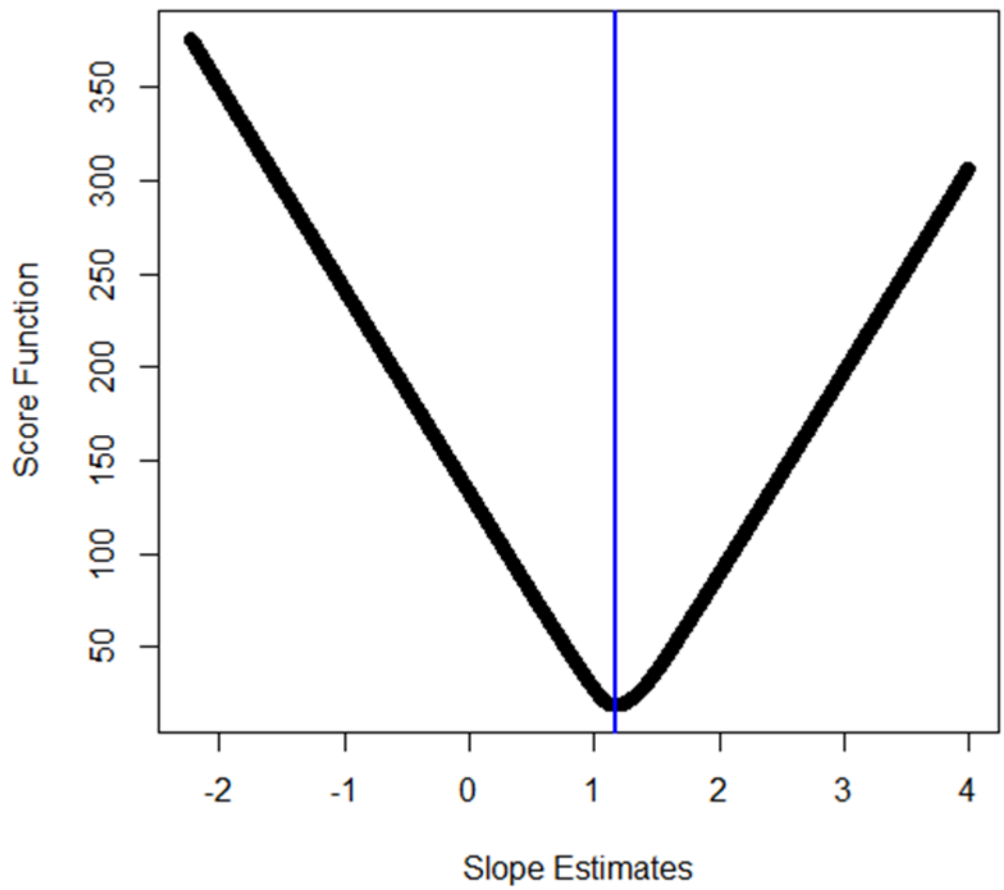
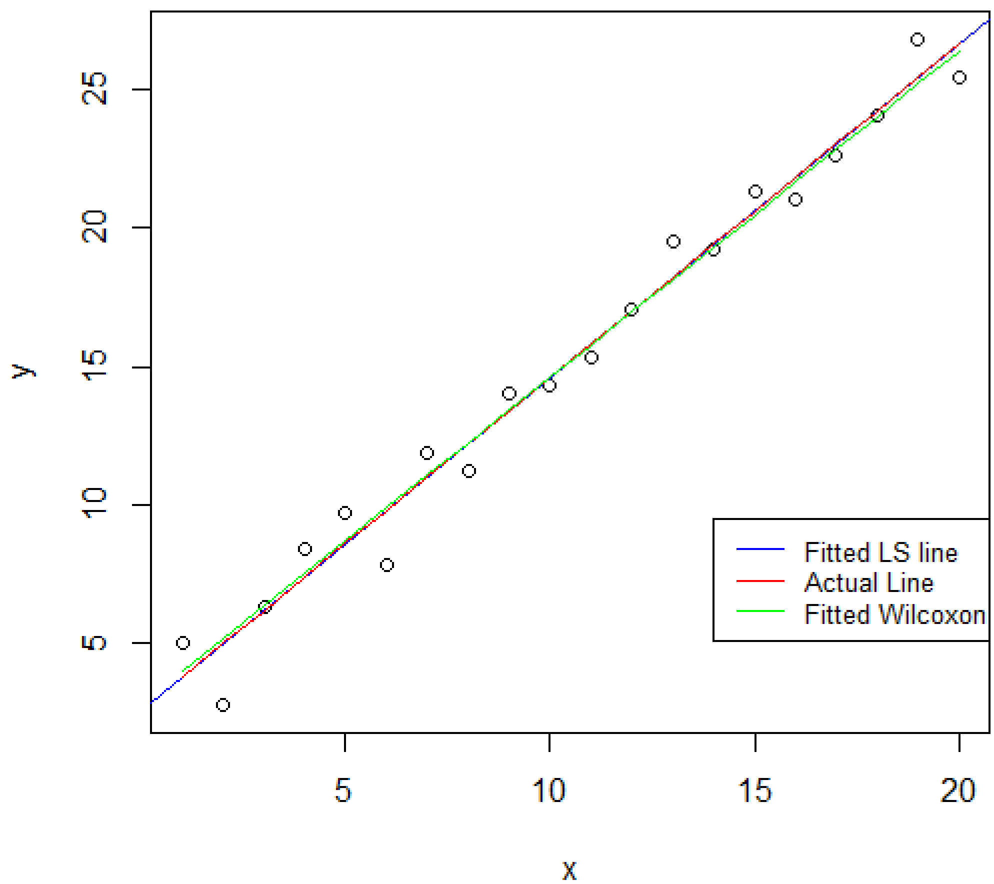


Table 4 below compares the Wilcoxon parameter estimates with the least-squares estimates and the actual parameter values. As observed with the Theil-Sen estimates, the Wilcoxon intercept estimate is within 10 percent of the actual parameter value, and the Wilcoxon slope estimate is within 3% of the true slope parameter. The Wilcoxon fit is given next to the least-squares fit and the actual line in Figure 5.

Table 4: Comparison of Least-Squares and Wilcoxon Estimators for Simple Example

Parameter	Actual	Least-Squares Estimator	Wilcoxon
Intercept	2.6	2.5954	2.8569
Slope	1.2	1.1996	1.1746

Figure 5: Wilcoxon fit vs LS fit for simple example, including actual fit.



1.4: Iteratively Reweighted Least Squares (IRLS) Regression

When the traditional least-squares assumptions on the error fail to hold, there is often a small number of observations which disproportionately influence the fitted regression line. The Iteratively Reweighted Least-Squares (IRLS) method is sometimes used to attempt to mitigate this problem (**#5**). This procedure involves using an algorithm to assign weights to the least-squares observations based on their least-squares residuals, and then iteratively re-calculates the weights throughout successive steps until convergence (to some level of accuracy) of parameter estimates is achieved. This regression method initially obtains NSLR estimates, then utilizes a weighting function applied to the least squares residuals to generate a set of weights for each of the “n” sample observations. Successive steps in the process of generating the IRLS estimates proceed by minimizing the squares of the weighted residuals below:

$$\|\boldsymbol{W} * \vec{\boldsymbol{r}}\| = \sum_i w_i r_i^2 = \vec{\boldsymbol{r}}^{\boldsymbol{T}} \boldsymbol{W}^{\boldsymbol{T}} \boldsymbol{W} \vec{\boldsymbol{r}}, \tag{20}$$

where $\vec{\boldsymbol{r}}$ is the nx1 column vector of residuals r_i as defined in (2), and $\boldsymbol{W} = \boldsymbol{W}^{\boldsymbol{T}}$ is a diagonal (nxn) matrix containing the square root of the weights for each observation. $\boldsymbol{W}_t$ will denote this matrix at the t^{th} iteration of the IRLS algorithm.

The weighted least-squares estimates obtained on the successive iterations are given as:

$$\vec{\boldsymbol{b}} = (\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}^{\boldsymbol{T}} \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}^{\boldsymbol{T}} \boldsymbol{W} \vec{\boldsymbol{Y}}, \tag{21}$$

where

$\vec{\boldsymbol{b}}$ is a px1 vector of the estimates for regression parameters $\vec{\boldsymbol{\beta}}$ as given in (9),

$\boldsymbol{X^T}$ = the pxn transpose of the nxp matrix $\boldsymbol{X}$ as given in (8),

and

$(\boldsymbol{X^T W^T W X})^{-1}$ = the inverse of the positive definite symmetric matrix

$\boldsymbol{X^T W^T W X}$.

There are many possible algorithms for determining the weights; however, two popular approaches are the Huber and the Bi-Square weighting functions.

1.4.1 IRLS Weighting Functions

1.4.1.1 Huber Weight Function:

The Huber Weight function calculates weights at the $(t+1)^{th}$ iteration in the following way:

$$w_{i,(t+1)} = \begin{cases} 1, & if\ \ |r_{s,i,t}| < 1.345 \\ \frac{1.345}{|r_{s,i,t}|}, & if\ \ |r_{s,i,t}|\ \geq 1.345 \end{cases} \tag{22}$$

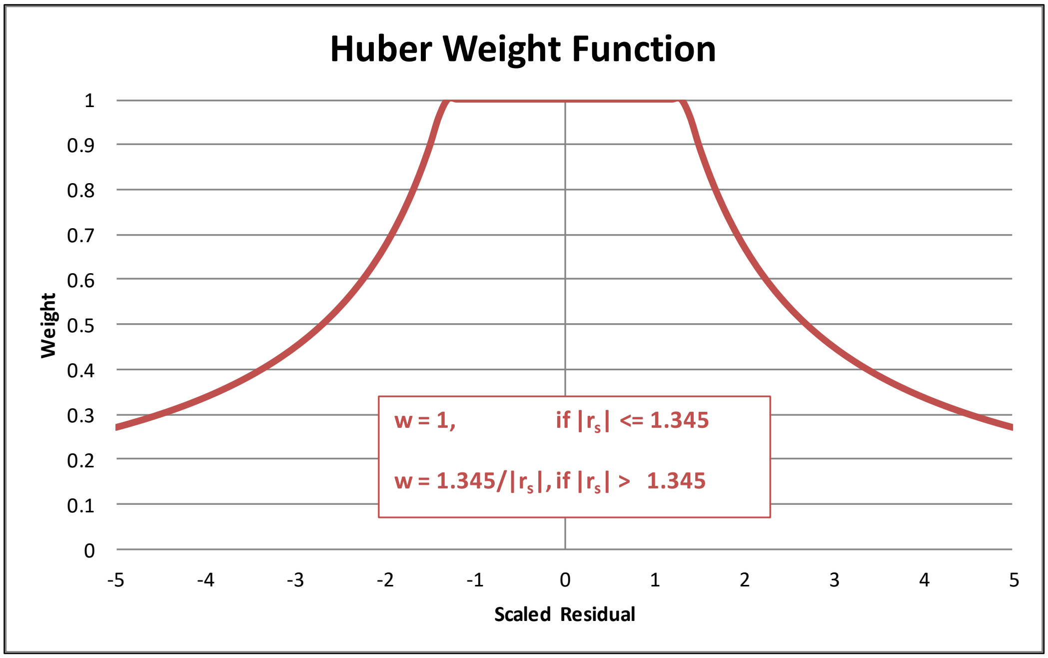
where the scaled residual $r_{s,i,t} = \frac{r_{i,t}}{median(|r_{i,t} - median(r_{i,t})|)/.6745}$, and $r_{i,t}$ = the raw residual as defined in (2) for the i^{th} observation at the t^{th} iteration. The denominator of the scaled residual is the popular Median Absolute Deviation (MAD) robust estimator of scale (**#6**).

The divisor of 0.6745 causes the MAD to be an unbiased estimate of the standard deviation of the response under normality. Scaling the raw residuals by an estimator of scale helps control for natural variation in the data, so that observations with high residuals are not penalized simply because of the overall variance in the response.

The residuals used for the first iteration (when t = 0) are acquired by NSLR (**#5**). As such, the initial weight for every observation at t = 0 is $w_{i,0}$ = 1, for i = 1, …, n, which means that the initial iteration is equivalent to regular least-squares regression, and minimizes the standard sum of squares.

The Huber weighting function (17) is displayed in Figure 6. Note that observations are only down-weighted when the absolute value of their scaled residuals exceed 1.345, and that the weights will always be larger than zero. As a result, the Huber weighting function will never entirely discard an observation. However, observations which are tied to “large” scaled residuals will be down-weighted in the next iteration, and will have less influence on that iteration’s regression line than observations with smaller scaled residuals. This is because in equation (20), the sum of weighted squared residuals is minimized when the larger residuals are paired with smaller weights. Therefore, the minimum sum of weighted squared residuals will fit a line which focuses more heavily on fitting those observations with smaller scaled residuals (since they have larger weights) than those with larger scaled residuals from previous iterations.

Figure 6: Huber Weighting Function



The weights for each observation are re-calculated at each iteration, and a new regression model is fit in the next iteration based on those weights. The iterative process proceeds until a desirable level of convergence is achieved. This usually occurs within 5 to 10 iterations.

1.4.1.2 The Bi-Square Weight Function:

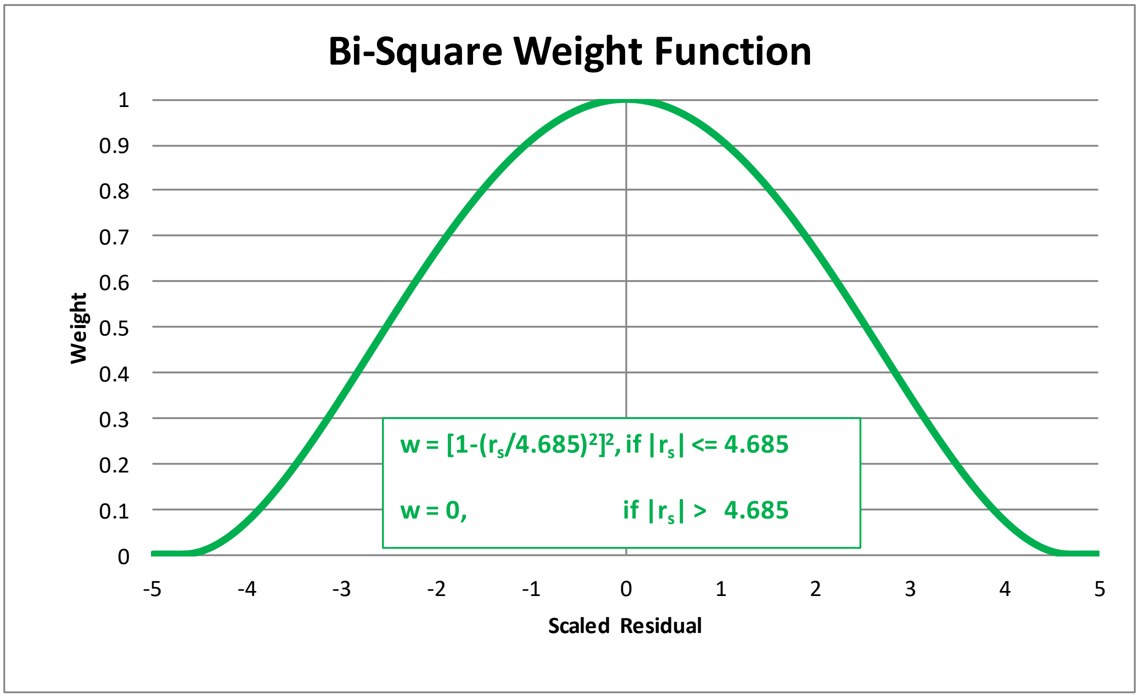
The Bi-Square weight function calculates weights as follows:

$$w_{i,(t+1)} = \begin{cases} \left[1 - \left(\frac{r_{s,i,t}}{4.685}\right)^2\right]^2, & if \quad |r_{s,i,t}| < 4.685 \\ 0, & if \quad |r_{s,i,t}| \geq 4.685 \end{cases} \qquad (23)$$

For the Bi-Square weight function, the first iteration (t = 0) can be acquired via NSLR. However, convergence can sometimes be achieved more quickly by using the residuals obtained by fitting the t=1 Huber IRLS estimates for $\vec{\boldsymbol{\beta}}$ (9).

Figure 7 displays the Bi-Weight function. Note that this function will down-weight all observations (i.e., all $w_{i,(t+1)}$ will be less than one, unless $r_{i,t}$ = 0); however, if the absolute value of the scaled residual for an observation is larger than 4.685, then that observation will receive a weight of zero, essentially discarding it from the data and any subsequent analysis. Furthermore, similar to the Huber weighting function, observations which are not completely thrown out will be down-weighted more heavily the larger their scaled residuals are. Again, as noted with the Huber example, the overall sum of weighted squared residuals will be minimized when larger residuals are associated with smaller weights, so the procedure will prioritize the fit to those observations with smaller scaled residuals rather than to those with larger ones.

Figure 7: Bi-Square Weighting Function



<u>1.4.2: IRLS – Simple Example.</u>

1.4.2.1 IRLS – Simple Example – Huber Weighting Function

Using the data displayed in Table 1, the IRLS approach utilizing the Huber weighting function produced the results observed in Tables 5 through 7. The initial least squares fit of the results produced the least squares residuals, which when scaled by the MAD = $median\left(\left|r_{i,t} - median\left(r_{i,t}\right)\right|\right)/.6745$ scale estimate, resulted in the values in the first column of Table 5 (iteration zero) below. Note that the only scaled residuals larger than 1.345 in magnitude were those for observations 2 and 6. As a result, these observations received the weights less than one that appear in the first column of Table 5. Using these new weights, the estimated intercept and slope parameters displayed on the first row of Table 6 were obtained using the approach described above. This process continues until a desired level of convergence is achieved.

Table 5: Scaled Residuals Across First Eight Iterations of Huber-Weighting Function

	Iteration Number								
Observation	0	1	2	3	4	5	6	7	8
1	1.0079	1.0137	0.9189	0.8812	0.8753	0.8744	0.8743	0.8743	0.8743
2	-1.8353	-1.8341	-2.0039	-2.0337	-2.0383	-2.0390	-2.0391	-2.0391	-2.0392
3	0.0979	0.1027	-0.0013	-0.0324	-0.0372	-0.0379	-0.0381	-0.0381	-0.0381
4	0.8371	0.8435	0.7700	0.7400	0.7354	0.7347	0.7345	0.7345	0.7345
5	0.9237	0.9306	0.8684	0.8408	0.8365	0.8358	0.8357	0.8357	0.8357
6	-1.6154	-1.6127	-1.7408	-1.7611	-1.7643	-1.7648	-1.7648	-1.7649	-1.7649
7	0.6963	0.7034	0.6521	0.6298	0.6264	0.6258	0.6258	0.6258	0.6258
8	-0.8090	-0.8043	-0.8910	-0.9080	-0.9106	-0.9110	-0.9111	-0.9111	-0.9111
9	0.5153	0.5226	0.4837	0.4667	0.4640	0.4636	0.4636	0.4636	0.4636
10	-0.2399	-0.2336	-0.2860	-0.2991	-0.3011	-0.3014	-0.3015	-0.3015	-0.3015

Table 5, Continued: Scaled Residuals Across First Eight Iterations of Huber-Weighting Function

	Iteration Number								
Observation	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
11	-0.4035	-0.3972	-0.4456	-0.4559	-0.4575	-0.4577	-0.4578	-0.4578	-0.4578
12	0.0483	0.0557	0.0293	0.0206	0.0193	0.0191	0.0190	0.0190	0.0190
13	1.1101	1.1196	1.1333	1.1251	1.1238	1.1236	1.1236	1.1236	1.1236
14	-0.1672	-0.1596	-0.1747	-0.1780	-0.1785	-0.1786	-0.1786	-0.1786	-0.1786
15	0.5861	0.5953	0.6111	0.6088	0.6085	0.6084	0.6084	0.6084	0.6084
16	-0.6297	-0.6223	-0.6335	-0.6310	-0.6306	-0.6305	-0.6305	-0.6305	-0.6305
17	-0.3344	-0.3262	-0.3199	-0.3155	-0.3148	-0.3147	-0.3147	-0.3147	-0.3147
18	-0.1149	-0.1060	-0.0844	-0.0780	-0.0769	-0.0768	-0.0768	-0.0768	-0.0768
19	1.1666	1.1780	1.2461	1.2526	1.2536	1.2538	1.2538	1.2538	1.2538
20	-0.9769	-0.9690	-0.9551	-0.9421	-0.9401	-0.9397	-0.9397	-0.9397	-0.9397

Table 6: Huber Weights in Simple Example Across First Eight Iterations

	Iteration Number							
Observation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1	1	1	1	1	1	1	1
2	0.7333	0.6712	0.6613	0.6598	0.6596	0.6595	0.6595	0.6595
3	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1
6	0.834	0.7726	0.7637	0.7623	0.76213	0.7621	0.7621	0.7621
7	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1

Table 6, Continued: Huber Weights in Simple Example Across First Eight Iterations

	Iteration Number							
Observation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
10	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	1

Note that convergence to three decimal places for the residuals, the weights, and the estimated parameters has been achieved after five or six iterations. Convergence to four decimal places is achieved for the intercept parameter estimate after six iterations and for the slope parameter estimate after only five iterations in this example. Additionally, the only observations which were ever down-weighted in the process were numbers 2 and 6.

Table 7: Parameter Estimates Using Huber Weighting Function

Iteration	Intercept	Slope
1	2.5954	1.1996
2	2.7516	1.18
3	2.7967	1.18
4	2.8038	1.186
5	2.8049	1.1859
6	2.805	1.1859
7	2.8051	1.1859
8	2.8051	1.1859
9	2.8051	1.1859
10	2.8051	1.1859

Example – Bi-Weight Weighting Function

In order to accelerate the process of convergence for the Bi-Weight function, the initial iteration was taken from the first iteration obtained using the Huber weighting function. Subsequent iterations weighted the scaled residuals using the Bi-Weight function (23). Ten iterations were performed, the results of which are given below in Tables 8, 9 and 10. Convergence to four decimal places was achieved after eight iterations for scaled residuals and parameter estimates, and after five iterations for weights. Note that the scaled residuals at iteration zero represent the Huber weights at iteration 1. This was done as stated before to promote faster convergence for the parameter estimates.

Convergence to three decimal places is achieved for the scaled residuals of every observation after no more than seven iterations, and for the weights of every observation after no more than four iterations. Convergence to four decimal places is achieved for the parameter estimates after seven iterations.

Unlike the Huber function, the Bi-Weight function down-weighted every observation by some amount. However, observations 2 (w_2 = 0.661), 6 (w_6=.741), 13 (w_{13}=0.885), and 19 (w_{19}=0.858) are the only observations with weights below 0.9. These first two observations are the same as those receiving weights less than one with the Huber weighting approach.

Table 8: Residuals Across First Seven Iterations Scaled of Bi-Squared Weighting Function

	Iteration Number							
Observation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1.0137	0.9318	0.8976	0.8903	0.8888	0.8885	0.8884	0.8884
2	-1.8341	-1.9910	-2.0172	-2.0228	-2.0240	-2.0242	-2.0243	-2.0243
3	0.1027	0.0116	-0.0159	-0.0218	-0.0230	-0.0233	-0.0233	-0.0233
4	0.8435	0.7829	0.7565	0.7509	0.7497	0.7494	0.7494	0.7494
5	0.9306	0.8813	0.8573	0.8522	0.8511	0.8508	0.8508	0.8508
6	-1.6127	-1.7279	-1.7445	-1.7481	-1.7489	-1.7490	-1.7491	-1.7491
7	0.7034	0.6650	0.6464	0.6424	0.6416	0.6414	0.6414	0.6414
8	-0.8043	-0.8781	-0.8913	-0.8942	-0.8948	-0.8949	-0.8949	-0.8949
9	0.5226	0.4966	0.4833	0.4805	0.4798	0.4797	0.4797	0.4797
10	-0.2336	-0.2731	-0.2824	-0.2844	-0.2849	-0.2850	-0.2850	-0.2850

Table 8: Residuals Across First Seven Iterations Scaled of Bi-Squared Weighting Function

	Iteration Number							
Observation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
11	-0.3972	-0.4327	-0.4392	-0.4406	-0.4409	-0.4410	-0.4410	-0.4410
12	0.0557	0.0422	0.0374	0.0363	0.0360	0.0360	0.0360	0.0360
13	1.1196	1.1463	1.1419	1.1409	1.1406	1.1406	1.1406	1.1406
14	-0.1596	-0.1617	-0.1612	-0.1612	-0.1612	-0.1612	-0.1612	-0.1612
15	0.5953	0.6241	0.6257	0.6259	0.6260	0.6260	0.6260	0.6260
16	-0.6223	-0.6205	-0.6141	-0.6128	-0.6126	-0.6125	-0.6125	-0.6125
17	-0.3262	-0.3069	-0.2985	-0.2969	-0.2965	-0.2965	-0.2965	-0.2965
18	-0.1060	-0.0715	-0.0610	-0.0589	-0.0585	-0.0584	-0.0584	-0.0584
19	1.1780	1.2590	1.2695	1.2716	1.2721	1.2722	1.2722	1.2722
20	-0.9690	-0.9422	-0.9250	-0.9216	-0.9208	-0.9207	-0.9207	-0.9206

Table 9: Bi-Squared Weights in Simple Example Across First Five Iterations

	Iteration Number					
Observation	0	1	2	3	4	5
1	1.0000	0.9245	0.9284	0.9292	0.9293	0.9294
2	0.7333	0.6676	0.6628	0.6618	0.6615	0.6615
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	1.0000	0.9467	0.9489	0.9494	0.9495	0.9495
5	1.0000	0.9325	0.9345	0.9350	0.9351	0.9351
6	0.8340	0.7429	0.7413	0.7408	0.7407	0.7407
7	1.0000	0.9616	0.9626	0.9628	0.9629	0.9629
8	1.0000	0.9290	0.9286	0.9284	0.9284	0.9284
9	1.0000	0.9788	0.9790	0.9791	0.9791	0.9791
10	1.0000	0.9926	0.9927	0.9926	0.9926	0.9926

Table 9: Bi-Squared Weights in Simple Example Across First Five Iterations

	Iteration Number					
Observation	0	1	2	3	4	5
11	1.0000	0.9820	0.9824	0.9824	0.9824	0.9824
12	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999
13	1.0000	0.8864	0.8850	0.8850	0.8850	0.8850
14	1.0000	0.9972	0.9976	0.9976	0.9976	0.9976
15	1.0000	0.9663	0.9648	0.9646	0.9646	0.9646
16	1.0000	0.9638	0.9658	0.9661	0.9661	0.9661
17	1.0000	0.9907	0.9918	0.9920	0.9920	0.9920
18	1.0000	0.9994	0.9997	0.9997	0.9997	0.9997
19	1.0000	0.8635	0.8587	0.8581	0.8580	0.8580
20	1.0000	0.9186	0.9235	0.9241	0.9242	0.9243

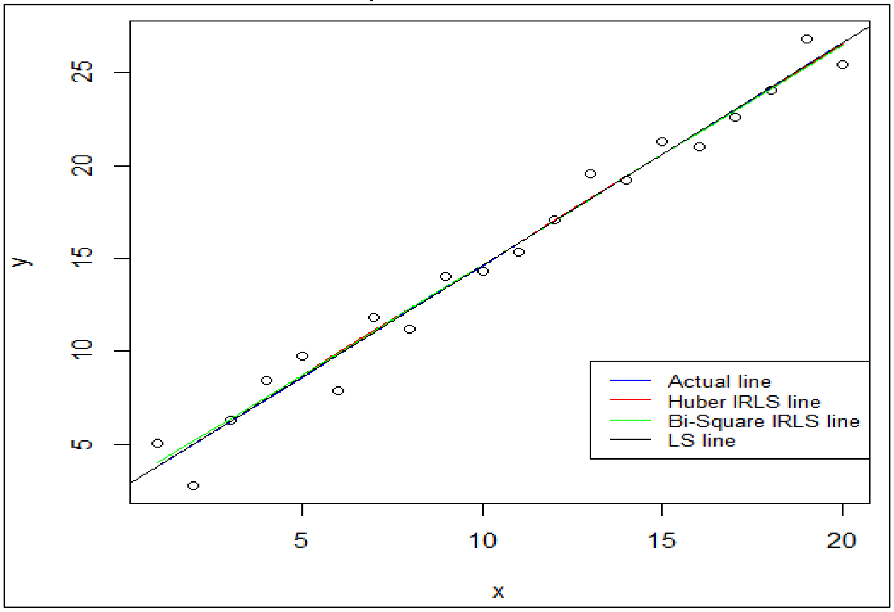
Table 10: Parameter Estimates Using Bi-Squared Weighting Function

Iteration	Intercept	Slope
1	2.5954	1.1996
2	2.7365	1.894
3	2.7776	1.864
4	2.7863	1.858
5	2.7881	1.857
6	2.7885	1.856
7	2.7886	1.856
8	2.7886	1.856
9	2.7886	1.856
10	2.7886	1.856

1.4.2.3 Comparison of IRLS Weighting Function Approaches

In Figure 8 below, the actual line is plotted with the IRLS lines obtained using the Huber and Bi-Weight functions, as well as the common NSLR line. Note that the fitted lines, as with the Theil-Sen and Wilcoxon lines, are again virtually indistinguishable in this plot.

Figure 8: IRLS Huber fit vs IRLS Bi-Square fit, including Actual Line and Least-Squares fit.



Parameter estimates obtained from the two weighting functions are given in Table 11 below.

Table 11: Comparison of NSLR with Huber and Bi-Weight function IRLS parameter estimates for Simple Example

Parameter	Actual	Huber	Bi-Square	Least-Squares
Intercept	2.6	2.8051	2.7886	2.5954
Slope	1.2	1.1859	1.1856	1.1996

Once again, the lines are nearly indistinguishable, the slope parameter estimates for both IRLS lines are within 1.2% of the actual slope, and the intercept estimates are within 8% of the actual parameter value. Moreover, the IRLS line parameter estimates are within 1% of each other, with the slope estimates being equal to three decimal places.

1.5: Bounded Influence Regression

Bounded Influence Regression utilizes the same procedure used in IRLS, except that the weighting function involved utilizes deleted residuals at each iteration instead of the raw residuals (**#5**). Recall that in least-squares, the fitted values for the response are given by:

$$\vec{\hat{\boldsymbol{Y}}} = \boldsymbol{X}\vec{\boldsymbol{b}}, \tag{24}$$

which after re-expressing $\vec{\boldsymbol{b}}$ as in equation (11), the above (24) can be given as:

$$\vec{\hat{\boldsymbol{Y}}} = [\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}]\vec{\boldsymbol{Y}} = \boldsymbol{H}\vec{\boldsymbol{Y}}. \tag{25}$$

The square, symmetric, and positive definite nxn "hat matrix" $\boldsymbol{H}$ is a linear transformation from $\boldsymbol{R}^n \rightarrow \boldsymbol{R}^n$ which maps the observed response values to their fitted values under the least-squares model. Note that the hat matrix is solely a function of the predictor matrix and does not depend on the response vector $\vec{\boldsymbol{Y}}$ at all.

One of the most useful aspects of the hat matrix involves its diagonal elements, which represent the “leverage values” of individual observations. Let h_{ii} for $1 \leq i \leq n$ represent the i^{th} diagonal element of the hat matrix, then h_{ii} represents the leverage value of the i^{th} observation (**#5**). The concept of leverage is related to how much impact or effect the i^{th} observation can have on the estimated parameters in the model being used to describe the relationship between the response and the predictor variables. The larger the leverage value, the more potential impact that observation can have on the model fit to the data and the overall conclusions suggested through an analysis based on that model. Generally, higher leverage values identify observations extreme (potentially to the point of being consider outliers) in the p-dimensional predictor variable space. Observations with low leverage values are those that are near the p-dimensional predictor variable mean vector.

Consider the raw residual $r_{i,t}$ as defined in (2), of the i^{th} observation after iteration “t”. The deleted residual $d_{i,t}$ represents a re-scaled raw residual given by $d_{i,t} = \frac{r_{i,t}}{1-h_{ii,t}}$, where $h_{ii,t}$ is the i^{th} diagonal element of the hat matrix at iteration “t”. Note that $0 < h_{ii,t} < 1$. Higher leverage values will cause the denominator of this expression to decrease, which increases the overall value of the deleted residual, resulting in a smaller weight associated with that observation via the Huber or Bi-Squared functions. Therefore, using deleted residuals in place of raw residuals provides a means to reduce the influence of extreme (possibly outlier) observations in the predictor variable space on the estimated parameters defining the regression model.

Equation (25) describes calculation of the hat-matrix for the regular least-squares approach. However, in a weighted least-squares approach, the vector of fitted response values is instead calculated by:

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\vec{\boldsymbol{b}} = \boldsymbol{X}(\boldsymbol{X^T}\boldsymbol{W^T}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X^T}\boldsymbol{W^T}\boldsymbol{W}\vec{\boldsymbol{Y}}. \tag{26}$$

Therefore, the hat matrix in this instance is given by:

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X^T}\boldsymbol{W^T}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X^T}\boldsymbol{W^T}\boldsymbol{W}. \tag{27}$$

The diagonal elements of this matrix (27) are used to calculate the leverages for the Bounded Influence approach. Once the deleted residuals, or $d_{i,t}$ values, are calculated, the MAD (Median Absolute Deviation) estimator is calculated on these deleted residuals to produce scaled deleted residuals, $d_{s,i,t}$, given by:

$$d_{s,i,t} = \frac{d_{i,t}}{median(|d_{i,t} - median(d_{i,t})|)/.6745}. \tag{28}$$

The weights for the next iteration are then calculated using the same weighting functions as used for the IRLS estimators, but now with the scaled deleted residuals as inputs rather than simply the scaled residuals.

1.5.1: Bounded Influence Regression – Simple Example.

1.5.1.1 Bounded Influence Regression – Simple Example – Huber Weighting Function

Using the data displayed in Table 1, the Bounded Influence approach utilizing the Huber weighting function produced the results observed in Tables 12 through 14 below. The initial least squares fit of the results produced the least squares residuals, which when divided by $(1 - h_{ii})$ (h_{ii} being the leverage value associated with the i^{th} observation) and scaled by the MAD estimate of the deleted residuals, resulted in the values in the first column of Table 12 (iteration zero).

Note that the only scaled deleted residuals in Table 12 larger than 1.345 in magnitude for iteration zero were those for observations 2 and 6. These were the same observations that received weights of less than one for the IRLS regression approach described in Section 1.4. However, note that the initial weights in Table 13 are not the same in this Bounded Influence case. The initial weight for observation 6 is similar, but for observation 2, the initial Bounded Influence weight (0.6705) is less than the IRLS weight (0.7333). This is because the relatively large residual for observation 2 has relatively high leverage in this data set (h_{22} = 0.159).

The final Bounded Influence weights are 0.6387 and 0.7741 for observations 2 and 6, respectively, while the corresponding final IRLS weights were 0.6595 and 0.7621. In addition, the Bounded Influence approach also down-weighted observation 19 (also with initial leverage = 0.159) to 0.9180, while this observation was not down-weighted at all in the IRLS approach.

Table 12: Scaled Deleted Residuals Across First Six Iterations of Huber-Weighting Function

		Iteration Number						
Observation	**Initial Leverage**	0	1	2	3	4	5	6
1	0.1857	1.1455	1.0254	0.9961	0.9945	0.9944	0.9944	0.9944
2	0.1586	-2.0059	-2.0968	-2.1058	-2.1058	-2.1059	-2.1059	-2.1059
3	0.1346	0.1092	-0.0226	-0.0456	-0.0468	-0.0468	-0.0469	-0.0469
4	0.1135	0.8755	0.7853	0.7637	0.7625	0.7625	0.7625	0.7625
5	0.0955	0.9466	0.8707	0.8511	0.8500	0.8500	0.8500	0.8500
6	0.0805	-1.6137	-1.7346	-1.7371	-1.7376	-1.7376	-1.7376	-1.7376
7	0.0684	0.6947	0.6325	0.6173	0.6165	0.6165	0.6165	0.6165
8	0.0594	-0.7868	-0.8831	-0.8929	-0.8934	-0.8934	-0.8934	-0.8934
9	0.0534	0.5080	0.4609	0.4498	0.4492	0.4492	0.4492	0.4492
10	0.0504	-0.2263	-0.2843	-0.2917	-0.2921	-0.2921	-0.2921	-0.2921

Table 12, Continued: Scaled Deleted Residuals Across First Six Iterations of Huber-Weighting Function

		Iteration Number						
Observation	**Initial Leverage**	0	1	2	3	4	5	6
11	0.0504	-0.3848	-0.4368	-0.4420	-0.4422	-0.4423	-0.4423	-0.4423
12	0.0534	0.0541	0.0243	0.0197	0.0195	0.0195	0.0195	0.0195
13	0.0594	1.0953	1.1035	1.0978	1.0976	1.0976	1.0976	1.0976
14	0.0684	-0.1576	-0.1729	-0.1734	-0.1733	-0.1733	-0.1733	-0.1733
15	0.0805	0.5956	0.6108	0.6101	0.6101	0.6101	0.6101	0.6101
16	0.0955	-0.6331	-0.6401	-0.6357	-0.6353	-0.6353	-0.6353	-0.6353
17	0.1135	-0.3386	-0.3267	-0.3212	-0.3208	-0.3207	-0.3207	-0.3207
18	0.1346	-0.1127	-0.0831	-0.0761	-0.0756	-0.0756	-0.0756	-0.0756
19	0.1586	1.2883	1.3695	1.3709	1.3711	1.3711	1.3711	1.3711
20	0.1857	-1.0949	-1.0680	-1.0544	-1.0535	-1.0535	-1.0535	-1.0535

Table 13: Bounded Influence Huber Weights in Simple Example Across First Five Iterations

Observation	Iteration Number				
	1	2	3	4	5
1	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.6705	0.6414	0.6387	0.6387	0.6387
3	1.0000	1.0000	1.0000	1.0000	1.0000
4	1.0000	1.0000	1.0000	1.0000	1.0000
5	1.0000	1.0000	1.0000	1.0000	1.0000
6	0.8335	0.7754	0.7743	0.7741	0.7741
7	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000

Table 13, Continued: Bounded Influence Huber Weights in Simple Example Across First Five Iterations

	Iteration Number				
Observation	1	2	3	4	5
11	1.0000	1.0000	1.0000	1.0000	1.0000
12	1.0000	1.0000	1.0000	1.0000	1.0000
13	1.0000	1.0000	1.0000	1.0000	1.0000
14	1.0000	1.0000	1.0000	1.0000	1.0000
15	1.0000	1.0000	1.0000	1.0000	1.0000
16	1.0000	1.0000	1.0000	1.0000	1.0000
17	1.0000	1.0000	1.0000	1.0000	1.0000
18	1.0000	1.0000	1.0000	1.0000	1.0000
19	1.0000	0.9821	0.9811	0.9810	0.9810
20	1.0000	1.0000	1.0000	1.0000	1.0000

Table 14: Bounded Influence Parameter Estimates Using Huber Weighting Function

Iteration	Intercept	Slope
1	2.5954	1.1996
2	2.7807	1.1873
3	2.8124	1.1852
4	2.8142	1.1850
5	2.8142	1.1850
6	2.8142	1.1850
7	2.8142	1.1850
8	2.8142	1.1850
9	2.8142	1.1850
10	2.8142	1.1850

Convergence to four decimal places for scaled deleted residuals, weights, and parameter estimates is achieved after no more than five iterations. Additionally, the only observations which were ever down-weighted in the process were observation numbers 2, 6 and 19. This is a slightly different result than that for the corresponding IRLS where only observations 2 and 6 were down-weighted.

