Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Xiaojing Wang Date

On Using Elemental and Non-Elemental Sets to Reproduce the OLS Estimator in Linear Regression

By

Xiaojing Wang

Master of Public Health

Biostatistics

_____

Robert H. Lyles, Ph.D.

Thesis Advisor

On Using Elemental and Non-Elemental Sets to Reproduce the OLS Estimator in Linear Regression

By

Xiaojing Wang
B.S., Nanjing Agricultural University, 2008

Advisor: Robert H. Lyles, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics
2012

Abstract

On Using Elemental and Non-Elemental Sets to Reproduce the OLS Estimator in Linear Regression

By Xiaojing Wang

It has been shown previously that Ordinary Least Squares (OLS) estimates based on a multiple linear regression model with p unknown parameters can be reproduced by combining the results from fitting the same model to all elemental sets, the unique subsets of size p from the total of n observations. In addition, it has been shown that the same goal of reproducing OLS estimates can be achieved by combining the results of the regressions on all unique non-elemental sets, i.e., subsets of size k where $p + 1 \leq k \leq n - 1$. We consider three new methods aimed at reproducing the overall OLS estimates of parameters. The three methods use the direct inverse-variance(INV) weights, the refined inverse-variance (REF) and the constrained optimal (CON) weights applied to each individual OLS estimator based on elemental or non-elemental sets. These methods are compared with the determinant-based weighting method which has previously been proven to reproduce the overall OLS estimates. The primary new insight gained by our study is the notion that the direct inverse-variance weighting essentially achieves the objective, while in theory there may be an infinitely large collection of different weights that can do so. We illustrate the use of the various weighting schemes using simulated data under various linear regression settings, including one-way and two-way ANOVA designs.

On Using Elemental and Non-Elemental Sets to Reproduce the OLS Estimator in Linear Regression

By

Xiaojing Wang

B.S., Nanjing Agricultural University, 2008

Advisor: Robert H. Lyles, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics
2012

Contents

List of Figures

List of Tables

Chapter1

Introduction

1.1 Elemental Sets and Elemental Regressions

In the context of multiple linear regression, an elemental set is a subset of the data containing the minimum number of points such that the unknown parameters in the model can be identified (Smyth & Hawkins (2000)). Assume we have n observations, and p is the number of unknown parameters (typically including an intercept) in the model. Consider the multiple linear regression model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $var(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$. Let $h = \{i_1, i_2, \dots, i_p\}$ be a set of p distinct observations. Define $\boldsymbol{X_h}$ to be the $p \times p$ submatrix consisting of the rows of $\boldsymbol{X}$ indexed by the subset h, and define $\boldsymbol{Y_h}$ to be the corresponding $p \times 1$ subvector of $\boldsymbol{Y}$. Any such set of p observations is an elemental set of the data. The solution to the system of p equations in p unknowns, $\boldsymbol{X_h\widehat{\beta}_h} = \boldsymbol{Y_h}$, is called an elemental regression and is given by

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{h}} = (\boldsymbol{X_h^t X_h})^{-1}\boldsymbol{X_h^t Y_h} = \boldsymbol{X_h^{-1} Y_h}$$

The $\widehat{\boldsymbol{\beta}}_{\boldsymbol{h}}$ can only be computed if $\boldsymbol{X_h}$ is nonsingular; if this is the case for all h, then the data set is said to be in "general position" (Hawkins, 1993).

Elemental set methods are centuries old in their origins. In 1755, Boscovich first considered a regression estimator based on a method of combining all possible elemental regressions, but it was never widely accepted due to its computational infeasibility in all but the smallest datasets (Sheynin, 1973). With the advent of modern computing power, researchers have become interested in the elemental set-based methods again. Theil (1950) and Sen (1968) used the elemental regressions to estimate simple linear regression coefficients. This work was extended to the multiple regression situation by Rubin (1980). Mayo and Gray (1997) introduced a new classification of regression estimators that generalizes a characterization of ordinary least squares (OLS) based on elemental regressions. Estimators in this class are a weighted average of the elemental regressions, where the weights are determined by leverage and residual information associated with the elemental sets. Rousseeuw (1984) and Hawkins, Bradu, and Kass (1984) used elemental sets to handle outlier problems in multiple linear regressions. Elemental sets also have been proposed as a computational device to approximate estimators in the areas of high breakdown regression (Stromberg, 1993) and multivariate location/scale estimation. Hawkins (1993) proposed that elemental set algorithms provide excellent approximations for the least median of squares, least trimmed squares and ordinary least squares criteria.

1.2 Existing Estimators as Functions of Elemental Regressions by OLS

In 1841, Jacobi showed that the least squares estimator is a weighted average of the elemental regressions in the set of data (Sheynin, 1973). If we let h denote an elemental

set, the least squares estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{OLS}}$ can be expressed in the following form:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{OLS}} = \frac{\sum_h |\boldsymbol{X}_h^t \boldsymbol{X}_h| \cdot \widehat{\boldsymbol{\beta}}_h}{\sum_h |\boldsymbol{X}_h^t \boldsymbol{X}_h|} = \frac{\sum_h |\boldsymbol{X}_h^t \boldsymbol{X}_h| \cdot \widehat{\boldsymbol{\beta}}_h}{|\boldsymbol{X}^t \boldsymbol{X}|} = \sum_h \frac{|\boldsymbol{X}_h^t \boldsymbol{X}_h|}{|\boldsymbol{X}^t \boldsymbol{X}|} \cdot \widehat{\boldsymbol{\beta}}_{\boldsymbol{h}} = \sum_h \omega_h \cdot \widehat{\boldsymbol{\beta}}_{\boldsymbol{h}}$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{h}} = (\boldsymbol{X}_{\boldsymbol{h}}^t \boldsymbol{X}_{\boldsymbol{h}})^{-1} \boldsymbol{X}_{\boldsymbol{h}}^t \boldsymbol{Y}_{\boldsymbol{h}} = \boldsymbol{X}_{\boldsymbol{h}}^{-1} \boldsymbol{Y}_{\boldsymbol{h}}$ is the solution based on the elemental set h, $|\boldsymbol{X}_{\boldsymbol{h}}^t \boldsymbol{X}_{\boldsymbol{h}}|$ is the determinant of the $p \times p$ matrix $\boldsymbol{X}_{\boldsymbol{h}}^t \boldsymbol{X}_{\boldsymbol{h}}$, and the sums are taken over all $\binom{n}{p}$ possible elemental sets h. Note that the expression for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{h}}$ assumes that $\boldsymbol{X}_{\boldsymbol{h}}^t \boldsymbol{X}_{\boldsymbol{h}}$ is invertible. Should this not be the case, we would term the h-th set "inadmissible" and leave it out of the weighted average when seeking to reproduce $\hat{\beta}_{OLS}$. Such inadmissible sets are common in ANOVA designs (see Section 3.4).

Hoerl and Kennard (1980) extended the previous result to show that, for any integer value of m such that $p < m \leq n$, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{OLS}}$ can also be expressed as a weighted average of all $\binom{n}{m}$ possible regressions based on m observations within the data. In summary, it is currently known that the OLS estimators can be reproduced using elemental or non-elemental sets by this determinant-based weighting method.

1.3 Outline

In this paper, we propose and illustrate three alternative weighting methods that can be used for the same objective targeted by Hoerl and Kennard (1980); namely, reproducing the OLS estimators using elemental or non-elemental sets. In several special cases, we compare the resulting weighted estimators with OLS and with the previous authors' determinant-based method presented in Section 1.2. The primary new insights gained

are the understanding that the direct inverse variance-based weights meet the desired objective and that in theory an infinite number of weighting schemes can be used to reproduce the OLS estimator.

Among the special cases considered for illustrations are typical simple and multiple linear regression settings, as well as one-way ANOVA and two-way ANOVA scenarios. In the latter case, we introduce the notion that only 'admissible' elemental or non-elemental sets must be included in the weighting process. This notion relates to that of 'general position' (Hawkins, 1993).

In this thesis, we are primarily interested in the methods of reproducing OLS estimators using elemental and non-elemental sets as an instructional tool, rather than for practical purposes. However, it has already been noted (Section 1.1) that techniques based on elemental sets have practical applications in the direction of "robust" regression. In particular, Ordinary Least Squares (OLS) estimates the model parameters by minimizing $\sum_{i=1}^{n} \hat{e}_i^2$, where $\hat{e}_i = y_i - x_i\beta$ is the fitting error. This criterion is sensitive to outliers. OLS estimation also suffers from the problem of masking, which occurs when a data set contains multiple outliers and, at the same time, these outliers are not detected by the usual LS diagnostic procedures (Agullo, 1997). To solve these problems, robust regression methods attempt to make the estimators less sensitive to outliers or influential observations. The notion of reproducing the OLS estimators using elemental or non-elemental sets via the direct inverse-variance weighting explored here might be useful toward the development of new ideas for robust estimation of regression

coefficients, as explored by Jin (2012, unpublished thesis). But firstly, we need to verify that our proposed the direct inverse variance-based weighting methods could indeed reproduce the OLS estimators using elemental and non-elemental sets.

Chapter 2

Methods

2.1 Methods to Reproduce the OLS Estimator in Linear Regression

To review, we assume a multiple linear regression model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors with $E(\boldsymbol{\varepsilon}) = 0$ and $var(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$. Here, n is the number of observations and p is the number of unknown parameters. Let $J = \{i_1, i_2, \dots, i_k\}$ be a set of k distinct observations for any k such that $p \leq k \leq n - 1$. Define $\boldsymbol{X_J}$ to be the $k \times p$ submatrix consisting of the rows of $\boldsymbol{X}$ indexed by the subset J, and define $\boldsymbol{Y_J}$ to be the corresponding $k \times 1$ subvector of $\boldsymbol{Y}$.

An interesting feature of the determinant-based weights is the fact that the weights used to combine elemental or non-elemental sets to reproduce OLS are identical for every regression coefficient in the model. In contrast, the three alternative approaches explored below produce different sets of weights corresponding to each individual OLS coefficient estimate that we try to reproduce.

