Distribution Agreement

New Thoughts on Using Elemental and Non-Elemental Sets to Produce Robust Estimates of Regression Coefficients

By

Chengyin Jin

Master of Science
Computer Science

Advisor: Dr. Robert H. Lyles

Approved for the Department by:

____
Dr. Robert H. Lyles
Advisor

____
Dr. Michael H. Kutner
Reader

____
Dr. James Lu
Reader

Accepted:

____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

____
Date

New Thoughts on Using Elemental and Non-Elemental Sets to Produce Robust Estimates of Regression Coefficients

By

Chengyin Jin

B.S., Fudan University

China, 2010

Advisor

Robert H. Lyles, Ph. D.

An Abstract of
A thesis submitted to the Faculty of the
James T. Laney Graduate School Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Computer Science

2012

Abstract

New Thoughts on Using Elemental and Non-Elemental Sets to Produce Robust Estimates of Regression Coefficients

By Chengyin Jin

The idea of utilizing elemental sets in linear regression was proposed in 1755, but was not widely accepted due to computational burdens. Renewed recent interest in this topic was inspired by the appeal of robust regression, along with the development of modern computational environments. In this paper, different weighting factors are developed and applied for elemental regression. These weights combine information from influence statistics together with variance information associated with elemental sets, to down-weight subsets with outliers with a view toward maintaining efficiency for estimating regression coefficients. A new approach, called the "Drop K" method, is proposed and assessed in simulation studies. Instead of selecting all unique elemental subsets containing the minimal number of observations sufficient to fit the desired model, k observations are dropped from the original dataset to form each subset, where k is the number of suspected outliers. Estimators were again calculated under different weights, including those based on influential statistics and the variance of estimators from each set. The performance of this method is compared with that of elemental regression, as well as with a popular robust regression technique (the Huber estimator) in simulation studies. The Drop K approach performed better than the least square estimators, and appears to provide an appealing alternative to elemental regression estimators.

New Thoughts on Using Elemental and Non-Elemental Sets to Produce Robust Estimates of Regression Coefficients

By

Chengyin Jin

B.S., Fudan University

China, 2010

Advisor

Robert H. Lyles, Ph. D.

A thesis submitted to the Faculty of the
James T. Laney Graduate School Studies of Emory University
In partial fulfillment of the requirements for the degree of
Master of Science
in Computer Science

2012

Acknowledgements:

First, I owe my deepest gratitude to my advisor, Prof. Robert H. Lyles. This thesis would not have been possible without his knowledge, generous encouragement and kind mentorship. His scientist intuition inspired and enriched my growth as a graduate student.

I would like to express my sincere thanks to Prof. James Lu, who guided me through the master program and provided his advice regarding the computer science part of the thesis.

It is my honor to have Prof. Michael Kutner as my thesis reader and I am grateful for his valuable advice.

I would also acknowledge Mrs. Heather Boldt for her constructive comments.

My special thanks to Dr Haowei Wang, who guided me with programming in Matlab and gave advice for algorithm improvement and debugging.

Many thanks go to my friends, Liyue Fan, Xia Lin, Baiyu Yang, Yize Zhao, Xiaojing Wang and Tian Dai for their kindness and support.

Finally, I would like to thank my parent for supporting me throughout my education, my uncle Tong Zhu for encouraging me to write a thesis, and my boyfriend Haowei, who companied and supported me through the whole process.

Table of Contents

List of Figures

List of Tables:

Chapter 1: Introduction

1.1 Multiple Linear Regression and Outliers

The standard multiple linear regression model is:

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \qquad (1.1)$$

where **Y** denotes an n×1 vector of random observations, **X** is an n×p design matrix of known predictors, **β** is a p×1 vector of unknown parameters, and **ε** is an n×1 vector of independent random errors with $E\{\boldsymbol{\varepsilon}\}=0$ and $Var\{\boldsymbol{\varepsilon}\}=\sigma^2\mathbf{I}$. The parameter vector **β** is typically estimated by Ordinary Least Squares (OLS), which minimizes the sum of squared residuals: $SSE(\boldsymbol{\beta})=(\mathbf{Y}-\mathbf{X}\ \boldsymbol{\beta})^t(\mathbf{Y}-\mathbf{X}\ \boldsymbol{\beta})=\boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon}=\sum_{i=1}^{n}\varepsilon_i^2$. The resulting OLS estimator is given by $\hat{\boldsymbol{\beta}}=(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$ assuming that the regression model is of full rank; otherwise a generalized inverse is required and the estimator is not unique.

OLS is very efficient under the usual assumptions above for linear regression; however, it is highly non-robust to outliers in the model. Outliers are residuals $Y_i-\hat{Y}_i=\varepsilon_i$, whose values are well separated from the reminder of the residuals. Outliers come from diverse sources, such data incompatible with model assumptions. When fitting linear regression models, outliers may dramatically affect the estimated regression coefficients obtained via the OLS solution. Figure 1.1 shows an example of how the fitted OLS line can be affected by an outlier.

A large number of methods have been developed for outlier detection. A traditional approach is the examination of standardized residuals and studentized residuals [1] .

Influential statistics, such as Cook's distance [2], Dffits [1] and Dfbetas [1], have also been developed to identify the outliers.

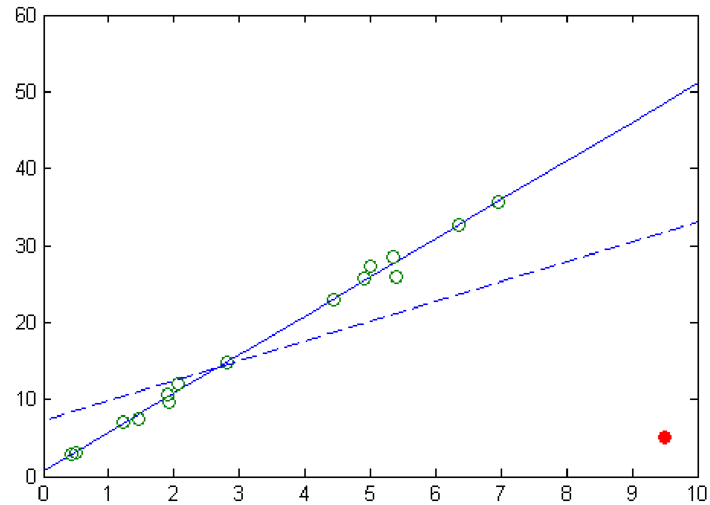

Figure 1.1: Fitted OLS lines with and without outlier. The Solid dot denotes the outlier.

Solid line: regression line based on observations without outlier;

Dashed line: regression line based on observations with outlier.

What is the best approach to deal with observation outliers? In some cases, outliers contain valuable information for fitted models. Simple deletion of outlying observations is never recommended except when they are due to verified recording errors and/or reflect implausible values. Robust regression methods have been proposed as alternatives to deal with outliers, and elemental regression has been identified as a potential facilitator for certain robust regression alternatives.

1.2 Robust Regression

Instead of objectively removing outliers, robust regression techniques consist of algorithms which provide more robust estimators than OLS. Statisticians have developed numerous robust regression methods, leading to estimators that can be roughly classified as M-estimators, R-estimators and L-estimators [3]. The M-estimators derive their name from "generalized maximum likelihood", which was proposed by Huber in 1964 [4]. M-estimators tend to perform well for outliers in the outcome variable. The R-estimators are derived from "rank tests" and were proposed by Jaeckel in 1972 [5, 6]. The L-estimators derive their name from the fact that they are "linear combinations of order statistics" [5]. Boscovich first suggested the Least Absolute Value (LAV) method in 1757 [7], and later Rousseeuw proposed Least Median Squares (LMS) [8] and Least Trimmed Squares procedures [9] in 1984.

As computing power increased, iteratively reweighted least squares (IRLS) became popular as an operational method for calculating robust regression estimators. In IRLS robust regression, weights are assigned to observations according to how outlying they are. These weights are revised until convergence is obtained for the robust fit, with the fitting process at each step resembling a weighted least squares strategy. Many weight functions have been developed to reduce the effect of outliers. Among the widely used approaches are those based on the so-called Huber and Bisquare weights [10]. The Huber weight function is defined as:

$$w = \begin{cases} 1 & |u| \le 1.345 \\ \dfrac{1.345}{|u|} & |u| > 1.345 \end{cases}$$

Here, w denotes the weight and u denotes the scaled residual:

$u_i = \dfrac{e_i}{MAD}$, where $MAD = \dfrac{1}{.6745}\ median\{|e_i - median\{e_i\}|\}$ denotes "median absolute difference" and where e_i is the ith sample residual [10].

1.3 Elemental Regression

The idea of elemental regression was proposed in 1775 by Boscovich [11], years before the least squares methodology became popular. However, the method was not widely accepted due to the computational burdens as n and p become larger. Modern computational efficiency makes elemental regression feasible, and has inspired renewed interest in this topic.

In regression model (1.1), the minimum number of observations required to obtain the OLS estimator is p. In elemental regression, p observations are taken from the dataset as a subset. The final estimator is calculated by combining the estimators from all possible unique subsets. Letting h={i_1, i_2, … i_p} denote a subset with p observations, the estimator for subset h (assuming that the h^{th} design matrix is full rank) is denoted:

$$\hat{\boldsymbol{\beta}}_h = (\mathbf{X}_h^t \mathbf{X}_h)^{-1} \mathbf{X}_h^t \mathbf{Y}_h \quad , \qquad (1.2)$$

where $\mathbf{X}_h$ is the p×p design matrix for subset h and $\mathbf{Y}_h$ is the p×1 subvector of $\mathbf{Y}$ [12].

As noted by several authors [12-14], the OLS estimator can be obtained as a weighted average of the estimators derived from each elemental subset:

$$\hat{\boldsymbol{\beta}}_{OLS} = \frac{\sum_h |\mathbf{X}_h^t \mathbf{X}_h| \hat{\boldsymbol{\beta}}_h}{\sum_h |\mathbf{X}_h^t \mathbf{X}_h|} = \sum_h \frac{|\mathbf{X}_h^t \mathbf{X}_h|}{|\mathbf{X}^t \mathbf{X}|} \hat{\boldsymbol{\beta}}_h = \sum_h w_h \hat{\boldsymbol{\beta}}_h \qquad (1.3)$$

When $\sum_h w_h = 1$, (1.3) is the same as the OLS estimator. For estimating the slope parameter in simple linear regression (SLR) with p=2, these determinant-based weights are equivalent to the following:

$$w_h = \frac{(x_i - x_j)^2}{\sum_{s \le t} (x_s - x_t)^2}, \qquad (1.4)$$

where i and j index the two observations contained in the h^th elemental set, x_i and x_j are the values of x for those two observations, and the summation in the denominator is over all possible elemental sets (see, e.g., [12]).