1.5.2.2 Bounded Influence Regression – Simple Example – Bi-Weight Weighting Function

In order to accelerate the process of convergence for the Bi-Weight function, the initial iteration was once again taken from the first iteration obtained using the Huber weighting function. Subsequent iterations weighted the scaled deleted residuals using the Bi-Weight function (23). Ten iterations were performed, the results of which are given below in Tables 15, 16 and 17. Note that the scaled residuals at iteration zero represent the first iteration Huber weights. This was done as stated before to promote faster convergence for the parameter estimates.

Table 15: Scaled Deleted Residuals Across First Six Iterations of Bi-Squared Weighting Function

	Iteration Number						
Observation	0	1	2	3	4	5	6
1	1.0254	1.0254	1.0046	1.0043	1.0042	1.0041	1.0041
2	-2.0968	-2.0968	-2.1007	-2.1009	-2.1010	-2.1010	-2.1010
3	-0.0226	-0.0226	-0.0276	-0.0282	-0.0283	-0.0284	-0.0284
4	0.7853	0.7853	0.7803	0.7798	0.7797	0.7797	0.7797
5	0.8707	0.8707	0.8657	0.8654	0.8653	0.8653	0.8653
6	-1.7346	-1.7346	-1.7218	-1.7227	-1.7227	-1.7227	-1.7227
7	0.6325	0.6325	0.6336	0.6335	0.6335	0.6335	0.6335
8	-0.8831	-0.8831	-0.8758	-0.8758	-0.8759	-0.8759	-0.8759
9	0.4609	0.4609	0.4656	0.4658	0.4658	0.4658	0.4658

Table 15, Continued: Scaled Deleted Residuals Across First Six Iterations of Bi-Squared Weighting Function

	Iteration Number						
Observation	0	1	2	3	4	5	6
10	-0.2843	-0.2843	-0.2774	-0.2771	-0.2771	-0.2771	-0.2771
11	-0.4368	-0.4368	-0.4281	-0.4277	-0.4276	-0.4276	-0.4276
12	0.0243	0.0243	0.0342	0.0348	0.0349	0.0349	0.0349
13	1.1035	1.1035	1.1079	1.1086	1.1087	1.1087	1.1087
14	-0.1729	-0.1729	-0.1598	-0.1589	-0.1588	-0.1588	-0.1588
15	0.6108	0.6108	0.6255	0.6264	0.6265	0.6265	0.6265
16	-0.6401	-0.6401	-0.6229	-0.6218	-0.6217	-0.6217	-0.6217
17	-0.3267	-0.3267	-0.3091	-0.3077	-0.3075	-0.3075	-0.3075
18	-0.2843	-0.2843	-0.2774	-0.2771	-0.2771	-0.2771	-0.2771

Table 16: Bi-Squared Weights in Simple Example Across First Four Iterations

	Iteration Number			
Observation	**1**	**2**	**3**	**4**
1	0.9065	0.9102	0.9102	0.9102
2	0.6395	0.6383	0.6383	0.6382
3	1.0000	0.9999	0.9999	0.9999
4	0.9446	0.9453	0.9454	0.9454
5	0.9321	0.9329	0.9329	0.9329
6	0.7446	0.7481	0.7479	0.7479
7	0.9639	0.9638	0.9638	0.9638
8	0.9302	0.9313	0.9313	0.9313
9	0.9807	0.9803	0.9803	0.9803
10	0.9927	0.9930	0.9930	0.9930
11	0.9827	0.9834	0.9834	0.9834
12	0.9999	0.9999	0.9999	0.9999
13	0.8921	0.8913	0.8911	0.8911
14	0.9973	0.9977	0.9977	0.9977
15	0.9663	0.9647	0.9646	0.9646
16	0.9630	0.9650	0.9651	0.9651
17	0.9903	0.9913	0.9914	0.9914
18	0.9994	0.9996	0.9997	0.9997
19	0.8364	0.8383	0.8378	0.8378
20	0.8988	0.9058	0.9059	0.9059

Table 17: Parameter Estimates Using Bi-Squared Weighting Function

Iteration	Intercept	Slope
1	2.7807	1.1873
2	2.7916	1.1854
3	2.7929	1.1853
4	2.7931	1.1853
5	2.7931	1.1853
6	2.7931	1.1853
7	2.7931	1.1853
8	2.7931	1.1853
9	2.7931	1.1853
10	2.7931	1.1853

Here, convergence to four decimal places is achieved for scaled deleted residuals, weights, and parameter estimates after no more than six iterations.

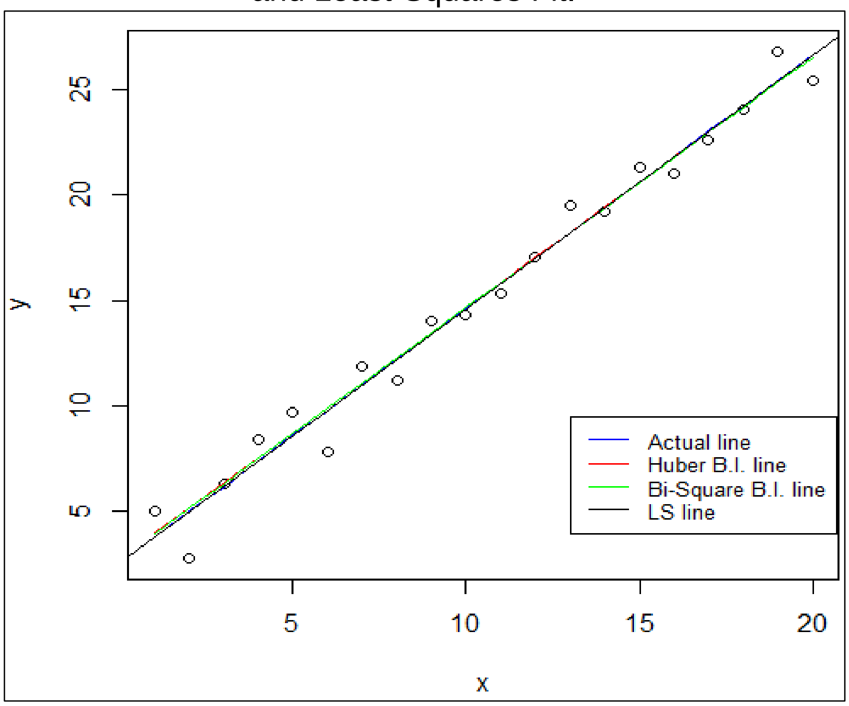
Unlike the Huber function, the Bi-Weight function once again down-weighted every observation. However, three of the four most down-weighted observations from the Bi-Squared function, which were weighted below 0.9 in the last iteration (namely, observations 2 (w_2 = 0.6382), 6 (w_6=.7479), 13

(w_{13}=0.8911), and 19 (w_{19}=0.8378)), were also the only observations down-weighted when the Huber function was used.

1.5.2.3 Comparison of Bounded Influence Weighting Function Approaches

In Figure 9 below, the actual line is plotted with the Bounded Influence lines obtained using the Huber and Bi-Weight functions, as well as the common NSLR line. Note that the fitted lines, as with the Theil-Sen and Wilcoxon lines, are again virtually indistinguishable in this plot.

Figure 9: B.I. Huber fit vs B.I. Bi-Square fit, including Actual Line and Least-Squares Fit.



Parameter estimates obtained from the two weighting functions are given in Table 18 below.

Table 18: Comparison of NSLR with Huber and Bi-Weight function IRLS parameter estimates for Simple Example

Parameter	Actual	Huber	Bi-Square	Least-Squares
Intercept	2.6	2.8142	2.7931	2.5954
Slope	1.2	1.1850	1.1853	1.1996

The slope parameter estimates for both Bounded Influence lines within 1.25% of the actual slope, and the intercept estimates are within 8.25% of the actual parameter value. Moreover, just as with the IRLS parameter estimates, the Bounded Influence line parameter estimates are within 1% of each other, with the slope estimates being equal to three decimal places.

1.6: Bootstrap Regression

When the number of observations in a sample is small, it is difficult to ascertain the distribution of the underlying population, and to make the assumptions of error normality and constant variance with acceptable certainty.

In these cases, bootstrapping methods can be used as a more robust alternative (**#7**) to least-squares. Bootstrapping regression first fits a simple linear model (1) to the data. From here, one of two procedures may be chosen.

In both procedures, a large number (say M, usually M = 300 to 3000) bootstrap samples (i.e., samples with replacement) are obtained. In the first procedure, these samples are comprised from the original set of residuals to the fitted NLSR model. These re-sampled residuals are then added to the original NLSR fitted values to create new bootstrap samples of the response. NLSR models are then fit to each of these bootstrapped samples of the response, and estimates for the regression parameters are obtained. The final estimate is obtained by taking the arithmetic mean of the parameter estimates for each of the bootstrapped samples. Clearly, this approach will result in a final bootstrap estimate essentially equal to the NLSR estimates. The value in the bootstrap approach using the residuals is in the formation of potentially more valid confidence interval estimates than those obtained when assuming the errors are NID(0, σ^2).

1.6.1: Residual Bootstrap

A description of this residual-only bootstrapping approach is initiated by consideration of model (1), given by:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ for } i = 1,2,\dots,n,$$

With $r_i = y_i - \hat{y}_i$, where $\hat{y}_i$ is the least-squares fit for observation $i, for\ i = 1,2, \ldots, n$, the residual-only bootstrapping procedure involves taking M bootstrap samples (i.e., samples with replacement) of the r_i – each residual bootstrap sample will be represented by a column vector $\vec{\boldsymbol{e}}_j$, with j = 1,2,…,M, whose components consist of sampled r_i (with replacement). Then, for each bootstrap sample j, define:

$$\vec{\boldsymbol{y}}^j = \vec{\hat{\boldsymbol{y}}} + \vec{\boldsymbol{e}}_j, \text{ for j = 1,2,…,M,} \tag{28}$$

where $\vec{\hat{\boldsymbol{y}}}$ is a column vector containing the original least-squares fitted values. Least-squares models are then fitted to each $\vec{\boldsymbol{y}}^j$ (with the original paired predictor values), which yields intercept and slope parameter estimates $\hat{b}_0^j$ and $\hat{b}_1^j$, respectively.

Finally, the overall parameter estimates are calculated as follows:

$$\hat{b}_0 = \frac{\sum_{j=1}^{M} \hat{b}_0^j}{M}, \quad \hat{b}_1 = \frac{\sum_{j=1}^{M} \hat{b}_1^j}{M}. \tag{29}$$

1.6.2: Observation Bootstrap

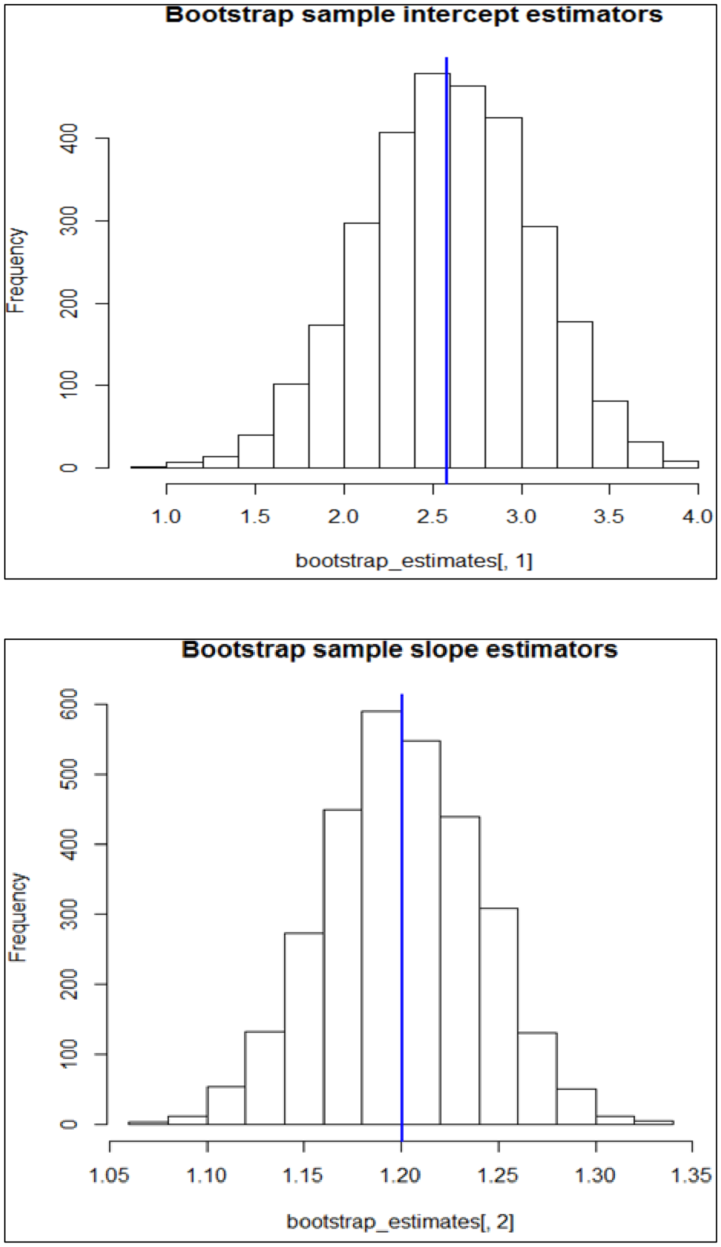
. The second procedure involves taking bootstrap samples of the multivariate observations themselves, rather than just the least-squares residuals, and then fitting least-squares models to each bootstrap sample of observations. The overall estimators for the regression parameters are calculated once again by equation (29), by taking the arithmetic mean of the estimators for all the samples. The value of this approach is that the final estimated parameters may not always simply return the original least squares estimates, but can also allow for more robust interval estimation of the parameters.

1.6.3: Bootstrap Regression – Simple Example

1.6.3.1: Bootstrap Regression – Residuals Only – Simple Example

Figure 10 below displays histograms for the M = 3,000 bootstrap intercept and slope estimates obtained through simply bootstrap sampling the original least squares residuals of the data displayed in Table 1, and using these to construct parameter estimates as described above. Note that both histograms are centered at the original least squares estimates of these parameters (2.5954 for the intercept and 1.1996 for the slope) as would be expected. In fact, to four decimal places, these bootstrap estimates are identical to the least squares estimates.

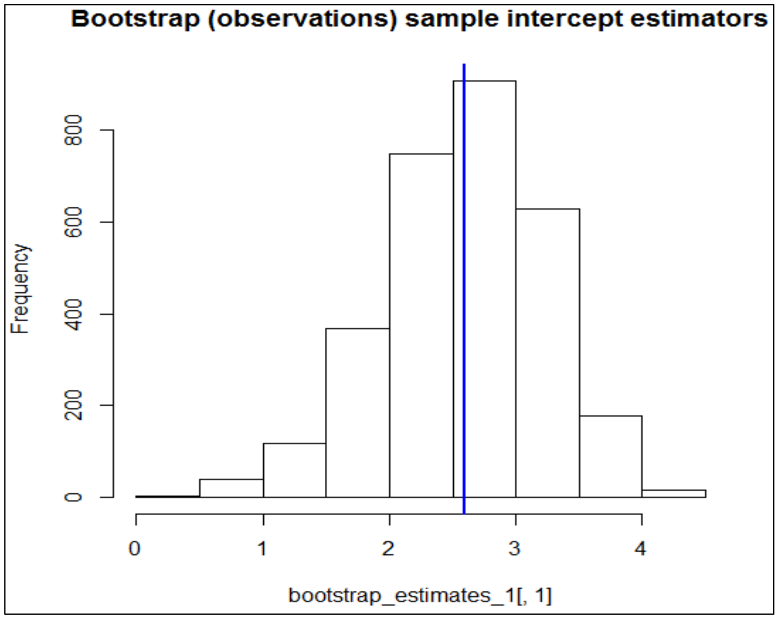
Figure 10: Histograms of Residual-Only Bootstrap Simple Example Parameter Estimates

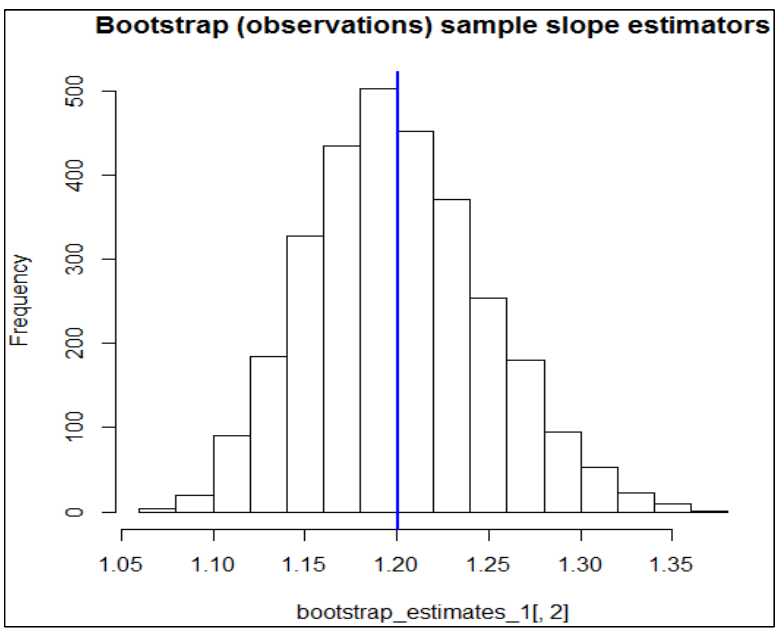


1.6.3.2: Bootstrap Regression – Complete Observations – Simple Example

Figure 11 below displays histograms for the M = 3,000 bootstrap intercept and slope estimates obtained through bootstrapping the observations, then fitting least squares estimates for each such bootstrap sample (i.e., the $\widehat{b}_0^j$ and $\widehat{b}_1^j$ for j = 1, …, M = 3000). In this simple example case, where the data is well-behaved, the results for this approach are nearly identical to those of the residual only bootstrap.

Figure 11: Histograms of Observation-Based Bootstrap Simple Example Parameter Estimates





1.6.1.3 Comparison of Bootstrap Approaches – Simple Example

A plot of the bootstrapped lines for both the residual-only and complete observation approaches, as well as the least-squares fit along and the actual model appears in Figure 12 below. As noted above, and as expected, the least squares and the bootstrap lines are nearly identical.

The respective parameter estimates are displayed in Table 19. The bootstrap estimators of slope are within 0.05% of the actual slope result, and the bootstrap estimators of the intercept are within 1% of the actual intercept value. This accuracy is primarily due to the least squares estimators for this data also having very small errors. Recall that these bootstrap estimators will converge to the least squares estimators, and depend on how well these estimate the actual parameters for the original set of data.

Figure 12: Bootstrapping (Residuals & Complete Observations) Regression vs Least Squares fit and Actual Model

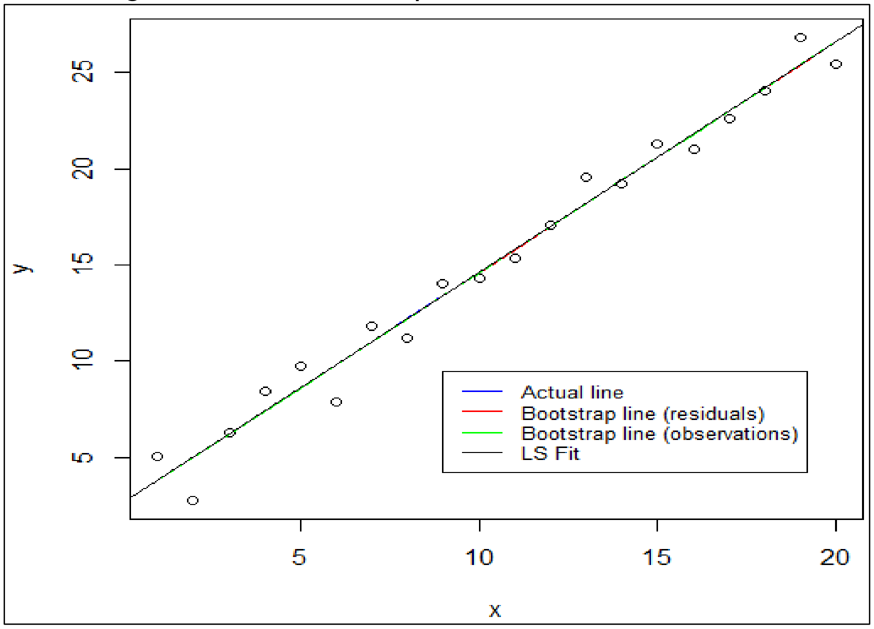


Table 19: Comparison of Bootstrapping (Residuals) Regression Model and Least Squares for Simple Example

Parameter	Actual	Bootstrap (Residuals)	Bootstrap (Observations)	Least-Squares
Intercept	2.6	2.5836	2.5922	2.5954
Slope	1.2	1.2003	1.2004	1.1996

1.7: Summary Comparison of Alternatives – Simple Example

In this simple example, fitting the data in Table 1, the data were obtained from a normally distributed error population. As a result, all of the alternative methods to NLSR still generate estimates very similar to those provided by the commonly applied NLSR. One way to compare estimators is to utilize confidence intervals as these essentially attempt to capture the inherent uncertainty in the related point estimates of the parameters of interest. More narrow confidence intervals suggest less uncertainty and a potentially better estimator; however, the most important aspect for any confidence interval is that it includes the true value of the parameter of interest. Any confidence interval will either include this parameter or it will not, but hopefully, the procedure for generating the interval will include it at a rate equal to, or very nearly equal to the stated confidence coefficient for the interval. Consequently, having some means to generate confidence intervals for all of the estimation methods will provide a frame of reference for cross-comparison of estimation approaches.

1.7.1 Traditional Least Squares Confidence Intervals

The NLSR model allows for construction of confidence intervals on the regression parameters by assuming that the model errors are normally and independently distributed with zero mean and constant variance. Under this assumption, the least squares parameter estimates divided by estimates of their standard errors follow a Student's t-distributions with (n-2) degrees of freedom. Consequently, NLSR confidence intervals are centered at the point estimates for the parameters (least-squares estimates) with a margin of error equal to the appropriate t-distribution percentile multiplied by an estimate of the respective parameter's standard error. This confidence interval method is particularly well-suited in the case of the simple example data in Table 1 because the errors in this data set come from a normally and independently distributed populations with zero means and the same variance.

Confidence intervals for $\beta_k, k = 0,1$ are given by (**#5**) as:

$$b_k \pm t_{\left(1-\alpha/_2,\ n-2\right)} * \hat{\sigma}\{b_k\},\ \text{k = 0,1,} \quad (30)$$

Where b_k = the respective least squares estimate of β_k, k = 0, 1; $t_{(p,\ df)}$ = the p^th^ percentile of a Student's t-distribution with df degrees of freedom; and estimates of the standard errors of the least squares estimates are given as:

$$\hat{\sigma}\{b_k\} = \sqrt{MSEc_{kk}}, \quad (31)$$

where

$$\text{MSE = Mean Square Error = } \frac{\sum_{i=1}^{n}(r_i)^2}{n-2}, \quad (32)$$

with r_i as given in (2), and c_{kk} the k^{th} diagonal element of $(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}$, with the matrix $\boldsymbol{X}$ as defined in (8) with p = 2 and k = 0, 1.

Performing the calculations shown in (30-32) with an α = 0.05 results in the 95% NLSR confidence intervals shown in Table 20 below:

Table 20: 95% Confidence Intervals on Least-Squares Parameters

Parameter	Lower Bound	Upper Bound
Intercept	1.5336	3.6573
Slope	1.111	1.2883

Not only do the actual parameter values of 2.6 and 1.2 lie inside their respective confidence intervals, but also all the results for the other estimation approaches considered do as well.

Section 1.7.1.2: Theil-Sen Confidence Intervals

Because the Theil-Sen slope estimator is the median of pairwise slopes, a confidence interval on the slope parameter can be constructed by using the percentile approach to finding a confidence interval for a median (**#8**). This procedure is initialized by finding $C_{\alpha/2}$, the $(\frac{\alpha}{2})^{th}$ lower percentile of the binomial distribution with parameters N = $\binom{n}{2} = \binom{20}{2} = 190$ (because there are $\binom{n}{2}$ total pairwise slopes) and p = 0.5. Then, define U and L as below:

$$U = N - C_{\alpha/2} \tag{33}$$

$$L = 1 + C_{\alpha/2} \tag{34}$$

Then a $(1-\alpha)\%$ confidence interval on the slope can be constructed by using the L^{th} and U^{th} order statistics of the pairwise slopes as lower and upper bounds of the confidence interval, respectively. Therefore, if x_L and x_U are the respective L^{th} and U^{th} order statistics of pairwise slopes, the $(1-\alpha)\%$ confidence interval for $\beta_1$is given by (x_L, x_U). It should be noted that because the binomial distribution is discrete, exact $(1-\alpha)\%$ confidence intervals often cannot be constructed for any given value of α. However, conservative approximations of $(1-\alpha)\%$ confidence intervals can be created by using the largest C_b such that $b \leq \frac{\alpha}{2}$ in place of $C_{\alpha/2}$ in equations (33) and (34) above.

For the binomial distribution with N = $\binom{20}{2}$ = 190 and p = 0.5, $C_{0.0249}$ = 81 (Note: $C_{0.0347}$ = 82) is the desired lower percentile value. Therefore, an approximate 95% confidence interval (with $\alpha \approx .05$) can be constructed by using 81 in place of $C_{\alpha/2}$ in equations (33) and (34) above. This interval is given by (x_{82}, x_{109}), which yields:

$$(\ 1.12568, 1.22094).$$

This ~95% confidence interval clearly captures the actual slope parameter of 1.2.

A $(1-\alpha)\%$ confidence interval on the intercept parameter can be constructed in a similar way because the intercept estimator is the median of observations regressed on the slope estimator. Recall from equation (14) that the Theil-Sen intercept estimator is given by:

$$b_0 = median(y_k - b_1 x_k), \text{ for } k = 1,2, \dots n.$$

Here, the median is taken from a sample of $n = 20$ observations. Therefore, a ~95% confidence interval can be constructed on the intercept parameter in a similar fashion to that described above; however, this time using binomial percentiles with $N = n = 20$ and $p = .5$. From this distribution, $C_{0.0207}$ = 5 and $C_{0.0577}$ = 6; therefore, a (1-2*0.0207)% = 95.86% confidence interval on the intercept can be computed as (x_6, x_{15}), where x_i is the i^{th} order statistic of the $(y_k - b_1 x_k)$, k = 1, …, n = 20, values. This yields the interval:

$$(2.404225, 3.704527),$$

which captures the true parameter of 2.6. Table 21 below summarizes the Theil-Sen confidence intervals:

Table 21: ~95% Confidence Intervals on Theil-Sen Parameters

Parameter	Lower Bound	Upper Bound
Intercept	2.4042	3.7045
Slope	1.1257	1.2209

Section 1.7.1.3: Wilcoxon Confidence Intervals

The Wilcoxon intercept estimator is a median, so a confidence interval for the intercept can be constructed in a similar manner to that used for the slope and the intercept in the Theil-Sen model. Let $C_{\alpha/2}$ be the $(\frac{\alpha}{2})^{th}$ lower percentile of the binomial distribution with parameters N = 20 and p = 0.5. Define U and L once again as given in equations (33) and (34). L and U give the order statistics of the twenty values $(y_i - b_1 x_i), i = 1,2,\dots,20,$ which serve as the lower and upper endpoints of the confidence interval, respectively. When calculating the Theil-Sen Intercept ~95% confidence interval in the previous section, we found L and U to be 6 and 15, respectively. Therefore the ~95% confidence interval for the Wilcoxon median will have endpoints equal to the 6th and 15th order statistics of the $(y_i - b_1 x_i)$ sample, which gives:

$$(2.3929,\ 3.6891).$$

For the slope parameter,(**#9**) proposes a confidence interval that uses the jackknife technique to estimate the variance of b_1. The $(1-\frac{\alpha}{2})$% confidence interval for β_1 is given by:

$$b_1 \pm t_{m-1,\alpha/2} * \sqrt{\widehat{var(b_1)}}. \quad (35)$$

In equation (35) above, "m" is the number of unique pseudo-values of b_1. Each of the n=20 psuedo-values of b_1, which we will denote by $\hat{b}_{1(i)}$ for $i\ =$

$1,2, \ldots. n$, is obtained by calculating the slope parameter estimate on the sample while excluding the i^{th} data point. Let $b_{1(i)}$ denote the slope estimate derived from excluding the i^{th} data point, for $i = 1,2, \ldots, n$; then the i^{th} pseudo-value is given by:

$$\hat{b}_{1(i)} = nb_1 - (n-1)b_{1(i)} \tag{36}$$

$\widehat{var(b_1)}$ is calculated as the sample variance of these pseudo-values.

This yields a 95% confidence interval on the slope parameter given below:

$$(1.0062, 1.3429).$$

Section 1.7.1.4: IRLS Confidence Intervals

Confidence Intervals can be constructed for the IRLS parameters in a similar fashion to the least-squares confidence intervals because of the similar nature of their calculation. Under normal least squares, confidence intervals were obtained using the estimators as interval midpoints, and margins of error – given as products of t-distribution percentiles and estimates of the standard errors of the parameter estimators – were added and subtracted from the midpoints, as given in equations (30), (31) and (32). With the inclusion of weights, the estimates of standard error differ slightly. Under IRLS, we have the following:

$$\hat{\sigma}\{b_{k,w}\} = \sqrt{MSE_w c_{kk,w}}, \tag{37}$$

where $b_{k,w}$ is the respective IRLS estimate of β_k, k = 0, 1, and $c_{kk,w}$ is the k[th] diagonal element of $(\boldsymbol{X}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{X})^{-1}$, with the matrix $\boldsymbol{X}$ as defined in (8) and the matrix $\boldsymbol{W}$ as defined immediately following equation (20).

To estimate the standard error under IRLS, equation (38) below was used in place of (31):

$$MSE_w = Mean\,Squared\,Error = \frac{\sum_{i=1}^{n} w_i (r_i)^2}{n-2}. \tag{38}$$

The w_i and r_i values are the weights and residuals (respectively) at the last iteration of the estimation process (i.e., the first iteration obtaining desired convergence).

Performing the calculations shown in equations (30 and 37-38) with α = 0.05 results in the 95% IRLS confidence intervals shown in Table 22 below:

Table 22: 95% Confidence Intervals on IRLS Parameters

Parameter	Lower Bound	Upper Bound
Intercept (Huber)	1.7778	3.8323
Slope (Huber)	1.1013	1.2705
Intercept (Bi-Squared)	1.7388	3.8385
Slope (Bi-Squared)	1.0988	1.2725

All four confidence intervals capture the respective true parameter values they are meant to estimate.

Section 1.7.1.5: Bounded Influence Confidence Intervals

Since both IRLS and Bounded Influence estimators are simply weighted least-squares estimators, confidence intervals for Bounded Influence parameters can be calculated using the same procedure used for IRLS intervals. Since the weights will differ, so will the residuals generated by each fit; however, all the same respective formulas given in equations (30 and 37-38) apply.

Confidence Intervals on the Bounded Influence parameter estimates are given Table 23 below:

Table 23: 95% Confidence Intervals on Bounded Influence Parameters

Parameter	Lower Bound	Upper Bound
Intercept (Huber)	1.7883	3.8402
Slope (Huber)	1.1005	1.2696
Intercept (Bi-Squared)	1.7413	3.845
Slope (Bi-Squared)	1.0982	1.2723

Section 1.7.1.6: Bootstrap Confidence Intervals

A 100$(1-\alpha)$% confidence interval can readily be constructed on the bootstrap parameters by using the $\frac{\alpha}{2}$ and $(1-\frac{\alpha}{2})$ percentiles of the parameter estimates obtained from the bootstrap samples. This is because the bootstrap method generates an empirical sampling distribution for each of the parameters. For a 95% confidence interval, the 2.5 and 97.5^{th} percentiles of the parameter estimates from the generated bootstrap samples provide the confidence intervals for intercept and slope parameters are given in Table 24 below.

Table 24: 95% Confidence Intervals on Bootstrap Parameters

Method	Residuals		Observations	
Parameter	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Intercept	1.6155	3.4946	1.2506	3.6940
Slope	1.1218	1.2772	1.1145	1.3026

The actual parameters of 2.6 (intercept) and 1.2 (slope) are captured in all four of the confidence intervals above.

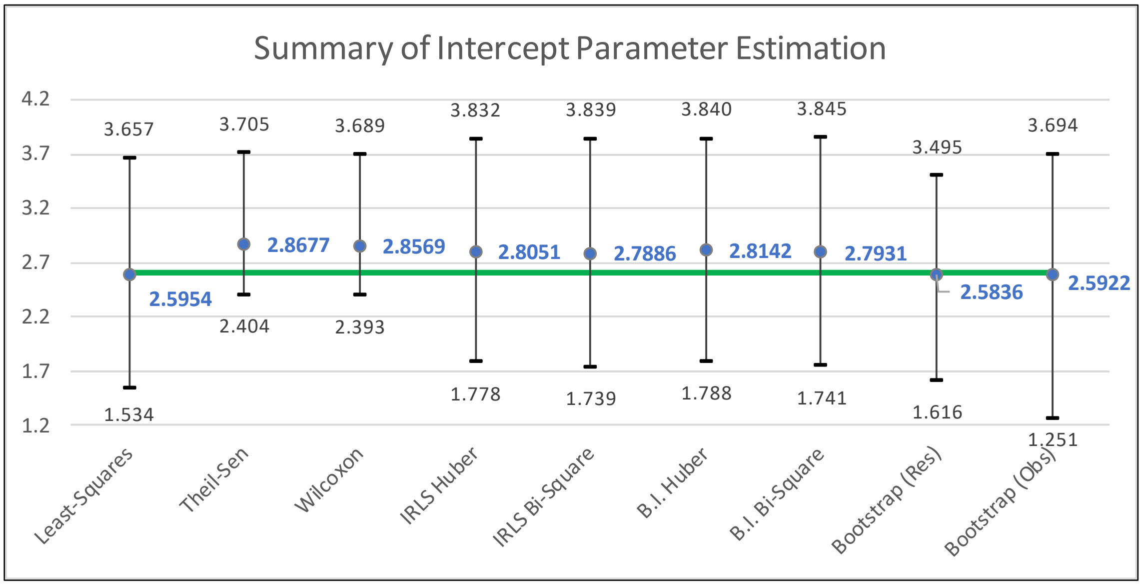
1.7.2 Summary and Comparison of Confidence Intervals

Charts showing the confidence interval bounds for both parameters using all methods are given in Figure 13 below. A table summarizing all confidence bounds is given below in Table 25 as well. Observe that every confidence interval captures the parameter that it is meant to estimate. This suggests that the alternative methods to least-squares are sufficient even in cases where NSLR (i.e., least squares) is supposed to be optimal, such as with this dataset, in which the errors are normally and identically distributed with mean zero and constant variance.

It is worth noting that the Theil-Sen and Wilcoxon confidence intervals are the narrowest for the intercept parameter. In addition, the Bootstrap-Residuals interval is also narrower than the least squares interval for the intercept parameter.

For the slope parameter, the Wilcoxon jackknife interval is clearly the widest. However, the Theil-Sen and Bootstrap-Residual intervals again are not as wide as the least squares interval.