2.1.1 The Direct Inverse-Variance Weighted Average Estimator

A natural approach to combining the k elemental or non-elemental set-based OLS estimates for a given regression parameter is to simply take the inverse-variance

weighted average of the OLS estimates from each of the $\binom{n}{k}$ sets. The resulting estimator can be expressed in the following form:

$$\hat{\beta}_i = \frac{\sum_J Var(\hat{\beta}_{iJ})^{-1} \cdot \hat{\beta}_{iJ}}{\sum_J Var(\hat{\beta}_{iJ})^{-1}}$$

where $i = 1, 2, \dots, p$, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{J}} = \left(\boldsymbol{X}_{\boldsymbol{J}}^t \boldsymbol{X}_{\boldsymbol{J}} \right)^{-1} \boldsymbol{X}_{\boldsymbol{J}}^t \boldsymbol{Y}_{\boldsymbol{J}} = \boldsymbol{X}_{\boldsymbol{J}}^{-1} \boldsymbol{Y}_{\boldsymbol{J}}$ is the solution based on the subset J and the sums are taken over all $\binom{n}{k}$ subsets J. $\hat{\beta}_{iJ}$ is the ith element of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{J}}$.

In Section 3, we demonstrate that the direct inverse variance weighted average does not exactly reproduce the overall OLS estimate in general, except in the special case of the slope estimate in simple linear regression. However, as we will see, the approximation to overall OLS based on this approach is quite close, and tends to get closer as n increases.

2.1.2 The Refined Inverse-Variance Weighted Average Estimator

The failure of the direct inverse-variance weights to exactly reproduce OLS in general is likely due to the fact that they ignore the covariances among the set-based OLS estimates that are being combined. Taking advantage of ideas based on generalized least squares theory (Arnold, 1981), we propose a refined version of the weights applied to the individual set-based OLS estimators. In this method, we take into account the covariances of these estimators in addition to their variances.

Let

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{*} = \begin{pmatrix} \hat{\beta}_{i1} \\ \hat{\beta}_{i2} \\ \vdots \\ \hat{\beta}_{iz} \end{pmatrix}$$

where $i = 1, 2, \ldots, p$, z = $\binom{n}{k}$ is the total number of possible subsets of size k. In theory (e.g. Weller et al., 2006), the optimal weights can be expressed in the following form:

$$\boldsymbol{\tau}_{\boldsymbol{\beta}_i} = \left[\mathbf{1}^t Var^{-1}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{*}\right)\mathbf{1}\right]^{-1}\mathbf{1}^t Var^{-1}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{*}\right)$$

where **1** is the z × 1 matrix with all its entries being 1.

The variance-covariance matrix of the z correlated estimates, $Var\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{*}\right)$, can be obtained by creating carefully chosen matrices that we will label as **A**, **B** and **C**. Let us take the simple linear regression model, $Y = \beta_0 + \beta_1 X + \varepsilon$ with n=3 and k=2, as an example. Denote the vector of responses as $\boldsymbol{Y}' = (y_1, y_2, y_3)$. All possible dependent variable subsets of size k are $\boldsymbol{Y_1'} = (y_1, y_2)$, $\boldsymbol{Y_2'} = (y_1, y_3)$ and $\boldsymbol{Y_3'} = (y_2, y_3)$. First we define a matrix **A** such that **AY**= $(\boldsymbol{Y_1}, \boldsymbol{Y_2}, \boldsymbol{Y_3})'$ is a column vector containing the dependent variable subset blocks . In this case the matrix **A** is (6×3 here) is as follows:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Next, define the matrices $\mathbf{B_j}$ (j=1,2,3), where $\mathbf{B_j} = (\boldsymbol{X_j}'\boldsymbol{X_j})^{-1}\boldsymbol{X_j}'$ and $\boldsymbol{X_j}$ is the design matrix for the jth subset. We then define a block diagonal matrix with the $\mathbf{B_j}$'s down the diagonal, in this case such that

$$\mathbf{B} = \begin{pmatrix} \mathbf{B_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B_3} \end{pmatrix}$$

For this example, **B** is a 6×6 matrix since each of the $\mathbf{B_j}$'s is 2×2. Now the matrix **BAY** will return the vector containing all of the individual OLS $\hat{\beta}_0$ and $\hat{\beta}_1$ estimates from the 3 separate regressions.

Finally, in order to sort the latter vector so as to group all of the $\hat{\beta}_0$'s together followed by the $\hat{\beta}_1$'s, we create matrix **C** as follows:

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Thus, we can obtain the vector $\widehat{\boldsymbol{\beta}}^* = \boldsymbol{CBAY}$ that contains the block of 3 individual $\hat{\beta}_0$'s followed by the block of 3 individual $\hat{\beta}_1$'s. Then the desired variance-covariance matrix $Var(\widehat{\boldsymbol{\beta}}^*) = Var(\boldsymbol{CBAY}) = \sigma^2\, \boldsymbol{CBAA'B'C'}$. We can then take the 2 separate (3×3) blocks along the diagonal of that matrix to give us the variance-covariance matrices $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^*)$, corresponding to the 3 correlated $\hat{\beta}_0$'s and the 3 correlated $\hat{\beta}_1$'s.

The above example is a simple special case, but the same process can be applied more generally (in theory, for arbitrary n, p, and k). Note that in general, the matrices **Y**, $\mathbf{Y_j}$, $\mathbf{X_j}$, $\mathbf{B_j}$, **A**, **B**, and **C** defined above are of dimensions (n×1), [(n−1)×1], [(n−1)×p], [p×(n−1)], [n(n−1)×n], [np×n(n−1)], and (np×np), respectively. This general technique was applied

to all the linear regression examples in this paper to get the variance-covariance matrix for each set of $\hat{\beta}_i$'s.

One problem with directly using the generalized least squares-based weights given previously is the fact that the variance-covariance matrices $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{*})$ are singular. As a result, we use generalized inverses in the above formula to calculate the optimal weights. The generalized inverse of a matrix is a matrix that shares some properties of the inverse matrix (Rao, 1971). Typically, the generalized inverse exists for an arbitrary matrix. When a matrix has an inverse, then its inverse and the generalized inverse are the same.

Having substituted a generalized inverse for $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{*})$, we offer the following observation: The overall OLS estimator for the ith regression coefficient is given by

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}} = \boldsymbol{\tau}_{\boldsymbol{\beta}_{\boldsymbol{i}}} \cdot \widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{*}$$

Although we offer the above observation without a formal proof, the result is demonstrated in several linear regression settings in Section 3. An important consequence of this result is the apparent fact that OLS can actually be reproduced by an infinite number of different weight combinations as applied to elemental or non-elemental sets (due to the non-uniqueness property of the generalized inverse). Note that if k=p (so that elemental sets are being used), it is not possible to estimate σ^2 in $Var(\widehat{\boldsymbol{\beta}}^{*}) = Var(\boldsymbol{CBAY})$ as outlined above. However, σ^2 factors out in the inverse variance-based weights and can thus be ignored.

2.1.3 The Constrained Optimal Weighted Average Estimator

Although it is not readily generalizable, a third approach to the weighting of set-based OLS estimates is to derive optimal weights that target the minimal variance for the resulting overall estimator. Due to complicated algebraic work, we only consider the simple linear regression with n=3, p=2 as an example. Then there are 3 possible subsets of observations. Let w_{i1}, w_{i2}, w_{i3} be weights for $\hat{\beta}_{i1}$, $\hat{\beta}_{i2}$, $\hat{\beta}_{i3}$ respectively, where we observe the constraint $w_{i3} = (1 - w_{i1} - w_{i2})$, where $i = 1,2$ since p=2. The weighted estimator for the ith coefficient is then

$$\hat{\beta}_i = w_{i1}\hat{\beta}_{i1} + w_{i2}\hat{\beta}_{i2} + (1 - w_{i1} - w_{i2})\hat{\beta}_{i3}$$

Suppressing the subscript i for simplicity, we use basic properties of linear combinations of random variables to determine that

$$Var(\hat{\beta}) = w_1^2 Var(\hat{\beta}_1) + w_2^2 Var(\hat{\beta}_2) + w_3^2 Var(\hat{\beta}_3) + 2w_1 w_2 \sigma_{12} + 2w_1 w_3 \sigma_{13} + 2w_2 w_3 \sigma_{23},$$

where the σ's represent covariances between the corresponding pairs of set-based estimates. Now replacing w_3 with $(1 - w_1 - w_2)$, we have

$$Var(\hat{\beta}) = w_1^2 Var(\hat{\beta}_1) + w_2^2 Var(\hat{\beta}_2) + (1 - w_1 - w_2)^2 Var(\hat{\beta}_3) + 2w_1 w_2 \sigma_{12} + 2w_1 (1 - w_1 - w_2)\sigma_{13} + 2w_2 (1 - w_1 - w_2)\sigma_{23}$$

To find the weights that minimize the variance, we take $\frac{\partial Var(\hat{\beta})}{\partial w_1} \overset{set}{=} 0$ and $\frac{\partial Var(\hat{\beta})}{\partial w_2} \overset{set}{=} 0$, obtaining

$w_1 = \frac{Var(\hat{\beta}_3) - \sigma_{13} - \{Var(\hat{\beta}_3) + \sigma_{12} - \sigma_{13} - \sigma_{23}\}w_2}{Var(\hat{\beta}_1) + Var(\hat{\beta}_3) - 2\sigma_{13}}$ and

$w_2 =$

$$\frac{\{Var(\hat{\beta}_3) + \sigma_{12} - \sigma_{13} - \sigma_{23}\}\{\sigma_{13} - Var(\hat{\beta}_3)\} - \{Var(\hat{\beta}_1) + Var(\hat{\beta}_3) - 2\sigma_{12}\}\{\sigma_{23} - Var(\hat{\beta}_3)\}}{\{Var(\hat{\beta}_2) + Var(\hat{\beta}_3) - 2\sigma_{23}\}\{Var(\hat{\beta}_1) + Var(\hat{\beta}_3) - 2\sigma_{13}\} - \{Var(\hat{\beta}_3) + \sigma_{12} - \sigma_{13} - \sigma_{23}\}\{Var(\hat{\beta}_3) + \sigma_{12} - \sigma_{13} - \sigma_{23}\}}$$

In this example, the required variances and covariances are very simple algebraically. In Section 3, we verify that this third approach (like the constrained optimal estimator in Section 2.1.3) directly reproduces OLS for the special case considered here.

Chapter 3

Simulation Studies

Simulation experiments were conducted to compare the performance of the estimates based on the proposed methods. Several special cases were listed to illustrate the methods we used to reproduce the OLS estimators. Uses of the weights based on four methods were compared. These methods utilize the previously published determinant-based weights (DET), along with the proposed direct inverse-variance weights (INV), refined inverse-variance weights (REF), and the constrained optimal weights (CON) (the latter are illustrated only for the special case of n=3, k=2).