Importantly, we note that this is the same as the inverse of the variance of $\hat{\boldsymbol{\beta}}_h$ divided by the sum of the inverse variances across all elemental sets.

Hoerl and Kennard [15] similarly used determinant-based weights and proved that in fact for any value of m, p ≤ m ≤ n, the OLS estimator can be reproduced by a weighted average of all $\binom{n}{m}$ regression coefficients based on all possible unique observation subsets of size m in the dataset. In Wang's thesis [14], it is shown that this extension of the elemental set result for reproducing OLS based on subsets of various sizes can also be accomplished by variations on inverse-variance weighting. Further, Wang [14] shows that direct inverse-variance weighting produces estimators very close to the OLS result, indicating that using inversed variances as weights works to ensure efficiency when combining regression estimates based on elemental or non-elemental subsets. Mayo and Gray [12, 13] showed that many regression estimators can be expressed in terms of elemental regressions, like OLS, LMS, weighted least squares and IRLS. They named this class the leverage-residual weighted elemental (LRWE) estimator. Each member of the LRWE class can be expressed in the form:

$$\hat{\boldsymbol{\beta}}(\lambda,\rho)=\frac{\sum_h w[\lambda(h),\rho(h)]\hat{\boldsymbol{\beta}}_h}{\sum_h w[\lambda(h),\rho(h)]}, \qquad (1.5)\ [12]$$

where λ(h) is a weighting factor based on leverage information from elemental subset h, ρ(h) is a weighting factor based on the residual or degree of fit information. For example, the weighting factors for OLS are $\lambda(h)=|\mathbf{X}_h^t\mathbf{X}_h|$, $\rho(h)=1$. [12, 13].

Mayo and Gray also proposed an approach, the trimmed elemental estimator (TEE), which changes the contributions of one or both weighting factors. TEE is a special case of (1.5), by setting

$$\lambda(h)=|\mathbf{X}_h^t\mathbf{X}_h|,$$
$$\rho(h)=I\Big(\sum_{i=1}^{n}|e_{hi}|\le(1-\alpha_p)\ 100th\ \ percentile\ \ of\ \ the\ \binom{n}{p}\ \sum_{i=1}^{n}|e_{hi}|\ \ values\Big). \qquad [13]$$

where I($\cdot$) is the indicator factor, $\sum_{i=1}^{n} | e_{hi} |$ is the sum of absolute residuals and α_p is the trimming constant ($0 \le \alpha_p \le 1$). For further details, refer to Mayo and Gray [12, 13].

1.4 Outline

In this thesis, different weighting factors are developed and applied for elemental regression. A new approach, called the "Drop K" method, is also proposed and assessed in simulation studies. The performance of this method is compared with that of elemental regression, as well as the Huber estimator.

Chapter 2: Methods

2.1 Introduction

In the first part of this chapter, a series of new weights for potential use in elemental regression is introduced. These weights combine information from influence statistics together with variance information associated with elemental sets, to down-weight subsets with outliers with a view toward maintaining efficiency for estimating regression coefficients.

The second part of the methods chapter proposes a new idea for subset construction. Instead of selecting all unique elemental subsets containing the minimal number of observations sufficient to fit the desired model, k observations are dropped from the original dataset to form each subset, where k is the number of suspected outliers detected in the dataset. Estimators are again calculated under different weights, including those based on influential statistics and the variance of estimators from each set.

2.2 Elemental Regression

2.2.1 Notation

Assuming there are n observations and p unknown parameters, then the total number of elemental subsets is $e = \binom{n}{p}$. Let E_j, j=1,2, …, e, represent the j^{th} unique elemental set with p observations.

2.2.2 β Estimator

As previously suggested by Mayo and Gray[12, 13], a potentially robust estimator of a given regression coefficient ($\hat{\boldsymbol{\beta}}$) using the elemental regression method can be defined as:

$$\hat{\boldsymbol{\beta}} = \sum_{j=1}^{e} w_{Ej}\, \hat{\boldsymbol{\beta}}_{Ej} \qquad (2.1)$$

$$\text{where} \quad w_{Ej} = \frac{\lambda(E_j)\rho(E_j)}{\sum_{j=1}^{e} \lambda(E_j)\rho(E_j)} \qquad (2.2)$$

Here, $\hat{\boldsymbol{\beta}}_{Ej}$ is the regression coefficient for subset E_j, $\lambda(E_j)$ is a weighting factor based on the elemental subset-specific design matrix $\mathbf{X}_{Ej}$, and $\rho(E_j)$ is a weighting factor based on influential statistics associated with the observations comprising the jth elemental set.

2.2.3 Weighting Factors

Here, we will use $\lambda(E_j)=1/\text{var}(\hat{\beta}_{Ej})$, in keeping with the fact (see Chapter 1) that inverse-variance weighting of elemental sets is essentially fully efficient under the usual linear regression assumptions. Thus, letting $\rho(E_j)=1$, we obtain the inverse-variance weights as:

$$W_{j,INV} = \frac{1/\text{var}(\hat{\beta}_{Ej})}{\sum_{j=1}^{e} 1/\text{var}(\hat{\beta}_{Ej})} \qquad (2.3)$$

The estimator $\hat{\beta}$ using $W_{j,INV}$ essentially reproduces the same result as OLS, ensuring the efficiency of our weighting method. Note that because of the assumed homoscedastic errors, the calculation in (2.3) can be made without estimating the residual variance; that is, we use only the $(\mathbf{X}^t\mathbf{X})^{-1}$ portion of $Var(\hat{\boldsymbol{\beta}}_{Ej})$ when working with elemental sets.

To obtain robust alternative estimators, we propose that the $\rho(E_j)$ be defined as the product of inversed influential statistics for each observation in the elemental subset E_j, i.e., $\rho(E_j) = \prod_{t=1}^{p} 1/F(tj)$, where F(tj) is the value of an influential statistic for the t^{th} observation within the jth elemental set. Cook's distance, absolute value of Dffits (|Dffits|), $Dffits^2$[16] and absolute value of Dfbetas (|Dfbetas|) are each used separately as influential statistics in our current study to produce "influence" weights(W_{INF}), as follows:

$$W_{j,INF} = \frac{\prod_{t=1}^{p} 1/F(tj)}{\sum_{j=1}^{e}\prod_{t=1}^{p} 1/F(tj)} \qquad (2.4)$$

For outlying observations, the influential statistics F(tj) are much larger than for normal observations; hence, the inversed influential weight can effectively down-weight the estimated coefficients from subsets that contain outliers.

There are many possible combinations for the final weight w_{Ej} based on $W_{j,INV}$ and $W_{j,INF}$. The following weighting factors were tested:

(1) $W_{j,INVF}$, which combines $W_{j,INV}$ and $W_{j,INF}$:

$$W_{j,INVF} = \frac{W_{j,INV} W_{j,INF}}{\sum_{j=1}^{e} W_{j,INV} W_{j,INF}} \qquad (2.5)$$

Note that (2.5) is algebraically equivalent to (2.2), with $\lambda(E_j)=1/var(\hat{\beta}_{Ej})$, and $\rho(E_j) = \prod_{t=1}^{p} 1/F(tj)$.

(2) $W_{j,MINVF}$, which takes the minimum of $W_{j,INV}$ and $W_{j,INF,}$ and scales the resulting weights so that their sum equals 1:

$$W_{j,MINVF} = \frac{Min(W_{j,INV}, W_{j,INF})}{\sum_{j=1}^{e} Min(W_{j,INV}, W_{j,INF})} \qquad (2.6)$$

(3) $W_{j,PINVF}$. As discussed in Chapter 1, Mayo and Gray [12, 13] considered some desired percentiles when building the weight factors for TEE. Using a similar idea as in the TEE method, we consider alternative weights for each elemental subset denoted as $W_{j,PINVF}$ and scaled so that the resulting weights sum to 1. First, we define

$$w_{j,\alpha} = \begin{cases} Min(w_{j,INV}, w_{j,INF}), & if \;\; w_{j,INF} < \alpha \times 100th \;\; percentile \;\; of \\ & the \binom{n}{p} \; w_{j,INF} \; values \\ w_{j,INV}, & otherwise \end{cases} \qquad (2.7)$$

The resulting weight $W_{j,PINVF}$ then becomes

$$W_{j,PINVF} = \frac{w_{j,\alpha}}{\sum_{j=1}^{e} w_{j,\alpha}} \qquad (2.8)$$

The idea here is to use the efficient inverse-variance weighting for all elemental sets except those that merit low influence weights.

2.3 Drop K Method

2.3.1 Notation

Assume again that we have a total of n observations. The Drop K method consists of two steps. First, standard regression diagnostics based on the assumed linear regression model are applied to detect how many suspected outliers there are in the full data set. We record this number of outliers as k. Then, the $d = \binom{n}{n-k}$ unique subsets containing n-k of the observations are constructed. Define D_j, j=1, 2, …, d, as the j^{th} "Drop K" subset of the original dataset in which k observations are dropped.

The idea behind the proposed Drop K method is similar to that of elemental regression, except that instead of selecting a small number of observations to build a subset, a certain number of points are dropped and the rest are used as a subset. As noted in Chapter 1, appropriate inverse-variance weighting of all unique subsets of arbitrary size m (≥ p) reproduces the OLS estimator [14, 15]. Thus, as before, the efficient aspect of the weighting can be accomplished via inverse-variance weights.

2.3.2 Outlier Detection

For our purposes, studentized (jackknife) residuals are used for outlier detection. These are defined as follows:

$$r_{(-i)} = \frac{e_i}{\sqrt{S^2_{(-i)}(1-h_i)}}$$

Here, $S^2_{(-i)}$ is the mean square error from the original regression of **Y** on the **X**'s after dropping the i^{th} observation, and h_i represents the leverage value for the i^{th} observation.