Figure 13: Confidence Interval Comparisons for Simple Example



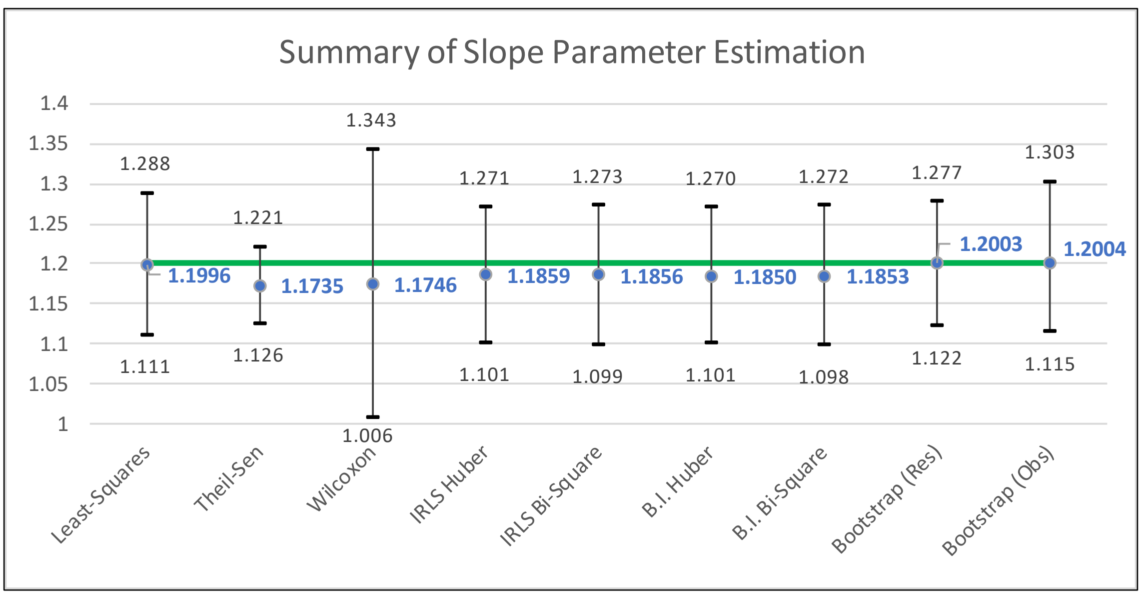


Table 25: Comparison of Confidence Intervals on Parameter Estimates for Simple Example

Regression Method	Intercept			Slope		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	1.5336	2.5954	3.8385	1.1110	1.1996	1.2883
Theil-Sen	2.4042	2.8677	3.7045	1.1257	1.1735	1.2209
Wilcoxon	2.3929	2.8569	3.6891	1.0062	1.1746	1.3429
IRLS (Huber)	1.7778	2.8051	3.8323	1.1013	1.1859	1.2705
IRLS (Bi-Square)	1.7388	2.7886	3.8385	1.0988	1.1856	1.2726
Bounded Influence (Huber)	1.7883	2.8142	3.8402	1.1005	1.1850	1.2696
Bounded Influence (Bi-Square)	1.7413	2.7931	3.8449	1.0982	1.1853	1.2723
Bootstrap (Residuals)	1.6155	2.5836	3.4946	1.1218	1.2003	1.2772
Bootstrap (Observations)	1.2506	2.5922	3.6940	1.1145	1.2004	1.3026

Section 2: Comparison of Alternatives in the Presence of Outliers

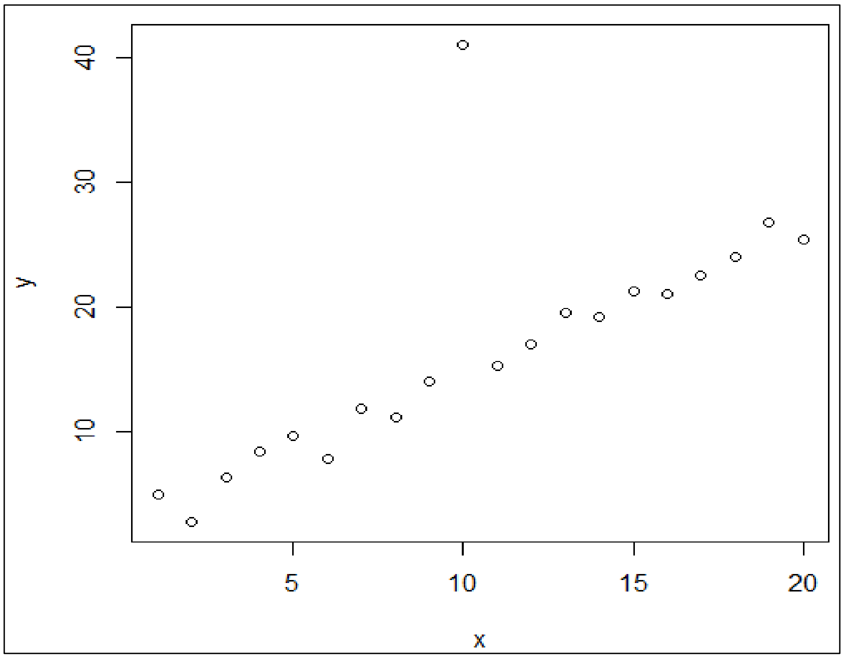
In Section 1, the various estimation methods were compared using a simple and well-behaved example. The NLSR estimators are known to be optimal in this situation; however, as was discussed in the previous section, all the alternatives produced results reasonably close to the NLSR result.

In this section, how NLSR and all the considered alternatives perform when the data is less well-behaved will be examined. One of the original observations will be manipulated to become an outlier in either the predictor space, the response variable space, or both.

2.1: Outliers in the Response Variable Space Only

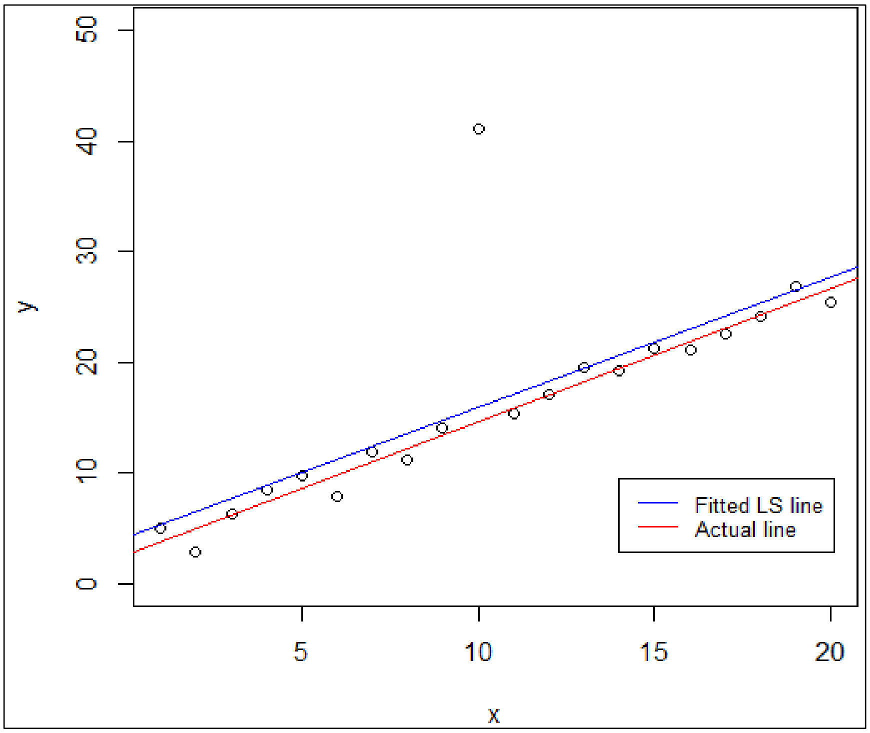
In order to examine the results of injecting outliers in the response variable only, the original observation (10, 14.31045) was altered by adding the maximum of the original response variables to the response value of this observation, resulting in the new paired observation (10, 41.1186). A scatterplot of the "new" (and now not so well-behaved) data is provided in Figure 14 below.

Figure 14: Scatterplot of Data with Response Outlier at X = 10



Using the methods described in Section 1, each estimation method was performed on this data set now including a single outlier in the response space. As displayed in Figure 15, the Least-Squares fit has a nearly identical slope to the actual line, but the fitted intercept (and overall line) has received a non-negligible upward shift. The similar slope is due to the outlier appearing at a very low leverage point – near the middle of the predictor variable values. Note that the general quality of the fit is poor as virtually all of the ordered pairs now fall below the fitted line.

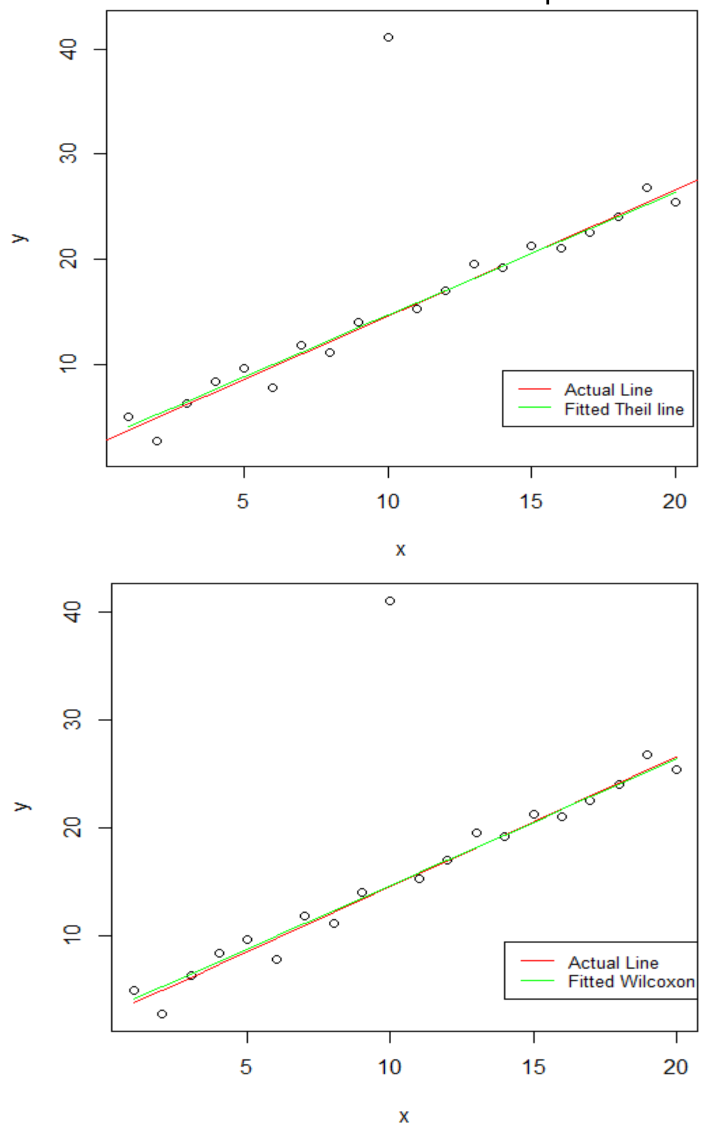
Figure 15: Least Squares fit with Response Outlier at X=10



As displayed in Figure 15, the Least-Squares fit has a nearly identical slope to the actual line, but the fitted intercept (and overall line) has received a non-negligible upward shift. The similar slope is due to the outlier appearing at a very low leverage point – near the middle of the predictor variable values. Note that the general quality of the fit is poor as virtually all of the ordered pairs now fall below the fitted line.

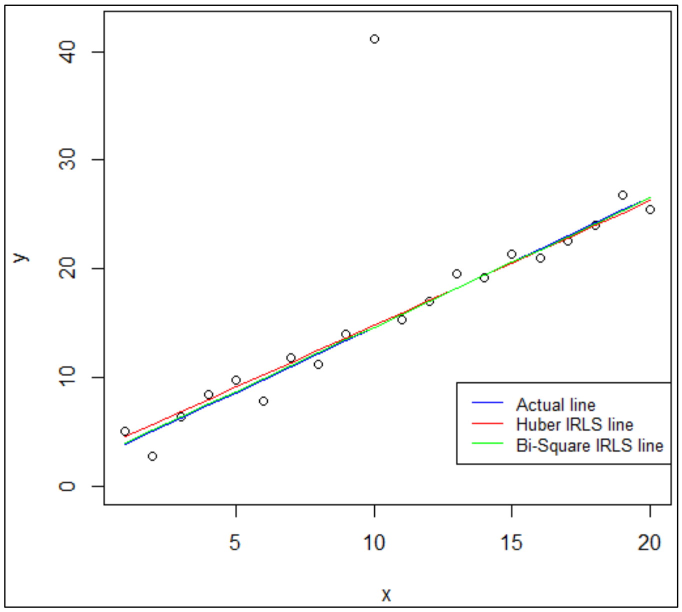
In contrast to this, Figure 16 shows that both the Theil-Sen and Wilcoxon lines are nearly indistinguishable from the actual line, and have hardly been influenced at all by the presence of the outlier. This observation demonstrates the value of these estimation approaches when such extreme observations might appear in the data.

Figure 16: Theil-Sen and Wilcoxon fits with Response Outlier at X=10



As shown in Figure 17, both IRLS and Bounded Influence lines, regardless of weighting function, also appear to provide fits similar to the Theil-Sen and Wilcoxon approaches. Table 26 below shows the final iteration weights for each of the twenty observations for all of the fits displayed in Figure 17, and "significantly down-weighted" observations (those with weights smaller than 0.9) are highlighted in red.

Figure 17: IRLS and Bounded Influence fits with Response Outlier at X=10



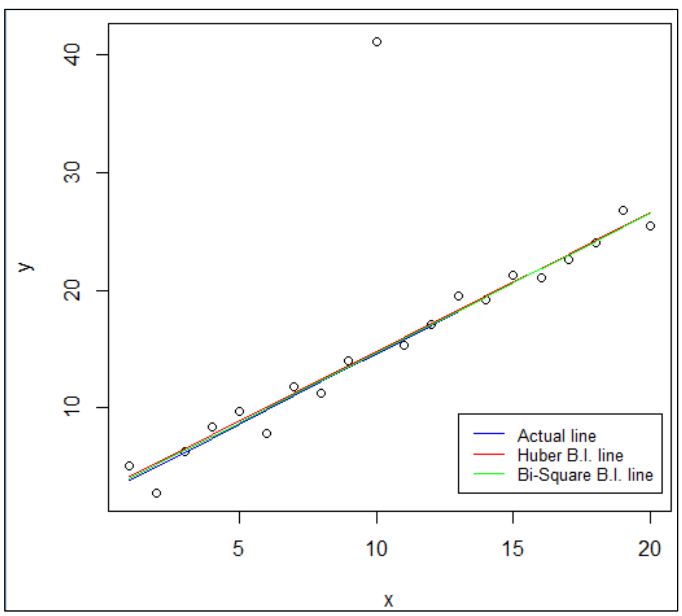


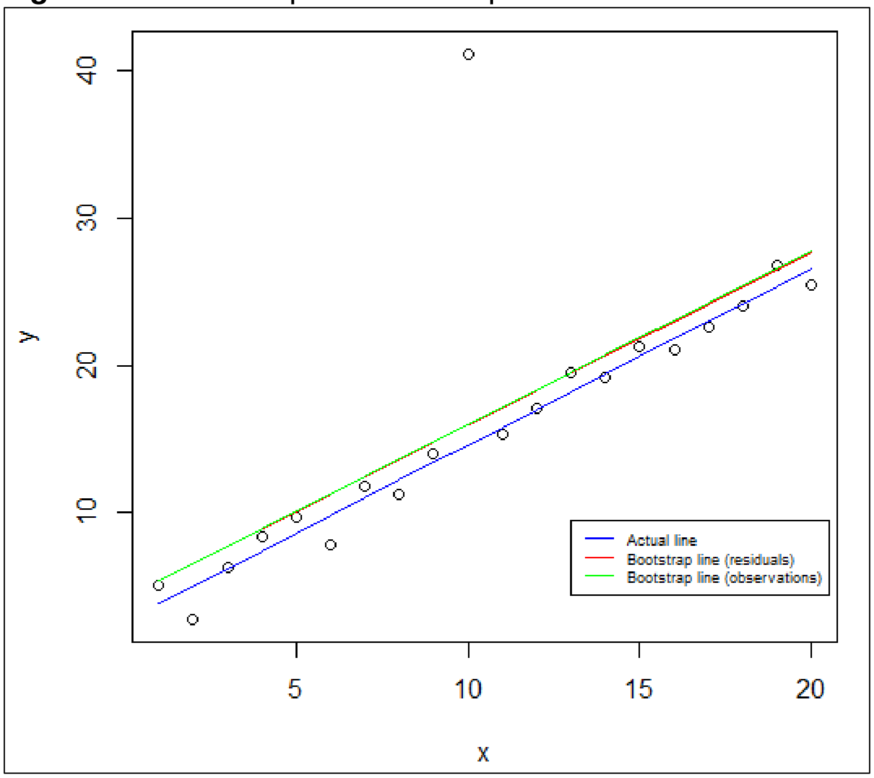
Table 26: IRLS and Bounded Influence Final Iteration Weights with Response Outlier at X = 10

	IRLS		Bounded Influence	
X	Huber	Bi-Square	Huber	Bi-Square
1	1	0.9347	1	0.9189
2	0.6267	0.6695	0.6251	0.6609
3	1	0.9999	1	0.9999
4	1	0.9537	1	0.9512
5	1	0.9402	1	0.94
6	0.7228	0.7464	0.7544	0.7645
7	1	0.9662	1	0.9681
8	1	0.9288	1	0.9352
9	1	0.9814	1	0.983
10	0.0605	0	0.0677	0
11	1	0.9818	1	0.9838
12	1	1	1	1
13	1	0.8928	1	0.9028
14	1	0.9972	1	0.9974
15	1	0.9679	1	0.9693
16	1	0.9657	1	0.9662
17	1	0.9915	1	0.9911
18	1	0.9995	1	0.9994
19	1	0.8672	1	0.8551
20	1	0.9247	1	0.9101

For both approaches, the Huber weighting function has down-weighted the observation at X = 10 significantly – especially in comparison to the rest of the observations – and the Bi-Square function has completely thrown out this observation (i.e., it has a weight of zero). An interesting note is that both the IRLS and Bounded Influence approaches have down-weighted the exact same observations under the Huber function – that is, observations with X = 2, 6 and 10. In the original dataset, observations with X = 2 and X = 6 were down-weighted by the Huber function, so the presence of the Response outlier at X = 10 has not caused any of the other observations to be down-weighted which were not originally down-weighted. A similar comparison cannot directly be made between the results of IRLS and Bounded Influence under the Bi-Square function, because the criteria used for identifying significant down-weights (those being smaller than 0.9) is arbitrary, however, the down-weighted observations with this weighting function are similar to those down-weighted with the original data, which were at X = 2, 6, 13, and 19.

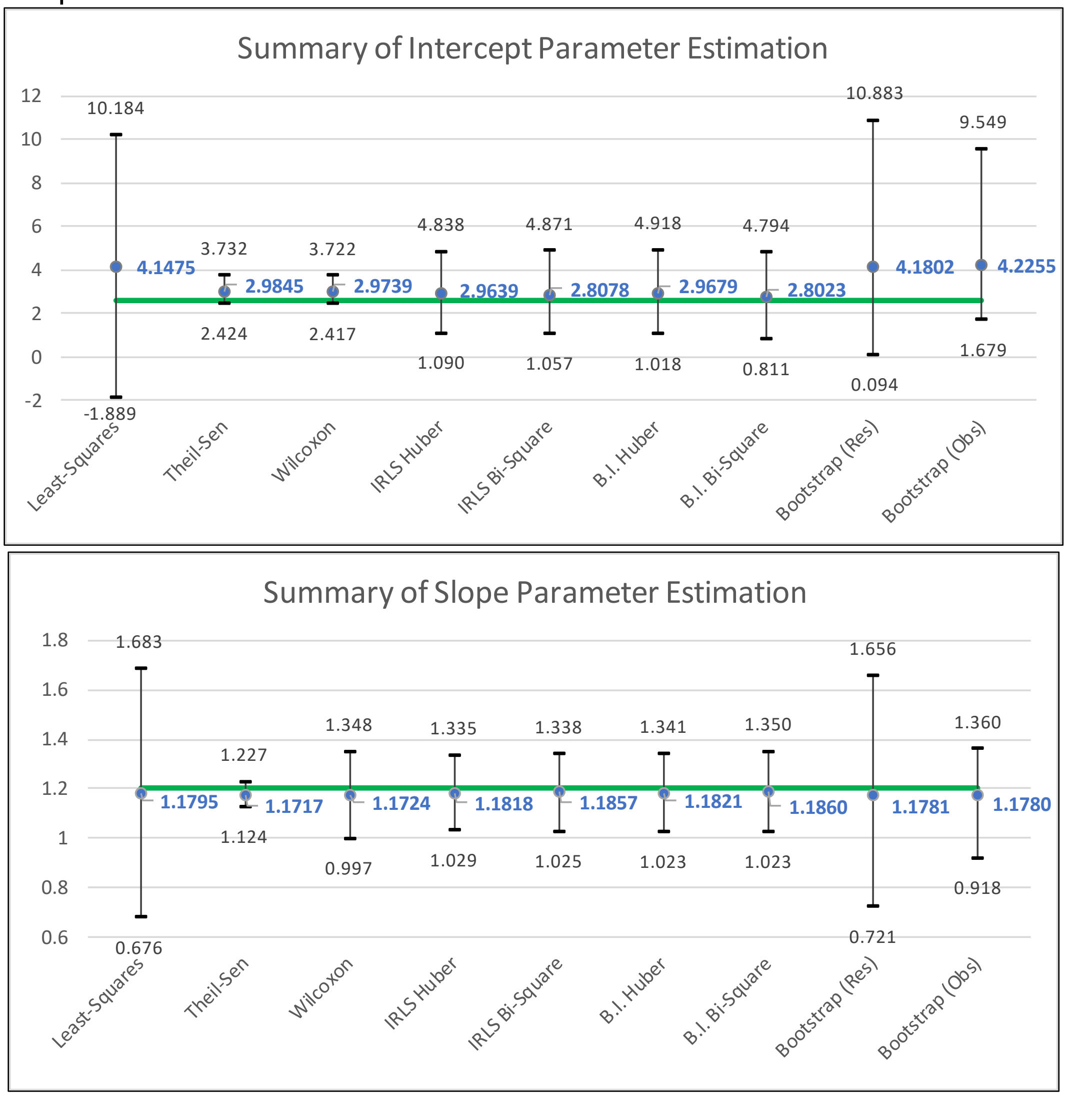
Figure 18 shows that the Bootstrap approaches, whether using residuals or the full observations for resampling, tend to provide fits similar to that of the least-squares approach. This result indicates that neither Bootstrap approach is particularly robust in the presence of such outliers.

Figure 18: Bootstrap fits with Response Outlier at X=10



All of the estimated parameters for the lines displayed in Figures 15-18 (and ~95% confidence intervals on those parameter estimates) are given in Tables A1 (intercept estimates) and A2 (slope estimates) in Appendix 1. All of these intervals are graphically displayed in Figure 19.

Figure 19: Confidence Intervals on Intercept and Slope Parameter with Response Outlier at X=10

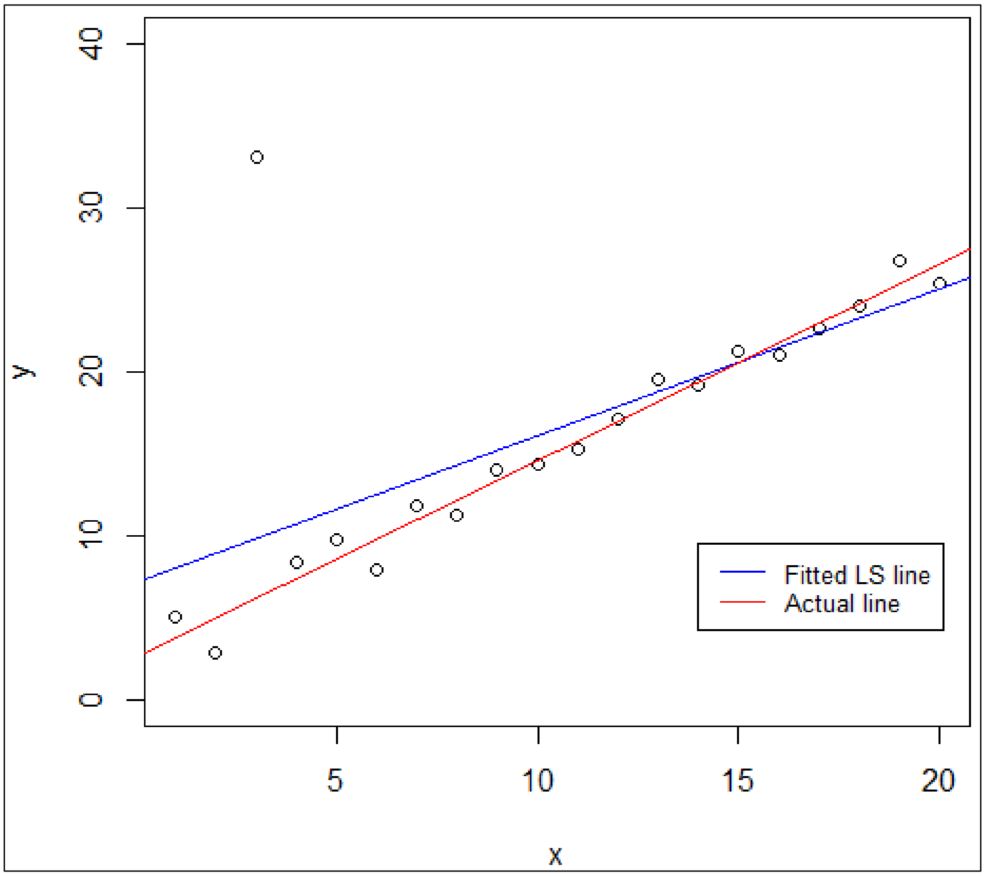


Observe that the slope estimates obtained once the outlier has been incorporated into the dataset have only changed minimally from those obtained in Section 1. The largest percentage change from the original slope estimator of any method is less than 2%, which occurred for the Bootstrap method on the observations. However, the confidence intervals for the slope estimate have noticeably widened for all methods except Theil-Sen.

The intercept point estimates have experienced larger changes.   For both bootstrapping methods and least-squares they have increased by more than 50% of their original values, while they have changed minimally for the other methods. The ~95% confidence intervals for least-squares and both bootstrap sets of estimators have widened much further than those for the other methods as well.

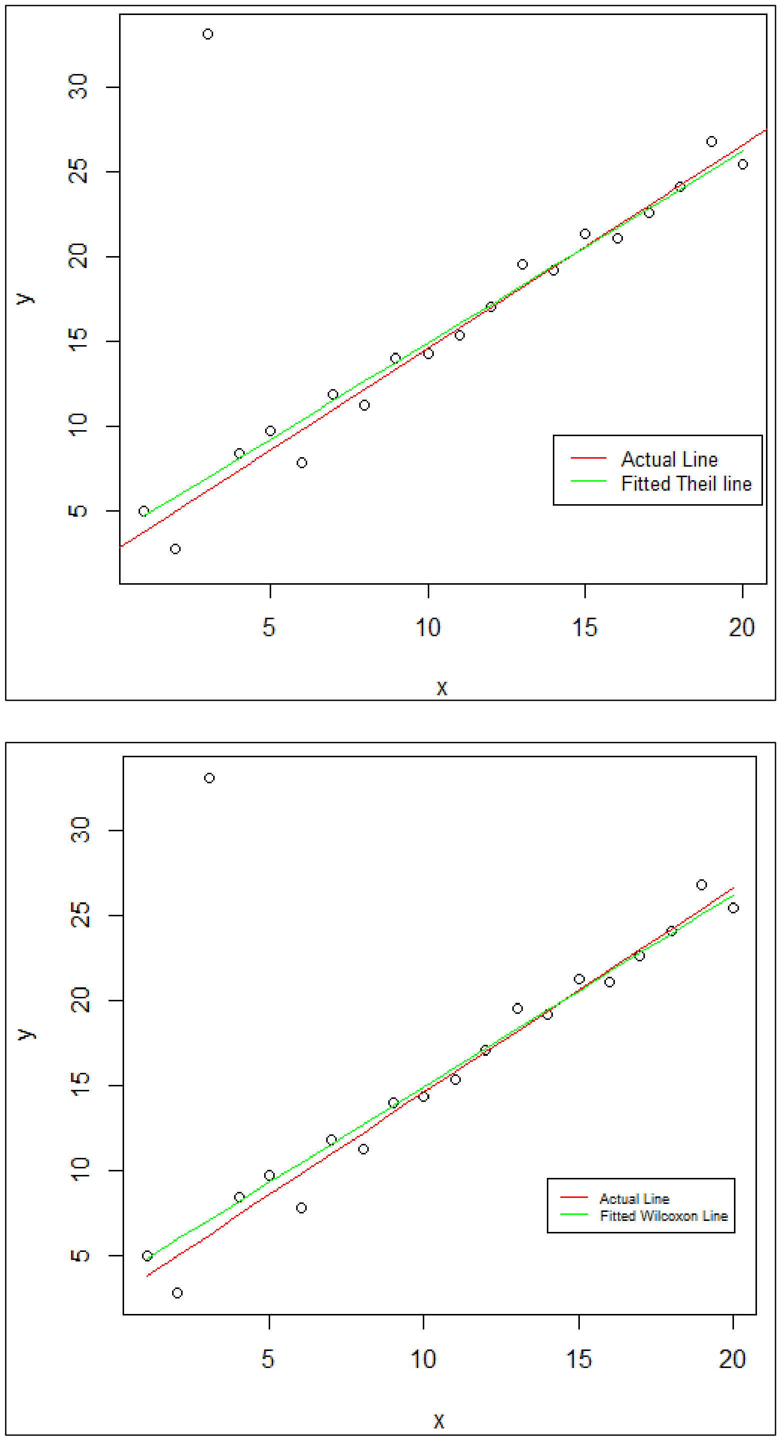
However, the robustness of the slope estimator for all methods in the presence of this outlier is primarily because the outlier was introduced near the mean of the predictor values ($\bar{X} = 10.5$). Recall that the predictor value associated with the outlying response is 10. If instead, an outlier is introduced at the observation with X = 3 by adding the maximum of all response values to the original response at X =3, producing a response value of 33.1263, then as can be seen in Figure 20 below, the least-squares line is now not merely shifted higher, but the slope is also noticeably flatter. The least-squares line again provides a relatively poor fit of the majority of the observations.

Figure 20: Least Squares Fit with Injected Response Outlier at X=3



As displayed in Figure 21, similar to when the response outlier was at X = 10, with it now at X =3, the Theil-Sen is notably robust and does not differ greatly from the actual line. The same can be said for the Wilcoxon line.

Figure 21: Theil-Sen and Wilcoxon fits with Response Outlier at X=3



As shown in Figure 22, IRLS lines under both Huber and Bi-Square schemes are barely distinguishable from the true line in the presence of this outlier. The Bounded Influence fits appear markedly robust as well, with the Bi-Square approach slightly outperforming the Huber approach.

The final iteration weights are shown in Table 27. Note that for both the IRLS and Bounded Influence estimators, the Bi-Square weighting function has again given the outlier a weight of zero, essentially removing it from the data set. The final Huber weight for this point is also near zero at ~0.05 for both IRLS and Bounded Influence approaches.

Figure 22: IRLS and Bounded Influence Fits with Response Outlier at X=3

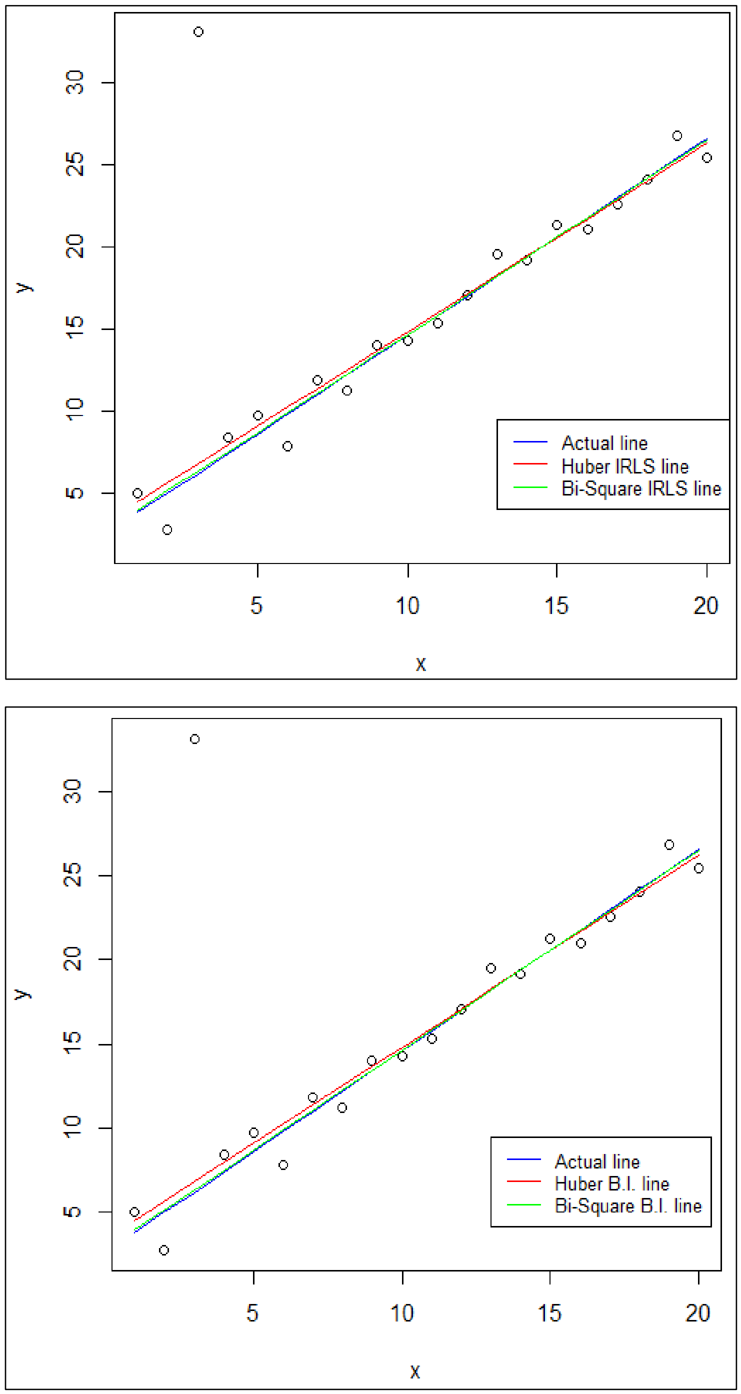


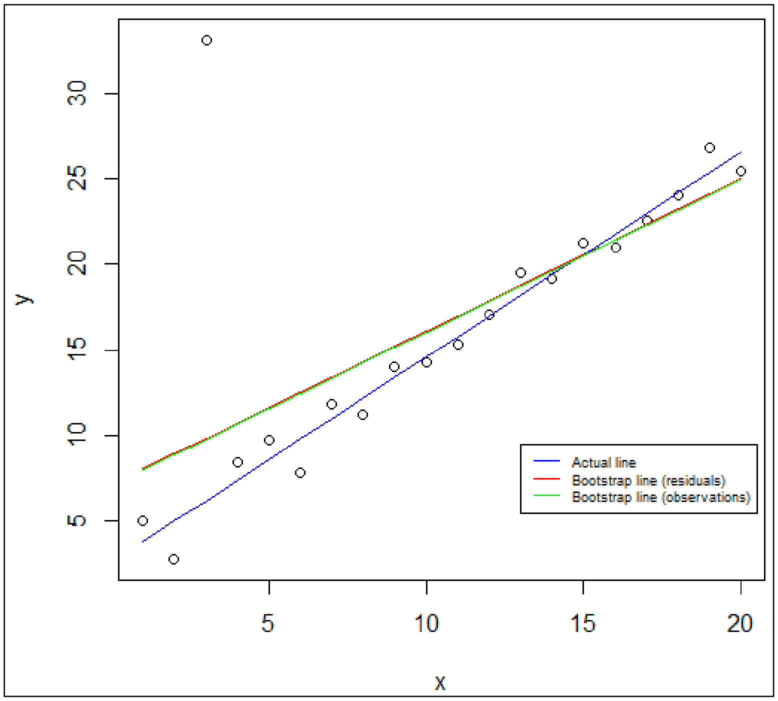
Table 27: IRLS and Bounded Influence Final Iteration Weights with Response Outlier at X = 3

X	IRLS		Bounded Influence	
	Huber	Bi-Square	Huber	Bi-Square
1	1	0.9354	1	0.9173
2	0.4487	0.7064	0.437	0.696
3	0.0487	0	0.0524	0
4	1	0.954	1	0.9516
5	1	0.9414	1	0.9418
6	0.5383	0.7750	0.5463	0.7951
7	1	0.9665	1	0.9690
8	0.9838	0.9387	0.9804	0.9461
9	1	0.9812	1	0.9834
10	1	0.9938	1	0.9948

Table 27: IRLS and Bounded Influence Final Iteration Weights with Response Outlier at X = 3

X	IRLS		Bounded Influence	
	Huber	Bi-Square	Huber	Bi-Square
11	1	0.9849	1	0.9871
12	1	0.9999	1	0.9999
13	1	0.8992	1	0.9127
14	1	0.9979	1	0.9982
15	1	0.9693	1	0.9722
16	1	0.9701	1	0.9717
17	1	0.9928	1	0.9928
18	1	0.9997	1	0.9996
19	0.7725	0.8773	0.7246	0.8710
20	1	0.9323	1	0.9214

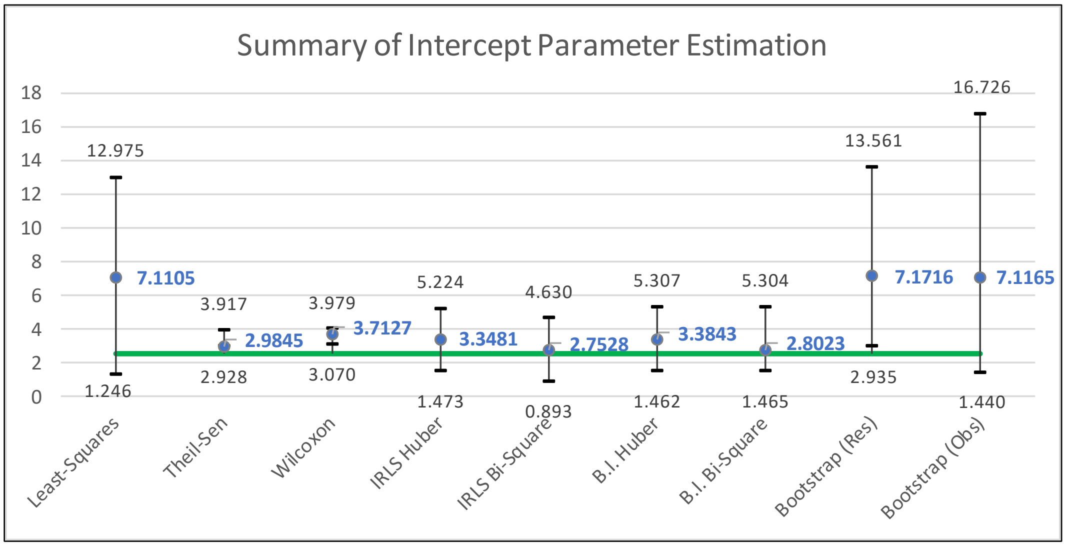
Figure 23: Bootstrap Fits with Response Outlier at X=3

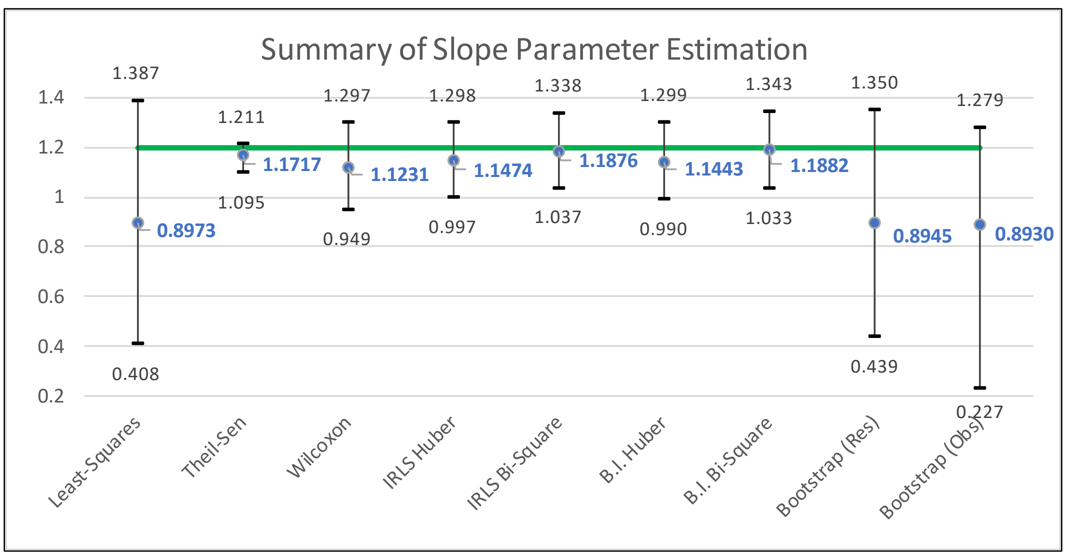


In comparing the fits in Figure 23 with the fit displayed in Figure 20, it becomes clear that both bootstrap methods produce nearly identical results with each other, as well as to least squares. Consequently, it appears bootstrapping methods fail to be robust estimators in this situation. This lack of robustness for the bootstrap estimators was suggested in consideration of the intercept estimates these approaches produced when the response outlier was at X = 10. These results with the outlier at X = 3 confirm that lack of robustness for these estimation approaches.