3.1 Special Case 1: Simple linear regression (SLR) with n=3, p=2

The model of simple linear regression is $Y = \beta_0 + \beta_1 X + \varepsilon$. From Table 3.1, we could see that the REF method, the DET method, and the CON method reproduce the OLS estimators exactly using elemental sets for a situation with n=3, k=2, and p=2. Note that the intercept estimator $\hat{\beta}_0$ based on the INV method is a little off relative to the overall OLS estimator, while the slope estimator $\hat{\beta}_1$ using the INV method reproduces the overall OLS estimator exactly (see Section 2.1.1).

Figure 3.1 and Figure 3.2 display the weights based on each of the four methods for estimating $\hat{\beta}_o$ and $\hat{\beta}_1$. The determinant (DET) weights for estimating $\hat{\beta}_o$ are the same as the DET weights for estimating $\hat{\beta}_1$, but the other methods' weights for estimating $\hat{\beta}_o$ are

different from those for estimating $\hat{\beta}_1$, as expected. The CON method weights are vastly different from the weights of the other methods, while the weights based on the other 3 methods are similar. However, though it is difficult to tell from Figures 3.1.1 and 3.1.2, the sets of weights used are different across all of the methods.

Table 3.1 Simulation results of reproducing SLR coefficients based on 500 simulated datasets (n=3, k=2, p=2)

Parameters	**OLS Coefficient**		**REF**		**DET**		**INV**		**CON**	
	Mean	SD	Mean Est.	SD Est.	Mean Est.	SD Est.	Mean Est.	SD Est.	Mean Est.	SD Est.
$\hat{\beta}_0$	1.0316	0.5257	1.0316	0.5257	1.0316	0.5257	1.0306	0.5405	1.0316	0.5257
$\hat{\beta}_1$	1.9914	0.2481	1.9914	0.2481	1.9914	0.2481	1.9914	0.2481	1.9914	0.2481

Figure 3.1.1 The weights of four methods for estimating $\hat{\beta}_0$ of SLR

(n=3, k=2, p=2)

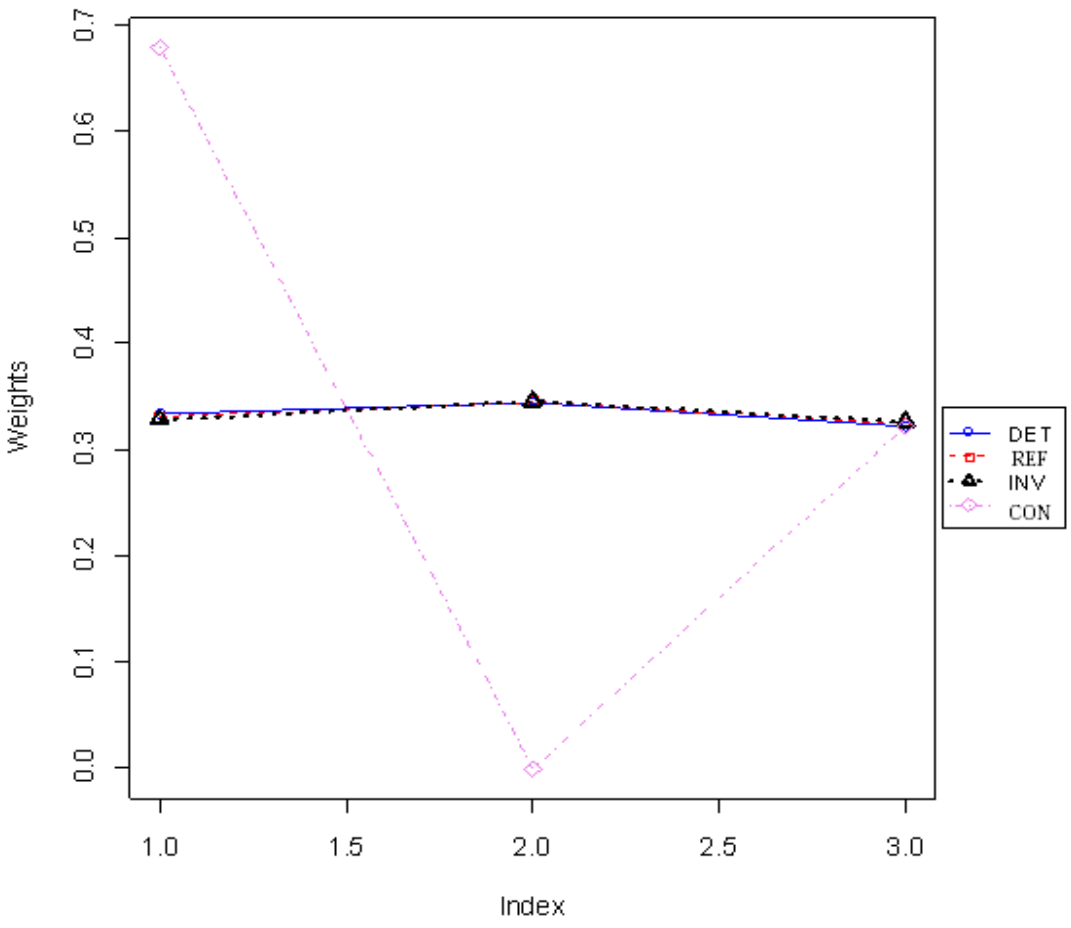
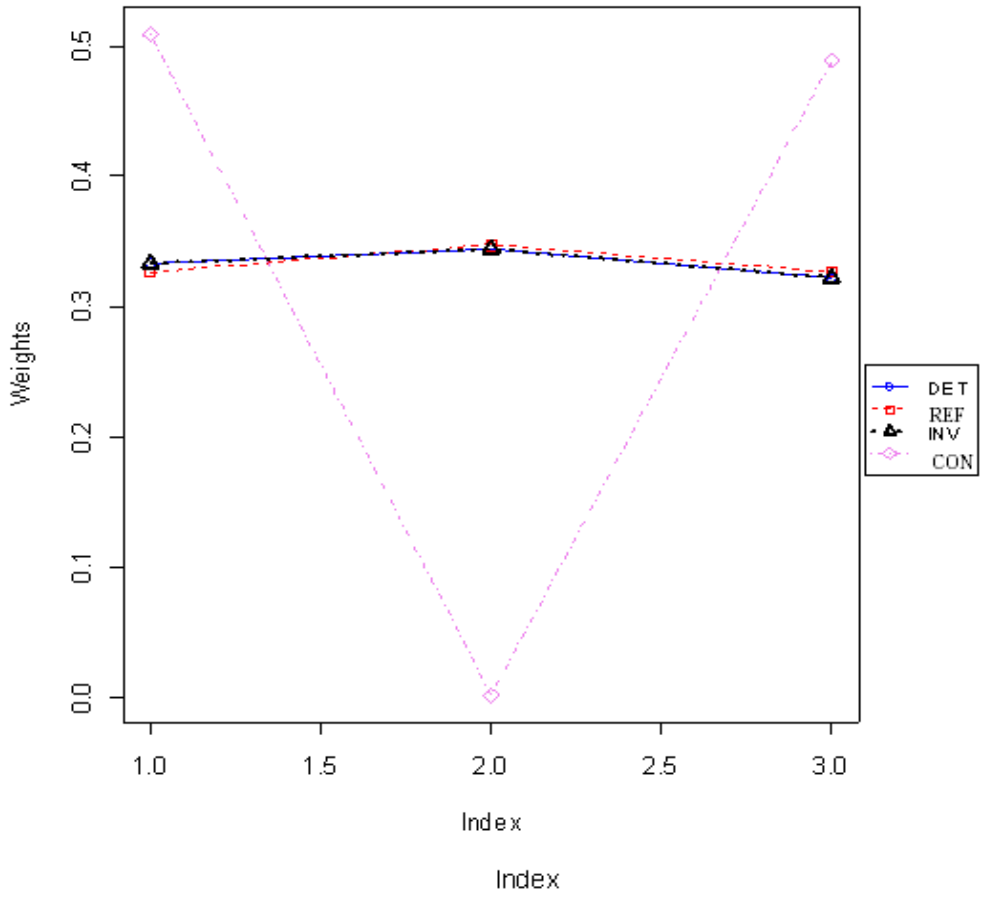


Figure3.1.2 The weights of four methods for estimating $\hat{\beta}_1$ of SLR (n=3, k=2, p=2)



3.2 Special Case 2: Multiple linear regression (MLR) with n=7, p=4

Consider the multiple linear regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$, where the total number of observations (n) is 7. So here the possible sizes of the subsets could be k where $4 \leq k \leq 6$. We provide the results based on weighting all possible subsets in Sections 3.2.1, 3.2.2, and 3.2.3.

The results demonstrate that the overall OLS estimates for a multiple linear regression can indeed be reproduced by weighting the elemental sets with k=4, or by weighting non-elemental sets with k=5 or k=6. In terms of the weighting methods, the coefficients of the linear regression can be exactly reproduced by the REF method and by the DET method using elemental and non-elemental sets. Note that the estimates produced by the direct inverse-variance weights method (INV) are not exactly the same as the OLS estimates, but are very close.

The DET weights for estimating each unknown parameter of a multiple linear regression are the same, but the REF weights and the INV weights are different for estimating each individual regression coefficient parameter. All three methods use different weights, yet all effectively reproduce the OLS estimators using elemental and non-elemental sets.