Note that $r_{(-i)}$ follows a t distribution with n-p-1 degrees of freedom under the typical assumptions for linear regression [1]. In the current study, observation i was flagged as an outlier if the absolute value of $r_{(-i)}$ was greater than the Bonferroni critical value t(1-α/2n; n-p-1), where we take α=0.1 [10]. The comparison was two-sided because we are concerned with extreme absolute values of the studentized residuals.

2.3.3 β Estimator

The proposed estimator $\hat{\boldsymbol{\beta}}$ using the Drop K method can be expressed in the following form:

$$\hat{\boldsymbol{\beta}} = \sum_{j=1}^{d} w_{Dj} \hat{\boldsymbol{\beta}}_{Dj} \qquad (2.9)$$

$$\text{where } w_{Dj} = \frac{\lambda(D_j)\rho(D_j)}{\sum_{j=1}^{d} \lambda(D_j)\rho(D_j)} \qquad (2.10)$$

Here, $\hat{\boldsymbol{\beta}}_{Dj}$ is the estimator based on the j^{th} drop k subset D_j, and $\rho(D_j)$ is a weight factor based on the influential statistics associated with the k observations that were dropped to obtain the subset D_j. The remaining component $\lambda(D_j)$ is the weighting factor incorporating leverage information based on the jth subset-specific design matrix ($\mathbf{X}_{Dj}$) associated with subset D_j. Note the similarity of (2.9) and (2.10) to the elemental set-based versions in (2.1) and (2.2), except that the leverage portion of the weight is now based on influence statistics corresponding to the observations that are not included in the jth set.

2.3.4 Weighting Factors

Here, we take $\lambda(D_j)$ to be the inversed variance of the $\boldsymbol{\beta}$ estimator from the j^{th} drop k set, i.e., $\lambda(D_j)=1/\text{var}(\hat{\beta}_{Dj})$. As we indicated in 2.2.3, $\lambda(D_j)$ acts to preserve efficiency. We define the inverse-variance weights as

$$W_{j,INV} = \frac{1\Big/\text{var}(\hat{\beta}_{Ej})}{\sum_{j=1}^{d} 1\Big/\text{var}(\hat{\beta}_{Ej})} \qquad (2.11)$$

Next, we take $\rho(D_j) = \prod_{t=1}^{k} F(tj)$ to be the influence weighting factor, where F(tj) is the influential statistic for the t^{th} observation dropped for the jth subset and in the current study F can represent Cook's distance, |Dffits|, Dffits^2 or |Dfbetas|. The "influence" weight for the drop k approach is then as follows:

$$W_{j,INF} = \frac{\prod_{t=1}^{k} F(tj)}{\sum_{j=1}^{d} \prod_{t=1}^{k} F(tj)} \qquad (2.12)$$

Note that (2.11) is the same as (2.3) for elemental regression; however, (2.12) is different from (2.4) in that the influential statistics are placed in the numerator instead of the denominator to blunt the effect of outliers. This is because the influential statistics are based on the observations deleted. Consider the extreme case in which all k outliers happen to be dropped for subset D_a, $0 \leq a \leq d$. Since the influential statistics for the k dropped outliers will be larger than those for other normal observations, $\hat{\boldsymbol{\beta}}_{Da}$ gains the largest influence weight based on $\rho(D_a)$.

Therefore for the Drop K method, $W_{j,INVF}$ is expressed by:

$$W_{j,INVF} = \frac{W_{j,INV} W_{j,INF}}{\sum_{j=1}^{d} W_{j,INV} W_{j,INF}} \qquad (2.13)$$

Note that (2.13) is algebraically equivalent to (2.10), with $\lambda(D_j)=1/\text{var}(\hat{\beta}_{Dj})$, and $\rho(D_j) = \prod_{t=1}^{k} F(tj)$. We focus primarily on (2.13) to obtain a robust estimator based on the Drop K approach, as we find in what follows that fewer adjustments to this natural incorporation of influence and inverse-variance weights are needed than in the elemental regression setting.

Chapter 3: Simulations

Simulation studies were conducted in a simple linear regression (SLR) situation to evaluate the proposed weighted estimators using elemental regression (with p=2), and using the Drop K method.

For elemental regression, W_{INVF}, W_{MINVF}, W_{PINVF} (α=0.05), W_{PINVF} (α=0.25), W_{PINVF} (α=0.50) and W_{PINVF} (α=0.75) were compared against regular OLS regarding robustness and relative performance. Weights with different influential statistics, including Cook's distance, |Dffits|, $Dffits^2$ and |Dfbetas| were tested for each form of weight factor.

For the Drop K method, only W_{INVF} (2.13) was used as the weight factor, with Cook's distance, |Dffits|, $Dffits^2$ and |Dfbetas| utilized to produce the influence weights (F), respectively.

Each simulation used 1000 iterations. For each regression situation, the sample size was set as n=25, 50, or 100. The predictor variable X was obtained by random number generation from the normal distribution N(3,3), and sorted in ascending order. The outcome variable Y_i was generated by:

$$Y_i=\beta_0+\beta_1 X_i+\varepsilon_i,\ i=1, \ldots, n \qquad (3.1)$$

where β_0 was set as 0 and $\beta_1$as 1. The error ε_i followed the normal distribution N(0,0.15) or N(0,1) for each situation.

First, two artificial outliers were created by setting Y_{n-1}=-10 and Y_n=-10. Under this setting, 2 outliers were expected for most cases. Around 3% of the data sets with n=25 and errors distributed as N(0,0.15) had less than 2 outliers identified by means of studentized residuals. About 10% of the data sets with n=25 and ε following N(0,0.1) had less than 2 outliers detected. The program was automatically adjusted for each simulated dataset according to the number of outliers detected for the Drop K method. Separately, datasets without any imposed outliers were used to assess the performance of our weighting factors under the usual SLR assumptions. For the Drop K approach, the regression diagnostic step still detected outliers for approximately 10% of the simulated datasets in this case, and the corresponding Drop K procedure was applied to each such dataset. For datasets with no outliers detected, the estimator used for the Drop K approach was the OLS estimator from the complete data.

As part of the comparison of regression coefficient estimators, we also included regular OLS after deleting the identified outliers in the dataset in addition to IRLS robust regression using the Huber weight function [10] for comparison.

Chapter 4: Results

4.1 Simulation Results

Table 4.1.1 and Table 4.1.2 summarize the β estimators from 1000 iterations with sample size n=25 and independently and identically distributed residual errors generated as N(0,0.15) and N(0,1). Tables 4.1.3-4.1.6 summarize similar simulations as in Tables 4.1.1 and 4.1.2, except that for Table 4.1.3 and Table 4.1.4 we used sample size n=50 and for Table 4.1.5 and Table 4.1.6 we used sample size n=100.

For the Drop K method, both Table 4.1.1 and Table 4.1.2 indicate that the proposed estimators yielded values much closer to the true value than did OLS. Among the four influential statistics used, weighting factors calculated based on Cook's distance and $Dffits^2$ performed slightly better than those utilizing |Dffits| and |Dfbetas|. The results with these two weighting factors in the Drop K setting were competitive in terms of bias with those based on the popular Huber weighting method. The situation regarding the Drop K approach was essentially the same for Tables 4.1.3 through 4.1.6. Comparing the results in Table 4.1.1, Table 4.1.3 and Table 4.1.5 together, the estimators got closer on average to the true value as the sample size n increased. The same conclusion held for Tables 4.1.2, 4.1.4 and 4.1.6.

For elemental regression Tables 4.1.1 and 4.1.2 suggest that using W_{MINVF}, which incorporates the minimum of the inversed variance and inversed influential statistics for each elemental set, yielded the best estimator compared to other weighting factors. Within the W_{MINVF} group, $Dffits^2$ and Cook's distance again appeared to outperform the rest of the influential statistics used in constructing the 'INF' weights. In Table 4.1.3, estimators using W_{MINVF} still performed well, but in Table 4.1.4, estimators using W_{PINVF} (α=0.25) performed best overall. All W_{PINVF} estimators with different α's produced relatively close mean estimates to the true value in Table 4.1.5 and Table 4.1.6, but W_{PINVF} (α=0.5) arguably performed best in Table 4.1.5 and W_{PINVF} (α=0.05) performed best in Table 4.1.6. As the residual variance increased from 0.15 to 1, most estimators in Table 4.1.2, Table 4.1.4 and Table 4.1.6 became further away from the true value on average and more scattered than the corresponding ones in Table 4.1.1, Table 4.1.3 and Table 4.1.5. However, when fixing the error distribution, there was not always a clear trend indicating that the elemental set-based estimates more closely approximated the true values as the sample size grew. Also, it was sometimes difficult to tell from these 6 tables, which influential statistic performed the best for constructing the weighting factors.

Tables 4.1.7 to Table 4.1.12 show the results for datasets without outliers under the same collection of settings that were studied in Tables 4.1.1 through 4.1.6. The estimated regression coefficients in all tables based on the proposed weighted estimators were very close to the OLS estimates, confirming the efficiency of the proposed robust estimators when data are consistent with the usual SLR assumptions.

4.2 Artificial Data Example for SLR

Here we used the same artificial dataset introduced from Mayo and Gray's paper [12]. The dataset contained 25 observations, with 24 regular observations and 1 induced outlier. Note that the authors set up the example intentionally so that the OLS intercept and slope would be exactly 0 and 1, respectively, after removal of the outlier. The data are listed in Table 4.2.1, and Figure 4.2 provides a scatterplot for the dataset.

Table 4.2.2 listed the estimates from Mayo's paper (see [12] for the definitions of the estimators compared in that paper, including various versions of the 'trimmed elemental estimators' (TEE)). Table 4.2.3 shows the results for our proposed elemental regression and Drop K approaches. Based on our outlier detection criteria, there was 1 outlier in the dataset and our program automatically chose the Drop 1 method. The estimate using the Drop 1 Method with Dffits^2 in the weight function appeared to perform the best for our method, and produced results closer to the ideal estimates of 0 and 1 relative to all methods listed in Mayo's paper except the TEE (50% trim, SAE) and the TEE (50% trim, SAE with 25% trim). Also, note in Table 4.2.3 that for this example most of the weighting approaches proposed here gave results closer to the desired values (0 and 1) than OLS.