Newly obtained parameter estimates with the introduction of this outlier at X = 3, along with their approximate 95% confidence intervals can be found both in Tables A3 (intercept) and A4 (slope) in Appendix 1.   These intervals are also displayed in Figure 24.

Figure 24: Confidence Intervals on Intercept and Slope Parameter with Response Outlier at X=3





The confidence intervals for all regression methods have captured the true intercept parameter except for the Theil-Sen and Residual Bootstrap estimators. Interestingly, for IRLS and Bounded Influence Regression, the intercept estimators are much closer to those from the original dataset when the Bi-Squared function is used rather than the Huber function.

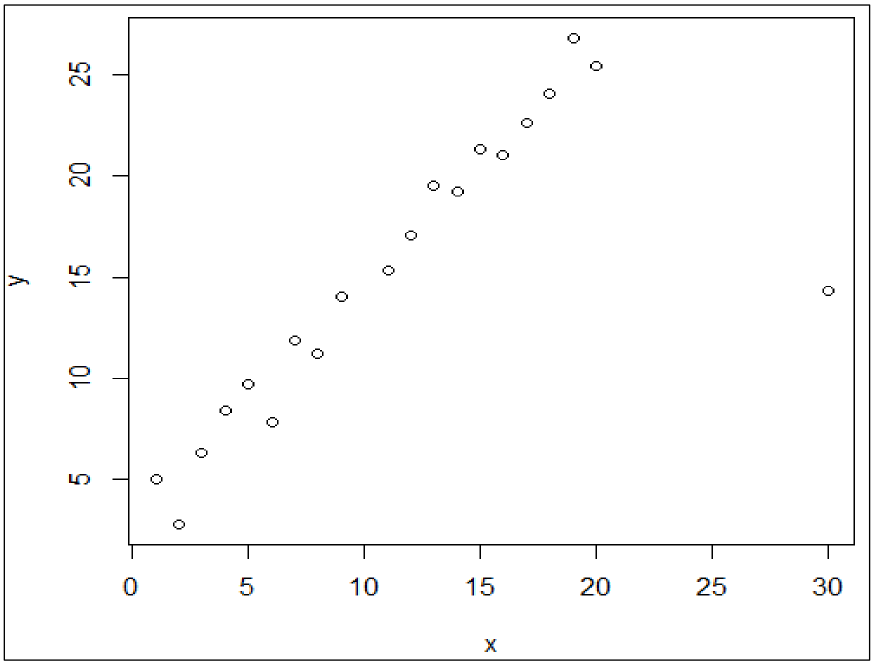
With this outlier at X = 3, the contaminated dataset yields slope estimates which vary significantly from the true slope of 1.2 for least-squares and bootstrap methods. This is in contrast to the dataset in which the response outlier was located at X = 10, for which all slope estimates were relatively robust. However, all confidence intervals still capture the true slope parameter value. Another interesting observation is the relative level of robustness of approaches that utilize the Bi-Squared weighting function (IRLS and Bounded Influence) in comparison to other approaches. The parameter estimates for these approaches change only minimally with the introduction of outliers, even compared to the other robust approaches.

Overall, both the least-squares and bootstrap sets of estimators can be unreliable in the presence of merely a single response outlier. Although the confidence intervals for these approaches have captured the true parameters they seek to estimate in these instances, they are excessively wide and the point estimates are far from the true value, both of which are potentially problematic for any further use of these models.

2.2: Outliers in the Predictor Variable Space Only

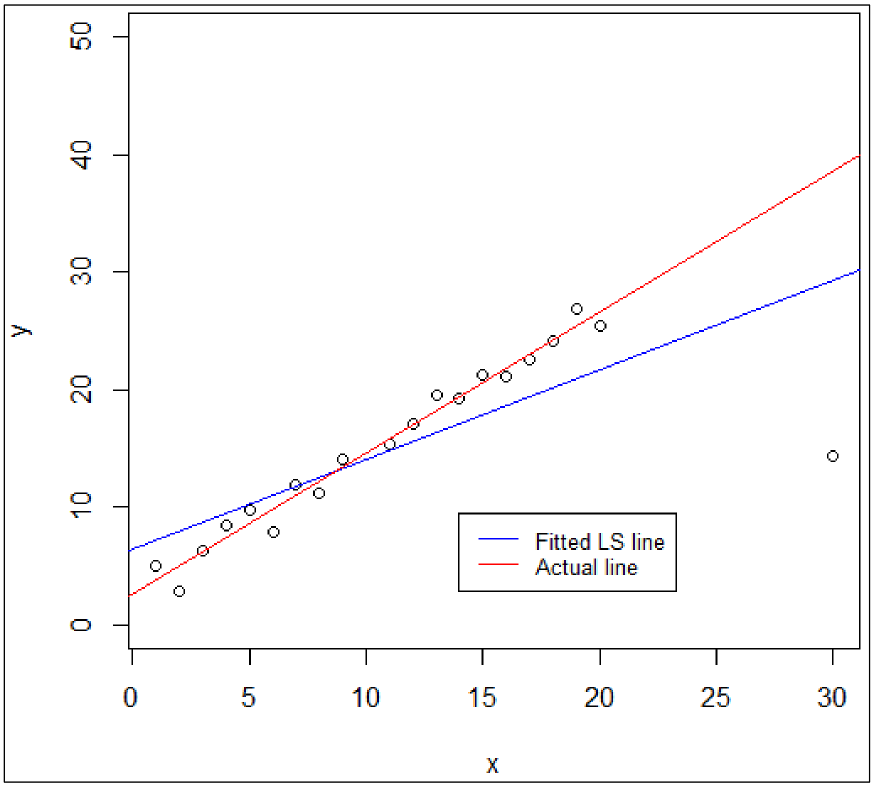
In this sub-section, the effects of an outlier in the predictor space only will be evaluated. Such an outlier will be constructed by changing the x-value of the paired observation (10, 14.3105) to (30, 14.3105). A plot of this altered data set is shown in Figure 25 below.

Figure 25: Scatterplot of Data with Injected Predictor Outlier



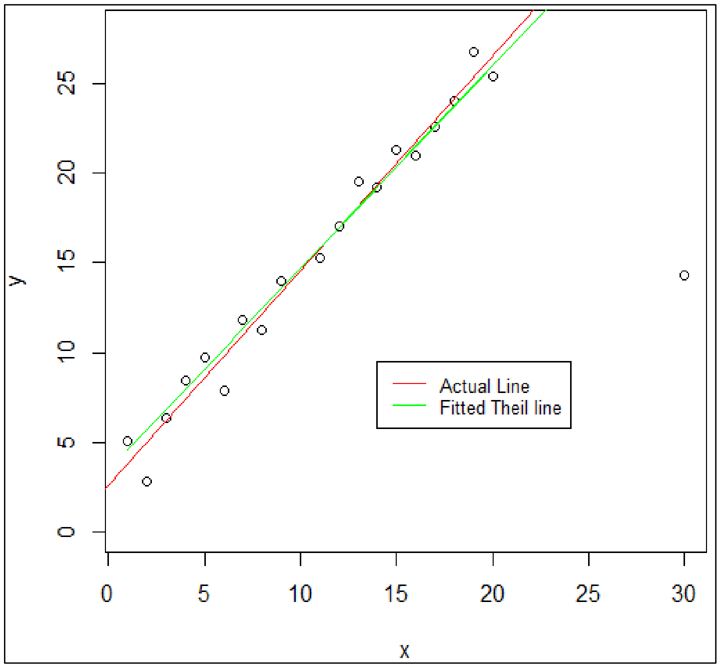
As demonstrated in Figure 26, this type of outlier causes serious problems for the least squares estimation approach. The high leverage associated with this outlier has driven the slope increasingly towards zero, and again produced a model that does a poor job of describing the trend exhibited by the majority of the data.

Figure 26: Least-Squares Fit with Predictor Space Outlier



On the other hand, Figure 27 shows that both the Theil-Sen and Wilcoxon estimation approaches have essentially ignored the outlier and continue to produce fits near to the actual relationship, and that, consequently, still provide reasonably useful summaries of a majority of the data.

Figure 27: Theil-Sen and Wilcoxon fits with Predictor Outlier (X = 10 -> 30)



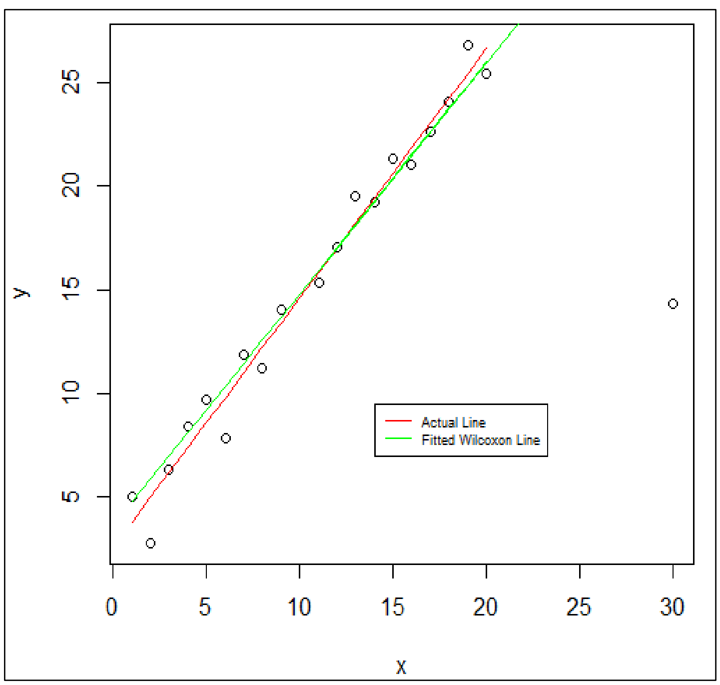


Figure 28 shows that like the Theil-Sen and Wilcoxon approaches, both the IRLS and Bounded Influence estimators also provide fits that are robust to the presence of this predictor space outlier. The final weights used for each of these approaches appear in Table 28 and show again that the Bi-Square weighting function removes the outlier from the dataset, while the Huber weights for the outlier are near 0.05.

Figure 28: IRLS and Bounded Influence Fits with Predictor Outlier (X = 10 -> 30)

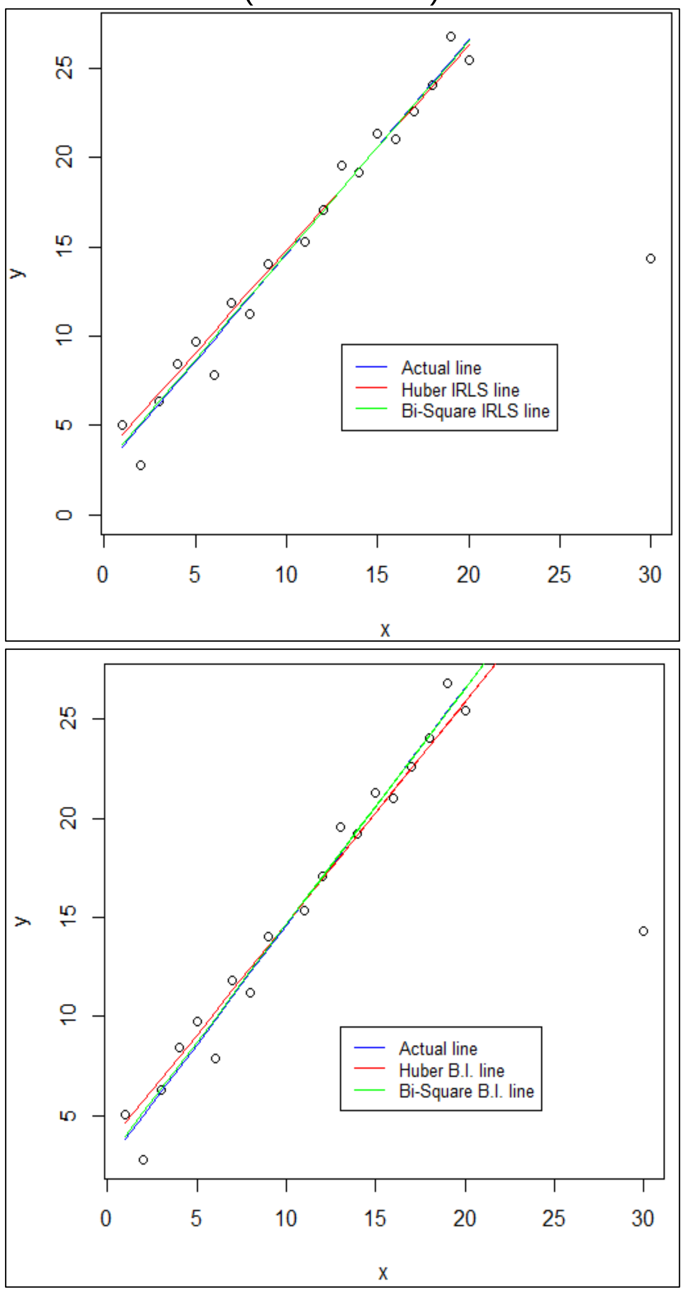


Table 28: IRLS and Bounded Influence Final Iteration Weights with Predictor Outlier (X = 10 -> 30)

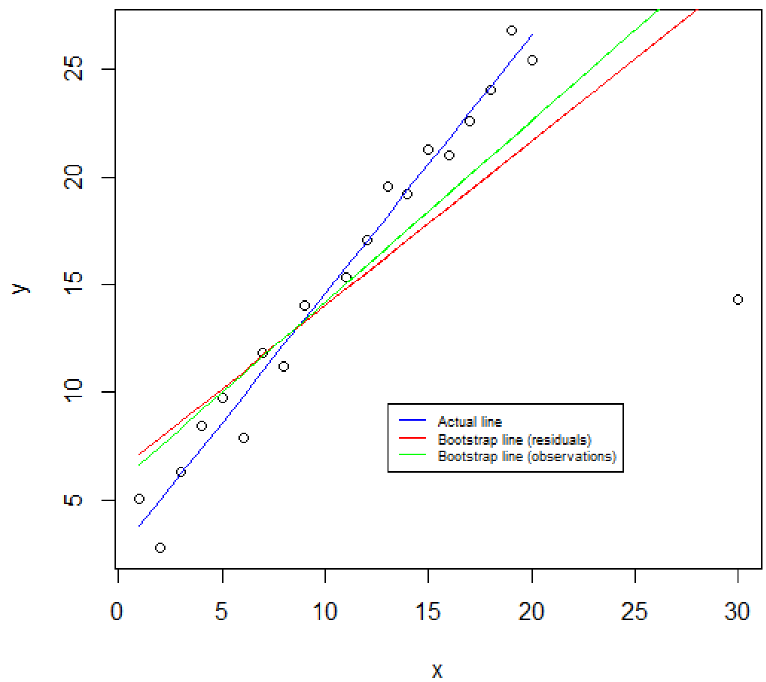
X	IRLS		Bounded Influence	
	Huber	**Bi-Square**	**Huber**	**Bi-Square**
1	1.0000	0.9407	1.0000	0.9259
2	0.3810	0.7215	0.3897	0.7029
3	1.0000	1.0000	1.0000	1.0000
4	1.0000	0.9581	1.0000	0.9556
5	1.0000	0.9465	1.0000	0.9458
6	0.4748	0.7861	0.4997	0.7955
7	1.0000	0.9698	1.0000	0.9711
8	0.9144	0.9410	0.9463	0.9448
9	1.0000	0.9834	1.0000	0.9846
10	0.0489	0.0000	0.0522	0.0000

Table 28, Continued: IRLS and Bounded Influence Final Iteration Weights with Predictor Outlier (X = 10 -> 30)

X	IRLS		Bounded Influence	
	Huber	**Bi-Square**	**Huber**	**Bi-Square**
11	1.0000	0.9850	1.0000	0.9862
12	1.0000	1.0000	1.0000	1.0000
13	0.7505	0.9080	0.7791	0.9148
14	1.0000	0.9977	1.0000	0.9978
15	1.0000	0.9728	1.0000	0.9734
16	1.0000	0.9705	1.0000	0.9702
17	1.0000	0.9924	1.0000	0.9920
18	1.0000	0.9995	1.0000	0.9994
19	0.5511	0.8883	0.5458	0.8742
20	1.0000	0.9341	1.0000	0.9197

In contrast to the robustness demonstrated by the least-squares alternatives considered above, the bootstrap approaches (similar to least squares) poorly fit the majority of the data. Comparison of Figures 29 and 26 indicate that the least squares and bootstrap fits are very similar, with the Bootstrap-Residual approach being virtually identical to the least squares fit.

Figure 29: Bootstrap Fits with Predictor Space Outlier

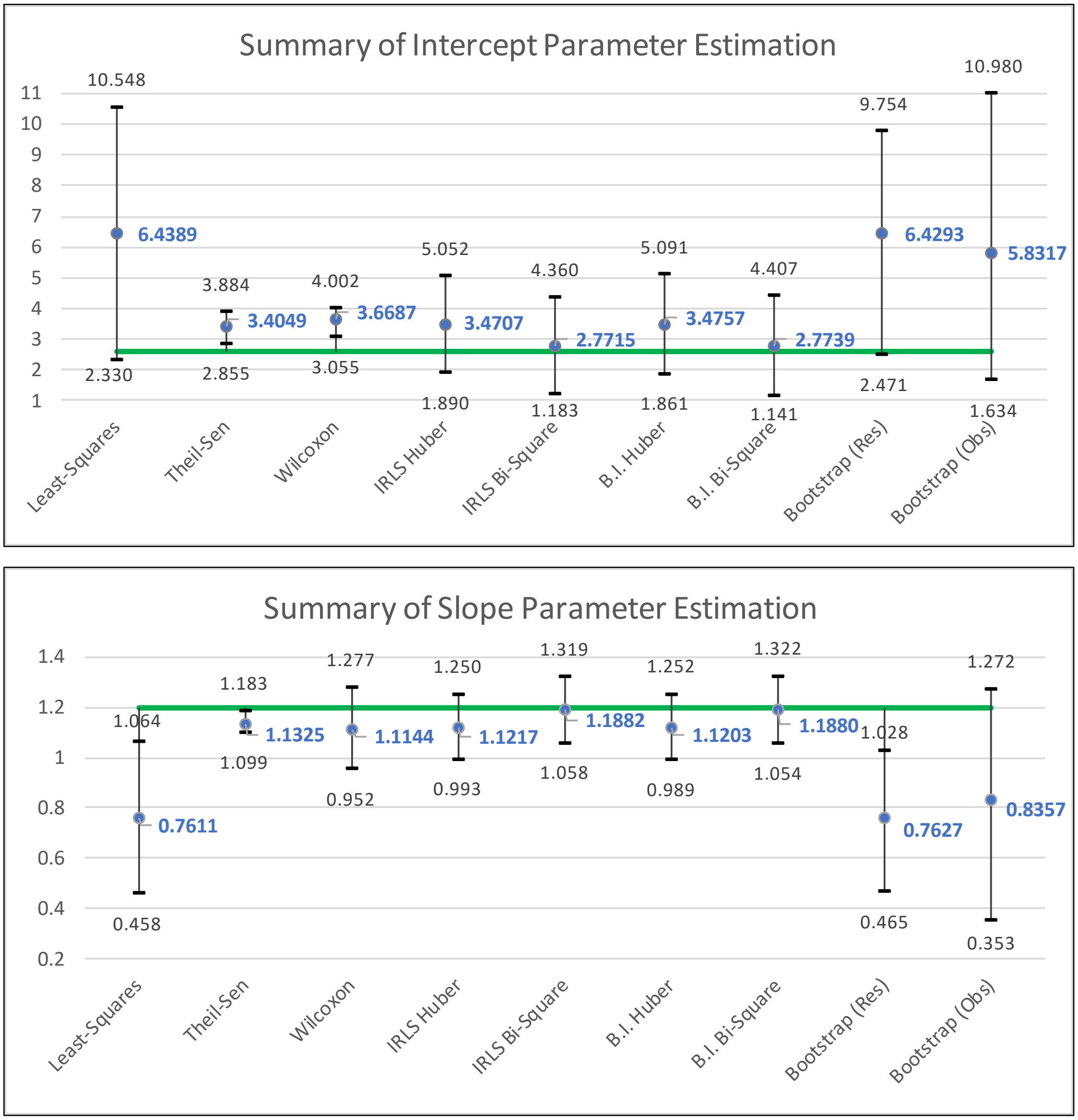


The parameter estimates for the models displayed in Figures 26-29 above (and their ~95% confidence intervals) can be found in Tables A5 (intercept) and A6 (slope) in Appendix 1, and are graphically displayed in Figure 30 below.

These results share similarities with the results from those obtained with the response outliers in sub-section 2.1, more so when the response outlier was located at a higher leverage location, which this predictor space outlier certainly is (i.e., leverage value = 0.384, over twice as large as the next largest of 0.158 at X =1). The least-squares and bootstrap approaches have the most deviant parameter estimates. However, their confidence intervals are much wider, which allowed them to capture the true intercept parameter; however, both least-squares and the residuals bootstrap intervals did not capture the true slope parameter.

The Theil-Sen approach, while producing a more accurate point estimate than least squares on the slope parameter, also failed to capture the slope parameter in a ~95% confidence interval because its confidence interval is so narrow compared to the confidence intervals from other approaches.

Figure 30: Confidence Intervals on Intercept and Slope Parameter with Predictor Space Outlier



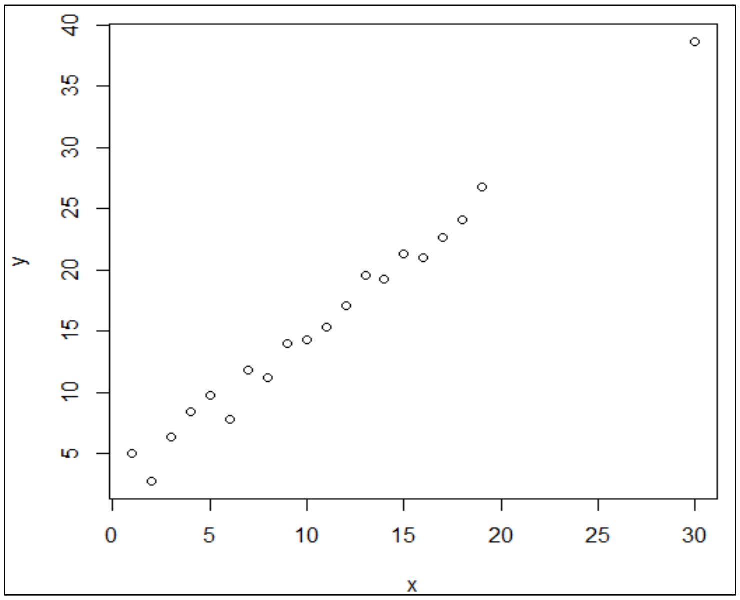
2.3: Outliers in Both the Response and Predictor Variable Spaces

In this section, estimation results with an outlier in both the predictor and response spaces will be evaluated. Within this category of outliers, two contrasting cases will be considered – one in which the outlier is consistent with the overall linear trend of the data, and one in which the outlier is inconsistent with the general linear trend.

2.3.1: Outlier in Both Spaces Consistent with Linear Trend

Here, the observation (20, 25.42093) has been changed to (30, 38.6), which is equivalent to Y = 1.2*(30) + 2.6. This is equivalent to an observation from the true linear relationship in the original data with an error term of zero. A scatterplot of the altered data appears in Figure 31 below.

Figure 31: Scatterplot of Data with Outlier Consistent with Linear Trend



The actual relationship is plotted against each of the fits above in Figures 32-35 below. As expected, since this type of outlier only reinforces the trend exhibited in the main body of the data, all methods came very close to fitting the actual line.

Figure 32: Least-Squares Fit with Outlier Consistent with Linear Trend

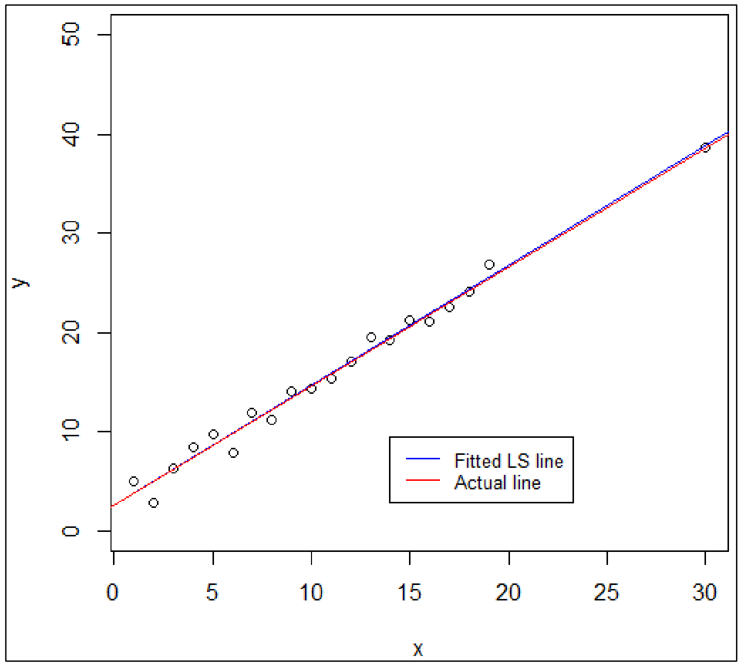
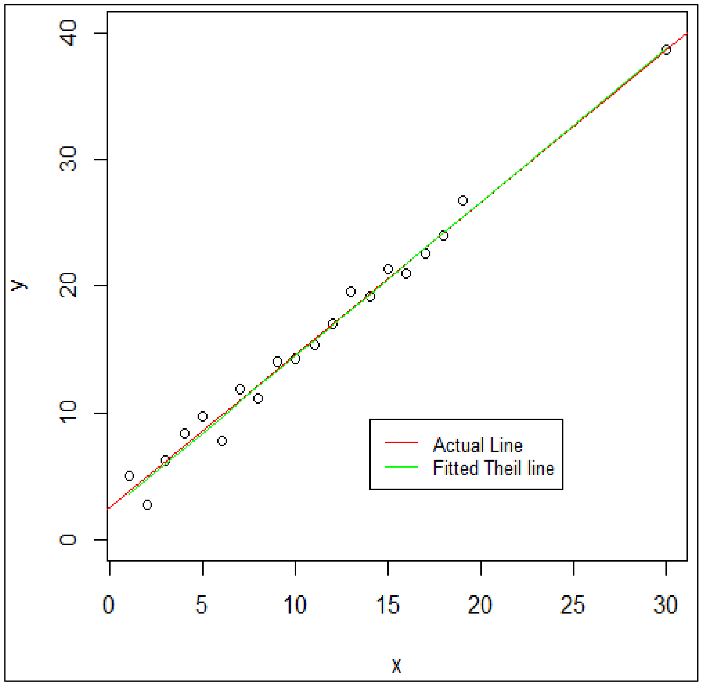


Figure 33: Theil-Sen and Wilcoxon Fits with Outlier Consistent with Linear Trend



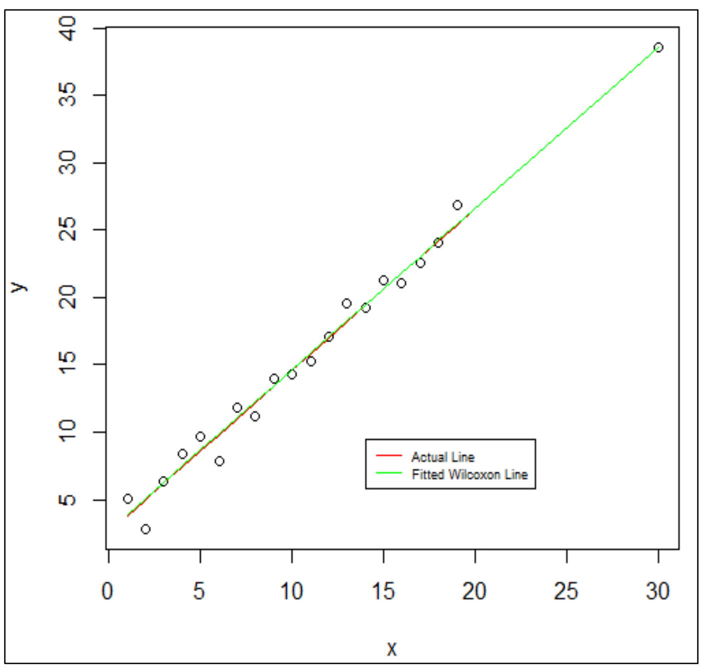
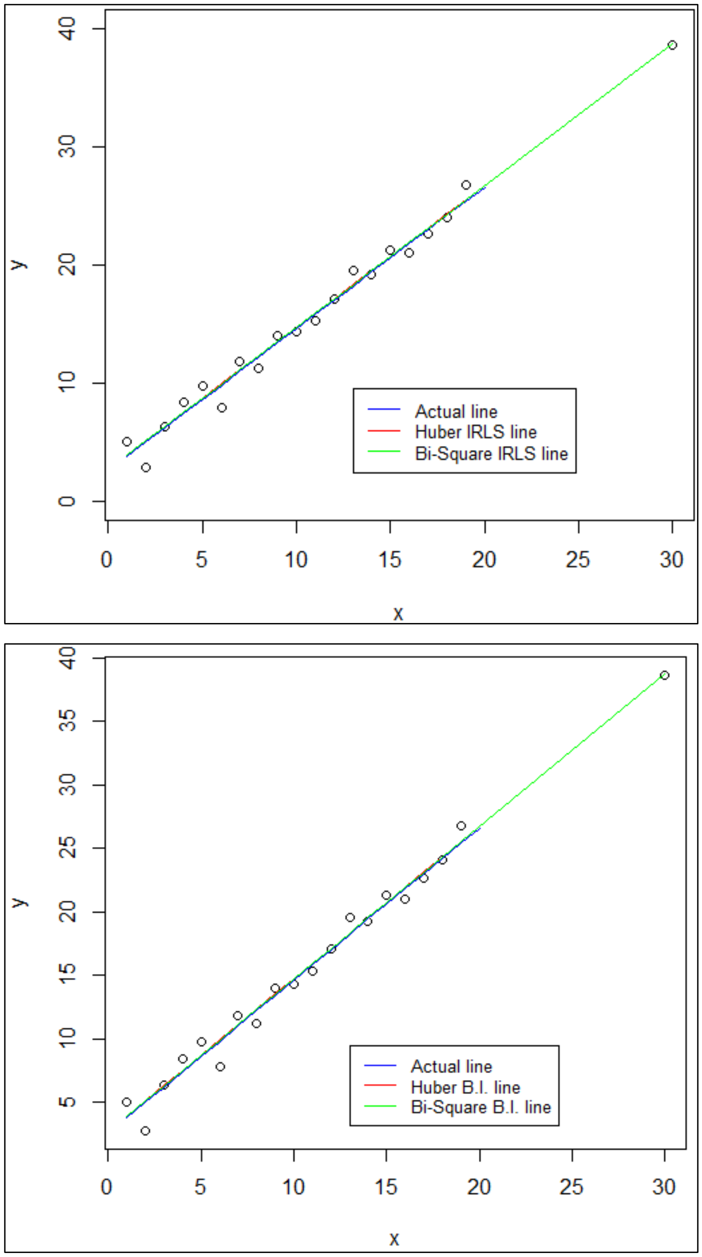


Figure 34: IRLS and Bounded Influence with Outlier Consistent with Linear Trend



The final iteration weights for the fits displayed in Figure 34 are shown in Table 29. Note that for this type of outlier (last row, when X = 30), the weights applied are 1 for both approaches for the Huber weights, and very near 1 for the Bi-Square weighting function.