3.2.1 Reproducing OLS coefficients via the elemental sets with n=7, k=4, p=4

Table3.2.1 Simulation results of reproducing MLR coefficients based on 500 simulated datasets (n=7, k=4, p=4)

Parameters	**OLS Coefficient**		**REF**		**DET**		**INV**	
	Mean	SD	Mean Est.	SD Est.	Mean Est.	SD Est.	Mean Est.	SD Est.
$\hat{\beta}_0$	1.0015	0.3195	1.0015	0.3195	1.0015	0.3195	1.0036	0.3241
$\hat{\beta}_1$	2.3834	1.6065	2.3834	1.6065	2.3834	1.6065	2.3727	1.6165
$\hat{\beta}_2$	1.2890	1.3933	1.2890	1.3933	1.2890	1.3933	1.2993	1.4140
$\hat{\beta}_3$	3.0916	0.2692	3.0916	0.2692	3.0916	0.2692	3.0873	0.2714

Figure 3.2.1.1 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_0$ of MLR (n=7, k=4, p=4)

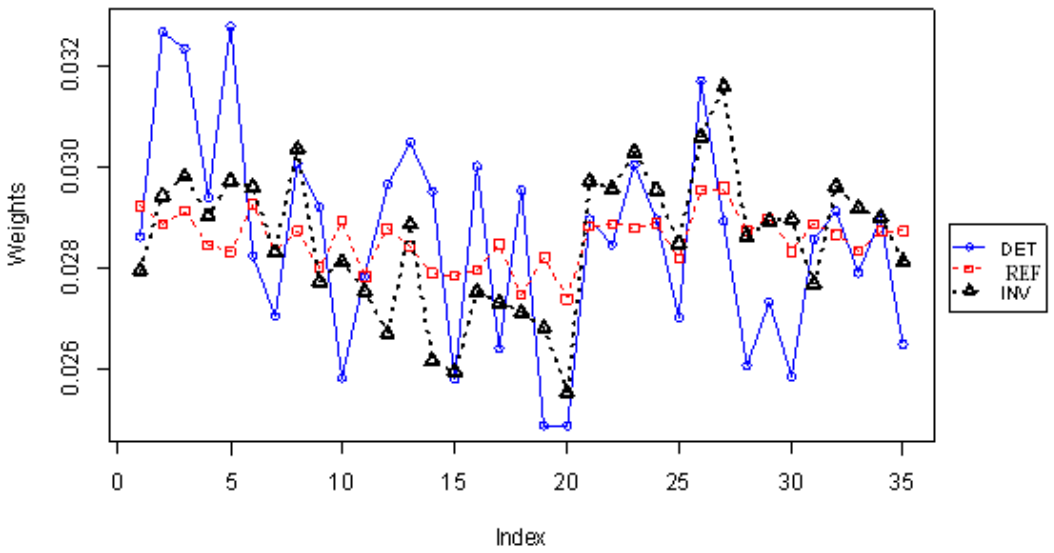


Figure 3.2.1.2 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_1$ of MLR (n=7, k=4, p=4)

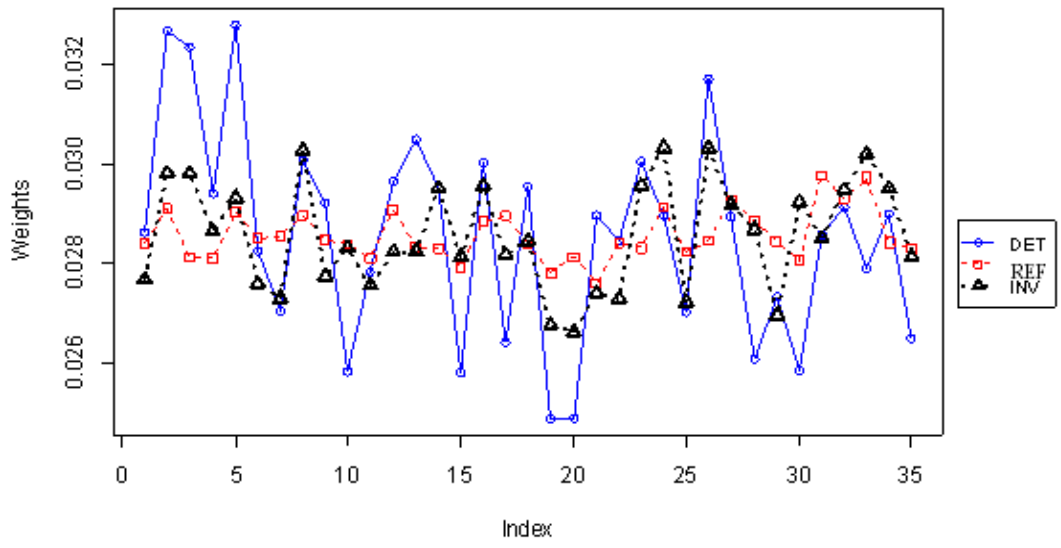


Figure3.2.1.3 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_2$ of MLR (n=7, k=4, p=4)

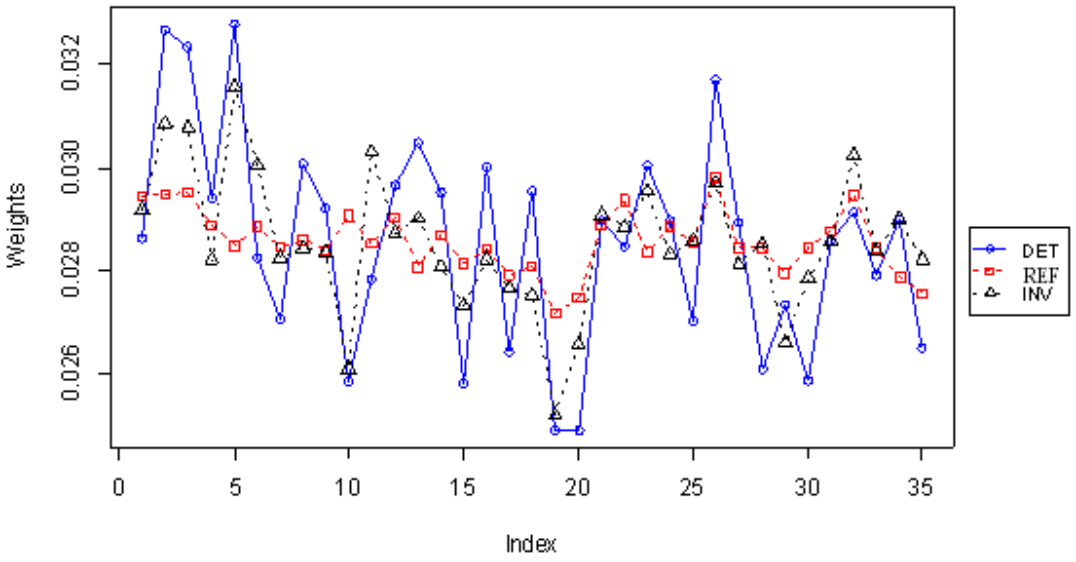
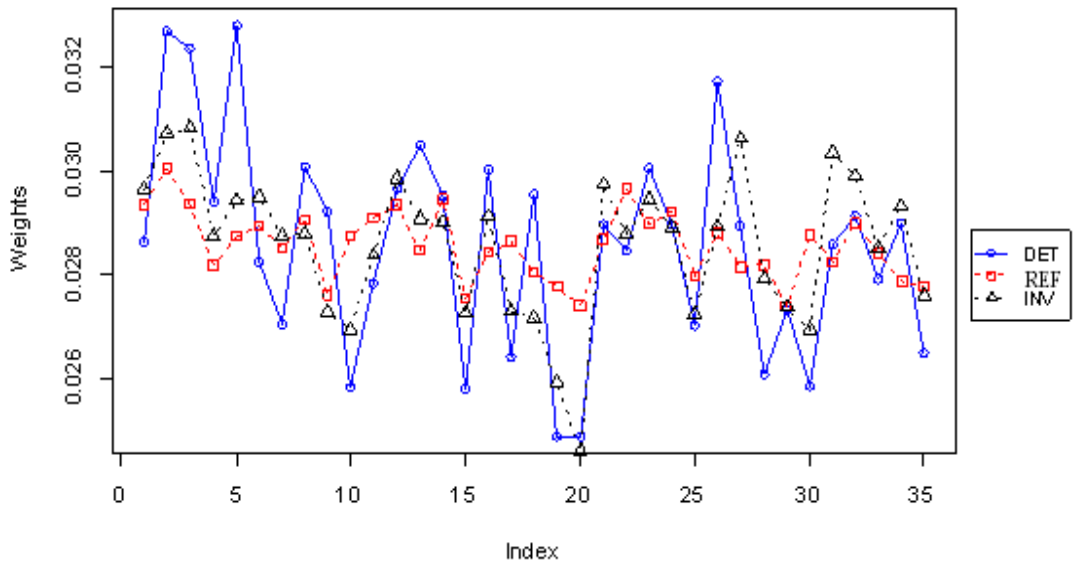


Figure3.2.1.4 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_3$ of MLR (n=7, k=4, p=4)



3.2.2 Reproducing OLS coefficients via the non-elemental sets with n=7, k=5, p=4

Table3.2.2 Simulation results of reproducing MLR coefficients based on 500 simulated datasets (n=7, k=5, p=4)

Parameters	Coefficient		REF		DET		INV	
	Mean	SD	Mean Est.	SD Est.	Mean Est.	SD Est.	Mean Est.	SD Est.
$\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$	1.0073	0.2682	1.0073	0.2682	1.0073	0.2682	1.0056	0.2729
$\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$	2.2381	1.3833	2.2381	1.3833	2.2381	1.3833	2.2276	1.3913
$\widehat{\boldsymbol{\beta}}_{\mathbf{2}}$	1.2967	1.1452	1.2967	1.1452	1.2967	1.1452	1.2986	1.1404
$\widehat{\boldsymbol{\beta}}_{\mathbf{3}}$	3.1088	0.2259	3.1088	0.2259	3.1088	0.2259	3.1080	0.2265

Figure3.2.2.1 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_0$ of MLR (n=7, k=5, p=4)

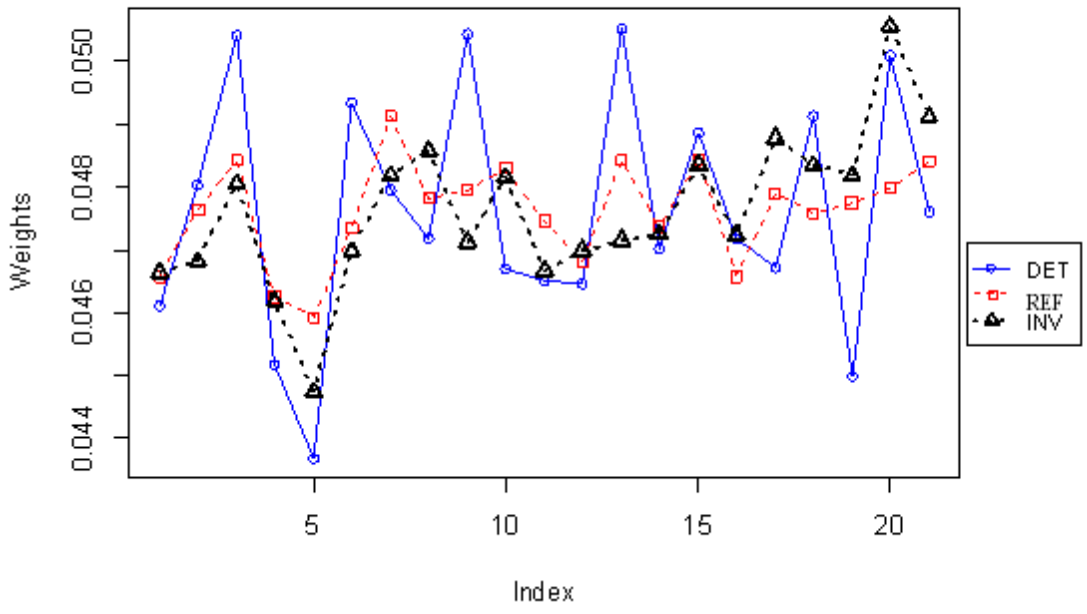


Figure3.2.2.2 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_1$ of MLR (n=7, k=5, p=4)