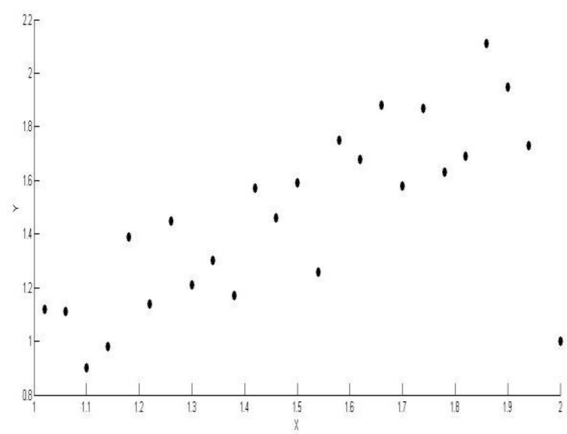

Figure 4.2: Scatter-plot of Artificial Data. The artificial sample consists of n=25 data points, 24 of which are from a simple linear regression model with a slope 1 and an intercept of 0. The 25th data point is an outlier located at (2,1) [12].

4.3 Data Set Example in MLR for Drop K Approach

While the elemental set-based approaches presented here are in theory applicable for any sample size (n) and number of predictors (p), the computational burden associated with identifying the potentially large number of elemental sets when n and p become large is an issue. However, for illustration we have extended the Drop K approach from SLR to multiple linear regression (MLR) with 2 predictor variables, and applied it to a previously-analyzed dataset. The dataset used here is Duncan's [17] occupational prestige data, with 45 observations and 4 variables. This dataset was also used by Fox [18], who detected two suspicious outliers. Thus, the Drop 2 approach was used here and the results are shown in Table 4.3.1. The estimators using the Drop 2 approach appear robust compared to the OLS estimators, and they yield values close to those obtained via the Huber estimators and the OLS estimators after discarding outliers.

4.4 Computing Time

The simulations were run on a Dell Vostro 3450 personal computer with 4 GB RAM running Matlab R2008a.

Considering the elemental regression and Drop K method in SLR, if the number of outliers detected for a given dataset was 2, then the number of subsets required for elemental regression was the same as the Drop 2 method. Figure 4.4.1 shows the computing time for elemental regression with p=2 compared with the Drop K method with k=2, using W_{INVF} ($Dffits^2$) as the weighting factor. As the sample size increased, the computing time for the Drop K method was higher than that for elemental regression. The longest time for the Drop K method with sample size n=200 was 35 seconds, compared with 27 seconds for elemental regression. Figure 4.4.2 provides the computing time for the Drop K method with k=1, 2 and 3. As k increased, the computing time grew dramatically. The time for sample size=125 was 0.13 seconds for k=1, 8.7 seconds for k=2 and 5,833 seconds for k=3. Analytically, the Drop K method has O(n choose k) complexity since for a given k, all possible subsets of size n-k need to be considered.

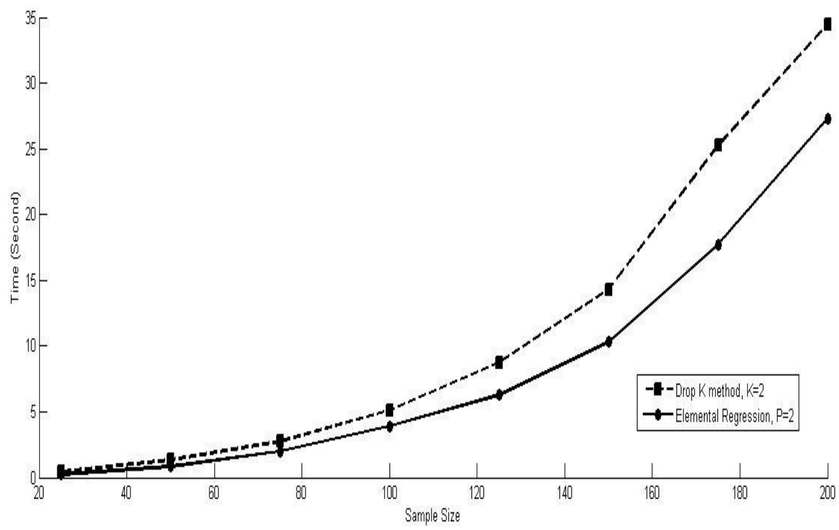

Figure 4.4.1: Computing time vs. Sample size for elemental regression and Drop K method. The solid line represents the seconds needed for elemental regression with p=2. The dashed line represents the seconds needed for Drop K method, k=2.

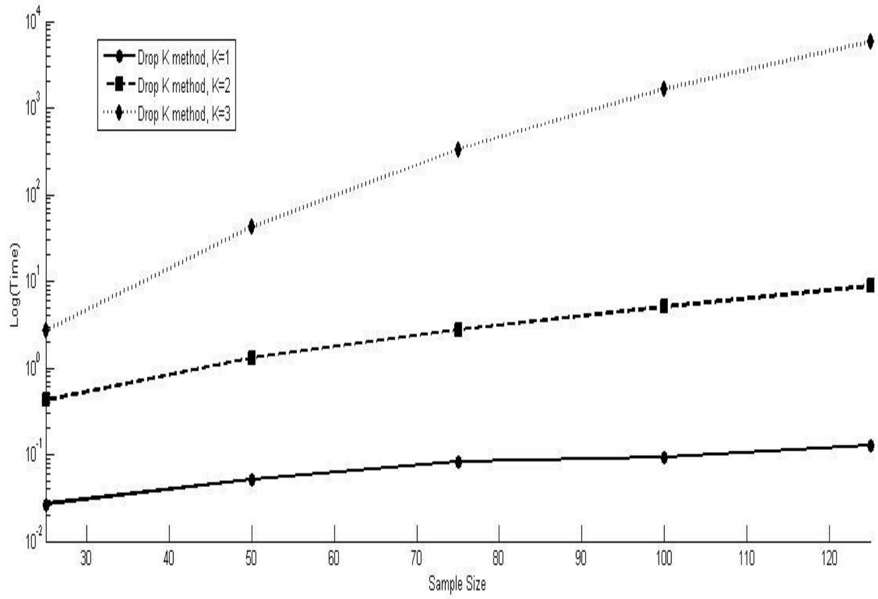

Figure 4.4.2: Computing time vs. Sample size for Drop K method. The solid line shows the log time required for Drop K method, k=1. The dashed line shows the log time needed for Drop K method, k=2. The dotted line represents the log time for Drop K method, k=3.

Chapter 5: Discussion

The appeal of elemental and non-elemental set-based weighted estimators is their potential for robust regression without the wholesale deletion of outliers or suspicious points.

From the simulation results, we can see that the Drop K method generally performed better than our elemental regression-based methods with regard to accuracy and therefore robustness. Within the Drop K method weighting factors, the best estimator was generally the one using Dffits^2 as influential statistics for all possible combinations of sample sizes and error distributions, and its result was competitive compared to the existing Huber method.

Elemental regression provided better estimators than OLS for most cases. However, it is harder to say which weighting factor was the best, since this varied across situation. The accuracy of the estimators was sensitive to the sample size and the choice of the influential statistic upon which to base the 'INF' component of the weights. For the percentile-adjusted weighting factors [see (2.7)], there was no clear trend between the selected α level and the estimators' performance. Mayo and Gray [13] found similar variations in performance depending on the α level selected for use with their proposed TEE estimators. Note that the proposal in (2.7) is similar in spirit to the TEE approach, although the weighting employed features some fundamental differences.

The results using the dataset from Mayo and Gray's paper were consistent with the conclusions drawn above. In particular, the Drop K method with Dffits^2 gave the best estimates and was close to the best result obtained in Mayo's paper.

We extended the Drop K method to MLR by using the data example of Duncan [17]. The dataset had two predictor variables and regression diagnostics found 2 outliers, so we applied the Drop 2 program. The results in Table 4.3.1 showed that the Drop 2 approach provided estimates which were much closer to those based on OLS without the detected outliers, compared to OLS using all data points.

The computing time required for elemental regression and the Drop K method with p=k was similar since the total numbers of subsets generated were the same. Elemental regression required a little bit less time than Drop K method. For the Drop K method, the computing time increased dramatically with k=3 compared to k=2. Therefore, in SLR, if k is larger than 2, elemental regression may be a better choice regarding the time for calculation. If k=p, then we recommend the Drop K method since both consume similar time for computing, while we found the results for the Drop K method to be much closer to the truth than results from elemental regression.

In this thesis, we listed one weighting factor for the Drop K approach and three forms of weight factors for elemental regression. We proposed the idea of adding influential statistics in the weighting factors to reduce the weight for subsets with outliers, and we utilize inverse-variance weighting in the interest of maintaining efficiency. There are many ways of constructing weighting factors combining the efficiency and influential information. Also there are other influential statistics that can be used in the weighting factors.

In conclusion, the Drop K approach appears more stable than elemental regression; however, it is limited based on the number of outliers detected and the sample size. One possible solution to reduce the computation time could be to sample the dataset into several small ones and then apply Drop K approach to each one and combine the results. Future research might include further evaluation of the elemental set-based weights proposed here under typical settings in which there are multiple suspicious (but not necessarily extreme) points. It would also be of great interest to directly compare the proposed estimators in this thesis to the TEE estimators [12, 13], and to develop appropriate and computationally feasible methods for estimating standard errors associated with our weighted estimators.