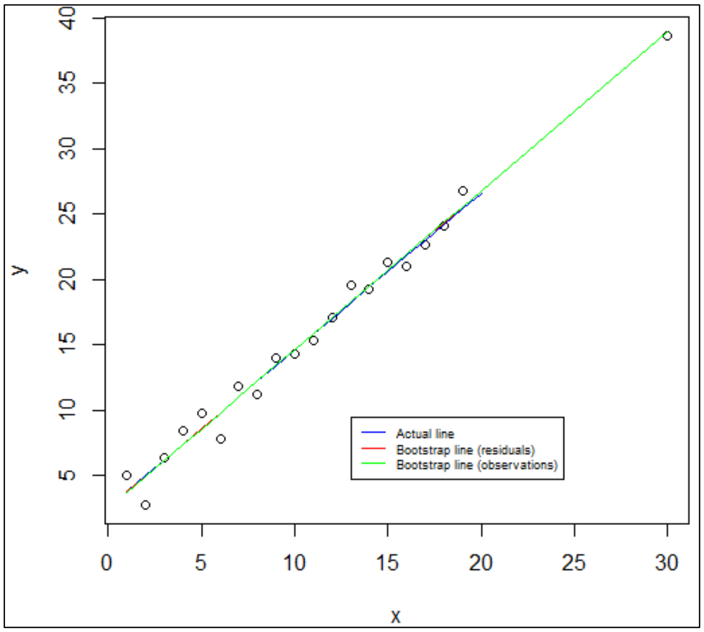
Table 29: IRLS and Bounded Influence Final Iteration Weights with Consistent Trend Outlier

	IRLS		Bounded Influence	
X	Huber	Bi-Square	Huber	Bi-Square
1	1.0000	0.9066	1.0000	0.8879
2	0.6233	0.6293	0.6117	0.6134
3	1.0000	1.0000	1.0000	1.0000
4	1.0000	0.9375	1.0000	0.9326
5	1.0000	0.9226	1.0000	0.9193
6	0.7036	0.7020	0.7139	0.7083
7	1.0000	0.9585	1.0000	0.9585
8	1.0000	0.9124	1.0000	0.9144
9	1.0000	0.9793	1.0000	0.9799
10	1.0000	0.9882	1.0000	0.9886

Table 29, Continued: IRLS and Bounded Influence Final Iteration Weights with Consistent Trend Outlier

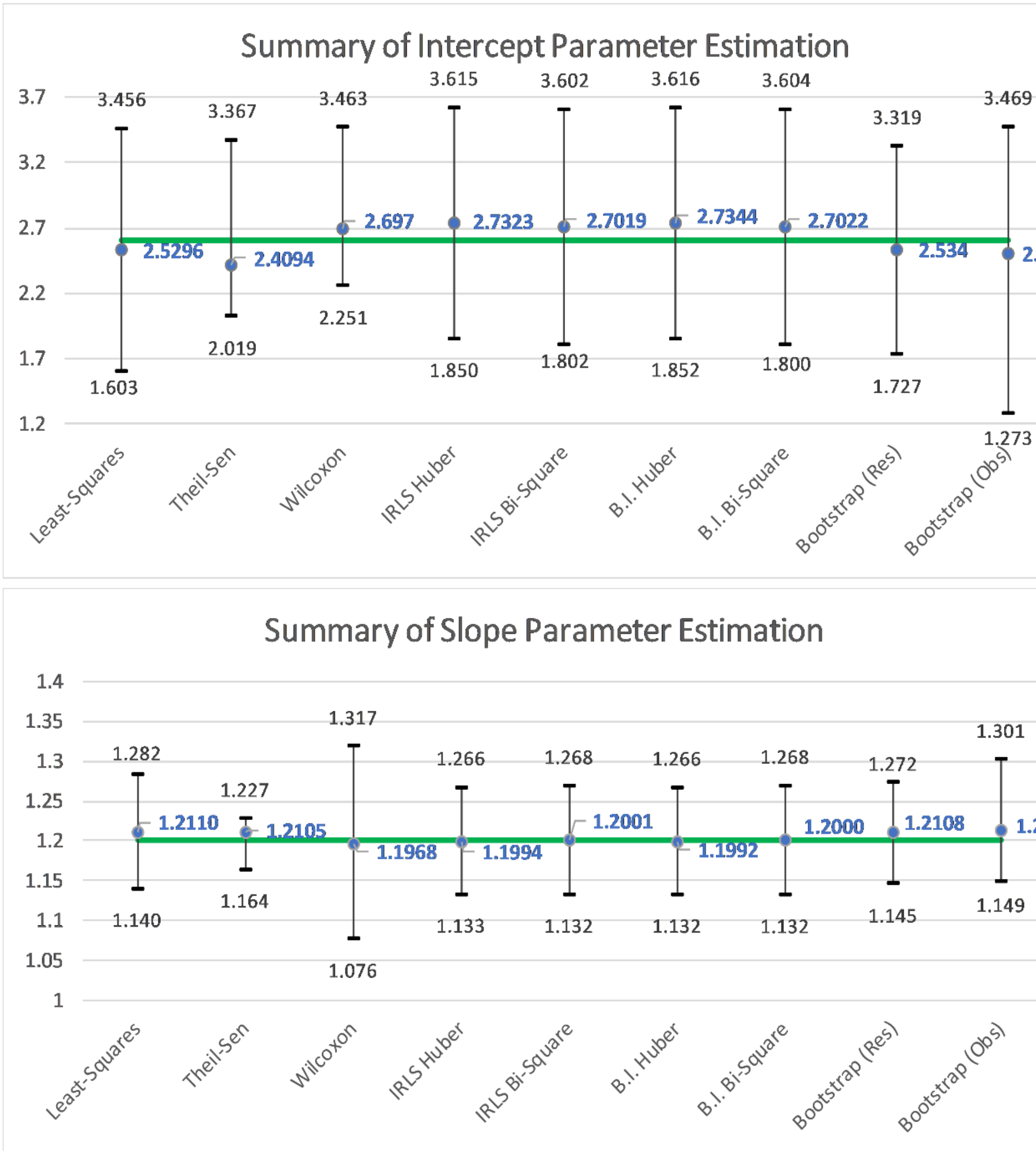
	IRLS		Bounded Influence	
X	**Huber**	**Bi-Square**	**Huber**	**Bi-Square**
11	1.0000	0.9734	1.0000	0.9745
12	1.0000	0.9998	1.0000	0.9999
13	1.0000	0.8857	1.0000	0.8903
14	1.0000	0.9929	1.0000	0.9930
15	1.0000	0.9721	1.0000	0.9722
16	1.0000	0.9434	1.0000	0.9432
17	1.0000	0.9803	1.0000	0.9796
18	1.0000	0.9955	1.0000	0.9952
19	1.0000	0.8733	1.0000	0.8637
20	1.0000	0.9992	1.0000	0.9977

Figure 35: Bootstrap Fit with Outlier Consistent with Linear Trend



The parameter estimates and associated ~95% confidence intervals for the fits in Figures 32-35 are displayed in Tables A7 (intercept) and A8 (slope) in Appendix 1, and also shown in Figure 35 below. The results obtained from all methods were relatively close, much like the original dataset from the Section 1 simple example. Since the outlier (30, 38.6) represented a random error of zero from the standard normal distribution (also its expected value), this created a situation in which the initial optimality of least squares did not change.

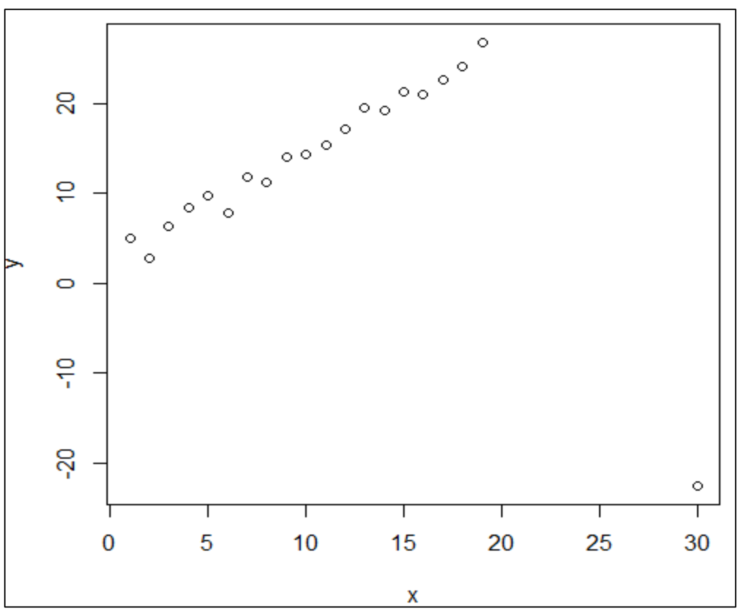
Figure 36: Confidence Intervals on Intercept and Slope Parameter with Consistent Trend Outlier



<u>2.3.2: Outlier in Both Spaces Inconsistent with Linear Trend</u>

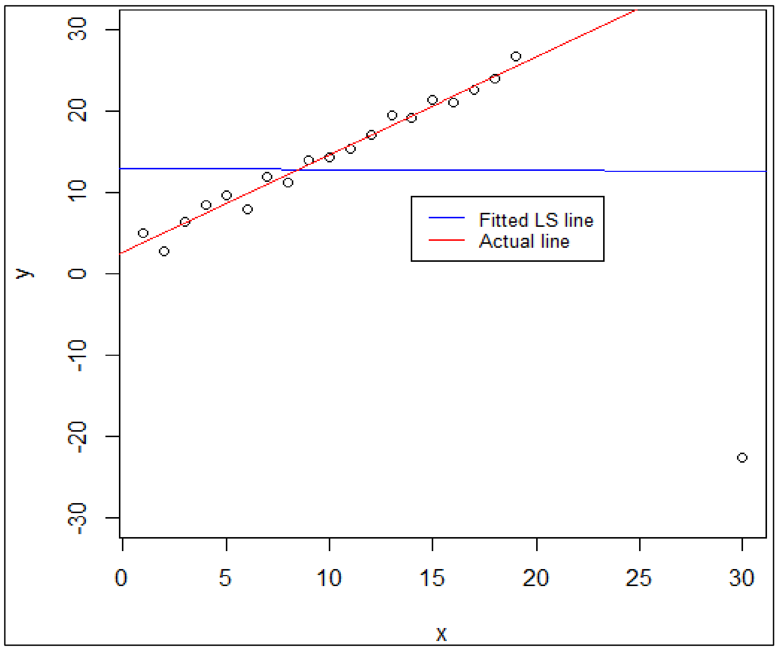
To create an outlier in both spaces which is inconsistent with the linear trend, the x-value of the original observation at x = 20 was changed to 30, and two times the range (max(y) – min(y)) of the original data was subtracted from the response of this observation. Hence, a new observation (30, -22.6261) was generated in place of the original observation (20, 25.4208). A scatterplot of this new data is displayed in Figure 37 below.

Figure 37: Scatterplot of Data with Outlier Inconsistent with Linear Trend



The actual relationship is plotted against each of the fits generated by the considered estimation approaches in Figures 38-41 below. Figure 38 clearly shows the problem an outlier of this nature poses when attempting to use least squares methods to model the trend in this data. The least squares fit is effectively a horizontal line suggesting no relationship between these two variables. This is obviously not true for a majority of the observations.

Figure 38: Least-Squares with Outlier Inconsistent with Linear Trend



In contrast to the extremely poor fit provided by the traditional least squares method, Figures 39 and 40 show that the Theil-Sen, Wilcoxon, IRLS, and Bounded Influence approaches provide estimates of the relationship parameters that are robust to this type of outlier. The final iteration weights used for these estimators are given in Table 30, and again show that the Bi-Square weighting function effectively removed the outlier from the dataset with the Huber weights also small at ~0.02.

Figure 39: Theil-Sen and Wilcoxon Fits with Outlier Inconsistent with Linear Trend

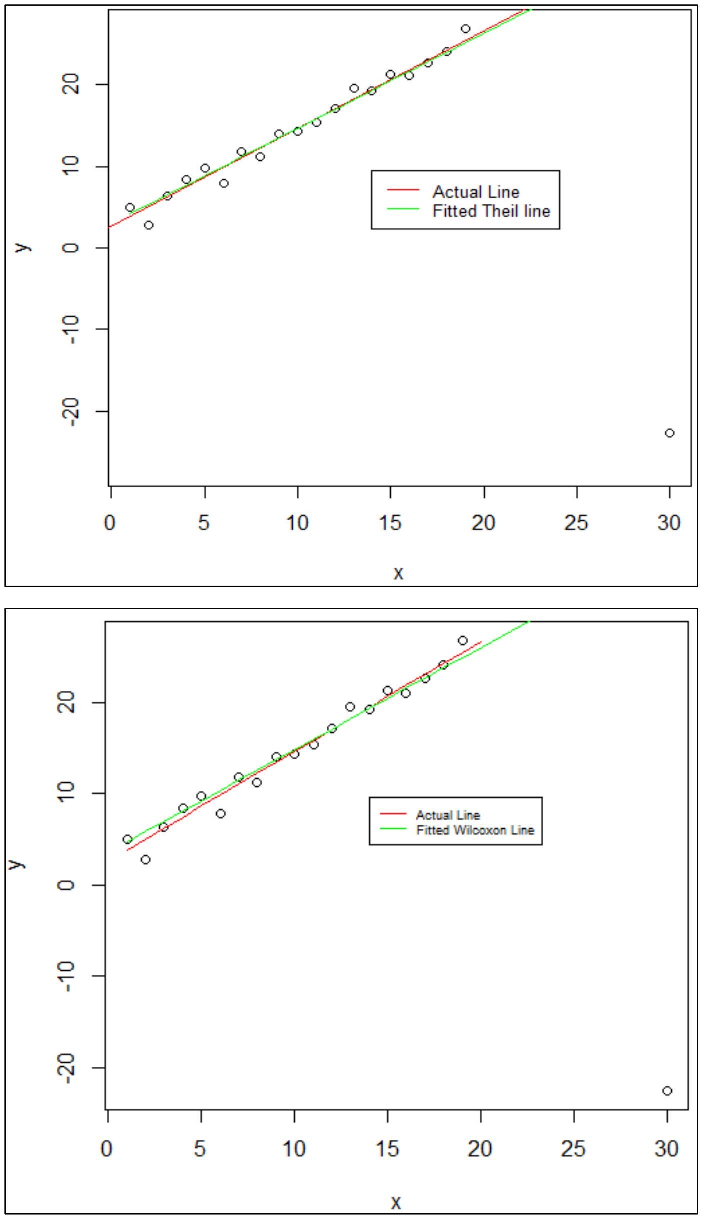


Figure 40: IRLS and Bounded Influence Fits with Outlier Inconsistent with Linear Trend

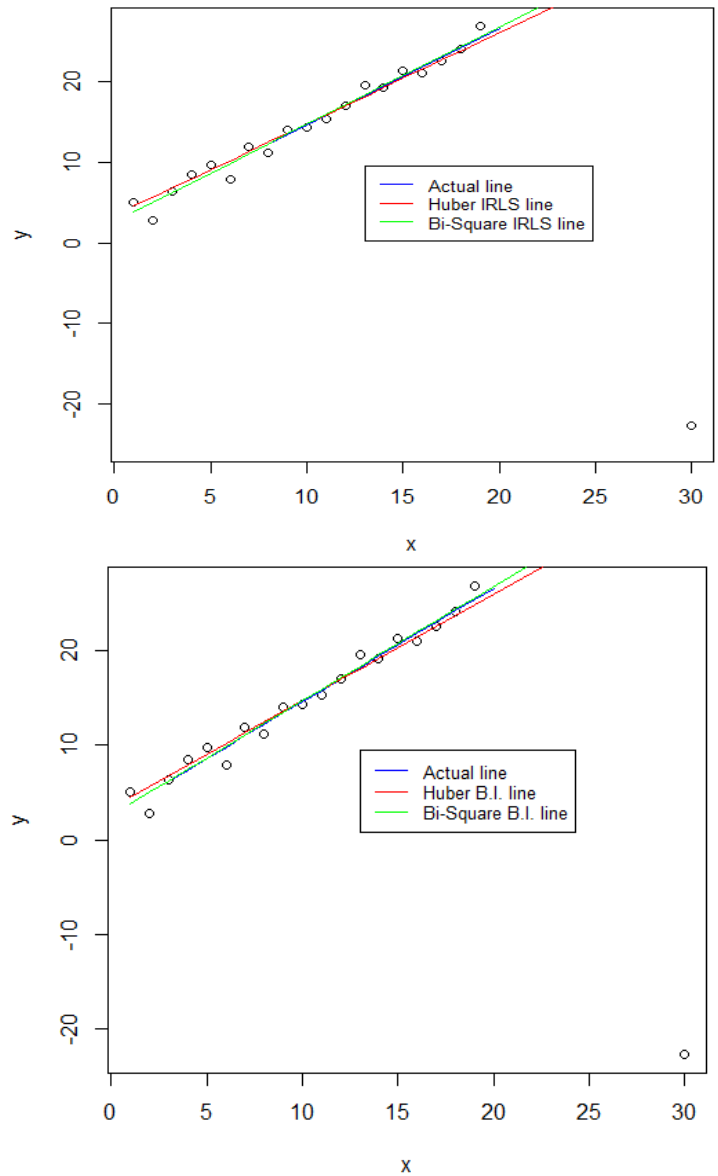


Table 30: IRLS and Bounded Influence Final Iteration Weights with Inconsistent Trend Outlier

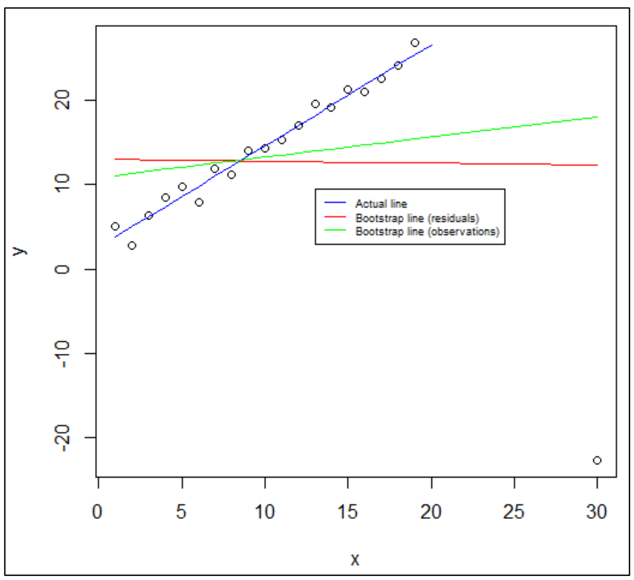
X	IRLS		Bounded Influence	
	Huber	**Bi-Square**	**Huber**	**Bi-Square**
1	1.0000	0.9113	1.0000	0.8861
2	0.3454	0.7093	0.3797	0.6778
3	1.0000	0.9995	1.0000	0.9995
4	1.0000	0.9418	1.0000	0.9367
5	1.0000	0.9299	1.0000	0.9271
6	0.4250	0.7622	0.4837	0.7658
7	1.0000	0.9628	1.0000	0.9634
8	0.8102	0.9315	0.9145	0.9337
9	1.0000	0.9819	1.0000	0.9828
10	1.0000	0.9908	1.0000	0.9912

Table 30, Continued: IRLS and Bounded Influence Final Iteration Weights with Inconsistent Trend Outlier

X	IRLS		Bounded Influence	
	Huber	**Bi-Square**	**Huber**	**Bi-Square**
11	1.0000	0.9783	1.0000	0.9793
12	1.0000	0.9998	1.0000	0.9998
13	0.6922	0.9085	0.7572	0.9106
14	1.0000	0.9932	1.0000	0.9932
15	1.0000	0.9796	1.0000	0.9784
16	1.0000	0.9491	1.0000	0.9455
17	1.0000	0.9805	1.0000	0.9780
18	1.0000	0.9943	1.0000	0.9933
19	0.5194	0.9061	0.5314	0.8802
20	0.0165	0.0000	0.0193	0.0000

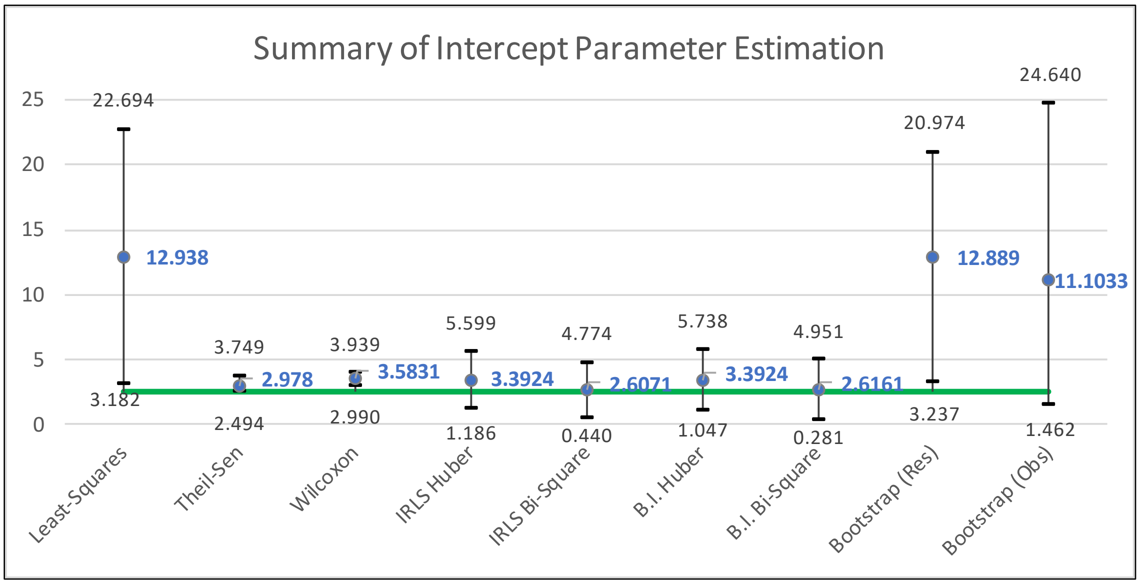
As seen in Figure 41, the bootstrap approaches again fail to effectively manage this type of outlier and provide fits similar to that generated by the least squares method. At least the Bootstrap-Observations approach provides a positive slope point estimate. Both least squares and the Bootstrap-Residuals approach return negative slope point estimates.

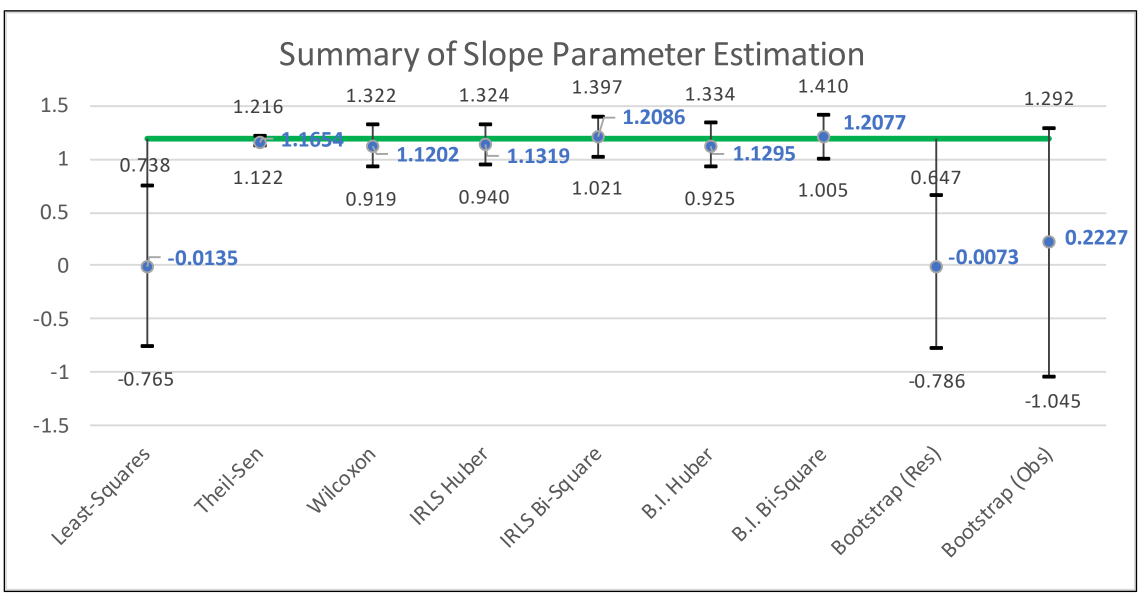
Figure 41: Bootstrap Fit with Outlier Inconsistent with Linear Trend



These parameter estimates and ~95% confidence intervals are shown in Tables A9 (intercept) and A10 (slope) in Appendix 1, and are displayed in Figure 42. These results show the most extreme differences between fits out of all datasets examined thus far, indicating that this type of outlier is the most problematic for the common least squares method of fitting a line to summarize an observed trend in a paired variable dataset.

Figure 42: Confidence Intervals on Intercept and Slope Parameter with Inconsistent Trend Outlier





Despite the great disparity in robustness between different methods for this dataset, the least-squares, the Wilcoxon, and Residual Bootstrap approaches did not capture the true intercept. The least squares and Residual Bootstrap approaches also failed to capture the true slope.

Also of note is that the Theil-Sen intervals are the narrowest for this situation for both parameters and successfully bound the true parameter values for both intercept and slope. As noted above, the least squares and Residual Bootstrap intervals fail to capture the true parameter values, but both are actually two of the widest intervals. The widest intervals are those for the Bootstrap Observations approach, and their width just allows them to successfully capture their respective parameter values.

Section 3: Comparison of Alternatives on Selected Data Sets

In this section, we will perform each of the discussed regression procedures on selected real datasets and compare their results. These data sets will allow for the evaluation of relative performance of the various estimation methods in the presence of more than one outlier.

3.1: Math Proficiency Data

The Educational Testing Service Study *America's Smallest School: The Family* **(#10)** evaluated relationships between student's educational results and their home environments. Earlier studies had investigated relationships between educational achievement and socio-economic status (e.g., educational level of parents, family income, parent's occupations, etc.), but this study attempted to use more direct measures of the situation within the student's home.

Table 31 below displays average math proficiency scores for eighth-grade students from the 1990 National Assessment of Educational Progress, as well as percentages of homes with both parents present for 37 U.S. states, Washington, D.C., and 2 U.S. territories (Virgin Islands and Guam). Potential outliers are those observations that are not states – Washington D.C., Guam, and the Virgin Islands – and are highlighted in red.

Table 31: Average Math Proficiency Scores and Percentage of Homes with 2 Parents by Location

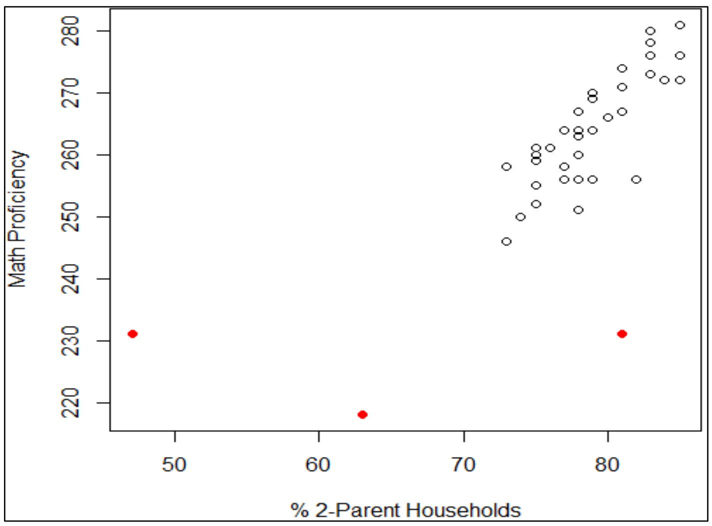
State	Percentage of 2-Parent Homes	Math Proficiency
Alabama	75	252
Arizona	75	259
Arkansas	77	256
California	78	256
Colorado	78	267
Connecticut	79	270
Delaware	75	261
D.C.	47	231
Florida	75	255
Georgia	73	258
Guam	81	231
Hawaii	78	251
Idaho	84	272
Illinois	78	260

Indiana	81	267
Iowa	83	278
Kentucky	79	256
Louisiana	73	246
Maryland	75	260
Michigan	77	264
Minnesota	83	276
Montana	83	280
Nebraska	85	276
New Hampshire	83	273
New Jersey	79	269
New Mexico	77	256
New York	76	261
North Carolina	74	250
North Dakota	85	281
Ohio	79	264
Oklahoma	78	263

Oregon	81	271
Pennsylvania	80	266
Rhode Island	78	260
Texas	77	258
Virgin Islands	63	218
Virginia	78	264
West Virginia	82	256
Wisconsin	81	274
Wyoming	85	272

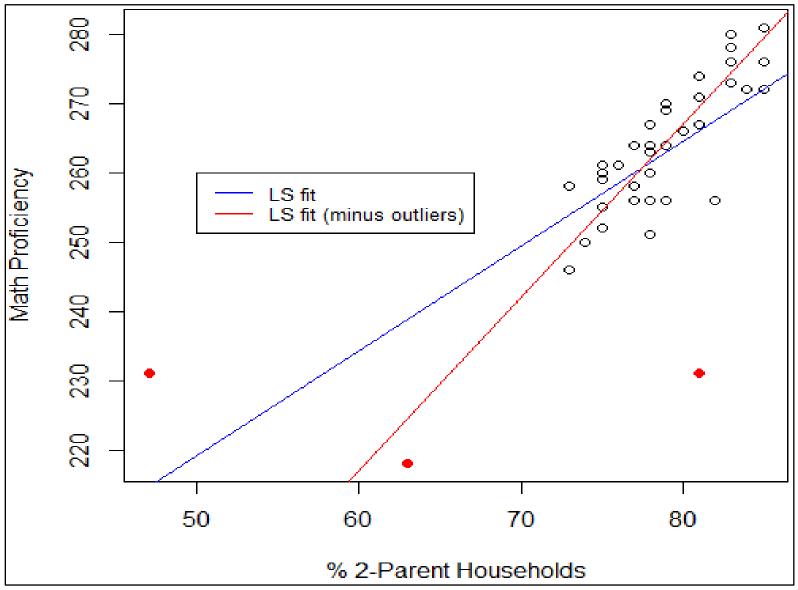
The scatterplot in Figure 43 suggests a general linear trend for the majority of the observations (those from the 37 states); however, the three outlier (red) are obvious. Washington D.C. is an outlier in both the predictor space with a very low percentage of two-parent families, as well as in the response space with a low average math proficiency score. The Virgin Islands is also low for both measures; however, this observation appears more consistent with the trend exhibited by the 37 states than the Washington D.C. result. The Guam data is only an outlier in the response space also with a low average math proficiency score. The percentage of two-parent families in Guam is, however, near the middle of these values for the 37 states, so would not be considered an outlier in the predictor variable space.

Figure 43: Math Proficiency vs. Percentage of 2-Parent Homes



The least squares fit of the data – both with and without the identified outliers (ie, D.C., Guam, and the Virgin Islands) – appear in Figure 44. It is readily observable that the outliers have an impact when using this most common approach to estimating the linear relationship between these two variables. When the outliers are removed, and only the state data is considered, the slope estimate is ~2.5. However, with the non-state outliers included, the slope is much lower, near 1.5. Moreover, this line when compared to the line fit without the outliers does not appear to as accurately reflect the general relationship between the Math Proficiency Test Score averages and the percentage of two-parent households.

Figure 44: Least-Squares Fit for Math Proficiency Data



When the outliers are removed from considerations, both the Theil-Sen and Wilcoxon fits displayed in Figure 45 provide a fit similar to the least-squares fit with slope estimates of ~2.4 and ~2.5, respectively. However, both of these approaches also show a lower slope estimate when the outliers are included in the data set than when they are not, with the Wilcoxon estimate having more of a downward bias. Neither approach provides an estimate as low as least squares (Theil-Sen ~2.1 and Wilcoxon ~1.8), but both show some marginal impact of these three outlying observations.

Figure 45: Theil-Sen and Wilcoxon Fits for Math Proficiency Data

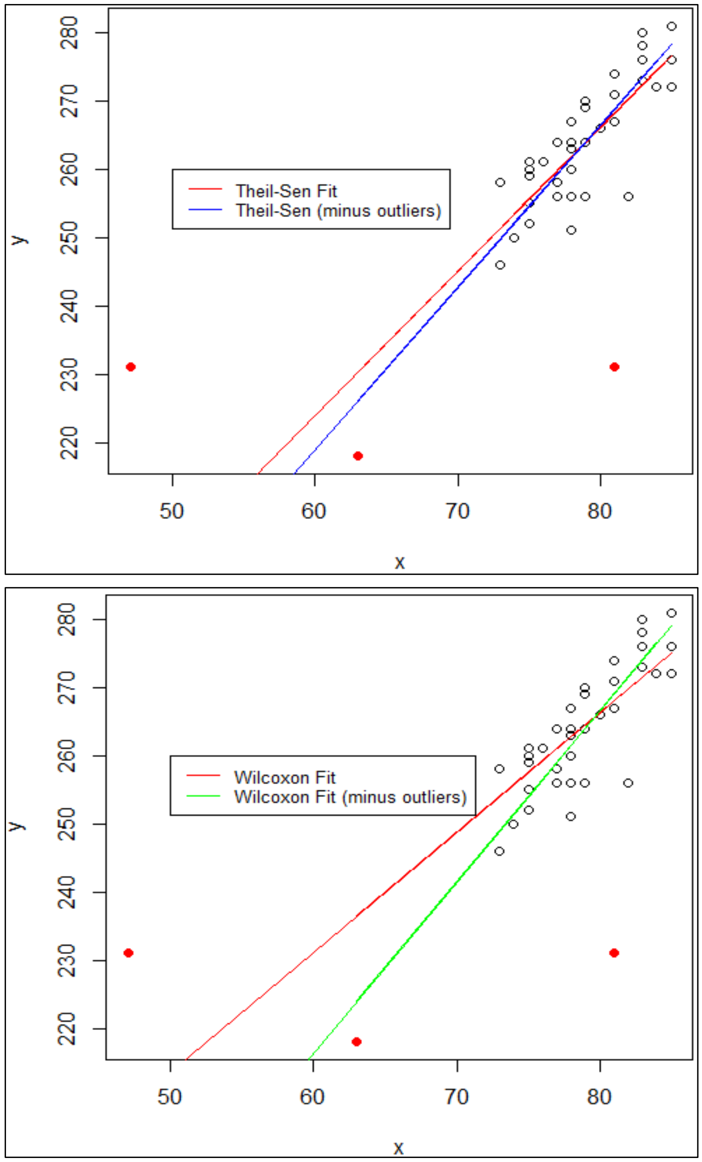
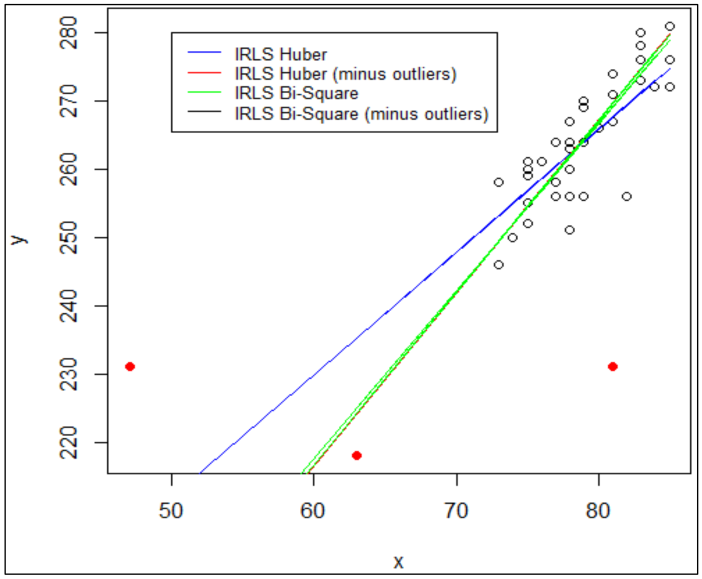


Figure 46 provides an interesting result. Only two of the IRLS and Bounded Influence fits appear to provide a lower slope estimate than the ~2.5

observed for least squares without the outliers. So while both these approaches appear to – regardless of weighting scheme – appear to mimic least squares when there are no outliers in the data set, the Huber weighting approach appears to provide a lower slope estimate similar to the Wilcoxon result when the outliers are present (Wilcoxon and IRLS-Huber ~1.8, Bounded Influence-Huber ~1.9). The Bi-Square weighting function, on the other hand, for both IRLS and Bounded Influence approaches, appears to return slope estimates reasonably close to 2.4.

This difference between weighting schemes seems to be most likely due to the Bi-Square weighting function actually providing weights of zero for both DC and Guam – or essentially removing them from the data set. The Huber weighting scheme does down-weight these observations (IRLS - ~0.35 for DC and ~0.23 for Guam; Bounded Influence – ~0.23 for DC and ~0.25 for Guam), but the scheme never allows the weight to be zero, so these observations are always retained in the data set.

Figure 46: IRLS and Bounded Influence Fits for Math Proficiency Data



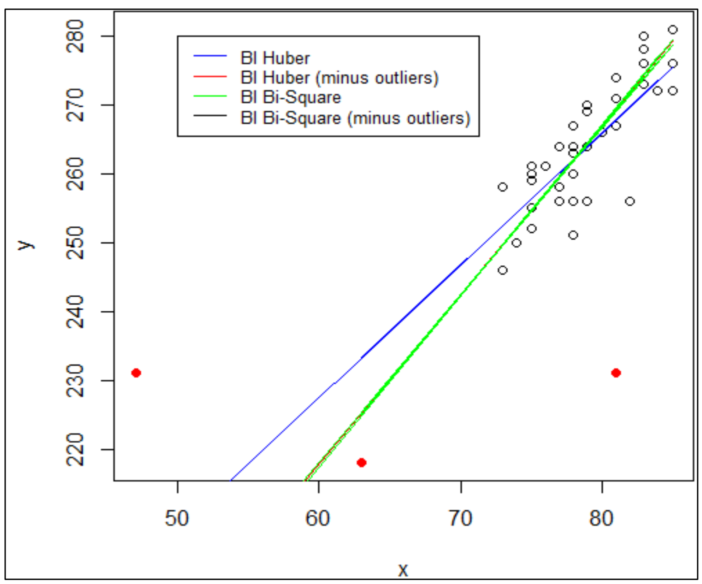


Figure 47 displays the Bootstrap fits – both when resampling residuals, as well as when resampling observations. As observed in Section 2, the bootstrap approaches tend to mimic what was observed for the least-squares approach. The approach using residuals provides a slope estimate very similar to the least squares estimates, and the approach using observations provides a result (~1.7) between the least-squares (~1.5) and the Wilcoxon (~1.8) results when the outliers are included in the data set.