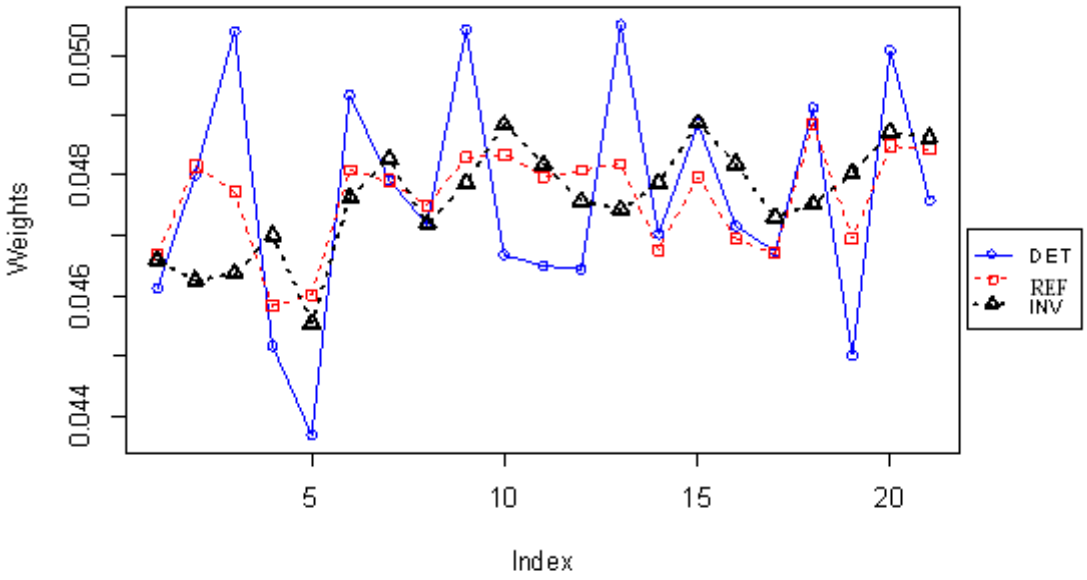


Figure3.2.2.3 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_2$ of MLR (n=7, k=5, p=4)

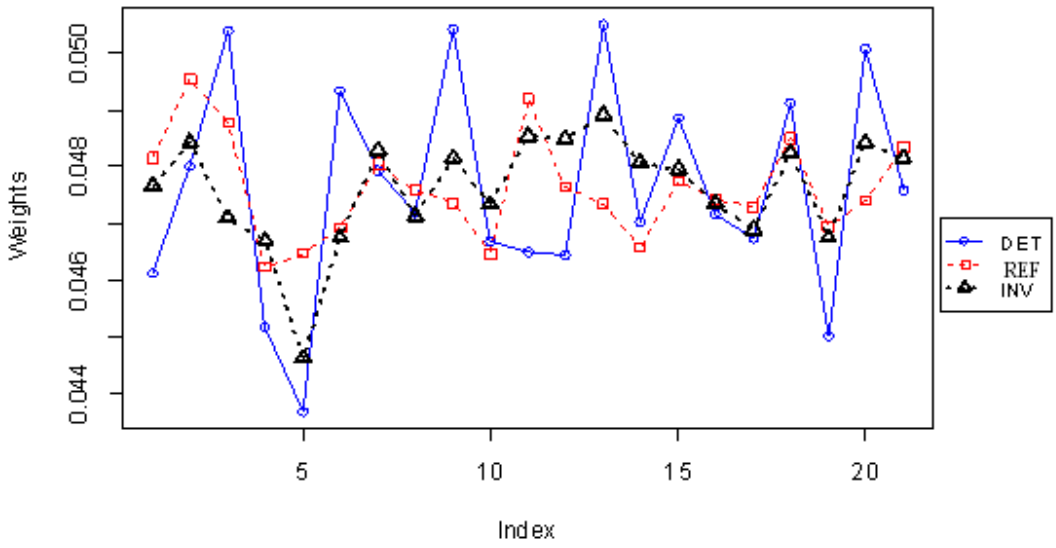
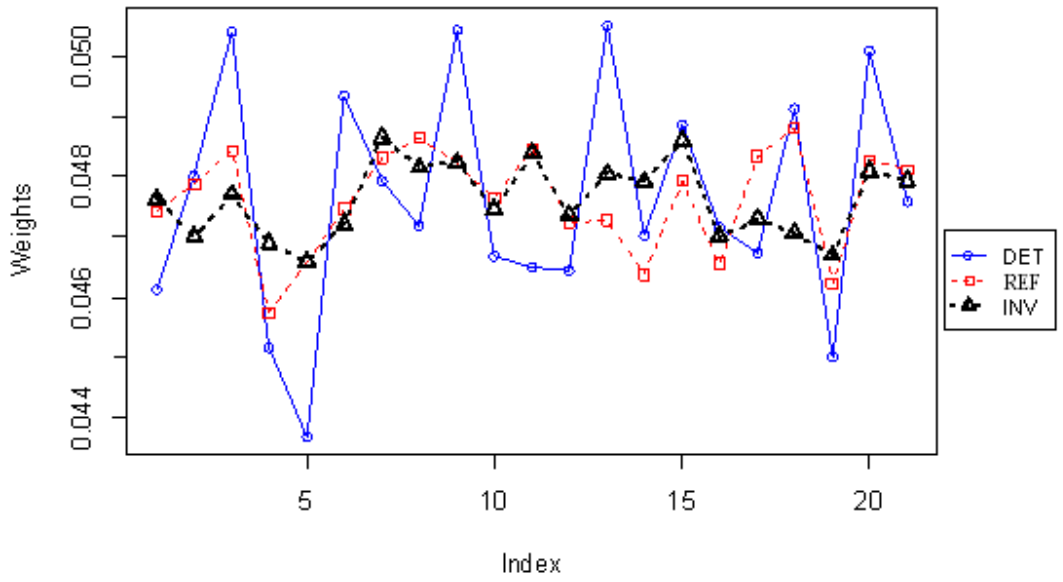


Figure3.2.2.4 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_3$ of MLR (n=7, k=5, p=4)



3.2.3 Reproduce the coefficients by the non-elemental sets with n=7, k=6, p=4

Table3.2.3 Simulation results of reproducing MLR coefficients based on 500 simulated datasets (n=7, k=6, p=4)

Parameters	Coefficient		REF		DET		INV	
	Mean	SD	Mean Est.	SD Est.	Mean Est.	SD Est.	Mean Est.	SD Est.
$\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$	1.0339	0.3091	1.0339	0.3091	1.0339	0.3091	1.0338	0.3090
$\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$	2.4225	1.3879	2.4225	1.3879	2.4225	1.3879	2.4212	1.4009
$\widehat{\boldsymbol{\beta}}_{\mathbf{2}}$	1.1744	1.2351	1.1744	1.2351	1.1744	1.2351	1.1598	1.2812
$\widehat{\boldsymbol{\beta}}_{\mathbf{3}}$	3.1143	0.2385	3.1143	0.2385	3.1143	0.2385	3.1157	0.2451

Figure3.2.3.1 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_0$ of MLR (n=7, k=6, p=4)

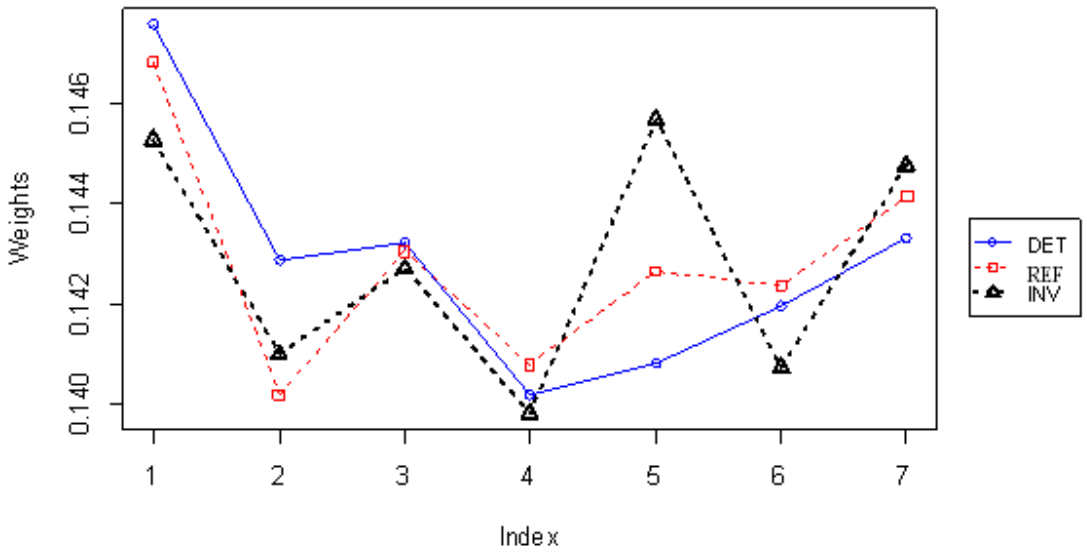


Figure3.2.3.2 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_1$ of MLR (n=7, k=6, p=4)

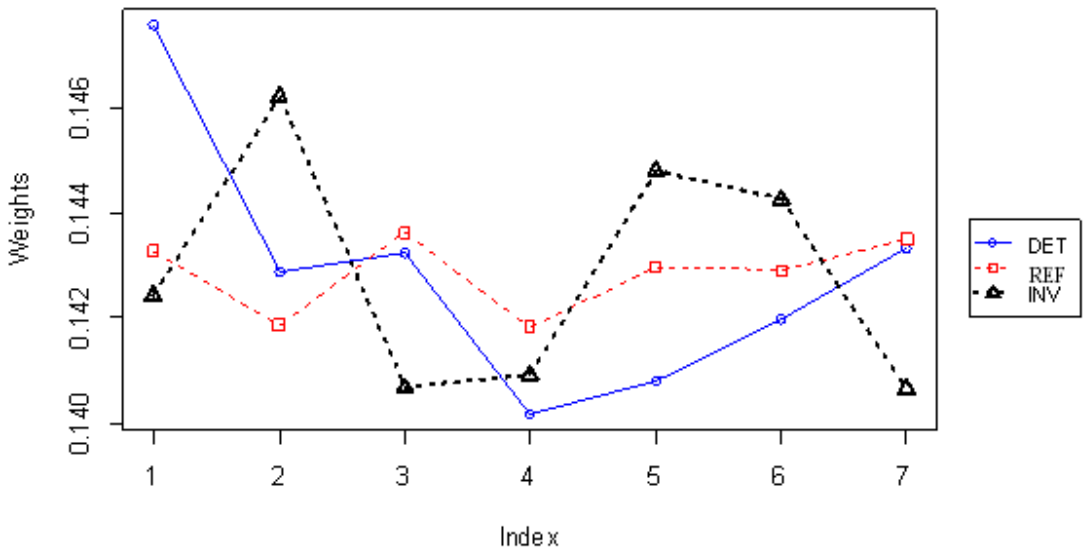


Figure3.2.3.3 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_2$ of MLR (n=7, k=6, p=4)