Tables:

Table 4.1.1: Regression coefficients when n=25, ε=N (0, 0.15) and 2 outliers

Method	Weighting Factor	Beta 0 (mean ±sd)	Beta1 (mean ±sd)
True Value		0.000	1.000
OLS (without outliers)		0.001 ±0.05	0.986 ±0.12
OLS		1.362 ±0.86	0.070 ±0.20
Huber		0.032 ±0.05	0.979 ±0.01
Drop K method	W_{INVF}(Cook's)	0.014 ±0.12	0.974 ±0.15
	W_{INVF}(|Dffits|)	0.025 ±0.14	0.965 ±0.16
	W_{INVF}($Dffits^2$)	0.013 ±0.12	0.975 ±0.14
	W_{INVF}(|Dfbetas|)	0.050 ±0.16	0.969 ±0.15
Elemental Regression	W_{INVF}(Cook's)	0.147 ±0.31	0.896 ±0.15
	W_{INVF}(|Dffits|)	0.282 ±0.22	0.823 ±0.07
	W_{INVF}($Dffits^2$)	0.138 ±0.31	0.902 ±0.15
	W_{INVF}(|Dfbetas|)	0.296 ±0.28	0.922 ±0.06
	W_{MINVF}(Cook's)	0.017 ±0.12	0.984 ±0.03
	W_{MINVF}(|Dffits|)	0.273 ±0.15	0.882 ±0.04
	W_{MINVF}($Dffits^2$)	0.010 ±0.12	0.988 ±0.03
	W_{MINVF}(|Dfbetas|)	0.861 ±0.36	0.935 ±0.03
	W_{PINVF}(Cook's),alpha=0.05	1.466 ±0.77	0.211 ±0.18
	W_{PINVF}(|Dffits|),alpha=0.05	1.459 ±0.76	0.216 ±0.18
	W_{PINVF}($Dffits^2$),alpha=0.05	1.447 ±0.76	0.218 ±0.18
	W_{PINVF}(|Dfbetas|),alpha=0.05	1.918 ±0.86	0.217 ±0.19
	W_{PINVF}(Cook's),alpha=0.25	0.575 ±0.43	0.631 ±0.16
	W_{PINVF}(|Dffits|),alpha=0.25	0.540 ±0.37	0.672 ±0.14
	W_{PINVF}($Dffits^2$),alpha=0.25	0.477 ±0.38	0.690 ±0.15
	W_{PINVF}(|Dfbetas|),alpha=0.25	0.903 ±0.36	0.489 ±0.17
	W_{PINVF}(Cook's),alpha=0.5	0.338 ±0.39	0.776 ±0.19
	W_{PINVF}(|Dffits|),alpha=0.5	0.407 ±0.32	0.776 ±0.14
	W_{PINVF}($Dffits^2$),alpha=0.5	0.276 ±0.36	0.817 ±0.17
	W_{PINVF}(|Dfbetas|),alpha=0.5	0.779 ±0.31	0.489 ±0.17
	W_{PINVF}(Cook's),alpha=0.75	0.206 ±0.42	0.860 ±0.23
	W_{PINVF}(|Dffits|),alpha=0.75	0.357 ±0.28	0.824 ±0.15
	W_{PINVF}($Dffits^2$),alpha=0.75	0.164 ±0.38	0.889 ±0.21
	W_{PINVF}(|Dfbetas|),alpha=0.75	0.841 ±0.34	0.891 ±0.15

Table 4.1.2: Regression coefficients when n=25, ε=N (0, 1) and 2 outliers

Method	**Weighting Factor**	**Beta 0(mean±sd)**	**Beta1 (mean±sd)**
True Value		0.000	1.000
OLS (without outliers)		-0.006±0.30	0.974±0.18
OLS		1.389±0.88	0.062±0.21
Huber		0.204±0.34	0.858±0.10
Drop K method	W_{INVF}(Cook's)	0.160±0.44	0.859±0.24
	W_{INVF}(|Dffits|)	0.378±0.50	0.695±0.24
	W_{INVF}($Dffits^2$)	0.128±0.43	0.882±0.24
	W_{INVF}(|Dfbetas|)	0.595±0.49	0.750±0.24
Elemental Regression	W_{INVF}(Cook's)	0.857±0.80	0.395±0.30
	W_{INVF}(|Dffits|)	0.548±0.55	0.611±0.19
	W_{INVF}($Dffits^2$)	0.857±0.80	0.396±0.30
	W_{INVF}(|Dfbetas|)	0.315±0.67	0.684±0.30
	W_{MINVF}(Cook's)	0.167±0.60	0.793±0.20
	W_{MINVF}(|Dffits|)	0.333±0.44	0.811±0.11
	W_{MINVF}($Dffits^2$)	0.164±0.60	0.794±0.20
	W_{MINVF}(|Dfbetas|)	0.966±0.54	0.856±0.13
	W_{PINVF}(Cook's),alpha=0.05	1.475±0.80	0.204±0.19
	W_{PINVF}(|Dffits|),alpha=0.05	1.466±0.80	0.209±0.19
	W_{PINVF}($Dffits^2$),alpha=0.05	1.454±0.80	0.211±0.19
	W_{PINVF}(|Dfbetas|),alpha=0.05	1.904±0.91	0.205±0.19
	W_{PINVF}(Cook's),alpha=0.25	0.628±0.52	0.611±0.18
	W_{PINVF}(|Dffits|),alpha=0.25	0.612±0.49	0.644±0.17
	W_{PINVF}($Dffits^2$),alpha=0.25	0.545±0.50	0.662±0.18
	W_{PINVF}(|Dfbetas|),alpha=0.25	1.035±0.47	0.530±0.18
	W_{PINVF}(Cook's),alpha=0.5	0.432±0.53	0.700±0.20
	W_{PINVF}(|Dffits|),alpha=0.5	0.508±0.47	0.707±0.17
	W_{PINVF}($Dffits^2$),alpha=0.5	0.376±0.51	0.735±0.20
	W_{PINVF}(|Dfbetas|),alpha=0.5	0.887±0.45	0.530±0.18
	W_{PINVF}(Cook's),alpha=0.75	0.330±0.64	0.729±0.24
	W_{PINVF}(|Dffits|),alpha=0.75	0.439±0.50	0.752±0.18
	W_{PINVF}($Dffits^2$),alpha=0.75	0.291±0.63	0.751±0.23
	W_{PINVF}(|Dfbetas|),alpha=0.75	0.946±0.50	0.795±0.20

Table 4.1.3: Regression coefficients when n=50, ε=N (0, 0.15) and 2 outliers

Method	**Weighting Factor**	**Beta 0 (mean ±sd)**	**Beta1 (mean ±sd)**
True Value		0.000	1.000
OLS (without outliers)		-0.002 ±0.03	1.000 ±0.01
OLS		0.885 ±0.39	0.452 ±0.10
Huber		0.013 ±0.03	0.991 ±0.01
Drop K method	W_{INVF}(Cook's)	-0.001 ±0.03	1.000 ±0.01
	W_{INVF}(|Dffits|)	0.008 ±0.03	0.994 ±0.01
	W_{INVF}($Dffits^2$)	-0.001 ±0.03	1.000 ±0.01
	W_{INVF}(|Dfbetas|)	0.015 ±0.03	0.996 ±0.01
Elemental Regression	W_{INVF}(Cook's)	0.242 ±0.27	0.847 ±0.14
	W_{INVF}(|Dffits|)	0.095 ±0.10	0.942 ±0.03
	W_{INVF}($Dffits^2$)	0.241 ±0.27	0.847 ±0.14
	W_{INVF}(|Dfbetas|)	0.048 ±0.10	0.968 ±0.05
	W_{MINVF}(Cook's)	0.015 ±0.11	0.986 ±0.03
	W_{MINVF}(|Dffits|)	0.050 ±0.06	0.976 ±0.01
	W_{MINVF}($Dffits^2$)	0.015 ±0.11	0.986 ±0.03
	W_{MINVF}(|Dfbetas|)	0.164 ±0.09	0.987 ±0.01
	W_{PINVF}(Cook's),alpha=0.05	0.570 ±0.23	0.696 ±0.07
	W_{PINVF}(|Dffits|),alpha=0.05	0.536 ±0.21	0.712 ±0.06
	W_{PINVF}($Dffits^2$),alpha=0.05	0.533 ±0.21	0.713 ±0.06
	W_{PINVF}(|Dfbetas|),alpha=0.05	0.848 ±0.28	0.677 ±0.07
	W_{PINVF}(Cook's),alpha=0.25	0.144 ±0.11	0.912 ±0.05
	W_{PINVF}(|Dffits|),alpha=0.25	0.121 ±0.09	0.929 ±0.04
	W_{PINVF}($Dffits^2$),alpha=0.25	0.111 ±0.09	0.932 ±0.04
	W_{PINVF}(|Dfbetas|),alpha=0.25	0.174 ±0.08	0.865 ±0.06
	W_{PINVF}(Cook's),alpha=0.5	0.096 ±0.11	0.940 ±0.06
	W_{PINVF}(|Dffits|),alpha=0.5	0.093 ±0.09	0.947 ±0.04
	W_{PINVF}($Dffits^2$),alpha=0.5	0.073 ±0.10	0.953 ±0.05
	W_{PINVF}(|Dfbetas|),alpha=0.5	0.162 ±0.08	0.865 ±0.06
	W_{PINVF}(Cook's),alpha=0.75	0.077 ±0.15	0.949 ±0.08
	W_{PINVF}(|Dffits|),alpha=0.75	0.084 ±0.11	0.955 ±0.06
	W_{PINVF}($Dffits^2$),alpha=0.75	0.059 ±0.14	0.961 ±0.07
	W_{PINVF}(|Dfbetas|),alpha=0.75	0.166 ±0.09	0.974 ±0.05

Table 4.1.4: Regression coefficients when n=50, ε=N (0, 1) and 2 outliers

Method	Weighting Factor	Beta 0(mean ±sd)	Beta1 (mean ±sd)
True Value		0.000	1.000
OLS (without outliers)		-0.001 ±0.21	1.000 ±0.06
OLS		0.896 ±0.45	0.448 ±0.12
Huber		0.095 ±0.22	0.940 ±0.06
Drop K method	W_{INVF}(Cook's)	0.046 ±0.21	0.970 ±0.06
	W_{INVF}(|Dffits|)	0.266 ±0.24	0.825 ±0.07
	W_{INVF}($Dffits^2$)	0.028 ±0.21	0.981 ±0.06
	W_{INVF}(|Dfbetas|)	0.380 ±0.24	0.866 ±0.07
Elemental Regression	W_{INVF}(Cook's)	0.773 ±0.42	0.518 ±0.12
	W_{INVF}(|Dffits|)	0.468 ±0.35	0.696 ±0.11
	W_{INVF}($Dffits^2$)	0.773 ±0.42	0.517 ±0.12
	W_{INVF}(|Dfbetas|)	0.183 ±0.49	0.683 ±0.16
	W_{MINVF}(Cook's)	0.344 ±0.40	0.742 ±0.13
	W_{MINVF}(|Dffits|)	0.171 ±0.29	0.874 ±0.08
	W_{MINVF}($Dffits^2$)	0.344 ±0.40	0.741 ±0.13
	W_{MINVF}(|Dfbetas|)	0.315 ±0.31	0.872 ±0.09
	W_{PINVF}(Cook's),alpha=0.05	0.567 ±0.32	0.700 ±0.09
	W_{PINVF}(|Dffits|),alpha=0.05	0.538 ±0.30	0.716 ±0.08
	W_{PINVF}($Dffits^2$),alpha=0.05	0.534 ±0.30	0.717 ±0.08
	W_{PINVF}(|Dfbetas|),alpha=0.05	0.871 ±0.35	0.670 ±0.10
	W_{PINVF}(Cook's),alpha=0.25	0.219 ±0.25	0.864 ±0.07
	W_{PINVF}(|Dffits|),alpha=0.25	0.200 ±0.24	0.879 ±0.07
	W_{PINVF}($Dffits^2$),alpha=0.25	0.187 ±0.25	0.882 ±0.07
	W_{PINVF}(|Dfbetas|),alpha=0.25	0.327 ±0.24	0.836 ±0.08
	W_{PINVF}(Cook's),alpha=0.5	0.238 ±0.28	0.834 ±0.08
	W_{PINVF}(|Dffits|),alpha=0.5	0.222 ±0.27	0.856 ±0.07
	W_{PINVF}($Dffits^2$),alpha=0.5	0.218 ±0.28	0.845 ±0.08
	W_{PINVF}(|Dfbetas|),alpha=0.5	0.303 ±0.26	0.836 ±0.08
	W_{PINVF}(Cook's),alpha=0.75	0.321 ±0.34	0.776 ±0.10
	W_{PINVF}(|Dffits|),alpha=0.75	0.223 ±0.29	0.849 ±0.09
	W_{PINVF}($Dffits^2$),alpha=0.75	0.309 ±0.34	0.782 ±0.10
	W_{PINVF}(|Dfbetas|),alpha=0.75	0.319 ±0.28	0.837 ±0.11