Figure 47: Bootstrap Fits for Math Proficiency Data

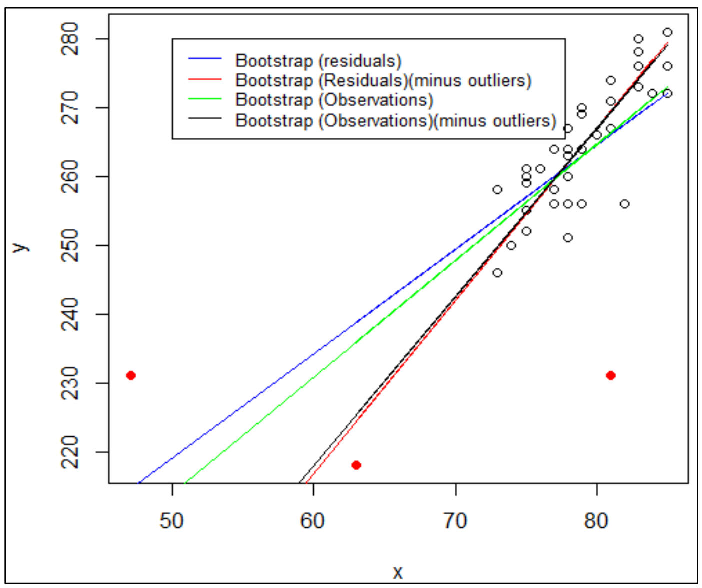
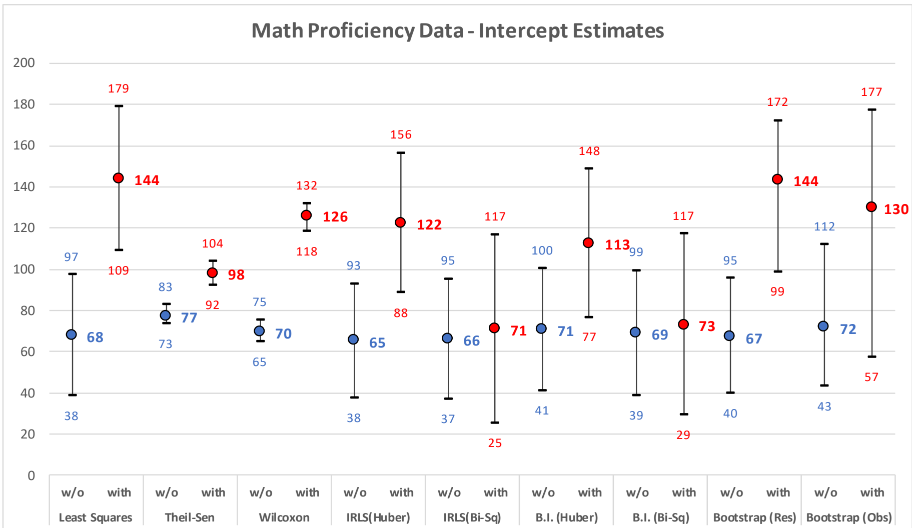
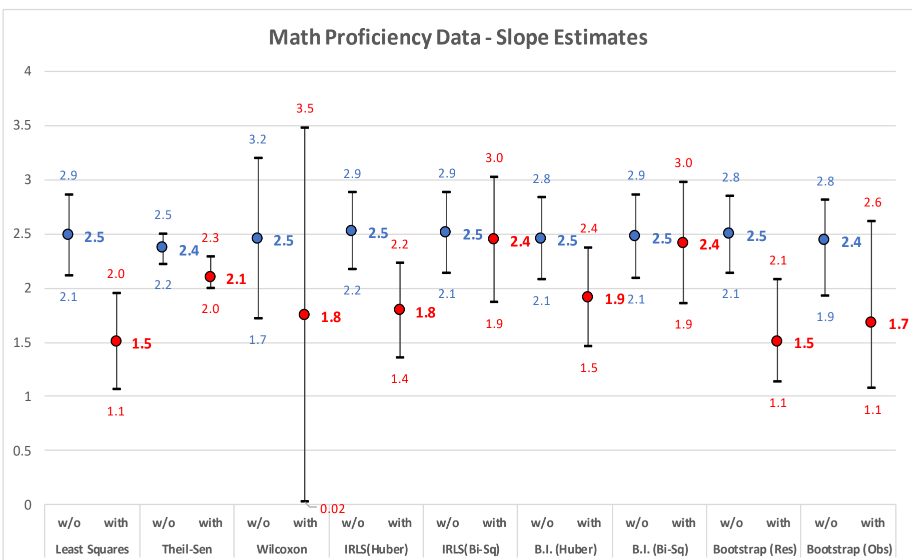


Figure 48 displays the approximate 95% confidence intervals for both the intercept and the slope parameter estimates for all of the fits plotted in Figures 43-47 above. The actual values are tables in Tables A11 (w/ outliers) and A12 (w/o outliers) in the Appendix.

Figure 48: Confidence Intervals on Intercept and Slope Parameters for the Math Proficiency Data





The most robust estimators for the slope and the intercept are clearly the Bi-Square weighting function IRLS and Bounded Influence approaches. The

Theil-Sen estimators are reasonably robust, but the intercept estimates for the fits with and without the outliers would be considered significantly different from each other.

None of the remaining approaches are very effective at managing this triumvirate of outliers. The least-squares and Bootstrap using Residuals are clearly the poorest approaches here. The Huber weighting function options, as well as the Wilcoxon and the Bootstrap using Observations approaches appear to make a marginal adjustment for the outliers. However, for the latter two, and especially the Wilcoxon approach, the widths of the respective interval estimates are much wider than those when no outliers were present.

While it would be reasonable to expect wider confidence intervals for the fits where the outliers were involved due to the inherent increase in uncertainty they introduce, their impact on the jackknife approach to estimate the variance of the Wilcoxon slope estimator was to increase the width of the interval by more than 2.3X. For comparison, the increase in width for the least-squares interval was about 20% (or 1.2X), and all the other approaches ranged between ~3% (or 1.03X for Theil-Sen) and ~75% (or 1.75X for Bootstrap with Observations).

While this increase is one of the most obvious features of Figure 48, exploration of robust estimators for the variance of the errors was not considered within the scope of this work. Reference **#9** suggests an alternative, bootstrap-based alternative estimator that appeared to provide intervals more narrow than those obtained with the jackknife approach used here. It was not utilized in this work due to the relative difficulty of calculations involved to produce it opposite that of the jackknife approach that was utilized.

This example – with multiple outliers of different types – has demonstrated that the use of a weighting function that allows for the actual elimination (i.e., weights equal to zero) can provide more robust estimators than weighting functions not allowing for such "trimming" of the data set. The Theil approach also provided reasonably robust parameter estimates, as well as the most stable confidence interval widths. Again, the Bootstrap approaches were poor performers in terms of robustness opposite these types of outliers, returning results most similar to those obtained with the common, but known to be non-robust, least-squares approach.

3.2: Belgian Telephone Volume Data

Between 1950 and 1973, the volume of telephone calls in Belgium was recorded for each year **(#4)**. This data is given in Table 32 and displayed in Figure 49 below:

Table 32: Number of Telephone Calls Made in Belgium (Tens of Millions) (1950-1973)

Year	# Telephone Calls (Tens of Millions)
1950	0.44
1951	0.47
1952	0.47
1953	0.59
1954	0.66
1955	0.73
1956	0.91
1957	0.88
1958	1.06
1959	1.2
1960	1.35
1961	1.49
1962	1.61
1963	2.12
1964	11.9
1965	12.4

1966	14.2
1967	15.9
1968	18.2
1969	21.2
1970	4.3
1971	2.4
1972	2.7
1973	2.9

It is widely known that the numbers of phone calls between the years of 1964-1969 were recorded erroneously. The plot of the data suggests that these numbers were grossly over-reported. If these points were excluded, it appears that the rest of the data can be fit well with a linear model. Therefore, to limit the influence of these recording errors on the fitted model, consideration of a model that is robust against the response outliers created by these errors could be of value.

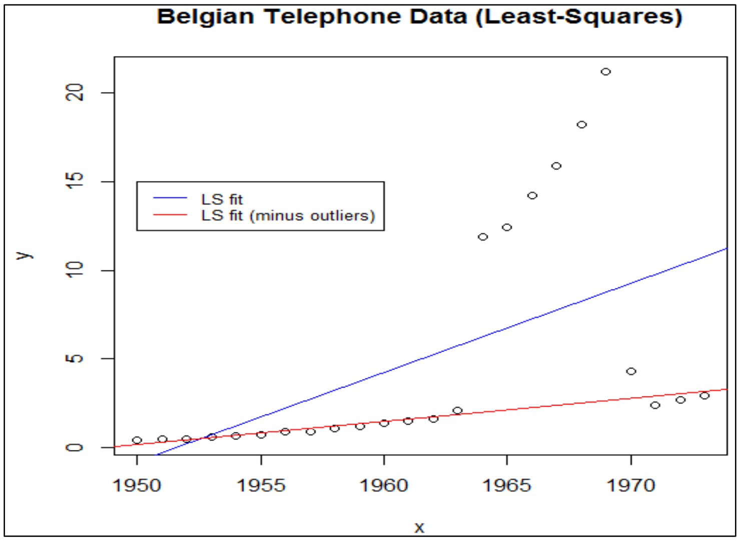
Figure 49: Number of Belgian Telephone Calls Made from 1950-1973

The multiple gross outliers in the response space in this data set present another type of challenge for the robust regression methods considered here. Consequently, each method will be evaluated by their performance on the original dataset (shown above) and the same data set with outliers removed.

As can be seen in Figure 50, the commonly utilized least-squares approach provides distinctly different results when the outliers are included in the data set than when they have been removed. The slope estimate is approximately 4X larger (~0.50 vs ~0.13) when the outliers are involved. It is certainly true that decisions made related to Belgian telephone call volume circa

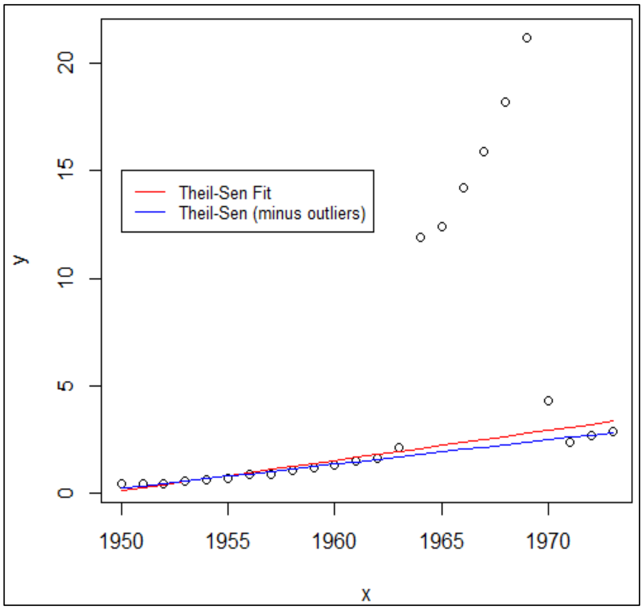
1974 would have been different if the belief had been that the volume was increasing at 5 million calls per year rather than 1.3 million calls per year.

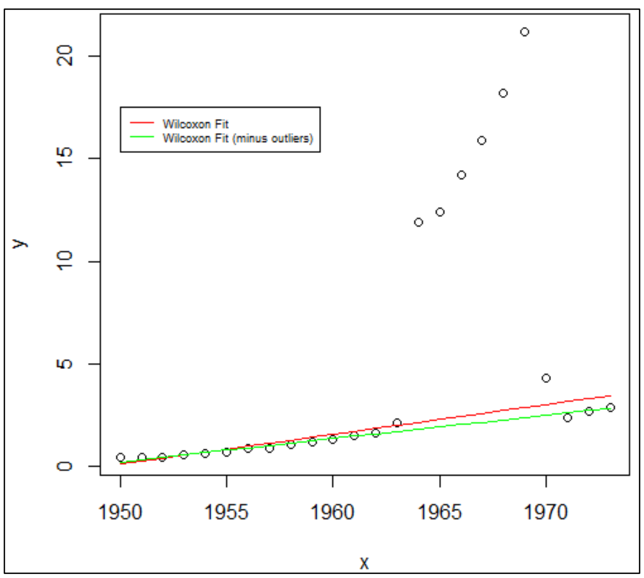
Figure 50: Least-Squares Fit for Belgian Telephone Data



In contrast to the least-squares fits in Figure 50, the Theil-Sen and Wilcoxon fits displayed in Figure 51 suggest only a relatively small increase in annual call volume when the outliers are involved in the estimation of the model parameters. The slope estimates without the 1964-1969 values are approximately 0.11 for both approaches, and these estimates only increase to ~1.4 (Theil-Sen) and ~1.5 (Wilcoxon) when the over-reported years are included in the data set.

Figure 51: Theil-Sen and Wilcoxon Fits
for Belgian Telephone Data

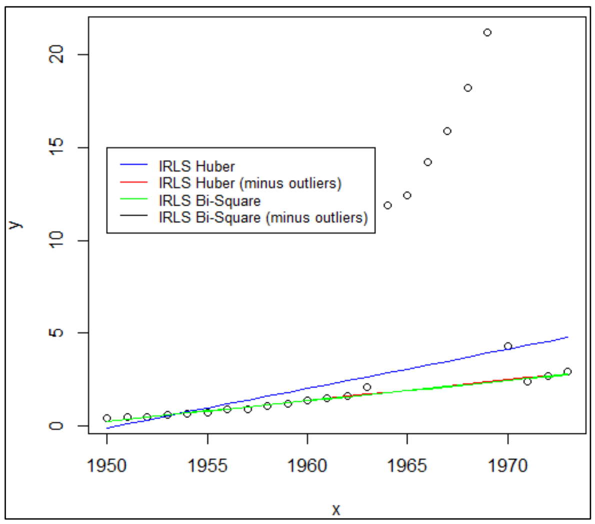


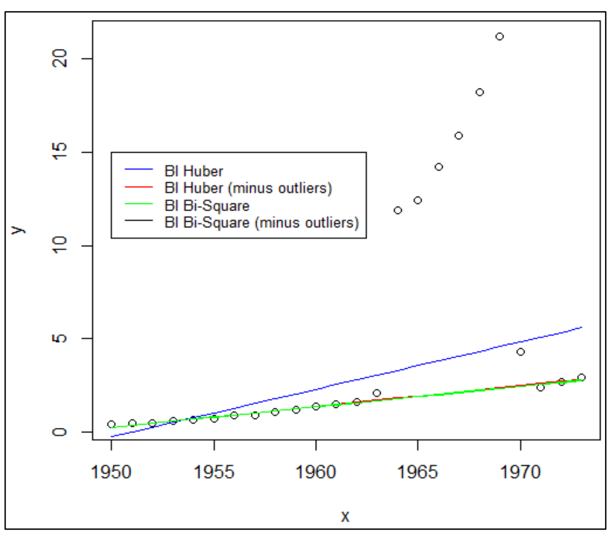


Similar to what was observed for the Math Proficiency data, Figure 52 indicates that among the IRLS and Bounded Influence approaches, the Bi-

Square weighting function provides a virtually identical fit both with and without the over-reported years. However, the Huber weighting function produced larger slopes with the outliers involved than when they were not.

Figure 52: IRLS and Bounded Influence Fits for Belgian Telephone Data





The Bi-Square weighting function provides slope estimates of ~0.11 both with and without the outliers involved, and for both IRLS and Bounded Influence approaches. This is effectively the same value as the Theil-Sen, Wilcoxon, and Huber weighting scheme for both IRLS and Bounded Influence approaches when the over-reported years are removed from the data set. On the other hand, the Huber weighting scheme provides slope estimates approximately 2X larger than this when the outliers are involved (~0.21 – IRLS and ~0.25 – Bounded Influence).

Again, the difference appears to be that the Bi-Square weighting function allows for much lower weights to be used than the Huber weighting function. The Bi-Square and Huber weights for the years 1963-1973 across both approaches appear in Table 33. This period includes the over-reported years of 1964-1969, as well as the year immediately preceding this period, as well as the years following it. The Huber weights for the years 1950-1962 are all one for both approaches, and the Bi-Square weights are generally close to one.

Note that for the Bi-Square weighting function, regardless of approach, all the over-reported years are essentially removed from the data set (i.e., weights are zero), as is the immediately following year 1970. Close observation of Figure 49 seems to suggest that the over-reporting might have slipped into early 1970 before being rectified, so this seems reasonable.

Table 33: Huber and Bi-Square Weights for Belgian Telephone Call Data Models

Year	IRLS		Bounded Influence	
	Huber	Bi-Square	Huber	Bi-Square
1963	1	0.481	1	0.543
1964	0.121	0	0.190	0
1965	0.117	0	0.184	0
1966	0.100	0	0.157	0
1967	0.088	0	0.137	0
1968	0.075	0	0.117	0
1969	0.063	0	0.098	0
1970	1	0	1	0
1971	0.561	0.908	0.538	0.877
1972	0.588	0.998	0.537	0.998
1973	0.584	0.957	0.519	0.938

However, the Huber weighting scheme, again, for both approaches gives full weight to 1970 and down-weight the years following it. As a result, the observed 1970 call volume virtually determines the Huber fits.

Again, consistent with the previous applications of the Bootstrap approaches, Figure 53 shows that they behave similarly to the least squares approach with this Belgian call volume data. The slope estimates for these approaches are both near to 0.13 least squares result when the over-reported years are not included in the data, and they increase to 0.50 (same as least squares) and 0.53 for the Residuals and Observations approaches, respectively, when the outliers are involved.

Figure 53: Bootstrap Fits for Belgian Telephone Data

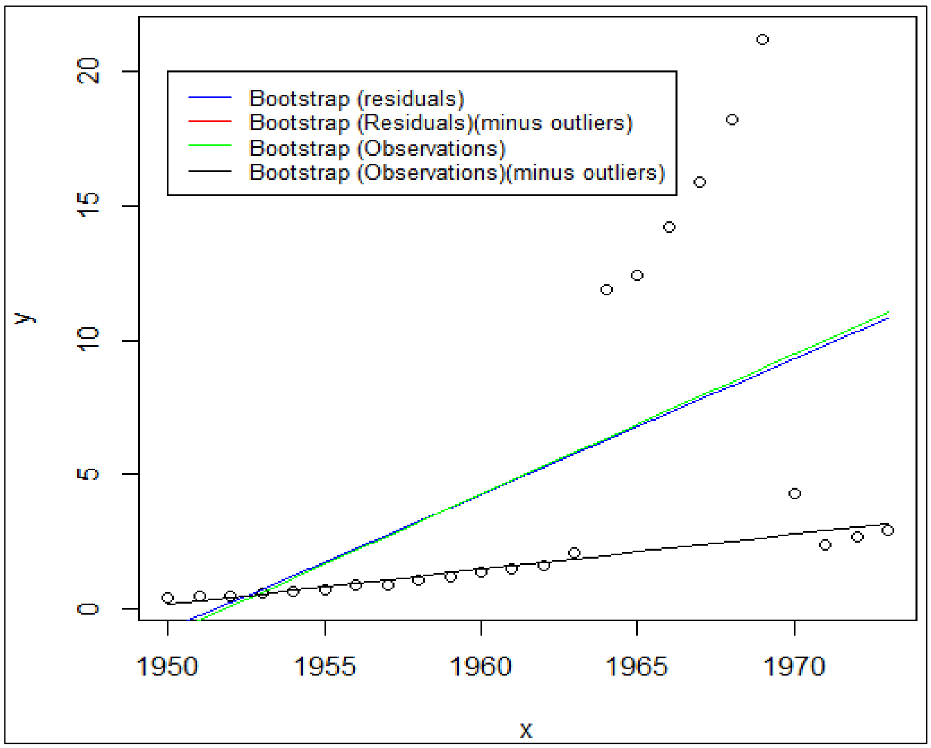
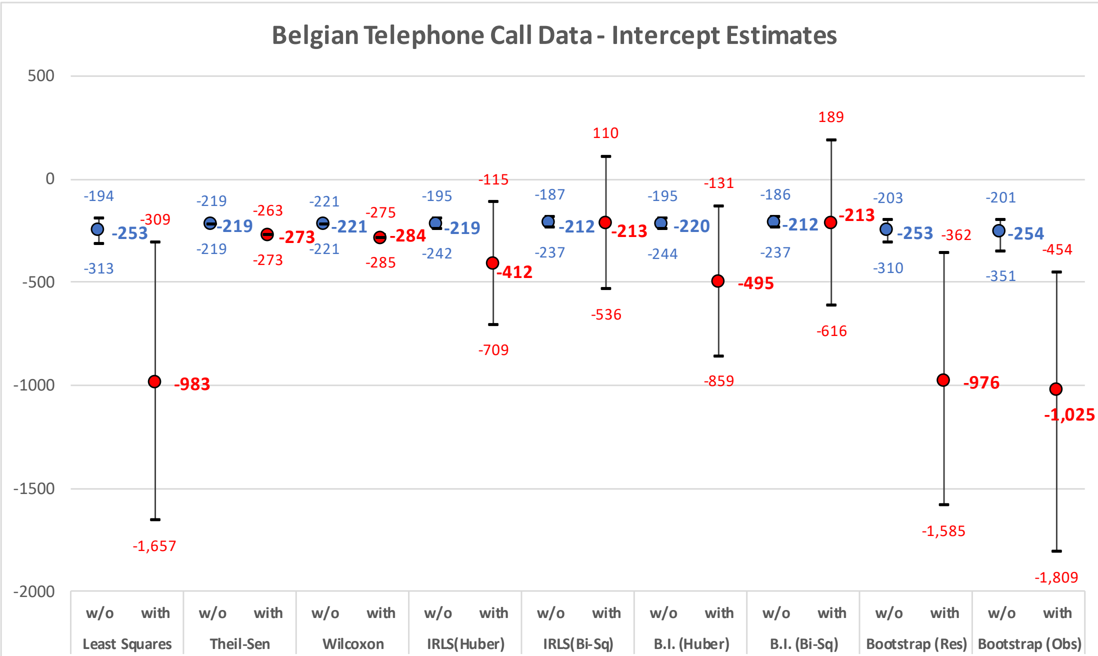
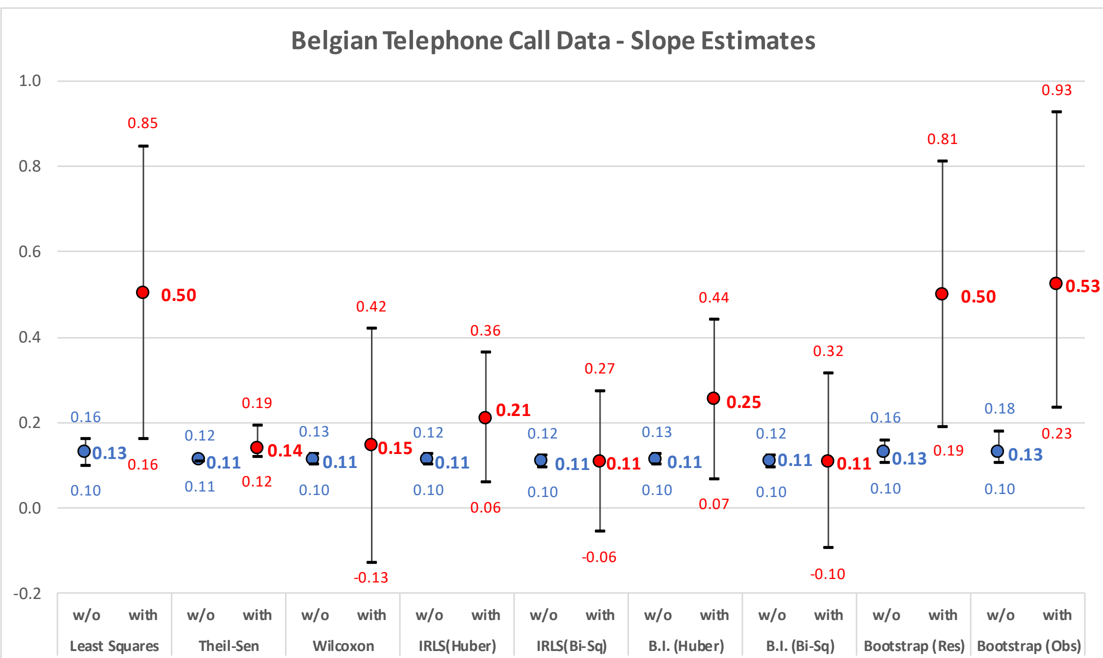


Figure 54 displays the intercept and slope estimates (point and approximate 95% confidence interval estimates) for the models displayed in Figures 50-53. Tables A13 (w outliers) and A14 (w/o outliers) in the Appendix display these results to more decimal places.

Figure 54: Confidence Intervals on Intercept and Slope Parameters for the Belgian Telephone Call Data





Without the over-reported years, the point estimates for slope are effectively the same for the approaches with least squares and the bootstrap approaches suggesting an ~1.3 million annual increase in call volume, and all other approaches suggesting a slightly lower rate at ~1.1 million. When the outliers are involved in the estimation process, only the approaches using the Bi-Square weighting function (both IRLS and Bounded Influence) retain nearly the same estimate. The Theil-Sen and Wilcoxon approaches suggest a modest increase in the rate at ~1.4 and ~1.5 million, respectively. The least squares and bootstrap approaches provide estimates indicating much larger increases at ~5 million (least squares and Bootstrap using Residuals) and ~5.3 million (Bootstrap using Observations).

The slope estimator confidence bounds are all relatively narrow when the outliers are not involved, and they all overlap to some degree. However, the width of these intervals increase by approximately an order of magnitude when the outliers are in the data set. The smallest increase is ~8.6X for the Theil-Sen approach, and the largest increase is ~21.2X for the Wilcoxon approach (this being the largest is expected, and why is discussed in Section 3.1 above). All the others lie between ~9.1X (Bootstrap on Observations) and ~15.6X (Bounded Influence – Bi-Square).

While unfortunate, such an increase is probably reasonable given the gross nature of the over-reporting for the years 1964-1969. Without knowledge that these are truly bogus results, it would be difficult to completely ignore their presence and the large uncertainty bounds would be one means to acknowledge that presence. Using a robust estimation approach for the fit parameters mitigates their impact on accurately describing the increase in annual call volume, but the wide intervals reflect the relative amount of certainty associated with those estimates. Without knowing the true nature of the very high values in the 1964-1969 timeframe, the uncertainty in the estimates would be larger.

For the Wilcoxon, and Bi-Square weighting function approaches (IRLS and Bounded Influence), this uncertainty would have been sufficient to seriously question the validity of a simple linear trend (the slope confidence intervals all include zero). Such a conclusion could reasonably be considered an appropriate one without further information clarifying the outlier results.

While the intercept parameter is of much less interest here (since it reflects the Belgian call volume in year zero), the results are reasonably consistent with those observed for the slope estimator.

3.3: University of South Florida Salary Data

At most major colleges and universities, administrators (i.e., presidents, vice-presidents, deans, and department chairs) are among the highest paid state employees. In the early 1990s, a group of members of the faculty union at the University of South Florida (USF) in Tampa, Florida was interested in determining if there was any relationship between salaries for such individuals and their job performance.

In order to attempt to evaluate this relationship, this group (which called itself the United Faculty of Florida or UFF) compared the ratings of 15 USF administrators to their subsequent raises in that year. The data appear in Table 34 below.

The ratings in Table 34 were measured on a 5-point scale with 1 = very poor and 5 = very good. These ratings were determined by surveying faculty at USF.

Table 34: University of South Florida Salary Data

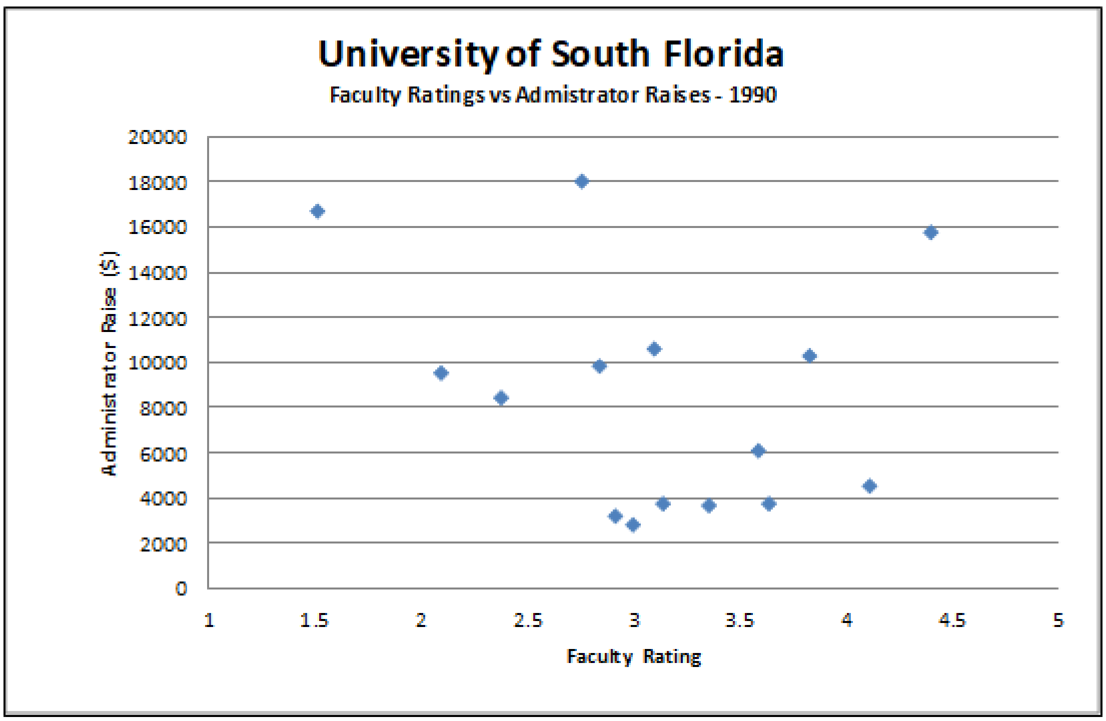
Administrator	RAISE($)[a]	Average Rating[b]
1	18000	2.76
2	16700	1.52
3	15787	4.4
4	10608	3.1
5	10268	3.83
6	9795	2.84
7	9513	2.1
8	8459	2.38
9	6099	3.59
10	4557	4.11
11	3751	3.14
12	3718	3.64
13	3652	3.36
14	3227	2.92
15	2808	3

[a] Faculty and A&P Salary Report, University of South Florida, Resource Analysis and Planning, 1990.

[b] Administrative Compensation Survey, *Chronicle of Higher Education*, Jan. 1991.

Figure 55 displays a plot of the data in Table 34 with the horizontal axis being the faculty rating and the vertical axis the respective raise for each administrator. Simple inspection of this plot suggests that the relationship is an inverse one. In other words, the lower the faculty rating, the higher the raise. This was the conclusion arrived the UFF arrived at, essentially leading them to summarize that poorly performing administrators were apparently more valuable than better performing administrators.

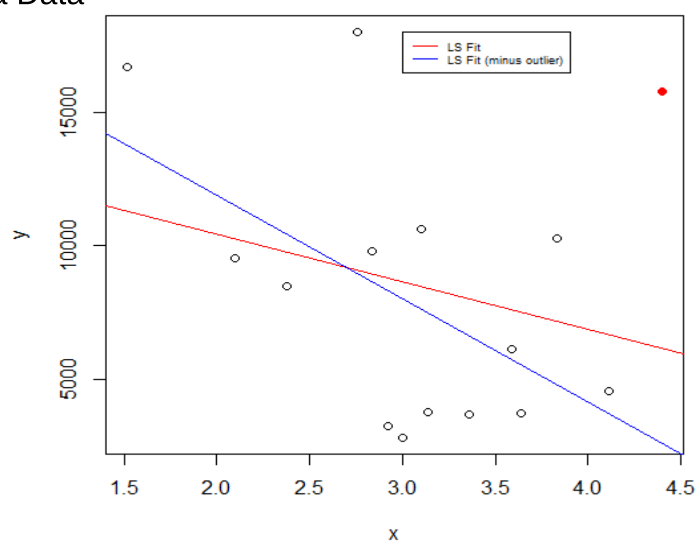
Figure 55: University of South Florida Salary Data



It happens that observation number 3 (4.4,1587) represents a faculty member who was promoted to a Dean position, which explains a portion of this individual's raise. This employee is distinguished from the rest in this manner and represents a potential outlier case. Therefore, linear models will be evaluated on this dataset with and without the potential outlier.

Figure 56 displays the least squares results for both the data sets – with and without Administrator 3. Note that the negative slope becomes more pronounced when the potential outlier is removed. For the UFF, this was desirable as they were interested in making a case that USF administrators were overpaid.

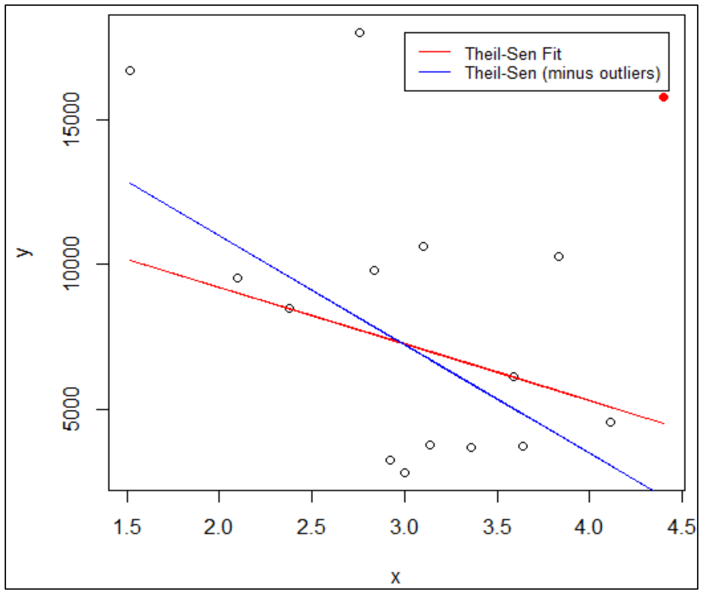
Figure 56: Least-Squares Fit for University of South Florida Data



The slope estimate when considering all 15 administrators is ~-$1,783, suggesting that the raise a USF administrator receives is reduced by this amount for every additional favorable rating point he or she receives from USF faculty. This estimate increases in magnitude to an approximate $3,887 reduction for every additional favorable rating point, or a more than 2X increase in magnitude, when the potential outlying administrator is removed from the data set.

Interestingly, as can be seen in Figures 57-59, none of the potentially more robust approaches seems to provide any result markedly different than the least squares results in Figure 56. It appears that this “outlier” is essentially not regarded as such by any of the more robust approaches.

Figure 57: Theil-Sen and Wilcoxon Fits for University of South Florida Data



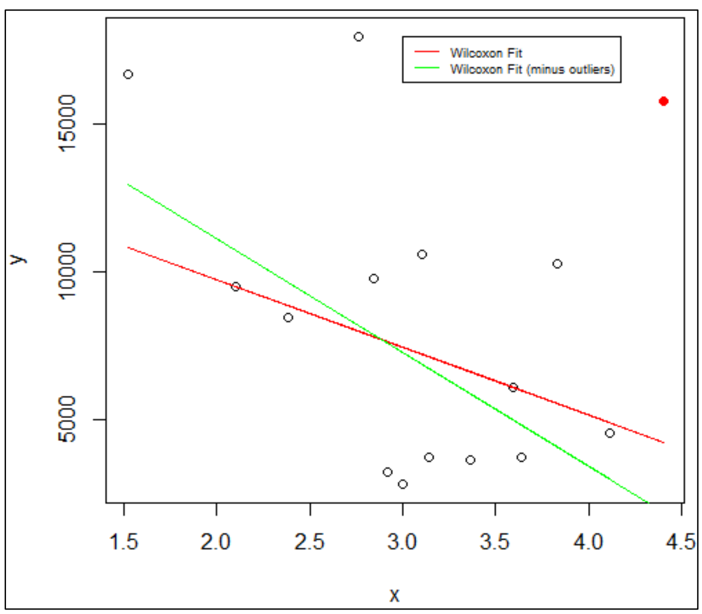


Figure 58: IRLS and Bounded Influence Fits for University of South Florida Data

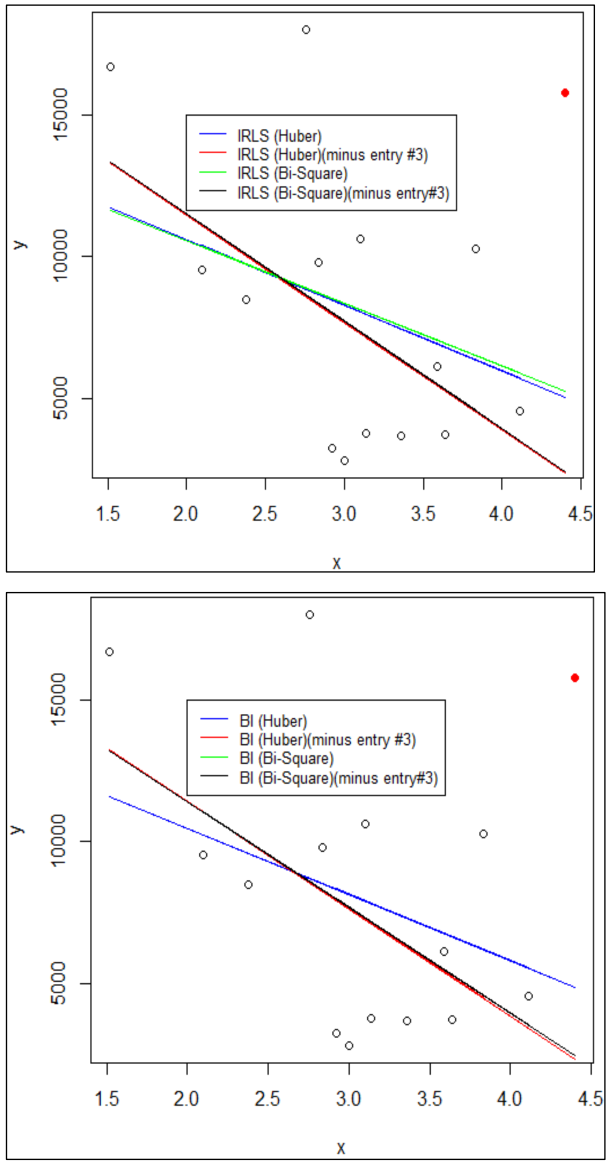


Figure 59: Bootstrap Fits for University of South Florida Data

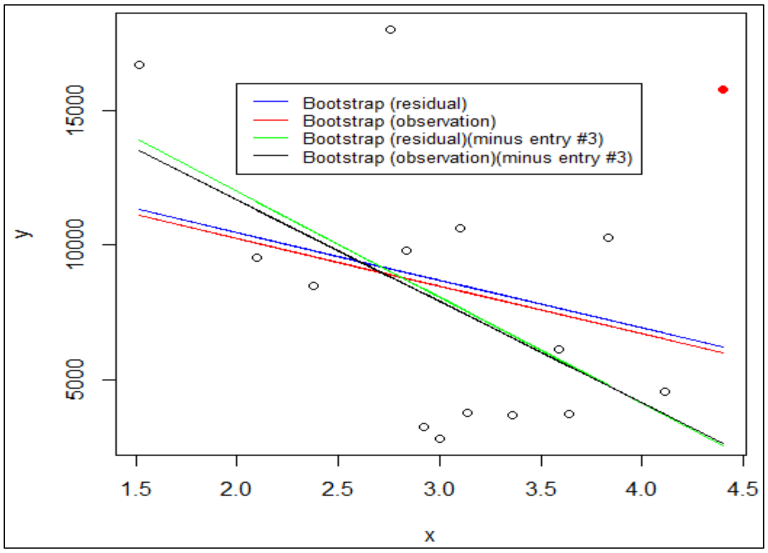


Figure 60 displays the estimated model parameters for the lines plotted in Figures 56-59. Tables A15 (w outlier) and A16 (w/o outlier) display these same results in tabular form and provide more decimal places.