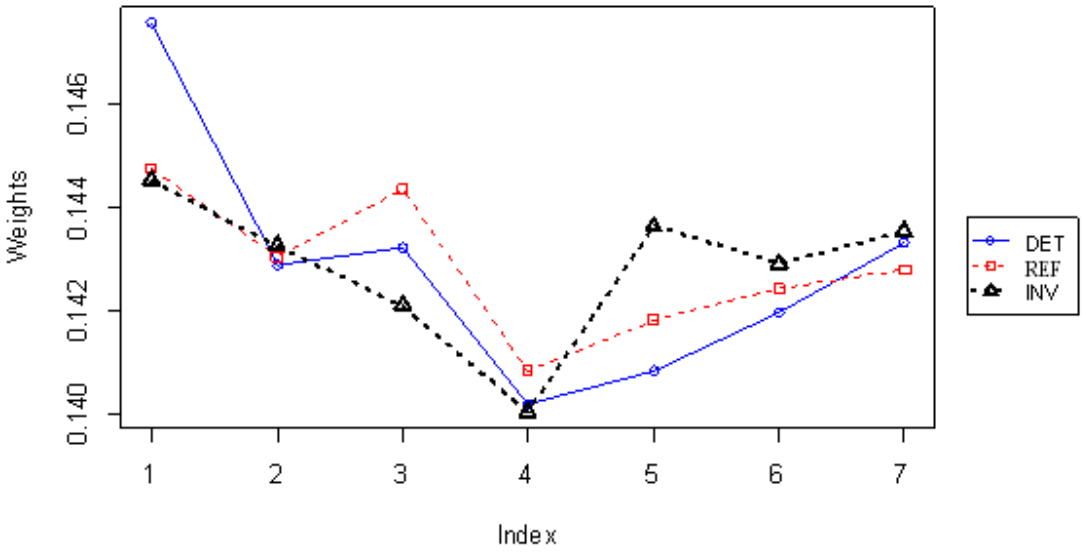
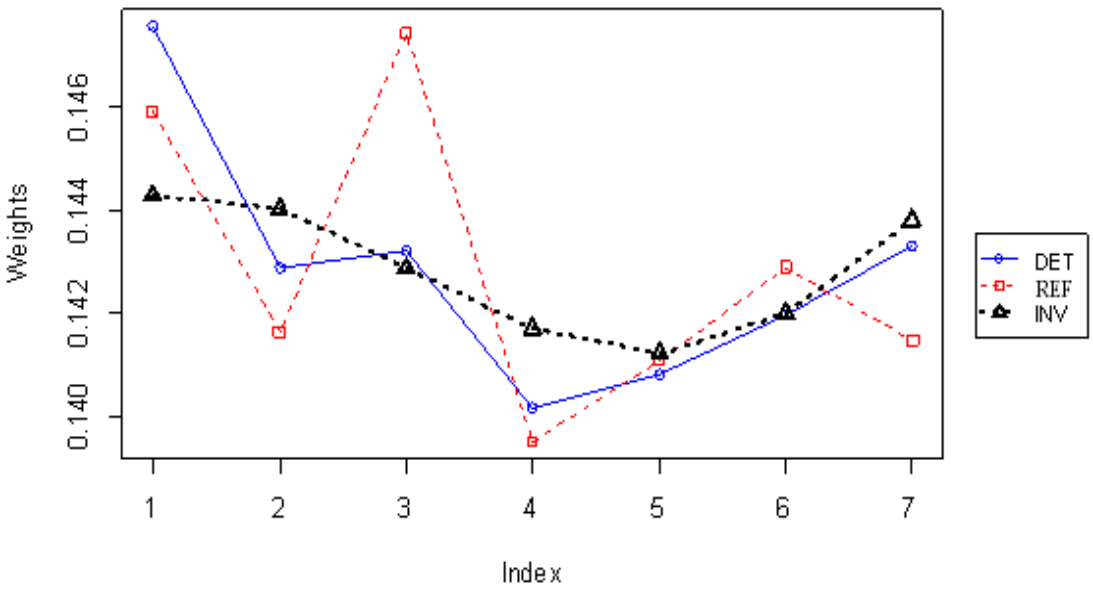


Figure3.2.3.4 The mean weights of three methods based on 500 simulated datasets for estimating $\hat{\beta}_3$ of MLR (n=7, k=6, p=4)



3.3 Evaluation of direct INV method as n increases

Suppose the multiple linear regression model is still $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$. Table 3.3 shows that the estimates produced by the direct inverse-variance (INV) weights method with k=n−1 are extremely close to the OLS estimates for a range of sample sizes (n). The refined inverse-variance (REF) weights method and the determinant (DET) weights method continue to reproduce exactly the OLS estimates in all conditions studied.

The potential significance of these results for the INV approach could be in the area of robust regression (see Section 1.3). In particular, some prior robust regression techniques utilize weighting of elemental sets with weights a function of leverage information corresponding to the observations in each set (e.g., Mayo and Gray, 1997). Our finding that direct inverse-variance weighting essentially reproduces OLS suggests that INV weights could be useful as one component in such a robust regression effort (for example, combined with other weights designed to provide robustness to outliers). The direct INV weights are more convenient that the REF or DET weights in this regard, and are utilized in this direction by Jin (2012, unpublished MS thesis). As discussed by Jin, the applicability of the INV weights to both elemental and non-elemental sets also opens up potentially new avenues for similar weighting approaches in robust regression.

Table 3.3 Simulation results for reproducing MLR coefficients based on 500 simulated datasets

(n=6, 50, 100; k=n−1; p=4)

No. of Observations (n)	**No. of Observations in Each Subset (k)**	**Parameters**	**OLS Coefficient**		**REF**		**DET**		**INV**	
			Mean	SD	Mean Est.	SD Est.	Mean Est.	SD Est.	Mean Est.	SD Est.
n=6	k=5	$\hat{\beta}_0$	1.000218	0.372262	1.000218	0.372262	1.000218	0.372262	0.999376	0.373623
		$\hat{\beta}_1$	2.236744	1.774762	2.236744	1.774762	2.236744	1.774762	2.226723	1.761088
		$\hat{\beta}_2$	1.322763	1.365702	1.322763	1.365702	1.322763	1.365702	1.324691	1.369867
		$\hat{\beta}_3$	3.10291	0.290713	3.10291	0.290713	3.10291	0.290713	3.102472	0.286573
n=50	k=49	$\hat{\beta}_0$	0.997059	0.068846	0.997059	0.068846	0.997059	0.068846	0.997057	0.068842
		$\hat{\beta}_1$	2.293481	0.322723	2.293481	0.322723	2.293481	0.322723	2.293473	0.32272
		$\hat{\beta}_2$	1.307413	0.248415	1.307413	0.248415	1.307413	0.248415	1.307414	0.248407
		$\hat{\beta}_3$	3.098996	0.048179	3.098996	0.048179	3.098996	0.048179	3.098999	0.048179
n=100	k=99	$\hat{\beta}_0$	1.001215	0.051865	1.001215	0.051865	1.001215	0.051865	1.001214	0.051866
		$\hat{\beta}_1$	2.306166	0.201351	2.306166	0.201351	2.306166	0.201351	2.306166	0.20135
		$\hat{\beta}_2$	1.298516	0.166794	1.298516	0.166794	1.298516	0.166794	1.298516	0.166792
		$\hat{\beta}_3$	3.099694	0.034296	3.099694	0.034296	3.099694	0.034296	3.099694	0.034295

3.4 Special Case 3: Reproducing OLS coefficients for one-way ANOVA

One issue that has not yet been addressed is the possibility that certain sets (elemental or non-elemental) may be "inadmissible" for weighting because the MLR model based on those sets is less than full rank. ANOVA designs provide obvious examples in which such an issue will arise, since many subsets may contain observations with identical predictor (X) values. Intuitively, we should not include in the weighting any subsets for which the OLS estimate for a particular coefficient of interest is not unique.

Assume the following MLR model corresponding to one-way ANOVA:

$$Y_{ij} = \mu + \sum_{i=1}^{r-1} \alpha_i X_i + \varepsilon_{ij} \; (i = 1,2,\dots,r-1; j = 1,2,\dots,n_i) \quad ,$$

where μ represents an overall mean, α_i corresponds to the "effect" of group i, and ε_{ij} is a random error term corresponding to the jth observation on the ith subject. Typically we assume the ε_{ij} 's are independent and identically distributed as Normal(0, σ^2), although normality is not required for OLS estimation. In one-way ANOVA, the criterion for an "admissible" subset is not only that it represents a set of k distinct observations for any k such that $p \leq k \leq n-1$. Besides this requirement, there must be at least one observation in each subset from all levels (i) of the group variable.

From the formulae for the three methods, we can see that they only involve the X's. Each subset regression should be weighted the same since the X_i's are dummy variables (0 or 1). All 3 methods (INV, REF, and DET) are found to reduce to using the same simple set of weights, namely, 1/(# of admissible sets) for each subset. So the $\hat{\mu}$ and $\hat{\alpha}$'s

simplify down to the unweighted average of the estimators from all possible subset regressions. Using the estimates from all subsets that meet these two requirements, one can thus reproduce the estimates for the one-way ANOVA via the simple average of the subset-specific OLS estimates.

For example, suppose there are two groups and 3 observations in each group, and the data are displayed in the following table:

X	
0	1
y=2 y=3 y=5	y=6 y=7 y=4

Then, although there are a total of $\binom{6}{2}$ possible elemental sets, the only "admissible" sets are [(0,2),(1,6)], [(0,2),(1,7)], [(0,2),(1,4)], [(0,3),(1,6)], [(0,3),(1,7)], [(0,3),(1,4)], [(0,5),(1,6)], [(0,5),(1,7)], [(0,5),(1,4)]. The weights of three methods are the same. Specifically, the weight is 1/9 for each elemental subset in this example.