Table 4.1.5: Regression coefficients when n=100, ε=N (0, 0.15) and 2 outliers

Method	**Weighting Factor**	**Beta 0 (mean ±sd)**	**Beta 1 (mean ±sd)**
True Value		0.000	1.000
OLS (without outliers)		0.000 ±0.02	1.000 ±0.01
OLS		0.558 ±0.19	0.683 ±0.06
Huber		0.008 ±0.02	0.996 ±0.01
Drop K method	W_{INVF}(Cook's)	0.001 ±0.02	1.000 ±0.01
	W_{INVF}(|Dffits|)	0.010 ±0.02	0.994 ±0.01
	$W_{INVF}(Dffits^2)$	0.001 ±0.02	1.000 ±0.01
	W_{INVF}(|Dfbetas|)	0.014 ±0.02	0.996 ±0.01
Elemental Regression	W_{INVF}(Cook's)	0.337 ±0.18	0.805 ±0.08
	W_{INVF}(|Dffits|)	0.079 ±0.07	0.953 ±0.03
	$W_{INVF}(Dffits^2)$	0.337 ±0.18	0.805 ±0.08
	W_{INVF}(|Dfbetas|)	0.014 ±0.09	0.955 ±0.05
	W_{MINVF}(Cook's)	0.038 ±0.09	0.972 ±0.02
	W_{MINVF}(|Dffits|)	0.016 ±0.04	0.990 ±0.01
	$W_{MINVF}(Dffits^2)$	0.038 ±0.09	0.972 ±0.02
	W_{MINVF}(|Dfbetas|)	0.036 ±0.04	0.991 ±0.01
	W_{PINVF}(Cook's),alpha=0.05	0.127 ±0.06	0.930 ±0.02
	W_{PINVF}(|Dffits|),alpha=0.05	0.099 ±0.05	0.946 ±0.02
	$W_{PINVF}(Dffits^2)$,alpha=0.05	0.098 ±0.05	0.946 ±0.02
	W_{PINVF}(|Dfbetas|),alpha=0.05	0.173 ±0.07	0.878 ±0.03
	W_{PINVF}(Cook's),alpha=0.25	0.044 ±0.04	0.975 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.25	0.036 ±0.03	0.980 ±0.01
	$W_{PINVF}(Dffits^2)$,alpha=0.25	0.034 ±0.03	0.981 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.25	0.045 ±0.03	0.960 ±0.02
	W_{PINVF}(Cook's),alpha=0.5	0.037 ±0.04	0.978 ±0.02
	W_{PINVF}(|Dffits|),alpha=0.5	0.033 ±0.04	0.981 ±0.01
	$W_{PINVF}(Dffits^2)$,alpha=0.5	0.030 ±0.04	0.982 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.5	0.043 ±0.04	0.960 ±0.02
	W_{PINVF}(Cook's),alpha=0.75	0.041 ±0.06	0.975 ±0.02
	W_{PINVF}(|Dffits|),alpha=0.75	0.030 ±0.05	0.983 ±0.02
	$W_{PINVF}(Dffits^2)$,alpha=0.75	0.036 ±0.06	0.978 ±0.02
	W_{PINVF}(|Dfbetas|),alpha=0.75	0.042 ±0.04	0.985 ±0.02

Table 4.1.6: Regression coefficients when n=100, ε=N (0, 1) and 2 outliers

Method	Weighting Factor	Beta 0 (mean ±sd)	Beta 1 (mean ±sd)
True Value		0.000	1.000
OLS (without outliers)		0.003 ±0.15	1.000 ±0.04
OLS		0.560 ±0.24	0.683 ±0.06
Huber		0.056 ±0.15	0.971 ±0.04
Drop K method	W_{INVF}(Cook's)	0.032 ±0.15	0.983 ±0.04
	W_{INVF}(|Dffits|)	0.227 ±0.16	0.868 ±0.04
	W_{INVF}($Dffits^2$)	0.021 ±0.15	0.990 ±0.04
	W_{INVF}(|Dfbetas|)	0.288 ±0.16	0.897 ±0.04
Elemental Regression	W_{INVF}(Cook's)	0.538 ±0.24	0.695 ±0.06
	W_{INVF}(|Dffits|)	0.385 ±0.21	0.776 ±0.06
	W_{INVF}($Dffits^2$)	0.538 ±0.24	0.695 ±0.06
	W_{INVF}(|Dfbetas|)	0.260 ±0.33	0.768 ±0.08
	W_{MINVF}(Cook's)	0.366 ±0.23	0.781 ±0.07
	W_{MINVF}(|Dffits|)	0.189 ±0.18	0.881 ±0.05
	W_{MINVF}($Dffits^2$)	0.366 ±0.23	0.780 ±0.07
	W_{MINVF}(|Dfbetas|)	0.210 ±0.20	0.879 ±0.05
	W_{PINVF}(Cook's),alpha=0.05	0.150 ±0.16	0.927 ±0.04
	W_{PINVF}(|Dffits|),alpha=0.05	0.125 ±0.15	0.940 ±0.04
	W_{PINVF}($Dffits^2$),alpha=0.05	0.123 ±0.15	0.940 ±0.04
	W_{PINVF}(|Dfbetas|),alpha=0.05	0.303 ±0.16	0.892 ±0.04
	W_{PINVF}(Cook's),alpha=0.25	0.155 ±0.16	0.912 ±0.04
	W_{PINVF}(|Dffits|),alpha=0.25	0.147 ±0.16	0.918 ±0.04
	W_{PINVF}($Dffits^2$),alpha=0.25	0.146 ±0.16	0.917 ±04
	W_{PINVF}(|Dfbetas|),alpha=0.25	0.185 ±0.16	0.911 ±04
	W_{PINVF}(Cook's),alpha=0.5	0.228 ±0.18	0.866 ±0.04
	W_{PINVF}(|Dffits|),alpha=0.5	0.204 ±0.18	0.881 ±0.04
	W_{PINVF}($Dffits^2$),alpha=0.5	0.221 ±0.18	0.869 ±0.04
	W_{PINVF}(|Dfbetas|),alpha=0.5	0.197 ±0.17	0.911 ±0.04
	W_{PINVF}(Cook's),alpha=0.75	0.300 ±0.20	0.822 ±0.05
	W_{PINVF}(|Dffits|),alpha=0.75	0.219 ±0.19	0.868 ±0.05
	W_{PINVF}($Dffits^2$),alpha=0.75	0.296 ±0.20	0.824 ±0.05
	W_{PINVF}(|Dfbetas|),alpha=0.75	0.220 ±0.18	0.864 ±0.05

Table 4.1.7: Regression coefficients when n=25, ε=N (0, 0.15) and no outliers

Method	Weighting Factor	Beta 0 (mean ±sd)	Beta 1 (mean ±sd)
True Value		0.000	1.000
OLS (without outliers)		0.004 ±0.05	0.999 ±0.01
OLS		0.000 ±0.02	1.000 ±0.31
Huber		0.000 ±0.04	1.000 ±0.01
Drop K method	W_{INVF}(Cook's)	0.003 ±0.04	0.999 ±0.01
	W_{INVF}(|Dffits|)	0.003 ±0.04	0.999 ±0.01
	W_{INVF}($Dffits^2$)	0.003 ±0.04	0.999 ±0.01
	W_{INVF}(|Dfbetas|)	0.003 ±0.04	1.000 ±0.01
Elemental Regression	W_{INVF}(Cook's)	0.000 ±0.05	1.000 ±0.01
	W_{INVF}(|Dffits|)	0.000 ±0.04	1.000 ±0.01
	W_{INVF}($Dffits^2$)	0.000 ±0.05	1.000 ±0.01
	W_{INVF}(|Dfbetas|)	-0.002 ±0.05	1.000 ±0.01
	W_{MINVF}(Cook's)	-0.001 ±0.05	1.000 ±0.01
	W_{MINVF}(|Dffits|)	-0.002 ±0.05	1.000 ±0.01
	W_{MINVF}($Dffits^2$)	-0.001 ±0.05	1.000 ±0.01
	W_{MINVF}(|Dfbetas|)	-0.002 ±0.05	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.05	-0.001 ±0.04	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.05	-0.001 ±0.04	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.05	-0.001 ±0.04	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.05	0.000 ±0.04	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.25	-0.001 ±0.05	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.25	-0.001 ±0.05	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.25	-0.001 ±0.05	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.25	0.000 ±0.05	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.5	-0.001 ±0.05	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.5	-0.001 ±0.05	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.5	-0.001 ±0.05	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.5	-0.001 ±0.05	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.75	0.000 ±0.05	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.75	-0.001 ±0.05	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.75	0.000 ±0.05	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.75	-0.001 ±0.05	1.000 ±0.01