Figure 60 clearly demonstrates the similarity of all the approaches for both situations – including and then not including the data for Administrator 3. It appears that this observation is not regarded as an “outlier” due to the large amount of residual error remaining when fitting any of these models.

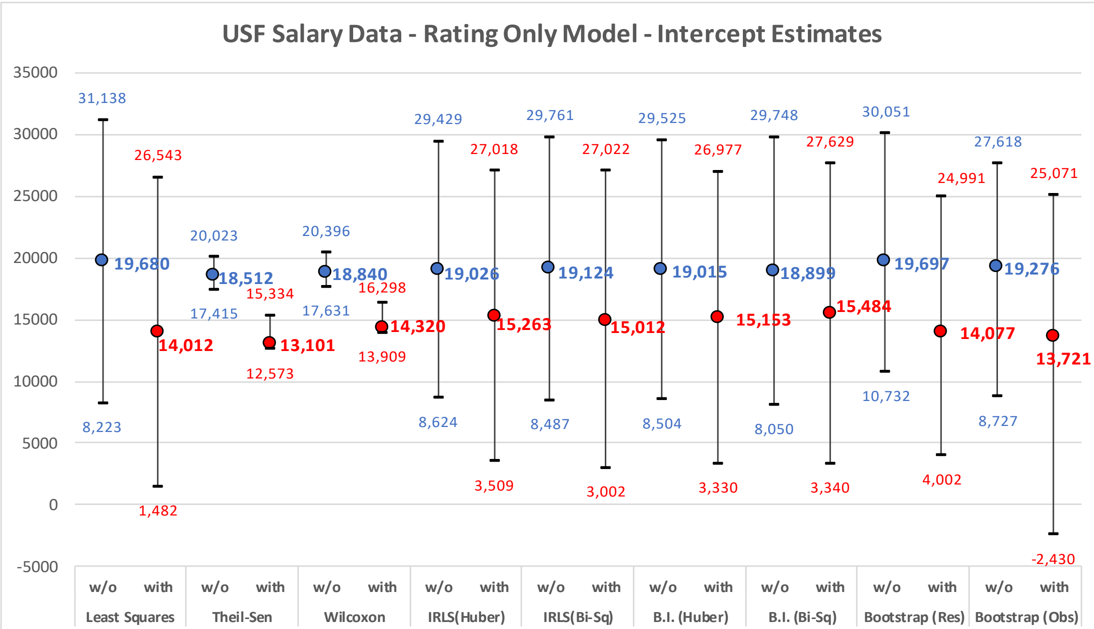
Of course, the important aspect of these charts for the UFF agenda is that the negative slope be considered significantly different from zero, and that it be negative. For the commonly applied least squares approach, this will be the result if Administrator 3 is removed from the data set. This is indeed how the

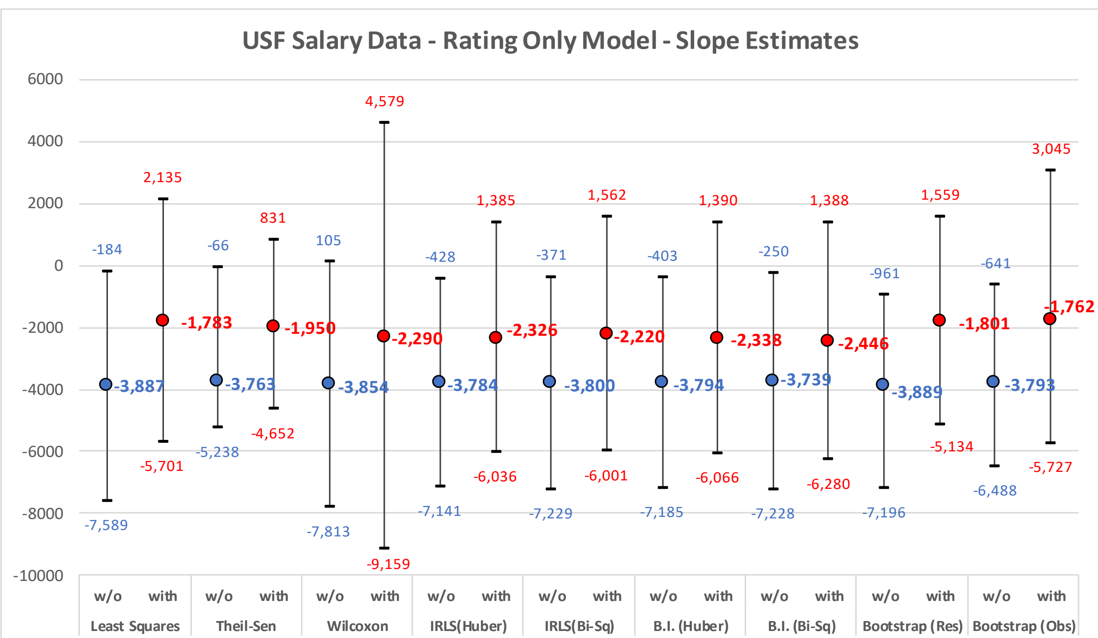
UFF decided to evaluate the data and present their results to the entire USF community.

When the entire data set is considered (including Administrator 3), all of the approaches indicate that there is no statistically significant relationship between the amount of a USF administrator's raise and the job performance rating he or she receives from USF faculty. All of the slope estimator confidence intervals include zero as a reasonable value for the true relationship parameter value.

On the other hand, when the data for Administrator 3 is removed from consideration, all of the approaches, except the Wilcoxon (where the variance estimate is known to tend to be large) suggest that there is a statistically significant relationship between a USF administrator's annual raise amount and the job performance rating that administrator receives from USF faculty members. In addition, the nature of this relationship is that apparently the better job the faculty perceives the administrator is doing, the lower his or her annual raise. Or stated differently, those administrators more poorly perceived by USF faculty tend to receive larger annual raises. Members of UFF dubbed this result the "son of a bitch" factor upon completing this analysis.

Figure 60: Confidence Intervals on Intercept and Slope Parameters for the USF Salary Data – Rating Only Model





What the UFF apparently failed to understand is that Human Resource organizations classify employees by title into salary ranges. Even among administrators there are likely to be multiple salary ranges with higher ranges for those individuals in the upper levels of administration, as successively higher levels carry successively more responsibility.

Consequently, in conjunction with awareness of the likely presence of multiple administrative salary levels, close observation of Figure 55 suggests there are at least three administrative levels represented.

When the administrators are grouped according to their respective salary level (High, Low, or Middle), there actually appears to be a positive linear relationship between average rating and salary raise within each of the respective groups. In Table 34 above, Administrators 1-3 would be considered members of the highest salary range, such as Deans, Provosts, and Assistant Provosts. Administrators 4-8 represent the mid-level salary range, such as Associate Deans and Directors. Administrators 9-15 represent the low-level salary range, such as those with the positions of Chair or Assistant Chair.

Once the administrators are grouped according to their respective salary ranges, the data may suggest a positive linear relationship between average salary raise and faculty rating, in contrast to the negative relationship observed thus far. Accordingly, an approach similar to a blocking scheme is proposed.

Therefore, we will fit attempt to fit the following model to the data using the proscribed methods:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i, \text{ for } i = 1,2, \dots, n, \qquad (39)$$

where y_i is the salary raise associated with the i^{th} administrator,

β_0 is an intercept parameter,

β_1 is the slope parameter for the rating variable x_{1i},

$$x_{2i} = \begin{cases} 1 & if \quad adminstrator \text{ "i" is in the "High" } group \\ 0 & otherwise \end{cases},$$

$$x_{3i} = \begin{cases} 1 & if \quad administrator \text{ "i" is in the "Low" } group \\ 0 & otherwise \end{cases},$$

β_j is an offset amount for salary group $j = 2, 3$,

and ε_i is a random variable error term associated with the i^{th} faculty member.

This model will create three different fits for the USF data, with one fit for each of the three groups based on common salary level. The coefficients β_2 and β_3 represent the differences in salary levels between three different groups. The β_1 model coefficient provides a common slope parameter for any effect faculty rating might have in relation to the raise amounts received once salary differences have been accounted for across the groups

. Thus, the equations for High-Level, Medium-Level and Low-Level faculty members' raises are given respectively by equations (40) through (42) below:

$$y_i = (\beta_0 + \beta_2) + \beta_1 x_i + \varepsilon_i, \text{ for } i = 1,2,3 \qquad (40)$$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ for } i = 4,5,6,7,8 \quad (41)$$

$$y_i = (\beta_0 + \beta_3) + \beta_1 x_i + \varepsilon_i, \text{ for } i = 9,10,11,12,13,14,15 \quad (42)$$

For all approaches except Theil-Sen and Wilcoxon, extending calculations to accommodate the two newly-introduced indicator variables is relatively straightforward, as only minor adjustments in the predictor matrix **X**, as given in equation (8), will be necessary. The Theil-Sen and Wilcoxon approaches, however, cannot be effectively handled with such a simple extension.

Instead, adjustments for administrative salary level will be made by subtracting group-wise medians for the observations when using these approaches. Group-wise medians are chosen rather than means to preserve consistency with the estimators of these non-parametric approaches.

When this is done, β_2 will be estimated by the difference between the medians of the High and Medium-level groups, respectively, and β_3 will be estimated by the difference between the medians of the Medium and Low-level groups. This interpretation for β_2 and β_3, as the general difference in raise amounts across salary levels, will be the same for all the other approaches where managing the presence of these different groups is more directly managed with the modeling approach.

Figures 61 displays plots of the least squares models both with and without the potential outlier observation (Administrator 3). Note that the slope estimate for the relationship between the raise amount and the faculty rating is larger at ~$1100 per rating point when the data for Administrator 3 is not considered. When this observation is included in the modeling effort, the slope estimate is ~$230 per rating point; still positive but less than a quarter of the magnitude of the estimate obtained with this observation removed from the data set.

Figure 61: Least-Squares Fits for Grouped USF Data

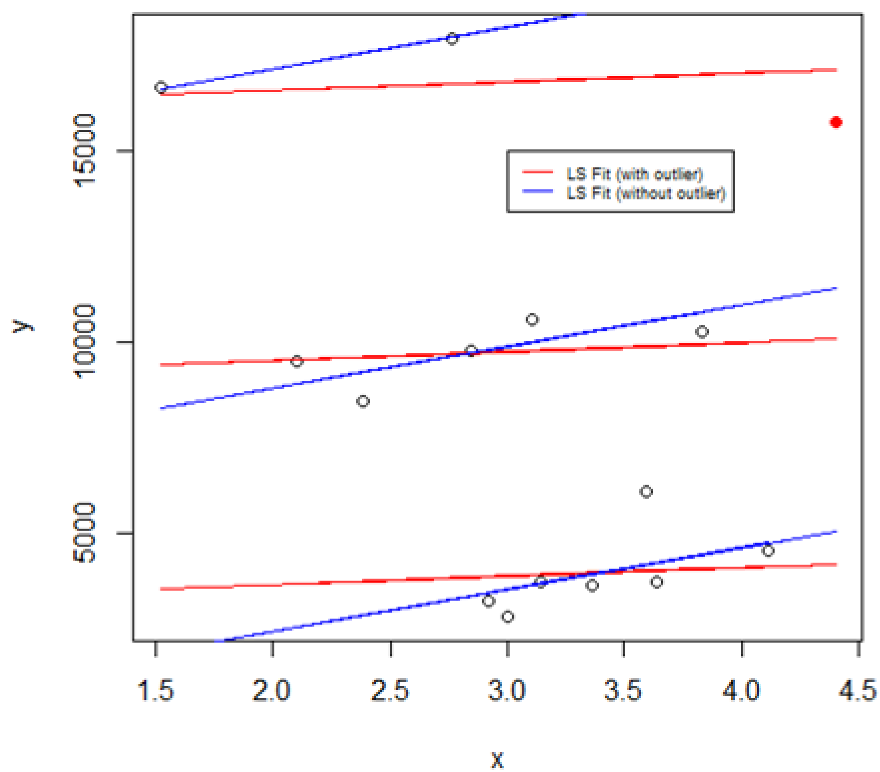


Figure 62 displays the Theil-Sen and Wilcoxon models for the grouped data. These models also show a larger slope estimate when the outlier is not involved; however, the increase versus when this observation is involved is not nearly as large. These slopes both suggest an ~$600 increase in raise amount per faculty rating point, or only about 3X larger than the ~$210 increase per rating point when the Administrator 3 data is used.

Figure 63 displays the models fit to the grouped data using the IRLS approach using both the Huber and the Bi-Square weighting functions. It appears that this approach recognizes that Administrator 3 is an unusual observation as its final Huber weight is ~0.215, and the Bi-Square weight is zero, effectively removing the observation from the data set. This approach also suggests that the observation for Administrator 10 is almost similarly unusual as it receives a Huber weight of ~0.233, and also receives a zero Bi-Square weight (i.e., it is also eliminated from the data set).

However, these weighting results are sufficiently different that the Bi-Square function essentially returns the same slope estimate (an ~$930 increase in raise per faculty rating point) both when Administrator 3 results are used and when they are not, while for the Huber weighting function, the slope estimate suggests a raise increase of only ~$530 per faculty rating point. However, this estimate does increase to ~$940 (i.e., near the Bi-Square function estimates) when the Administrator 3 data is not considered.

Figure 62: Theil-Sen and Wilcoxon fits for Grouped USF Data

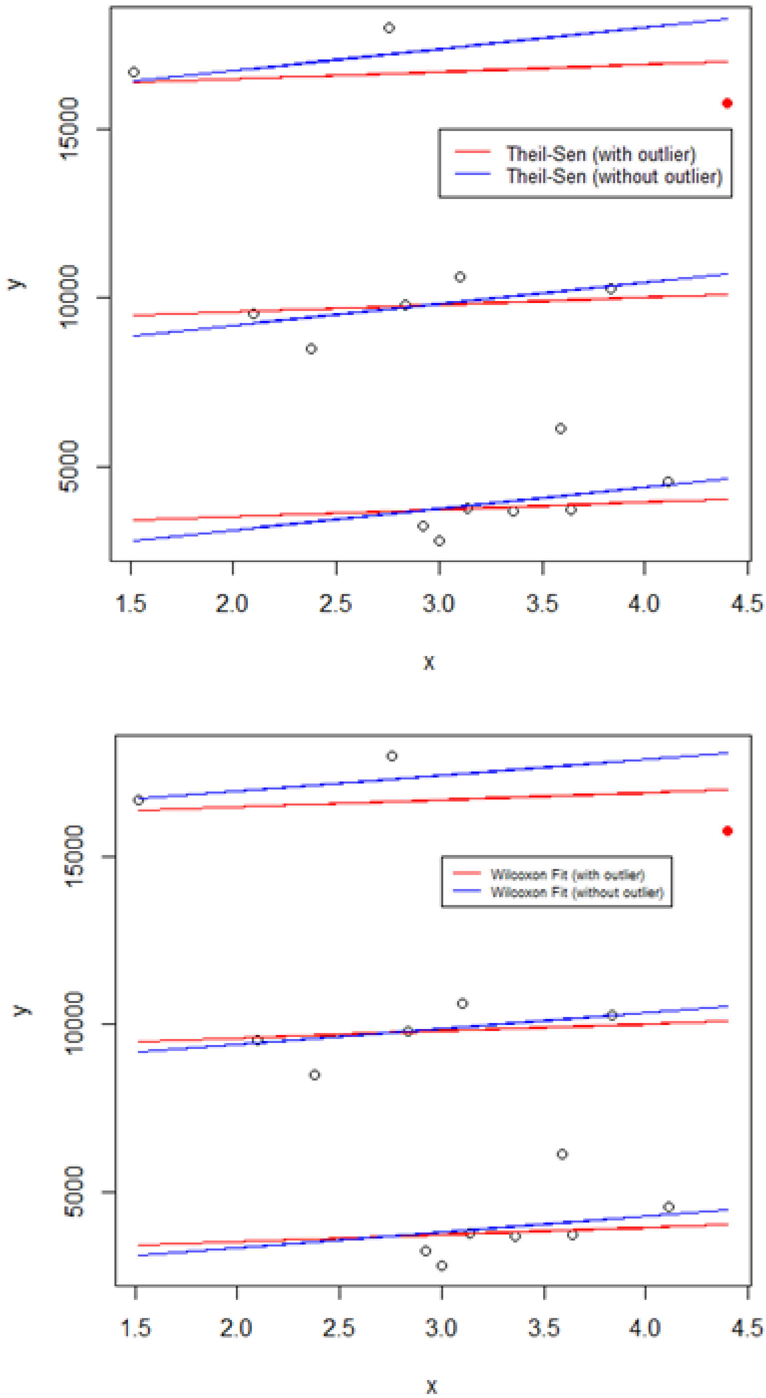
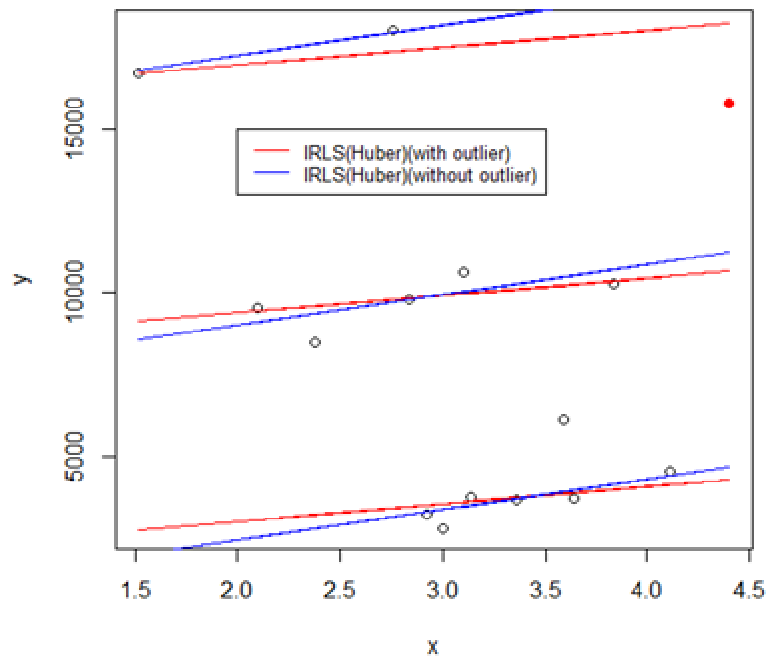


Figure 63: IRLS Fits to Grouped USF Data



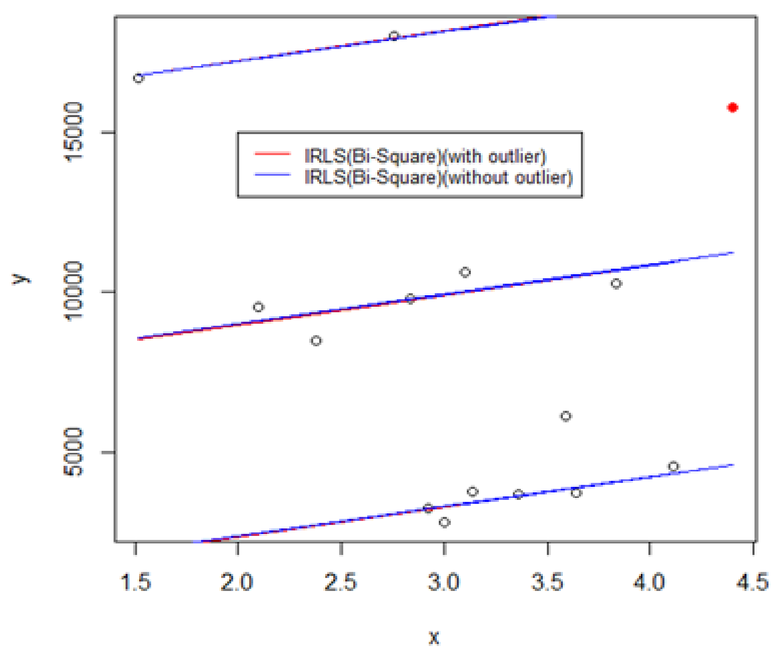
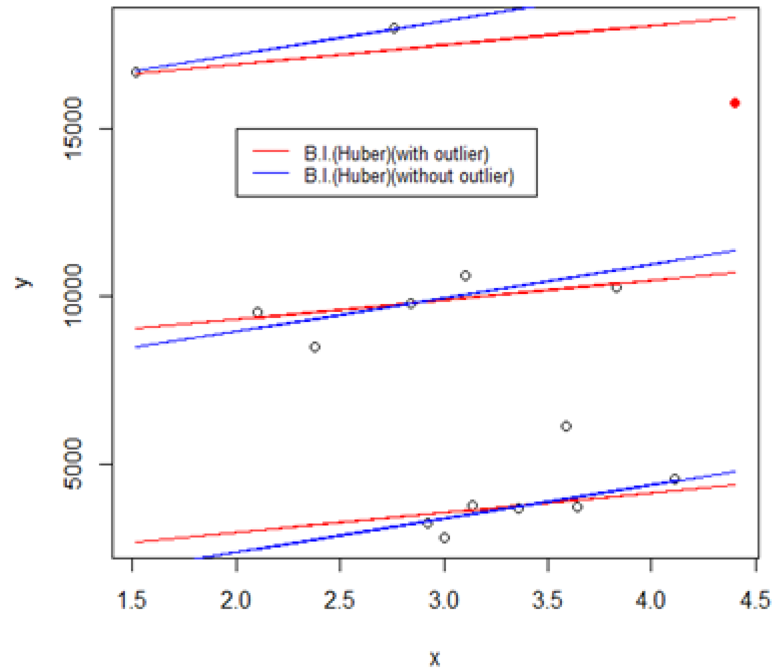
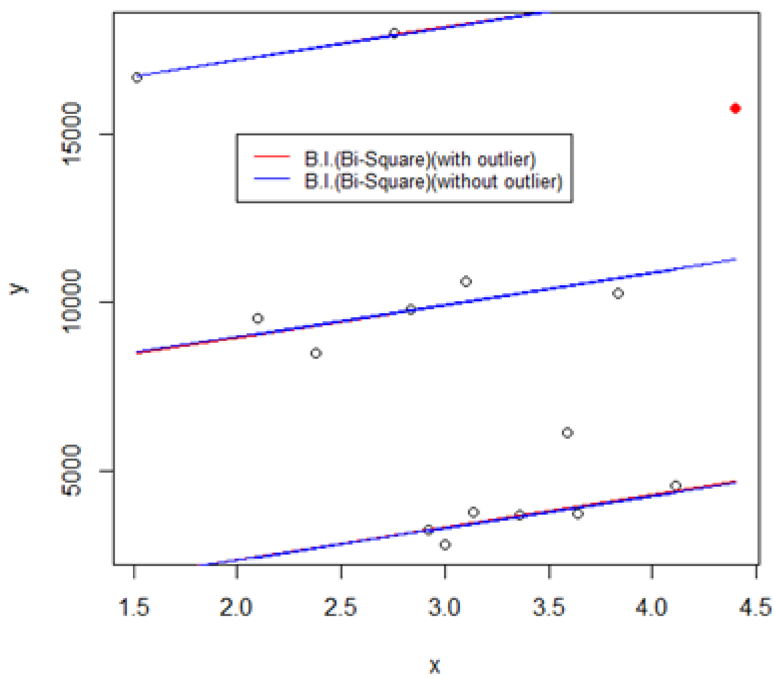


Figure 64 is analogous to Figure 63, except the Bounded Influence approach is used. The results for this approach are similar to those for the IRLS approach; however, the estimated slopes are marginally larger in each case. The Bounded Influence Huber weight applied to the outlier is lower at ~0.183 than the corresponding IRLS weight (~0.215). The Bounded Influence Bi-Square weight for Administrator 3 is again zero as observed for the IRLS Bi-Square weight. However, for Administrator 10 (thrown out by IRLS Bi-Square), the Bounded Influence Bi-Square weight is ~0.068 (> 0, so not eliminated from the data set). The Bounded Influence Huber weight for this observation is also larger than the corresponding IRLS weight (at ~0.286 vs ~0.233).

Figure 65 displays the Bootstrap fits for the grouped data, both the fit resampling residuals, as well as the fit resampling entire observations. When all the data is involved, the residuals approach is similar to least squares, which is consistent with applications to other data sets investigated previously in this work. However, the observations approach seems to provide a slope estimate that is nearly twice as large as the least squares and bootstrap-residuals approaches.

Figure 64: Bounded Influence Fits to Grouped USF Data



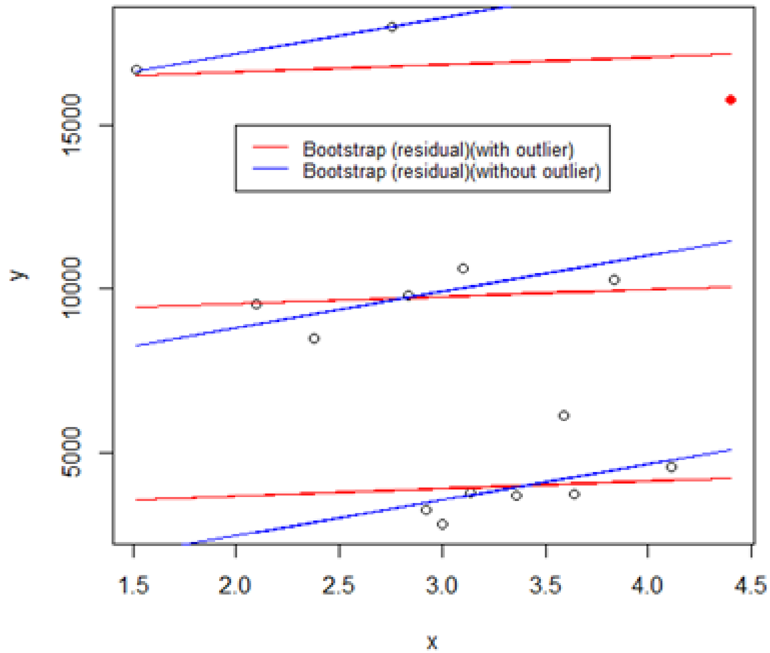


When removing Administrator 3, the bootstrap-residuals slope estimate (~$460) does not mimic the corresponding least squares estimate (~$1100), but the bootstrap-observations slope estimate is again about twice the magnitude of the bootstrap-residuals estimate (~$870). So, while the increase for both of these bootstrap slope estimators is ~2X when the outlier is removed from the data, this increase is much smaller than the almost 5X increase observed for the least squares approach.

As can be seen in Figure 66, much of the discussion of these differences is likely to be of little significance given the relative magnitude of the associated confidence intervals for these slope parameters. All of them are sufficiently wide to encompass all the other point estimates. However, the intervals for the Bi-Square weighting function approaches are sufficiently narrow to suggest a significantly positive relationship between the raise amount received by a USF administrator and their corresponding faculty rating, even when the data for Administrator 3 is involved (although, the Bi-Square weight throws it back out).

When the Administrator 3 data is initially removed from the data set, all the approaches, except the Wilcoxon and for those using the bootstrap, suggest a significantly positive relationship between raises and ratings. Note that this is exactly the opposite direction of the relationship between these two variables suggested when the data was not grouped by salary levels.

Figure 65: Bootstrap Fits for Grouped USF Data



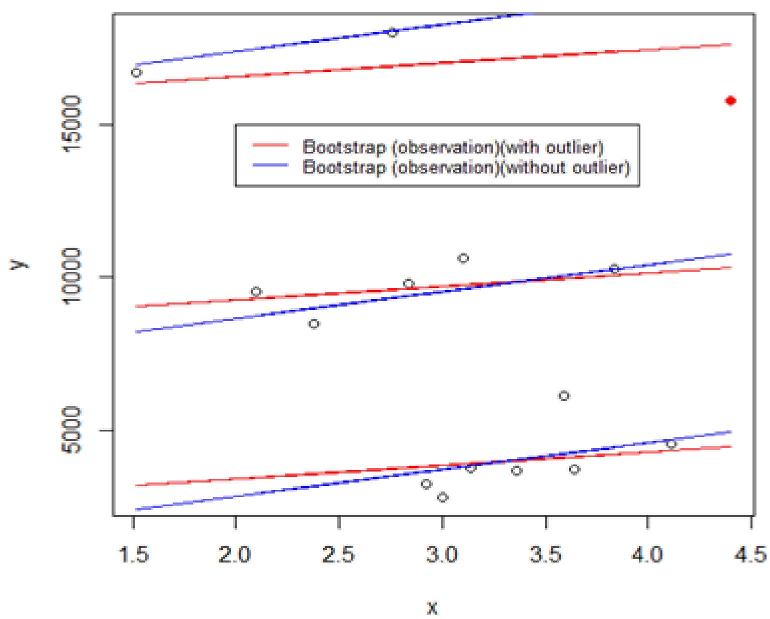
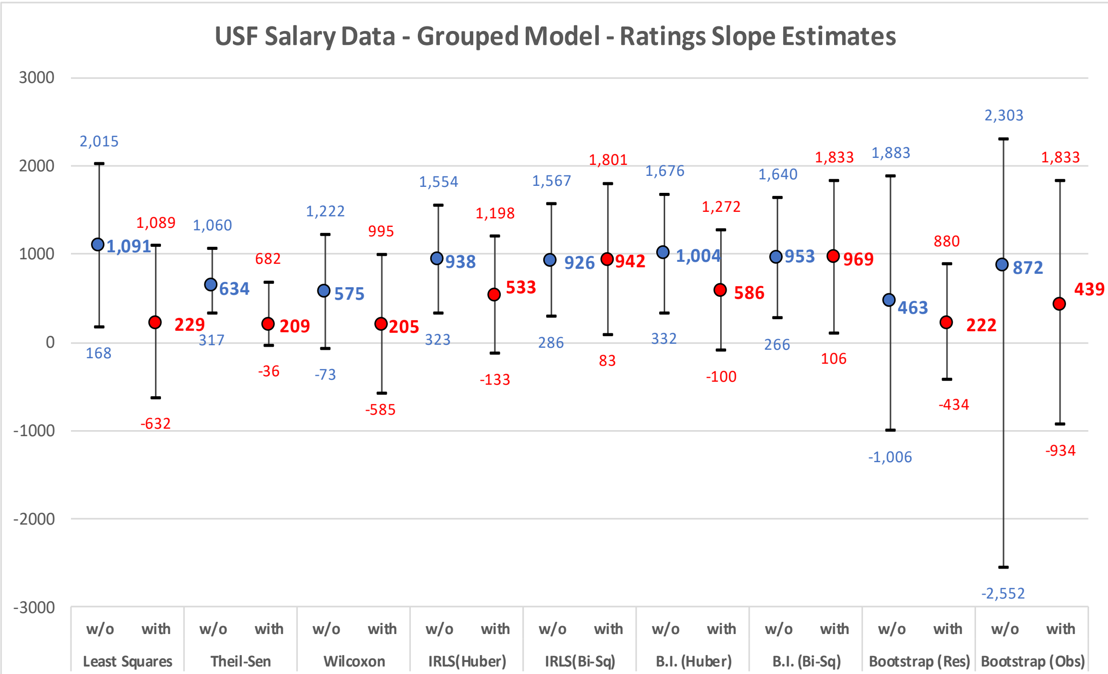


Figure 66: Confidence Intervals on Ratings Slope Parameter for the USF Salary Data – Grouped Model



The consistency for the Bi-Square weighting approaches (IRLS and Bounded Influence) again suggests that these approaches are the most robust of those evaluated here. However, the clear message related to the analysis of this USF salary data is that the success of virtually any modeling approach is much more dependent upon having sufficient contextual knowledge to fit an appropriate model than it is related to the statistical approach used to fit the chosen model. Also, the words of George Box are again relevant: “All models are wrong, but some are more useful than others” (**#13**). The model fit without grouping the data by salary level and not using the Administrator 3 data was indeed “useful” to the UFF in promoting their agenda of USF administrators being overpaid. However, it was not a “useful” model for accurately describing a valid relationship between raise amounts and faculty ratings.

The lower ratings for administrators more removed from the level of faculty (i.e., those at higher levels of administration) would be expected. It reasonable to expect faculty to have less interaction with individuals at higher levels, and less understanding of their roles and responsibilities, and to, therefore, be more likely to assign such individuals lower ratings than the level of administrators with which they have more interaction and understanding. However, within the respective administrative levels, it appears faculty ratings and performance are positively related. Faculty apparently does have some idea of relative performance within level.

The value of context and understanding that context when conducting any statistical analysis can rarely be underestimated. Without this, the choice of statistical methods to apply in any situation is almost irrelevant.

Section 4: Simulation Study

A simulation study was performed to provide a more general comparison of the performances of the regression approaches considered in this work. The responses of interest in this study were defined as the SLR parameter estimates, confidence interval widths, and capture rates among the selected regression approaches. The simulations were generated data from contaminated normal distributions with varying sample sizes, contamination probabilities, contamination variance multipliers, and predictor space outlier locations. These samples formed the error terms for the model given below:

$$y_i = 2 + 3x_i + \varepsilon_i, for\ i = 1,2, \dots \boldsymbol{N}, \quad (43)$$

...where $\varepsilon_i \sim N(0,1)$ with probability (1-**L**), and $\varepsilon_i \sim N(0, \boldsymbol{V})$ with probability "**L**" for each $i = 1,2, \dots N$.

Consequently, the simulation design involved a total of four variables, which are listed below:

L, the contamination frequency,

N, the sample size,

V, the contaminated variance multiplier, and

K, which determined the location of the predictor space outlier (explanation below).

Here, the predictor values ($x_i's$) were chosen to be an equally partitioned sequence from $\frac{1}{N+1}$ to $\frac{N}{N+1}$. These were uniformly the same for all simulations, except that the middle predictor value ($\frac{N}{2(N+1)}$) was modified to become ($\frac{K*N}{2(N+1)}$).

This was done so that for certain values of K, the original median point of the predictor values would shift to form an outlier in the predictor space. Values of 1, 4 and 8 were chosen for K, so that the original median predictor value enlarged to become twice or four times as large as the original maximum predictor value ($\frac{N}{N+1}$) when K = 4 or 8. This allowed for comparisons in the presence of predictor space outliers as well as response outliers.

Note that when **K** =1, this middle predictor value has a minimum leverage value for the respective sample size (for **N** = 10, $h_{5,5}$ ≈ 0.103, for **N** = 20, $h_{10,10}$ ≈ 0.050, and for **N** = 40, $h_{20,20}$ ≈ 0.025). When **K** = 4, the new predictor value has much larger leverage; for **N** = 10, $h_{5,5}$ ≈ 0.726, for **N** = 20, $h_{10,10}$ ≈ 0.576, and for **N** = 40, $h_{20,20}$ ≈ 0.408. For all these sample sizes, these leverage values are now the largest in the predictor variable space, and are approximately 3, 4, and 5 times as large as the next largest leverage value, respectively.

When **K** = 8, the leverage values are now given as: for **N** = 10, $h_{5,5}$ ≈ 0.936, for **N** = 20, $h_{10,10}$ ≈ 0.880, and for **N** = 40, $h_{20,20}$ ≈ 0.786. These values are now much larger than the next largest leverage value at ~6X, ~11X, and ~19X larger, respectively.

The impact of this predictor space outlier is likely to be mitigated to some degree in the actual simulation since for many of the simulation iterations the presence of an outlier in the response may not occur at this point. When a standard normal error occurs for this predictor variable value, it is likely that this point will effectively determine the slope estimate. However, if a contaminated error occurs at this predictor variable value, then it is likely that least squares regression will experience issues. It is expected that the more robust approaches would be less impacted in such a situation.

The levels for each of the four simulation design variables are presented below:

L = (0, 0.1, 0.2)

N = (10, 20, 40)

V = (2, 6, 10)

K = (1, 4, 8).

When **L** (the contamination frequency) is zero, the contamination variance multiplier **V** becomes obsolete because the simulated error terms no longer come from a contaminated normal distribution, but a standard normal one. Therefore, when **L** = 0, a full factorial design was evaluated with the two remaining variables **N** (sample size) and **K** (predictor space outlier location). Since there are three levels of each factor, simulations are performed on a total of 3x3 = 9 paired combinations when **L** = 0.