3.4.1 Reproducing the coefficients of the balanced one-way ANOVA

As a simple example, suppose there are 3 groups (A, B and C) and 2 observations in each group. An appropriate MLR model for this case is $Y = \mu + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$. Here group A was set as the reference group. So the total number of observations is $n = 6$ and the number of unknown mean parameters is $p = 3$. We can consider subsets containing k observations, where $3 \leq k \leq 5$. For a subset to be admissible, there must

be at least one observation in each group. The admissible subsets can be combined to reproduce the overall one-way ANOVA coefficients.

All possible situations characterizing possible subsets that are elemental or non-elemental are listed in Table 3.4.1. The results show that the coefficients for one-way ANOVA with a balanced dataset can be reproduced exactly based on the elemental or non-elemental sets. Also the weight for each subset is the same for each of the three weighting methods (INV, REF, and DET), and that common weight is 1/(# of admissible sets).

Table 3.4.1 Simulation results of reproducing one-way ANOVA coefficients based on 500 simulated balanced datasets

k	No. of Obs. taken in each group	No. of Admiss-ible sets	Paramet-ers	OLS Coefficients		Estimates by averaging all subset estimators	
				Mean	SD	Mean Est.	SD Est.
3	A=1 B=1 C=1	8	$\hat{\mu}$ $\hat{\alpha}_1$ $\hat{\alpha}_2$	0.05207 -0.04236 2.95248	0.73444 1.59258 0.84194	0.05207 -0.04236 2.95248	0.73444 1.59258 0.84194
4	A=2 B=1 C=1	4	$\hat{\mu}$ $\hat{\alpha}_1$ $\hat{\alpha}_2$	0.07162 -0.04622 2.91471	0.70955 1.71136 0.80993	0.07162 -0.04622 2.91471	0.70955 1.71136 0.80993
	A=1 B=2 c=1	4	$\hat{\mu}$ $\hat{\alpha}_1$ $\hat{\alpha}_2$	0.01308 -0.03811 2.98167	0.65205 1.53333 0.75435	0.01308 -0.03811 2.98167	0.65205 1.53333 0.75435
	A=1 B=1 C=2	4	$\hat{\mu}$ $\hat{\alpha}_1$ $\hat{\alpha}_2$	0.00339 -0.07937 3.00873	0.69931 1.58048 0.80089	0.00339 -0.07937 3.00873	0.69931 1.58048 0.80089
5	A=2 B=2 C=1	2	$\hat{\mu}$ $\hat{\alpha}_1$ $\hat{\alpha}_2$	0.04019 -0.03321 2.96357	0.68845 1.56381 0.79344	0.04019 -0.03321 2.96357	0.68845 1.56381 0.79344
	A=1 B=2 C=2	2	$\hat{\mu}$ $\hat{\alpha}_1$ $\hat{\alpha}_2$	0.00301 -0.00655 3.00200	0.73036 1.58438 0.83730	0.00301 -0.00655 3.00200	0.73036 1.58438 0.83730
	A=2 B=1 C=2	2	$\hat{\mu}$ $\hat{\alpha}_1$ $\hat{\alpha}_2$	0.03576 -0.06058 2.97227	0.70473 1.57554 0.78023	0.03576 -0.06058 2.97227	0.70473 1.57554 0.78023

3.4.2 Reproducing the coefficients of the unbalanced one-way ANOVA

The method can also be applied to unbalanced datasets. We take the same example as that in 3.4.1 except that group A was changed to contain one observation. So the total number of observations is $n = 5$ and the number of unknown mean parameters remains $p = 3$. The subsets contain k observations such that $3 \leq k \leq 4$. Also there must be at least one observation in each group for a subset to be admissible.

All possible situations characterizing the subsets that are elemental or non-elemental are listed in Table 3.4.2. Again, the results show that the coefficients of one-way ANOVA with unbalanced dataset can be reproduced exactly based on the elemental or non-elemental sets. The weights of the three methods (DET, INV, REF) are again the same, namely, the weight is 1/ (# of admissible sets) for each subset.

Table 3.4.2 Simulation results of reproducing one-way ANOVA coefficients with based on 500 simulated unbalanced datasets

k	**No. of Obs. taken in each group**	**No. of Admiss-ible sets**	**Paramet-ers**	**OLS Coefficients**		**Estimates by averaging all subset estimators**	
				Mean	SD	Mean Est.	SD Est.
3	A=1 B=1 C=1	4	$\hat{\mu}$	0.05679	1.02981	0.05679	1.02981
			$\hat{\alpha}_1$	-0.11910	1.75473	-0.11910	1.75473
			$\hat{\alpha}_2$	2.97401	1.07256	2.97401	1.07256
4	A=1 B=2 c=1	2	$\hat{\mu}$	-0.0157	0.9884	-0.0157	0.9884
			$\hat{\alpha}_1$	-0.0253	1.7023	-0.0253	1.7023
			$\hat{\alpha}_2$	3.0149	1.0428	3.0149	1.0428
	A=1 B=1 C=2	2	$\hat{\mu}$	-0.0611	1.0189	-0.0611	1.0189
			$\hat{\alpha}_1$	0.0495	1.7315	0.0495	1.7315
			$\hat{\alpha}_2$	3.0508	1.0605	3.0508	1.0605

3.5 Special Case 4: Reproducing the coefficients of two-way ANOVA

Assume the model corresponding to two-way ANOVA is

$$Y_{ijk} = \mu + \sum_{i=1}^{r-1} \alpha_i X_i + \sum_{j=1}^{c-1} \beta_j Z_j + \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} \gamma_{ij} X_i Z_j + \varepsilon$$

$$i = 1, \dots, c-1; j = 1, \dots, c-1 \ ,$$

where μ represents an overall mean for the reference cell, α_i is the effect due to the ith level of factor A relative to its reference level , β_j is the effect due to the jth level of factor B relative to its reference, the γ_{ij} 's capture any interaction between the ith level of A and the jth level of B, and ε is a random error term. In the two-way ANOVA, the formulae for the three methods for reproducing the estimators again only involved the X's and Z's. Every subset regression should be weighted the same, since the X_i and Z_j are dummy variables (0 or 1). The weights produced by all 3 methods (DET, INV, REF) are identical across the methods, and once again the common weight reduces to 1/ (# of admissible sets) for each subset. So the $\hat{\mu}$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$'s can be simplified to be the simple weighted average of their estimates of all possible admissible subset regressions.

Regarding the admissibility of subsets, again each subset including elemental and non-elemental sets to reproduce the estimates for the two-way ANOVA is not only a set of k distinct observations for any k such that $p \leq k \leq n-1$. There must also be at least one observation in the ith level of factor A and the jth level of factor B in every admissible subset of the data so that the subset produces a unique estimate of each

parameter. For example, suppose there are 2 levels of factor A and 2 levels of factor B. The following table provides hypothetical data for the two-way model:

		Factor A	
		Level 1	Level 2
Factor B	Level 1	y=3 y=4	y=2 y=6
	Level 2	y=8 y=9	y=1 y=5

Then all possible elemental sets should be [(1,1,3), (1,2,2), (2,1,8), (2,2,1)], [(1,1,3), (1,2,2), (2,1,8), (2,2,5)], [(1,1,3), (1,2,2), (2,1,9), (2,2,1)], [(1,1,3), (1,2,2), (2,1,9), (2,2,5)], [(1,1,3), (1,2,6), (2,1,8), (2,2,1)], [(1,1,3), (1,2,6), (2,1,8), (2,2,5)], [(1,1,3), (1,2,6), (2,1,9), (2,2,1)], [(1,1,3), (1,2,6), (2,1,9), (2,2,5)], [(1,1,4), (1,2,2), (2,1,8), (2,2,1)], [(1,1,4), (1,2,2), (2,1,8), (2,2,5)], [(1,1,4), (1,2,2), (2,1,9), (2,2,1)], [(1,1,4), (1,2,2), (2,1,9), (2,2,5)], [(1,1,4), (1,2,6), (2,1,8), (2,2,1)], [(1,1,4), (1,2,6), (2,1,8), (2,2,5)], [(1,1,4), (1,2,6), (2,1,9), (2,2,1)], [(1,1,4), (1,2,6), (2,1,9), (2,2,5)]. The number of possible elemental sets should not be simply $\binom{8}{4}$, since many of them are inadmissible.

3.5.1 Reproducing the coefficients of the balanced two-way ANOVA

As in the previous table, suppose there are two factors A and B and two levels in each factor. There are 2 observations in the ith level of factor A and in the jth level of factor B, for $i = 1, 2$ and $j = 1, 2$. The model in this case is $Y = \mu + \alpha X + \beta Z + \gamma XZ + \varepsilon$. The level 1 of factor A and the level 1 of factor B were considered as reference levels in a

standard reference cell coding. The total number of observations is n=8 and the number of unknown parameter is p=4. So each potential subset is a set of k distinct observations, where $4 \leq k \leq 7$. There must be at least one observation in the cell of ith level of factor A and the jth level of factor B in the subset to correctly reproduce the overall two-way ANOVA coefficients (i.e., each admissible subset must have this characteristic).

All possible situations characterizing the subsets that are elemental or non-elemental are listed in Table 3.5.1. The results confirm that the coefficients of two-way ANOVA with a balanced dataset can be reproduced exactly based on the elemental or non-elemental sets. Each subset is weighted by 1/ (# of admissible sets), based on each of the 3 methods (INV, REF, DET).