Table 4.1.8: Regression coefficients when n=25, ε=N (0, 1) and no outliers

Method	Weighting Factor	Beta 0 (mean±sd)	Beta 1 (mean±sd)
True Value		0.000	1.000
OLS (without outliers)		-0.032±0.33	1.015±0.08
OLS		0.005±0.30	1.002±0.07
Huber		0.008±0.30	1.002±0.07
Drop K method	W_{INVF}(Cook's)	-0.007±0.29	1.009±0.08
	W_{INVF}(|Dffits|)	-0.006±0.29	1.007±0.08
	W_{INVF}($Dffits^2$)	-0.011±0.30	1.010±0.08
	W_{INVF}(|Dfbetas|)	-0.007±0.29	1.007±0.08
Elemental Regression	W_{INVF}(Cook's)	0.010±0.31	1.001±0.07
	W_{INVF}(|Dffits|)	0.007±0.30	1.002±0.07
	W_{INVF}($Dffits^2$)	0.010±0.31	1.001±0.07
	W_{INVF}(|Dfbetas|)	0.011±0.35	1.002±0.08
	W_{MINVF}(Cook's)	0.005±0.32	1.002±0.08
	W_{MINVF}(|Dffits|)	0.008±0.32	1.002±0.08
	W_{MINVF}($Dffits^2$)	0.005±0.32	1.002±0.08
	W_{MINVF}(|Dfbetas|)	0.007±0.34	1.001±0.08
	W_{PINVF}(Cook's),alpha=0.05	0.010±0.30	1.001±0.07
	W_{PINVF}(|Dffits|),alpha=0.05	0.010±0.30	1.001±0.07
	W_{PINVF}($Dffits^2$),alpha=0.05	0.010±0.30	1.001±0.07
	W_{PINVF}(|Dfbetas|),alpha=0.05	0.008±0.30	1.002±0.07
	W_{PINVF}(Cook's),alpha=0.25	0.008±0.31	1.001±0.07
	W_{PINVF}(|Dffits|),alpha=0.25	0.008±0.31	1.001±0.07
	W_{PINVF}($Dffits^2$),alpha=0.25	0.008±0.31	1.001±0.07
	W_{PINVF}(|Dfbetas|),alpha=0.25	0.009±0.31	1.002±0.07
	W_{PINVF}(Cook's),alpha=0.5	0.006±0.31	1.002±0.07
	W_{PINVF}(|Dffits|),alpha=0.5	0.008±0.31	1.002±0.07
	W_{PINVF}($Dffits^2$),alpha=0.5	0.007±0.31	1.002±0.07
	W_{PINVF}(|Dfbetas|),alpha=0.5	0.011±0.31	1.002±0.07
	W_{PINVF}(Cook's),alpha=0.75	0.006±0.31	1.002±0.07
	W_{PINVF}(|Dffits|),alpha=0.75	0.007±0.31	1.002±0.07
	W_{PINVF}($Dffits^2$),alpha=0.75	0.006±0.31	1.002±0.07
	W_{PINVF}(|Dfbetas|),alpha=0.75	0.009±0.32	1.001±0.08

Table 4.1.9: Regression coefficients when n=50, ε=N (0, 0.15) and no outliers

Method	Weighting Factor	Beta 0 (mean ±sd)	Beta 1 (mean ±sd)
True Value		0.000	1.000
OLS (without outliers)		-0.003 ±0.04	1.002 ±0.01
OLS		0.001 ±0.03	1.000 ±0.01
Huber		0.001 ±0.03	1.000 ±0.01
Drop K method	W_{INVF}(Cook's)	-0.003 ±0.03	1.001 ±0.01
	W_{INVF}(|Dffits|)	-0.003 ±0.03	1.001 ±0.01
	W_{INVF}($Dffits^2$)	-0.003 ±0.03	1.001 ±0.01
	W_{INVF}(|Dfbetas|)	-0.003 ±0.03	1.001 ±0.01
Elemental Regression	W_{INVF}(Cook's)	0.000 ±0.03	1.000 ±0.01
	W_{INVF}(|Dffits|)	0.000 ±0.03	1.000 ±0.01
	W_{INVF}($Dffits^2$)	0.000 ±0.03	1.000 ±0.01
	W_{INVF}(|Dfbetas|)	0.000 ±0.04	1.000 ±0.01
	W_{MINVF}(Cook's)	0.000 ±0.03	1.000 ±0.01
	W_{MINVF}(|Dffits|)	0.000 ±0.03	1.000 ±0.01
	W_{MINVF}($Dffits^2$)	0.000 ±0.03	1.000 ±0.01
	W_{MINVF}(|Dfbetas|)	0.000 ±0.04	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.05	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.05	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.05	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.05	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.25	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.25	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.25	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.25	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.5	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.5	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.5	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.5	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.75	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.75	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.75	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.75	0.000 ±0.03	1.000 ±0.01

Table 4.1.10: Regression coefficients when n=50, ε=N (0, 1) and no outliers

Method	Weighting Factor	Beta 0 (mean ±sd)	Beta 1 (mean ±sd)
True Value		0.000	1.000
OLS (without outliers)		-0.004 ±0.22	1.002 ±0.06
OLS		-0.007 ±0.20	1.000 ±0.05
Huber		-0.006 ±0.21	1.000 ±0.05
Drop K method	W_{INVF}(Cook's)	-0.021 ±0.20	1.005 ±0.05
	W_{INVF}(|Dffits|)	-0.024 ±0.20	1.005 ±0.05
	W_{INVF}($Dffits^2$)	-0.020 ±0.20	1.004 ±0.06
	W_{INVF}(|Dfbetas|)	-0.023 ±0.20	1.005 ±0.05
Elemental Regression	W_{INVF}(Cook's)	-0.006 ±0.21	1.000 ±0.05
	W_{INVF}(|Dffits|)	-0.006 ±0.21	1.000 ±0.05
	W_{INVF}($Dffits^2$)	-0.006 ±0.21	1.000 ±0.05
	W_{INVF}(|Dfbetas|)	-0.009 ±0.27	0.998 ±0.06
	W_{MINVF}(Cook's)	-0.005 ±0.22	0.999 ±0.05
	W_{MINVF}(|Dffits|)	-0.005 ±0.22	0.999 ±0.05
	W_{MINVF}($Dffits^2$)	-0.005 ±0.22	0.999 ±0.05
	W_{MINVF}(|Dfbetas|)	-0.007 ±0.23	0.999 ±0.06
	W_{PINVF}(Cook's),alpha=0.05	-0.006 ±0.21	1.000 ±0.05
	W_{PINVF}(|Dffits|),alpha=0.05	-0.006 ±0.21	1.000 ±0.05
	W_{PINVF}($Dffits^2$),alpha=0.05	-0.006 ±0.21	1.000 ±0.05
	W_{PINVF}(|Dfbetas|),alpha=0.05	-0.007 ±0.20	1.000 ±0.05
	W_{PINVF}(Cook's),alpha=0.25	-0.006 ±0.21	1.000 ±0.05
	W_{PINVF}(|Dffits|),alpha=0.25	-0.007 ±0.21	1.000 ±0.05
	W_{PINVF}($Dffits^2$),alpha=0.25	-0.006 ±0.21	1.000 ±0.05
	W_{PINVF}(|Dfbetas|),alpha=0.25	-0.007 ±0.21	1.000 ±0.05
	W_{PINVF}(Cook's),alpha=0.5	-0.006 ±0.21	1.000 ±0.05
	W_{PINVF}(|Dffits|),alpha=0.5	-0.006 ±0.21	1.000 ±0.05
	W_{PINVF}($Dffits^2$),alpha=0.5	-0.006 ±0.21	1.000 ±0.05
	W_{PINVF}(|Dfbetas|),alpha=0.5	-0.007 ±0.21	1.000 ±0.05
	W_{PINVF}(Cook's),alpha=0.75	-0.006 ±0.21	1.000 ±0.05
	W_{PINVF}(|Dffits|),alpha=0.75	-0.006 ±0.22	0.999 ±0.05
	W_{PINVF}($Dffits^2$),alpha=0.75	-0.006 ±0.21	1.000 ±0.05
	W_{PINVF}(|Dfbetas|),alpha=0.75	-0.007 ±0.22	0.999 ±0.05

Table 4.1.11: Regression coefficients when n=100, ε=N (0, 0.15) and no outliers

Method	Weighting Factor	Beta 0 (mean ±sd)	Beta 1 (mean ±sd)
True Value		0.000	1.000
OLS (without outliers)		0.004 ±0.02	0.999 ±0.01
OLS		0.000 ±0.02	1.000 ±0.01
Huber		0.001 ±0.02	1.000 ±0.01
Drop K method	W_{INVF}(Cook's)	0.003 ±0.02	0.999 ±0.01
	W_{INVF}(|Dffits|)	0.003 ±0.02	0.999 ±0.01
	W_{INVF}($Dffits^2$)	0.003 ±0.02	0.999 ±0.01
	W_{INVF}(|Dfbetas|)	0.003 ±0.02	0.999 ±0.01
Elemental Regression	W_{INVF}(Cook's)	0.000 ±0.02	1.000 ±0.01
	W_{INVF}(|Dffits|)	0.000 ±0.02	1.000 ±0.01
	W_{INVF}($Dffits^2$)	0.000 ±0.02	1.000 ±0.01
	W_{INVF}(|Dfbetas|)	0.000 ±0.03	1.000 ±0.01
	W_{MINVF}(Cook's)	0.000 ±0.02	1.000 ±0.01
	W_{MINVF}(|Dffits|)	0.000 ±0.02	1.000 ±0.01
	W_{MINVF}($Dffits^2$)	0.000 ±0.02	1.000 ±0.01
	W_{MINVF}(|Dfbetas|)	0.000 ±0.03	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.05	0.001 ±0.02	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.05	0.001 ±0.02	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.05	0.001 ±0.02	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.05	0.001 ±0.02	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.25	0.000 ±0.02	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.25	0.000 ±0.02	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.25	0.000 ±0.02	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.25	0.001 ±0.02	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.5	0.001 ±0.02	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.5	0.000 ±0.02	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.5	0.001 ±0.02	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.5	0.000 ±0.02	1.000 ±0.01
	W_{PINVF}(Cook's),alpha=0.75	0.000 ±0.02	1.000 ±0.01
	W_{PINVF}(|Dffits|),alpha=0.75	0.000 ±0.02	1.000 ±0.01
	W_{PINVF}($Dffits^2$),alpha=0.75	0.000 ±0.02	1.000 ±0.01
	W_{PINVF}(|Dfbetas|),alpha=0.75	0.000 ±0.02	1.000 ±0.01