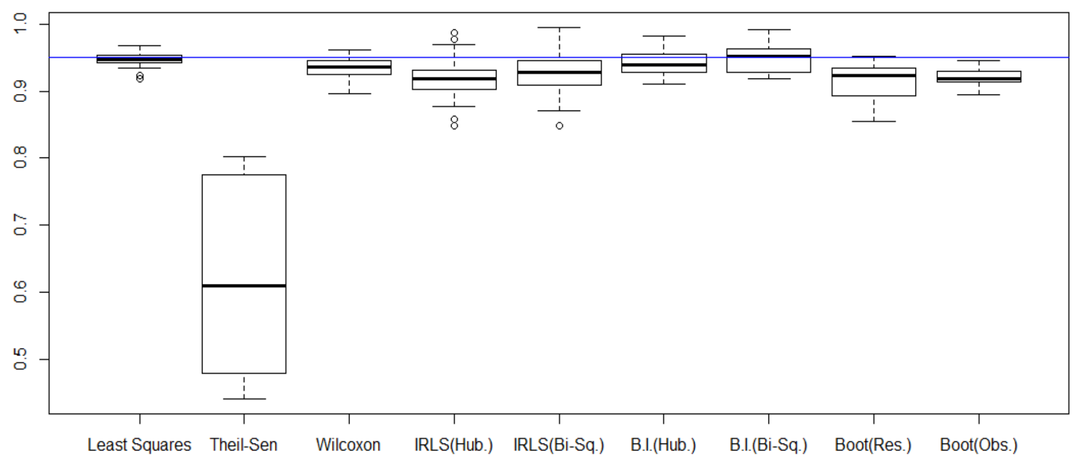
When **L** = 0.1 or 0.2, the contamination variance multiplier "**V**" becomes relevant, and a face-centered cube design is used with three variables (or factors). Let "**N**", "**V**" and "**K**" represent the axes of a cube, with the levels of each respective factor ordered from one side of the cube to the other side. Since there are three levels to each of the three factors, the cube is effectively partitioned into 8 equally-sized smaller cubes, with the vertices of each smaller cube representing a specific three-way factor combination. The face-centered cube design performs simulations and analysis for each factor-combination that represents either a corner of the cube, a center of one of its faces, or the center of the cube. A cube has 8 corners, 6 faces and one "center point", so this amounts to 15 different factor combinations for **L** = 0.1 and the same for **L** = 0.2. Therefore, adding up the number of factor combinations for each level of **L** results in 9 + 15 + 15 = 39 factor combinations for the simulation effort.

For each of the 39 factor combinations, 1,000 data simulations were performed (under each regression approach) and all regression approaches were applied to these. For each regression performance, the errors in parameter estimates (defined as $b_i - \beta_i, for\ i = 0{,}1$) were recorded, as well as confidence interval widths for the parameter estimates, and their capture rates of the true parameters (β_0 = 2 for intercept and β_1 =3 for slope).

Capture Rates (Slope):

Figure 67 displays boxplots of the distributions of capture rates for the slope parameter across all 39 factor-level combinations across all methods. Except for the Theil-Sen estimator, the capture rate distributions are centered marginally below the nominal 95% level, with all median capture rates greater than 0.90.

Figure 67: Capture Rates Across All Factor-Level Combinations



The only approach that appears to compete favorably with least squares for capture rate is the Bounded Influence (Bi-Square) approach. This approach even had a larger mean capture rate than least squares (0.949 vs. 0.947); although, a smaller median capture rate (0.948 vs. 0.952).

Among the various factor-level combinations, the Bounded Influence (Bi-Square) approach experienced lower slope capture rates under the opposite conditions than those of least squares. The Bounded Influence (Bi-Square) capture rates were always above 0.918, and 7 of the 11 instances where the observed capture rate was below 0.93 were on simulations with no contamination present. This suggests that this approach tended to down-weight observations that were in no need of being down-weighted.

By contrast, least squares capture rates were lowest (below 0.93) when sample sizes were small (**N** = 10), the contamination level was high (**L** = 0.20), there was a large predictor outlier (**K** = 4 or 8), contamination variance was larger (**V** = 6 or 10).

The most notable result was the generally poor coverage for the Theil-Sen slope estimator where the capture rates were much lower than the nominal ~95%. The bulk of this capture rate deficit appears to be related to sample size. When N = 10, the Theil-Sen capture rate was between 70% and 80%, and when N = 40, the capture rate was reduced to between 40% and 50%.

This phenomenon indicates that the common approach of finding a confidence for a median does not perform as expected. This may be related to the distribution of the pairwise slopes used in Theil-Sen confidence intervals being decidedly heavy-tailed; so much so that their percentiles cannot be trusted to accurately reflect the ~95% confidence bounds on the pairwise slope median.

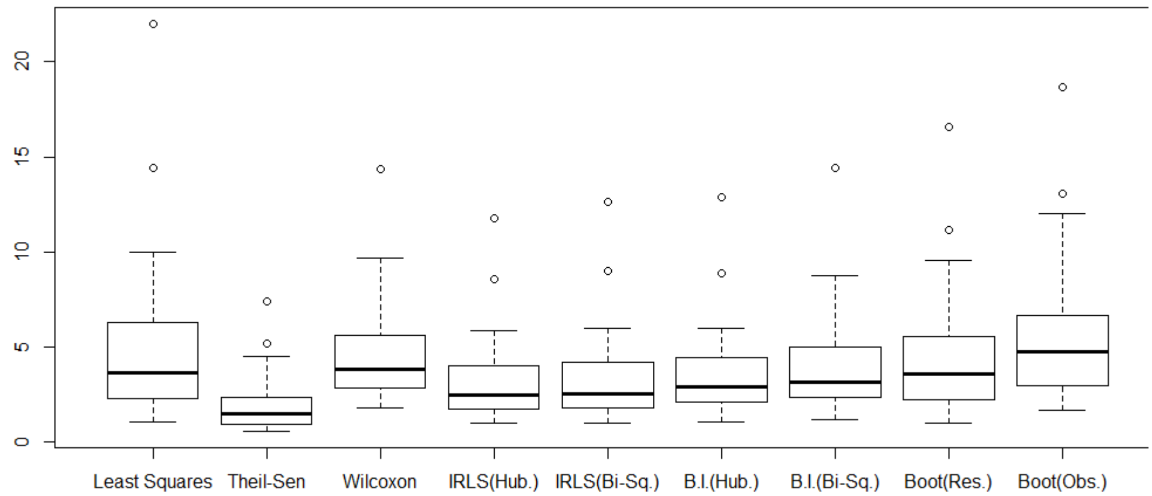
Another consideration is that the order statistics that form the confidence interval bounds for Theil-Sen become proportionally closer and closer to the mean/median order statistic as the sample size becomes larger. For example, there are 45 pairwise slopes when N = 10, which results in confidence bounds which use the 16th and 30th order statistics of these pairwise slopes. These order statistics differ by 14, which is a little less than one third of the overall percentile range. When N = 40, there are 780 pairwise slopes, and this results in order statistics 363 and 418 being used as the confidence bounds. Clearly, the interval from 363 to 418 represents a much narrower proportion (less than 10%) of the overall range of the order statistics than in the case with N = 10. This may provide some insight into why the Theil-Sen capture rates decrease so heavily with larger sample sizes.

Confidence Interval Width (Slope):

Figure 68 displays the distributions of mean (slope) confidence interval widths across all 39 factor-level combinations. It is readily apparent that the Theil-Sen confidence interval widths are noticeably lower than those for the other approaches. Given the much lower capture rates of this approach, this is unsurprising.

These confidence interval width distributions across simulation conditions for all the other approaches are similarly skewed to the high side. However, the distribution of least squares widths has a marginally larger skew, and larger extreme values. Both of these extreme values occur with the maximum magnitude of contamination variance (**V** = 10) and the narrowest design matrix (**K** = 1). These would indeed be the conditions where least squares estimators would be expected to experience the most difficulty.

Figure 68: Confidence Interval Widths Across All Factor-Level Combinations



Disregarding the poor coverage Theil-Sen intervals, the weighted approaches (IRLS and Bounded Influence) have generally narrower (on average) interval widths across the simulation conditions. For the Bounded Influence (Bi-Square) approach, the largest mean width is when **N** = 10, **L** = 0.20, **V** = 10, and **K** = 1 (same as for least squares); however, this mean width is ~2/3 of the width of the least squares intervals at this combination of conditions. Actually, for all of the approaches, the two largest average width results occur when **N** = 10, **V** = 10, **K** = 1, and **L** = 0.20 (largest) and **L** = 0.10 (next largest).

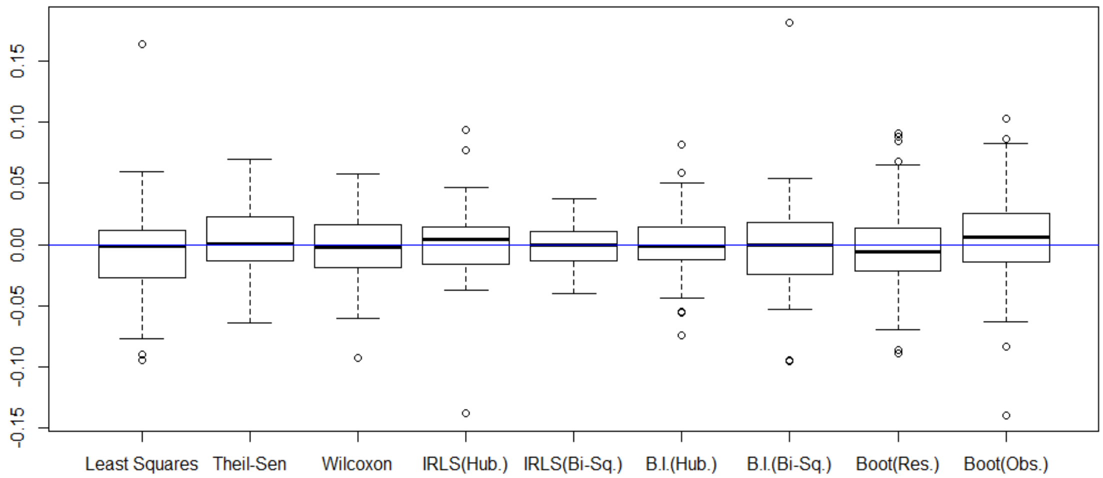
While none of the estimation methods employed a particularly robust estimator of error variance, it does appear that the mean confidence interval width results are related to robustness of that estimate. The least squares, bootstrap and Wilcoxon methods generally have larger mean widths, and the estimators employed for each of these approaches were known to be marginally less robust to extreme observations than those employed for the other methods. It should be noted again, however, that the even with apparently wider intervals, all of these approaches still had coverages that were marginally lower than 95%, on average.

Bias/Error (Slope Estimates):

Figure 69 presents boxplots of the distributions of mean slope estimation bias across all factor level combinations. Consequently, each observation contributing to this boxplot represents the mean bias in slope estimation across all 1,000 iterations for a specific factor-level combination.

As seen in the figure, the mean bias distributions are centered near zero for all approaches, which is consistent with all of these estimators being unbiased. The two largest mean bias results appear for the Bounded Influence (Bi-Square) approach (at **N** = 10, **V** = 10, **L** = 0.20, and **K** = 1 – same conditions as for largest confidence interval width using this approach), and for least squares (at N = 40, V = 10, L = 0.20, and K = 1); however, these results are still less than 0.20 in magnitude. This is less than 7% of the value of the parameter (i.e., 3) and less than 1/5th of the magnitude of the uncontaminated error distribution (i.e., 1).

Figure 69: Slope Mean Bias Distributions Across All Factor-Level Combinations



In summary, all the approaches have reasonably narrow mean slope estimate bias distributions across the space of conditions considered in the simulation. Almost all the sets of conditions considered produced mean slope bias results within ±2% of the magnitude of the slope parameter value.

Inter-Quartile Ranges of Slope Estimation Bias:

Those approaches with more robust estimates would generally be expected to have less variance in their bias distributions. To examine this expectation, Figure 70 displays boxplots of distributions of inter-quartile ranges for the slope estimates across all 39 sets of simulated conditions. The IQR (inter-quartile range) of slope estimator bias results across the 1000 simulated results for each set of conditions provides a robust estimate of the variance or spread of the slope bias distribution.

Figure 70: Inter-Quartile Ranges of Slope Estimation Bias Across All Factor-Level Combinations

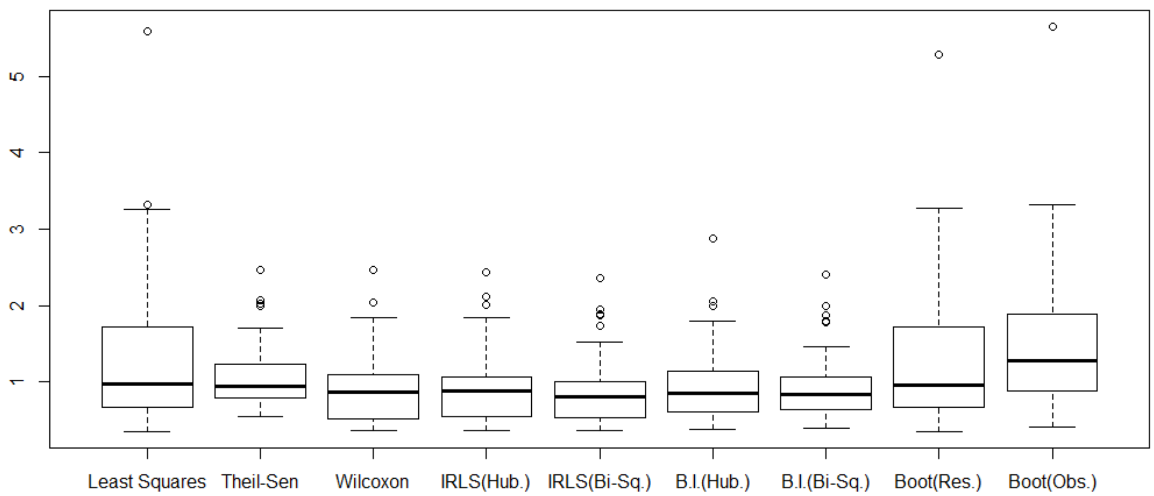


Figure 69 suggests that the slope mean bias distributions are all centered near zero, and roughly symmetric. Examination of inter-quartile ranges should provide additional insight into the spread of slope estimators. Larger variation among for specific approach estimates across the simulation conditions would suggest less general robustness for the approach.

Figure 70 shows that most IQR medians are centered near one, with perhaps a significantly larger median for the Bootstrap (Observations) approach. However, least squares and both Bootstrap approaches also have highly right-skewed IQR distributions, which is consistent with previous observations that these approaches produce non-robust parameter estimates. The IQR distributions for all other approaches look relatively similar and are smaller in comparison to Least-Squares and Bootstraps overall.

Section 5: Conclusions

The most robust parameter estimates for a simple linear regression model come from the Theil-Sen, IRLS (Bi-Square) and Bounded Influence (Bi-Square) approaches. The Bootstrap (Residual) and Least-Squares approaches are associated with the least robust parameter estimates in the presence of outliers.

In terms of robustness, among the iterative approaches (IRLS and Bounded Influence), there appears to be more separation between the selected weighting functions than between the use of scaled residuals (IRLS) versus the use of scaled deleted residuals (Bounded Influence). The Bi-Square weighting function appears to produce significantly more robust results than when using the Huber weighting function. Investigation into the reasons for this occurrence would be of further interest; however, this result suggests that a weighting function that allows for an extreme observation to essentially be thrown out of the analysis (i.e, receive a weight of zero) would likely be preferred over a weighting function that always retains observations even with very small weight attached. Further evaluation of other weighting functions might be of some value.

While bootstrapping approaches are distribution-free, this work demonstrates that they do not necessarily produce robust estimators. Especially, if the estimator used for each re-sample is non-robust itself. This is why throughout this work, the Bootstrap approaches appeared to behave essentially in concert with the common, known to be non-robust least squares estimators. Another avenue for further study might include the use of bootstrapping approaches with the utilization of one of the more robust estimation approaches considered here being used on each re-sample rather than least squares.

In addition to providing non-robust parameter estimates, least squares and both bootstrap methods generally appear to provide larger parameter confidence intervals than the other methods in the presence of outliers. This appears primarily due to the use of a non-robust estimator of the error distribution variance. Determination and utilization of such an estimator was beyond the scope of this work, but this observation again suggests a path for further study.

The seriously poor lack of coverage observed for the Theil-Sen slope parameter confidence interval estimators suggests yet another area for potential future study. Actually, this poor coverage was observed for all the interval estimation approaches that used the same basic approach (i.e, the Theil-Sen and Wilcoxon intercept intervals). This suggests that this relatively common approach for finding a confidence interval for a population median does not generally produce valid intervals. This indicates that perhaps other approaches to this problem might be of value.

The simulation design essentially utilized a contaminated error distribution approach to the introduction of potential outliers into the problem. In addition, this contaminating distribution was only distinguished by its larger variance. Other approaches to introducing potential outliers in simulation work might include a contaminating distribution that was essentially a point mass at an extreme value (i.e, instead of zero mean with large variance as was considered here, a contaminating distribution with a large mean and zero variance), or might even include specifically non-normal distributions (e.g., double-exponential, Cauchy, etc.). Again, this might provide an area for further study in the future.

Finally, the performance of the robust estimators on actual data sets again underlined the fact that regardless of the estimation approach utilized, whether robust to extreme observations or not, if the statistical model being applied has no validity, then results will generally fail to adequately summarize the true information carried by the data. Essentially, the choice of estimation approach is only important if the chosen statistical model is sufficiently useful in the context of the problem being considered. Recognition of the over-riding importance of understanding the context surrounding the data opposite virtually any other specific estimation or analysis approach in any statistical evaluation is perhaps the most important realization for any statistician. Without appropriate understanding of the context surrounding a problem, it is difficult for any statistician to provide a meaningful analysis of any value to those involved.

APPENDIX: TABLES OF PARAMETER ESTIMATES

Table A1: Comparison of Intercept Estimates for Original Data and with Response Outlier at X =10

Regression Method	Intercept New Data (with outlier)			Original Data		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	-1.8892	4.1475	10.1842	1.5336	2.5954	3.6573
Theil-Sen	2.4244	2.9845	3.7321	2.4042	2.8677	3.7045
Wilcoxon	2.4167	2.9739	3.7215	2.3929	2.8569	3.6891
IRLS (Huber)	1.09	2.9639	4.8378	1.7778	2.8051	3.8323
IRLS (Bi-Square)	.9006	2.8078	4.715	1.7388	2.7886	3.8385
Bounded Influence (Huber)	1.0181	2.9679	4.9177	1.7883	2.8142	3.8402
Bounded Influence (Bi-Square)	0.811	2.8023	4.7937	1.7413	2.7931	3.8449
Bootstrap (Residuals)	0.0937	4.1802	10.8826	1.6155	2.5836	3.4946
Bootstrap (Observations)	1.6789	4.2255	9.5490	1.2506	2.5922	3.694

Table A2: Comparison of Slope Estimates for Original Data and with Response Outlier at X = 10

Regression Method	New Data (with outlier)			Original Data		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	0.6756	1.1795	1.6834	1.1072	1.1996	1.2921
Theil-Sen	1.1237	1.1717	1.2265	1.1257	1.1735	1.2209
Wilcoxon	0.9969	1.1724	1.3479	1.0062	1.1746	1.3429
IRLS (Huber)	1.0289	1.1818	1.3347	1.1013	1.1859	1.2706
IRLS (Bi-Square)	1.0254	1.1857	1.3383	1.0991	1.1856	1.2728
Bounded Influence (Huber)	1.0229	1.1821	1.3413	1.1005	1.185	1.2696
Bounded Influence (Bi-Square)	1.0225	1.186	1.3495	1.098	1.1852	1.2721
Bootstrap (Residuals)	0.7213	1.1781	1.6558	1.1218	1.2003	1.2772
Bootstrap (Observations)	0.918	1.178	1.3595	1.1145	1.2004	1.3026

Table A3: Comparison of Intercept Estimates for Original Data and with Response Outlier at $X = 3$

Regression Method	New Data (with outlier)			Original Data		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	1.2463	7.1105	12.9747	1.5336	2.5954	3.6573
Theil-Sen	2.9282	2.9845	3.9166	2.4042	2.8677	3.7045
Wilcoxon	3.0697	3.7127	3.9786	2.3929	2.8569	3.6891
IRLS (Huber)	1.4726	3.3481	5.2235	1.7778	2.8051	3.8323
IRLS (Bi-Square)	0.8934	2.7615	4.6296	1.7388	2.7886	3.8385
Bounded Influence (Huber)	1.4615	3.3844	5.3073	1.7883	2.8142	3.8402
Bounded Influence (Bi-Square)	.8333	2.7528	4.6723	1.7413	2.7931	3.8449
Bootstrap (Residuals)	2.9354	7.1716	13.5611	1.6155	2.5836	3.4946
Bootstrap (Observations)	1.4396	7.1165	16.7262	1.2506	2.5922	3.694

Table A4: Comparison of Slope Estimates for Original Data and with Response Outlier at X = 3

Regression Method	New Data (with outlier)			Original Data		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	0.4078	0.8973	1.3868	1.111	1.1996	1.2883
Theil-Sen	1.0952	1.1717	1.2106	1.1257	1.1735	1.2209
Wilcoxon	0.9492	1.1231	1.2971	1.0062	1.1746	1.3429
IRLS (Huber)	0.9968	1.1474	1.298	1.1013	1.1859	1.2705
IRLS (Bi-Square)	1.0368	1.1876	1.3384	1.0988	1.1856	1.2725
Bounded Influence (Huber)	0.9896	1.1443	1.299	1.1005	1.185	1.2696
Bounded Influence (Bi-Square)	1.0332	1.1882	1.3432	1.0982	1.1853	.2723
Bootstrap (Residuals)	0.4388	0.8945	1.3497	1.1218	1.2003	1.2772
Bootstrap (Observations)	0.2272	0.8930	1.2794	1.1145	1.2004	1.3026

Table A5: Comparison of Intercept Estimates for Original Data with Predictor Outlier (X = 10 -> 30)

Regression Method	New Data (with outlier)			Original Data		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	2.3301	6.4389	10.5478	1.5336	2.5954	3.6573
Theil-Sen	2.8547	3.4049	3.8840	2.4042	2.8677	3.7045
Wilcoxon	3.0547	3.6687	4.0024	2.3929	2.8569	3.6891
IRLS (Huber)	1.8898	3.4707	5.0515	1.7778	2.8051	3.8323
IRLS (Bi-Square)	1.1832	2.7715	4.3599	1.7388	2.7886	3.8385
Bounded Influence (Huber)	1.8608	3.4757	5.0906	1.7883	2.8142	3.8402
Bounded Influence (Bi-Square)	1.1410	2.7739	4.4069	1.7413	2.7931	3.8449
Bootstrap (Residuals)	2.4713	6.4293	9.7535	1.6155	2.5836	3.4946
Bootstrap (Observations)	1.6341	5.8317	10.9802	1.2506	2.5922	3.694

Table A6: Comparison of Slope Estimates for Original Data with Predictor Outlier (X = 10 -> 30)

Regression Method	New Data (with outlier)			Original Data		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	0.4578	0.7611	1.0644	1.1072	1.1996	1.2921
Theil-Sen	1.0987	1.1325	1.1834	1.1257	1.1735	1.2209
Wilcoxon	0.9521	1.1144	1.2766	1.0062	1.1746	1.3429
IRLS (Huber)	0.993	1.1217	1.2504	1.1013	1.1859	1.2706
IRLS (Bi-Square)	1.0577	1.1882	1.3187	1.0991	1.1856	1.2728
Bounded Influence (Huber)	0.9887	1.1203	1.2519	1.1005	1.1850	1.2696
Bounded Influence (Bi-Square)	1.0538	1.1880	1.3222	1.098	1.1852	1.2721
Bootstrap (Residuals)	0.4649	0.7627	1.0275	1.1218	1.2003	1.2772
Bootstrap (Observations)	0.3528	0.8357	1.2721	1.1145	1.2004	1.3026

Table A7: Comparison of Intercept Estimates for Original Data and with Consistent Trend Outlier

Regression Method	New Data (with outlier)			Original Data		
	~95% C.I. Lower Bound	**Point Estimate**	**~95% C.I. Upper Bound**	**~95% C.I. Lower Bound**	**Point Estimate**	**~95% C.I. Upper Bound**
Least-Squares	1.6027	2.5296	3.4564	1.5336	2.5954	3.6573
Theil-Sen	2.0185	2.4094	3.3673	2.4042	2.8677	3.7045
Wilcoxon	2.2513	2.6970	3.4631	2.3929	2.8569	3.6891
IRLS (Huber)	1.8502	2.7323	3.6145	1.7778	2.8051	3.8323
IRLS (Bi-Square)	1.8023	2.7019	3.6016	1.7388	2.7886	3.8385
Bounded Influence (Huber)	1.8523	2.7344	3.6164	1.7883	2.8142	3.8402
Bounded Influence (Bi-Square)	1.8003	2.7022	3.6042	1.7413	2.7931	3.8449
Bootstrap (Residuals)	1.7269	2.534	3.3192	1.6155	2.5836	3.4946
Bootstrap (Observations)	1.2728	2.5043	3.4687	1.2506	2.5922	3.694

Table A8: Comparison of Slope Estimates for Original Data and Consistent Trend Outlier

Regression Method	New Data (with outlier)			Original Data		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	1.1396	1.2110	1.2824	1.111	1.1996	1.2883
Theil-Sen	1.1635	1.2105	1.2271	1.1257	1.1735	1.2209
Wilcoxon	1.0762	1.1968	1.3173	1.0062	1.1746	1.3429
IRLS (Huber)	1.1325	1.1994	1.2663	1.1013	1.1859	1.2705
IRLS (Bi-Square)	1.1321	1.2001	1.2681	1.0988	1.1856	1.2725
Bounded Influence (Huber)	1.1323	1.1992	1.2662	1.1005	1.185	1.2696
Bounded Influence (Bi-Square)	1.1318	1.2000	1.2682	1.0982	1.1853	1.2723
Bootstrap (Residuals)	1.1452	1.2108	1.2723	1.1218	1.2003	1.2772
Bootstrap (Observations)	1.1488	1.2135	1.3009	1.1145	1.2004	1.3026

Table A9: Comparison of Intercept Estimates for Original Data and Inconsistent Trend Outlier

Regression Method	New Data (with outlier)			Original Data		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	3.1821	12.9380	22.6939	1.5336	2.5954	3.6573
Theil-Sen	2.4936	2.9780	3.7488	2.4042	2.8677	3.7045
Wilcoxon	2.9898	3.5831	3.9394	2.3929	2.8569	3.6891
IRLS (Huber)	1.1857	3.3924	5.5990	1.7778	2.8051	3.8323
IRLS (Bi-Square)	0.4401	2.6071	4.7741	1.7388	2.7886	3.8385
Bounded Influence (Huber)	1.0471	3.3924	5.7378	1.7883	2.8142	3.8402
Bounded Influence (Bi-Square)	0.281	2.6162	4.9514	1.7413	2.7931	3.8449
Bootstrap (Residuals)	3.2370	12.889	20.9735	1.6155	2.5836	3.4946
Bootstrap (Observations)	1.4622	11.1033	24.6404	1.2506	2.5922	3.694

Table A10: Comparison of Slope Estimates for Original Data and Inconsistent Trend Outlier

Regression Method	New Data (with outlier)			Original Data		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	-0.7651	-0.0135	0.7381	1.111	1.1996	1.2883
Theil-Sen	1.1216	1.1654	1.2158	1.1257	1.1735	1.2209
Wilcoxon	0.919	1.1202	1.3216	1.0062	1.1746	1.3429
IRLS (Huber)	0.9401	1.1319	1.3237	1.1013	1.1859	1.2705
IRLS (Bi-Square)	1.0208	1.2086	1.3965	1.0988	1.1856	1.2726
Bounded Influence (Huber)	0.9252	1.1295	1.3339	1.1005	1.185	1.2696
Bounded Influence (Bi-Square)	1.0053	1.2077	1.4101	1.0982	1.1853	1.2723
Bootstrap (Residuals)	-0.7855	-0.0073	0.647	1.1218	1.2003	1.2772
Bootstrap (Observations)	-1.0450	0.2227	1.2920	1.1145	1.2004	1.3026

Table A11: Comparison of Parameter Estimates for Math Proficiency Data

Regression Method	Intercept			Slope		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	108.9773	143.8897	178.8021	1.0588	1.5066	1.9544
Theil-Sen	92.2	98.05	103.7	2.0000	2.1000	2.2857
Wilcoxon	118.2500	126.25	131.7500	0.0211	1.75	3.4789
IRLS (Huber)	88.4227	122.4037	156.3848	1.3591	1.7922	2.2252
IRLS (Bi-Square)	25.4747	71.0490	116.6234	1.8643	2.4446	3.0249
Bounded Influence (Huber)	76.7239	112.6045	148.4851	1.4590	1.9158	2.3725
Bounded Influence (Bi-Square)	29.4068	73.2230	117.0393	1.8596	2.4171	2.9747
Bootstrap (Residuals)	100.7496	144.3326	172.6881	1.1251	1.5011	2.0557
Bootstrap (Observations)	58.0351	129.7882	177.7460	1.0591	1.6850	2.6026

Table A12: Comparison of Parameter Estimates for Math Proficiency Data (Outliers Removed)

Regression Method	Intercept			Slope		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	38.45	67.8086	97.1673	2.114	2.4883	2.8626
Theil-Sen	73.4545	77.2727	82.6364	2.2222	2.3636	2.5
Wilcoxon	64.5552	69.6464	75.192	1.7173	2.4544	3.1916
IRLS (Huber)	37.5456	65.2479	92.9502	2.1699	2.523	2.8762
IRLS (Bi-Square)	37.0326	66.1999	95.3672	2.1389	2.5104	2.882
Bounded Influence (Huber)	41.0737	70.7056	100.3376	2.0767	2.4538	2.8309
Bounded Influence (Bi-Square)	38.7636	69.0862	99.4088	2.0878	2.4738	2.8598
Bootstrap (Residuals)	39.631	67.2577	95.4011	2.1383	2.4953	2.8455
Bootstrap (Observations)	43.1398	71.7348	111.961	1.9269	2.4389	2.8062

Table A13: Comparison of Parameter Estimates for Belgian Telephone Call Data

Regression Method	Intercept			Slope		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	-1657.2486	-982.9445	-308.6405	0.1599	0.5037	0.8474
Theil-Sen	-273.0600	-272.8650	-263.0600	0.1200	0.1400	0.1930
Wilcoxon	-284.5258	-284.2969	-274.5533	-0.1276	0.1459	0.4193
IRLS (Huber)	-709.4339	-412.0278	-114.6217	0.0594	0.2112	0.3631
IRLS (Bi-Square)	-536.1098	-213.0847	109.9403	-0.0555	0.1094	0.2743
Bounded Influence (Huber)	-858.5958	-494.8856	-131.1754	0.0680	0.2537	0.4393
Bounded Influence (Bi-Square)	-616.1551	-213.4747	189.2058	-0.0959	0.1096	0.3152
Bootstrap (Residuals)	-1585.0908	-976.2587	-362.4977	0.1878	0.5003	0.8103
Bootstrap (Observations)	-1808.6686	-1024.7791	-453.8008	0.2327	0.5250	0.9253

Table A14: Comparison of Parameter Estimates for Belgian Telephone Call Data (Outliers Removed)

Regression Method	Intercept			Slope		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	-312.9090	-253.3078	-193.7066	0.0996	.13	0.1604
Theil-Sen	-219.1650	-219.1350	-219.1150	0.1086	0.1125	0.1171
Wilcoxon	-220.7556	-220.7256	-220.7121	0.1004	0.1133	0.1262
IRLS (Huber)	-242.3479	-218.7498	-195.1518	0.1003	0.1123	0.1244
IRLS (Bi-Square)	-236.6968	-211.8418	-186.9867	0.0961	0.1088	0.1215
Bounded Influence (Huber)	-244.0503	-219.6389	-195.2274	0.1003	0.1128	0.1252
Bounded Influence (Bi-Square)	-237.4349	-211.6576	-185.8804	0.0955	0.1087	0.1218
Bootstrap (Residuals)	-310.4515	-253.4968	-202.8849	0.1041	0.1301	0.1591
Bootstrap (Observations)	-350.9175	-254.4091	-201.3121	0.1033	0.1306	0.1798

Table A15: Comparison of Parameter Estimates for University of South Florida Data

Regression Method	Intercept			Slope		
	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound	~95% C.I. Lower Bound	Point Estimate	~95% C.I. Upper Bound
Least-Squares	1481.688	14012.167	26542.646	-5701.003	-1782.834	2135.336
Theil-Sen	12573.198	13100.98	15334.1736	-4651.587	-1950.413	830.9278
Wilcoxon	13908.98	14319.768	16298.337	-9158.633	-2289.908	4578.818
IRLS (Huber)	3508.58	15263.213	27017.847	-6036.308	-2325.641	1385.026
IRLS (Bi-Square)	3001.728	15011.661	27021.594	-6001.308	-2219.633	1562.043
Bounded Influence (Huber)	3329.72	15153.44	26977.168	-6066.322	-2338.232	1389.858
Bounded Influence (Bi-Square)	3339.525	15484.31	27629.104	-6279.723	-2445.667	1388.391
Bootstrap (Residuals)	4002.125	14076.96	24990.883	-5133.964	-1801.19	1559.31
Bootstrap (Observations)	-2430.374	13721.212	25070.967	-5727.488	-1761.526	3044.837

Table A16: Comparison of Parameter Estimates for University of South Florida Data (Outlier Removed)

Regression Method	Intercept			Slope		
	~95% C.I. Lower Bound	**Point Estimate**	**~95% C.I. Upper Bound**	**~95% C.I. Lower Bound**	**Point Estimate**	**~95% C.I. Upper Bound**
Least-Squares	8223.059	19680.334	31137.608	-7589.339	-3886.727	-184.114
Theil-Sen	17415.27	18511.7	20022.88	-5237.5	-3762.987	-66
Wilcoxon	17631.049	18839.995	20396.1272	-7812.879	-3853.802	105.2745
IRLS (Huber)	8624.103	19026.342	29428.5802	-7141.278	-3784.494	-427.7104
IRLS (Bi-Square)	8486.721	19123.688	29760.6549	-7228.62	-3799.833	-371.0461
Bounded Influence (Huber)	8504.432	19014.634	29524.8362	-7185.263	-3794.341	-403.4193
Bounded Influence (Bi-Square)	8049.934	18898.74	29747.5517	-7227.536	-3738.6	-249.6639
Bootstrap Residuals)	10732.212	19696.917	30051.027	-7195.936	-3888.706	-961.1796
Bootstrap (Observations)	8726.948	19275.748	27618.4585	-6487.635	-3793.04	-640.9637

BIBLIOGRAPHY

1. "Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors." *Journal of Statistics Education, V9N3: Stanton,* ww2.amstat.org/publications/jse/v9n3/stanton.html. Accessed 14 July 2017.

2. "The Comprehensive R Archive Network." *The Comprehensive R Archive Network*, cran.r-project.org/. Accessed 14 July 2017.

3. Dang, Xin, et al. "Theil-Sen Estimators in a Multiple Linear Regression Model." *Olemiss.edu,* The University of Mississippi, home.olemiss.edu/~xdang/papers/MTSE.pdf. Accessed 13 July 2017.

4. Hogg, Robert V., et al. *Introduction to mathematical statistics*. Upper Saddle River, NJ, Pearson Education, 2013.

5. Kutner, Michael H., et al. *Applied linear regression models*. 4th ed., Boston, McGraw-Hill/Irwin, 2004.

6. Hollander, Myles, et al. *Nonparametric statistical methods*. 3rd ed., Hoboken, N.J, John Wiley and Sons Ltd, 2014.

7. Efron, B., and R. J. Tibshirani. *An introduction to the bootstrap*. Chapman & Hall/CRC, Boca Raton, Florida, 2000.

8. "Lesson 47: Distribution-Free Confidence Intervals for Percentiles." *Lesson 47: Distribution-Free Confidence Intervals for Percentiles | STAT 414 / 415*, Penn State Eberly College of Science, onlinecourses.science.psu.edu/stat414/node/231. Accessed 13 July 2017.

9. George, Karen J., et al. "A Comparison of Confidence Intervals from R-Estimators in Regression." *www.smu.edu,* Southern Methodist University, www.smu.edu/-/media/Site/Dedman/Departments/Statistics/TechReports/TR277.ashx?la=en. Accessed 13 July 2017.

10. "The Family: America's Smallest School." *Ets.org*, Educational Testing Services, www.ets.org/Media/Education_Topics/pdf/5678_PERCReport_School.pdf. Accessed 13 July 2017.

11. Faculty and A&P Salary Report, University of South Florida, Resource Analysis and Planning, 1990.

12. Administrative Compensation Survey, *Chronicle of Higher Education*, Jan. 1991.

13. Box, George and Norman Draper. Empirical Model Building and Response Surfaces. John Wiley, 1987, p. 74.

VITA

Gregory Frank Malek graduated from Ruston High School, Ruston, LA in May 2010. He enrolled at the University of Houston beginning in August 2010, pursuing a Bachelor of Science degree in Mathematics. He graduated Cum Laude in December 2014 with a Bachelor of Science in Mathematics and a minor in Economics. In August 2015, he enrolled at Stephen F. Austin State University, Nacogdoches, Texas in pursuit of a Master of Science degree in Statistics. He graduated from Stephen F. Austin State University in August 2017, having taught seven remedial algebra classes and assisted with three Calculus labs during his enrollment.

Permanent Address: 211 Bear Creek Rd.
Ruston, LA 71270

Modern Language Association Handbook, 8th Edition

This thesis was written by Gregory Frank Malek.