Table 4.1.12: Regression coefficients when n=100, ε=N (0, 1) and no outliers

Method	Weighting Factor	Beta 0(mean ±sd)	Beta 1(mean ±sd)
True Value		0.000	1.000
OLS (without outliers)		-0.025 ±0.15	1.002 ±0.04
OLS		-0.002 ±0.14	1.000 ±0.03
Huber		-0.004 ±0.15	1.000 ±0.03
Drop K method	W_{INVF}(Cook's)	-0.017 ±0.15	1.002 ±0.03
	W_{INVF}(|Dffits|)	-0.015 ±0.14	1.002 ±0.03
	W_{INVF}(Dffits^2)	-0.017 ±0.15	1.002 ±0.03
	W_{INVF}(|Dfbetas|)	-0.015 ±0.15	1.002 ±0.03
Elemental Regression	W_{INVF}(Cook's)	-0.004 ±0.15	1.000 ±0.03
	W_{INVF}(|Dffits|)	-0.003 ±0.15	1.000 ±0.03
	W_{INVF}(Dffits^2)	-0.004 ±0.15	1.000 ±0.03
	W_{INVF}(|Dfbetas|)	-0.007 ±0.19	1.000 ±0.04
	W_{MINVF}(Cook's)	-0.005 ±0.15	1.000 ±0.04
	W_{MINVF}(|Dffits|)	-0.004 ±0.15	1.000 ±0.04
	W_{MINVF}(Dffits^2)	-0.005 ±0.15	1.000 ±0.04
	W_{MINVF}(|Dfbetas|)	-0.007 ±0.16	1.000 ±0.04
	W_{PINVF}(Cook's),alpha=0.05	-0.004 ±0.15	1.000 ±0.03
	W_{PINVF}(|Dffits|),alpha=0.05	-0.005 ±0.15	1.000 ±0.03
	W_{PINVF}(Dffits^2),alpha=0.05	-0.005 ±0.15	1.000 ±0.03
	W_{PINVF}(|Dfbetas|),alpha=0.05	-0.004 ±0.15	1.000 ±0.03
	W_{PINVF}(Cook's),alpha=0.25	-0.005 ±0.15	1.000 ±0.03
	W_{PINVF}(|Dffits|),alpha=0.25	-0.005 ±0.15	1.000 ±0.03
	W_{PINVF}(Dffits^2),alpha=0.25	-0.005 ±0.15	1.000 ±0.03
	W_{PINVF}(|Dfbetas|),alpha=0.25	-0.006 ±0.15	1.000 ±0.03
	W_{PINVF}(Cook's),alpha=0.5	-0.004 ±0.15	1.000 ±0.03
	W_{PINVF}(|Dffits|),alpha=0.5	-0.004 ±0.15	1.000 ±0.03
	W_{PINVF}(Dffits^2),alpha=0.5	-0.004 ±0.15	1.000 ±0.03
	W_{PINVF}(|Dfbetas|),alpha=0.5	-0.006 ±0.15	1.000 ±0.03
	W_{PINVF}(Cook's),alpha=0.75	-0.003 ±0.15	1.000 ±0.03
	W_{PINVF}(|Dffits|),alpha=0.75	-0.004 ±0.15	1.000 ±0.03
	W_{PINVF}(Dffits^2),alpha=0.75	-0.003 ±0.15	1.000 ±0.03
	W_{PINVF}(|Dfbetas|),alpha=0.75	-0.005 ±0.15	1.000 ±0.04

Table 4.2.1: Artificial Simple Linear Regression Data Set [12]

X	Y
1.02	1.12
1.06	1.11
1.1	0.9
1.14	0.98
1.18	1.39
1.22	1.14
1.26	1.45
1.3	1.21
1.34	1.3
1.38	1.17
1.42	1.57
1.46	1.46
1.5	1.59
1.54	1.26
1.58	1.75
1.62	1.68
1.66	1.88
1.7	1.58
1.74	1.87
1.78	1.63
1.82	1.69
1.86	2.11
1.9	1.95
1.94	1.73
2	1

Table 4.2.2: Regression Estimates for Artificial Dataset [12]

Estimator	Intercept	Slope
True Value	0	1
OLS	.316	.763
LAV	.301	.763
LMS	-.655	1.500
LMS – BEE	-.821	1.607
Chebyshev	1.341	.111
Chebyshev – BEE	1.095	.250
TEE (25% trim, SAE)	.154	.877
TEE (50% trim, SAE)	.010	.995
TEE (75% trim, SAE)	.090	.922
TEE (50% trim, SAE with 25% trim)	-.012	1.019

Table 4.2.3: Regression coefficients using data example in Mayo (1997) paper

Method	**Weighting Factor**	**Beta 0**	**Beta 1**
True Value		0.000	1.000
OLS (without outliers)		0.000	1.000
OLS		0.316	0.763
Huber		0.113	0.915
Drop 1 method	W_{INVF}(Cook's)	0.058	0.956
	W_{INVF}(|Dffits|)	0.148	0.887
	W_{INVF}($Dffits^2$)	0.027	0.979
	W_{INVF}(|Dfbetas|)	0.129	0.908
Elemental Regression	W_{INVF}(Cook's)	0.328	0.750
	W_{INVF}(|Dffits|)	0.287	0.784
	W_{INVF}($Dffits^2$)	0.328	0.750
	W_{INVF}(|Dfbetas|)	0.274	0.821
	W_{MINVF}(Cook's)	0.318	0.764
	W_{MINVF}(|Dffits|)	0.223	0.841
	W_{MINVF}($Dffits^2$)	0.319	0.763
	W_{MINVF}(|Dfbetas|)	0.200	0.880
	W_{PINVF}(Cook's),alpha=0.05	0.165	0.877
	W_{PINVF}(|Dffits|),alpha=0.05	0.142	0.892
	W_{PINVF}($Dffits^2$),alpha=0.05	0.139	0.894
	W_{PINVF}(|Dfbetas|),alpha=0.05	0.175	0.871
	W_{PINVF}(Cook's),alpha=0.25	0.199	0.851
	W_{PINVF}(|Dffits|),alpha=0.25	0.141	0.900
	W_{PINVF}($Dffits^2$),alpha=0.25	0.145	0.895
	W_{PINVF}(|Dfbetas|),alpha=0.25	0.107	0.886
	W_{PINVF}(Cook's),alpha=0.5	0.204	0.852
	W_{PINVF}(|Dffits|),alpha=0.5	0.193	0.864
	W_{PINVF}($Dffits^2$),alpha=0.5	0.204	0.852
	W_{PINVF}(|Dfbetas|),alpha=0.5	0.215	0.886
	W_{PINVF}(Cook's),alpha=0.75	0.318	0.762
	W_{PINVF}(|Dffits|),alpha=0.75	0.232	0.832
	W_{PINVF}($Dffits^2$),alpha=0.75	0.318	0.761
	W_{PINVF}(|Dfbetas|),alpha=0.75	0.205	0.788

Table 4.3.1: Regression Coefficients using Duncan's data with Drop 2 Approach

Method	**Weighting Factor**	**Beta 0**	**Beta 1**	**Beta 2**
OLS		-6.065	0.599	0.546
OLS without outliers		-6.628	0.732	0.433
Huber		-7.026	0.710	0.477
Drop 2 method	W_{INVF}(Cook's)	-6.582	0.705	0.464
	W_{INVF}(|Dffits|)	-6.361	0.731	0.427
	W_{INVF}(Dffits^2)	-6.608	0.711	0.460
	W_{INVF}(|Dfbetas|)	-6.326	0.666	0.492

References:

1. Welsch, R.E. and E. Kuh, *Linear Regression Diagnostics*. 1977, Cambridge, Mass.: Massachusetts Institute of Technology.
2. Cook, R.D., *Detection of Influential Observation in Linear-Regression.* Technometrics, 1977. **19**(1): p. 15-18.
3. Huber, P.J., *Robust Statistics*. 1981, New York: John Wiley.
4. Huber, P.J., *Robust Estimation of a Location Parameter.* The Annals of Mathematical Statistics 1964. **35**(1): p. 73-101.
5. Hampel, F.R., *Robust statistics : The Approach Based on Influence Functions*. 1986, New York: John Wiley.
6. Jaeckel, L.A., *Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals.* The Annals of Mathematical Statistics, 1972. **43**(5): p. 1449-1458.
7. Dielman, T.E., *Least absolute value estimation in regression models: an annotated bibliography.* Communications in Statistics - Theory and Methods, 1984. **13**(4): p. 513-541.
8. Rousseeuw, P.J., *Least Median of Squares Regression.* Journal of the American Statistical Association, 1984. **79**(388): p. 871-880.
9. Rousseeuw, P.J. and A.M. Leroy, *Robust Regression and Outlier Detection*. 1987, New York: John Wiley.
10. Kutner, M.H., Nachtsheim, C. J., Neter, J. and W. Li, *Applied Linear Statistical Models*. 5th ed. 2005, Boston: McGraw-Hill Irwin.
11. Farebrother, R.W., *Notes on the Early History of Elemental Set Methods.* Lecture Notes-Monograph Series, 1997. **31**: p. 161-170.
12. Mayo, M.S. and J.B. Gray, *Elemental Subsets: The Building Blocks of Regression.* American Statistician, 1997. **51**(2): p. 122-129.
13. Mayo, M.S. and J.B. Gray, *The Robustness and Efficiency of Trimmed Elemental Estimation in Regression Analysis: a Monte Carlo Simulation Study.* Probabilistic Engineering Mechanics, 2001. **16**(4): p. 323-330.
14. Wang, X., *On Using Elemental and Non-Elemental Sets to Reproduce the OLS Estimator in Linear Regression.* Unpublished Master Thesis, 2012.
15. Hoerl, A.E. and R.W. Kennard, *A Note on Least-Squares Estimates.* Communications in Statistics Part B-Simulation and Computation, 1980. **9**(3): p. 315-317.
16. Krasker, W.S. and R.E. Welsch, *Efficient Bounded-Influence Regression Estimation.* Journal of the American Statistical Association, 1982. **77**(379): p. 595-604.
17. Duncan, O.D., *A socioeconomic index for all occupations.* In Reiss, A. J., Jr. (Ed.) 1961.
18. Fox, J., *An R and S-PLUS Companion to Applied Regression*. 2002: Sage Publications.