Table 3.5.1 Simulation results of reproducing two-way ANOVA coefficients with based on 500 simulated balanced datasets

K	No. of Obs. taken in each cell	No. of Admiss-ible sets	Paramet-ers	OLS Coefficients		Estimates by averaging all subset estimators	
				Mean	SD	Mean Est.	SD Est.
4	a=1 b=1 c=1 d=1	16	$\hat{\mu}$	0.1555	2.7934	0.1555	2.7934
			$\hat{\alpha}$	0.1209	3.8719	0.1209	3.8719
			$\hat{\beta}$	-0.0795	3.7807	-0.0795	3.7807
			$\hat{\gamma}$	-0.3007	5.0680	-0.3007	5.0680
5	a=1 b=1 c=1 d=2	8	$\hat{\mu}$	0.1345	3.0118	0.1345	3.0118
			$\hat{\alpha}$	-0.2806	4.0075	-0.2806	4.0075
			$\hat{\beta}$	-0.1732	3.9685	-0.1732	3.9685
			$\hat{\gamma}$	0.6713	5.5127	0.6713	5.5127
	a=1 b=1 c=2 d=1	8	$\hat{\mu}$	0.0107	2.8425	0.0107	2.8425
			$\hat{\alpha}$	0.0373	3.9974	0.0373	3.9974
			$\hat{\beta}$	0.0870	4.0699	0.0870	4.0699
			$\hat{\gamma}$	-0.1508	5.4751	-0.1508	5.4751
	a=1 b=2 c=1 d=1	8	$\hat{\mu}$	-0.0103	2.8120	-0.0103	2.8120
			$\hat{\alpha}$	0.0410	3.9653	0.0410	3.9653
			$\hat{\beta}$	0.0414	3.8927	0.0414	3.8927
			$\hat{\gamma}$	0.1579	5.6456	0.1579	5.6456
	a=2 b=1 c=1 d=1	8	$\hat{\mu}$	-0.0818	2.9232	-0.0818	2.9232
			$\hat{\alpha}$	-0.1356	3.9696	-0.1356	3.9696
			$\hat{\beta}$	0.0573	4.0017	0.0573	4.0017
			$\hat{\gamma}$	0.2121	5.6489	0.2121	5.6489
6	a=2 b=1 c=1 d=2	4	$\hat{\mu}$	7.34E-05	2.7705	7.34E-05	2.7705
			$\hat{\alpha}$	-0.0407	3.9368	-0.0407	3.9368
			$\hat{\beta}$	0.1882	4.1234	0.1882	4.1234
			$\hat{\gamma}$	0.0543	5.8008	0.0543	5.8008
	a=2 b=1 c=2 d=1	4	$\hat{\mu}$	0.0703	2.9174	0.0703	2.9174
			$\hat{\alpha}$	0.0973	4.0659	0.0973	4.0659
			$\hat{\beta}$	0.1842	4.2310	0.1842	4.2310
			$\hat{\gamma}$	-0.3641	5.9959	-0.3641	5.9959
	a=2 b=2		$\hat{\mu}$	-0.0068	2.8241	-0.0068	2.8241
			$\hat{\alpha}$	0.0248	3.8454	0.0248	3.8454

	c=1 d=1	4	$\hat{\beta}$ $\hat{\gamma}$	0.1148 -0.064	3.9979 5.5172	0.1148 -0.064	3.9979 5.5172
	a=1 b=2 c=2 d=1	4	$\hat{\mu}$ $\hat{\alpha}$ $\hat{\beta}$ $\hat{\gamma}$	0.0956 -0.0164 -0.1708 0.1824	2.7659 3.9723 4.0537 5.6700	0.0956 -0.0164 -0.1708 0.1824	2.7659 3.9723 4.0537 5.6700
	a=1 b=2 c=1 d=2	4	$\hat{\mu}$ $\hat{\alpha}$ $\hat{\beta}$ $\hat{\gamma}$	-0.0303 0.1604 0.0765 -0.1215	2.7810 4.2050 3.9507 5.8527	-0.0303 0.1604 0.0765 -0.1215	2.7810 4.2050 3.9507 5.8527
	a=1 b=1 c=2 d=2	4	$\hat{\mu}$ $\hat{\alpha}$ $\hat{\beta}$ $\hat{\gamma}$	0.0689 0.0097 0.1039 -0.1067	2.8894 4.2500 4.1444 6.0224	0.0689 0.0097 0.1039 -0.1067	2.8894 4.2500 4.1444 6.0224
7	a=2 b=2 c=1 d=2	2	$\hat{\mu}$ $\hat{\alpha}$ $\hat{\beta}$ $\hat{\gamma}$	0.0964 -0.0018 -0.1964 0.0631	2.8087 3.9673 3.9553 5.4865	0.0964 -0.0018 -0.1964 0.0631	2.8087 3.9673 3.9553 5.4865
	a=2 b=2 c=2 d=1	2	$\hat{\mu}$ $\hat{\alpha}$ $\hat{\beta}$ $\hat{\gamma}$	-0.0387 0.1117 -0.0632 0.1211	2.8647 3.8966 3.9459 5.5211	-0.0387 0.1117 -0.0632 0.1211	2.8647 3.8966 3.9459 5.5211
	a=1 b=2 c=2 d=2	2	$\hat{\mu}$ $\hat{\alpha}$ $\hat{\beta}$ $\hat{\gamma}$	-0.0312 -0.0284 0.0223 0.1137	2.7940 4.0783 3.9989 5.5034	-0.0312 -0.0284 0.0223 0.1137	2.7940 4.0783 3.9989 5.5034
	a=2 b=1 c=2 d=2	2	$\hat{\mu}$ $\hat{\alpha}$ $\hat{\beta}$ $\hat{\gamma}$	0.0422 -0.0999 -0.0135 0.2315	2.8226 4.0117 3.8747 5.5370	0.0422 -0.0999 -0.0135 0.2315	2.8226 4.0117 3.8747 5.5370

3.5.2 Reproducing the coefficients of the unbalanced two-way ANOVA

The example is the same as that in 3.5.1 except that there is only one observation in the first level of factor A and the first level of factor B. So the total number of observations is $n = 7$ and the number of unknown parameters is $p = 4$. The subsets contain k observations such that $4 \leq k \leq 6$. There must be at least one observation in the cell of the ith level of factor A and the jth level of factor B to obtain an admissible subset, and these subsets are weighted equally to correctly reproduce the overall two-way ANOVA coefficients. Again, the weight is 1/ (# of admissible sets) for each admissible subset based on all of the 3 weighting methods (DET, INV, REF).

All possible situations characterizing the subsets that are elemental or non-elemental are listed in Table 3.5.2. The coefficients of two-way ANOVA with the unbalanced dataset can be also reproduced exactly based on the elemental or non-elemental sets.

Table 3.5.2 Simulation results of reproducing two-way ANOVA coefficients with based on 500 simulated unbalanced datasets

K	No. of Obs. taken in each cell	No. of Admiss-ible sets	Paramet-ers	Coefficients		Estimates by averaging all subset estimators	
				Mean	SD	Mean Est.	SD Est.
4	a=1 b=1 c=1 d=1	8	$\hat{\mu}$	-0.0476	3.7700	-0.0476	3.7700
			$\hat{\alpha}$	0.1612	4.5789	0.1612	4.5789
			$\hat{\beta}$	0.0163	4.6731	0.0163	4.6731
			$\hat{\gamma}$	-0.0177	6.0265	-0.0177	6.0265
5	a=1 b=1 c=1 d=2	4	$\hat{\mu}$	0.0178	3.6748	0.0178	3.6748
			$\hat{\alpha}$	-0.1422	4.5087	-0.1422	4.5087
			$\hat{\beta}$	-0.1114	4.6988	-0.1114	4.6988
			$\hat{\gamma}$	0.1065	6.4397	0.1065	6.4397
	a=1 b=1 c=2 d=1	4	$\hat{\mu}$	0.1962	4.0519	0.1962	4.0519
			$\hat{\alpha}$	-0.2678	4.8482	-0.2678	4.8482
			$\hat{\beta}$	-0.1420	5.1289	-0.1420	5.1289
			$\hat{\gamma}$	0.1327	6.5053	0.1327	6.5053
	a=1 b=2 c=1 d=1	4	$\hat{\mu}$	0.0832	3.9810	0.0832	3.9810
			$\hat{\alpha}$	-0.2006	4.9213	-0.2006	4.9213
			$\hat{\beta}$	0.0674	4.8111	0.0674	4.8111
			$\hat{\gamma}$	-0.0025	6.4622	-0.0025	6.4622
6	a=1 b=2 c=2 d=1	2	$\hat{\mu}$	-0.0114	4.0148	-0.0114	4.0148
			$\hat{\alpha}$	0.0293	4.6329	0.0293	4.6329
			$\hat{\beta}$	0.0871	4.6875	0.0871	4.6875
			$\hat{\gamma}$	-0.0843	5.8695	-0.0843	5.8695
	a=1 b=2 c=1 d=2	2	$\hat{\mu}$	-0.0080	4.0600	-0.0080	4.0600
			$\hat{\alpha}$	0.2520	5.0039	0.2520	5.0039
			$\hat{\beta}$	-0.2070	5.1758	-0.2070	5.1758
			$\hat{\gamma}$	0.0191	6.5687	0.0191	6.5687
	a=1 b=1 c=2 d=2	2	$\hat{\mu}$	0.0625	4.1084	0.0625	4.1084
			$\hat{\alpha}$	-0.0235	5.0131	-0.0235	5.0131
			$\hat{\beta}$	-0.0502	5.0665	-0.0502	5.0665
			$\hat{\gamma}$	0.1801	6.3033	0.1801	6.3033

Chapter 4

SUMMARY AND FUTURE WORK

4.1 Summary

In this paper, we discussed three methods that could reproduce the OLS estimators of the linear regression based on elemental and non-elemental sets. Both the refined inverse-variance (REF) and the constrained optimal (CON) weights method, just as with the determinant (DET; Hoerl and Kennard, 1980) weights method, lead to exact replication of the OLS estimators. However, we assumed only n=3 observations when illustrating the CON method. If we increase n, the number of equations that we need to solve will correspondingly increase. Normally, there are $\binom{n}{k}$ equations to solve, which will result in computational infeasibility when the dataset is large. So the CON method is not readily generalized. However, the REF method, which refines direct inverse-variance weighting by accounting for covariances among individual subset-based OLS estimates, is generalizable and represents the primary original contribution of this thesis.

The estimators produced by the direct inverse-variance (INV) method based on combining results of all the elemental or non-elemental regressions are found to be very close to the OLS estimators. As mentioned in Section 3.3, this may have applications in the area of robust regression analysis (see Jin, 2012).

Among the examples considered, we demonstrated that the coefficients of one-way or two-way ANOVA can also be reproduced exactly based on elemental or non-elemental sets. This required discussion of the notion of “admissible” subsets, i.e., those for which the desired model can be fit to produce unique OLS estimates. In these ANOVA cases, we find that the overall OLS coefficients can be reproduced by simply taking an average of all estimators based on the admissible the elemental or non-elemental sets.

4.2 Future Research

In this paper, only special cases were listed to illustrate the fact that our methods can reproduce the OLS estimators of the linear regressions based on the elemental and non-elemental sets. A formal proof that the REF method can reproduce OLS in the general case might represent a worthwhile educational contribution to the literature.

In addition, we know that Ordinary Least Squares (OLS) estimation is sensitive to outliers, hence not robust. We could use non-elemental sets to produce robust estimators which are more robust to outliers or influential observations by incorporating the inverse-variance weights method, along with other weight components based on outlier diagnostics (see Jin, 2012).

References

Agullo Jose (1997). Exact Algorithms for Computing the Least Median of Squares Estimate in Multiple Linear Regression. *Lecture Notes-Monograph Series*, 31, 133-146

Arnold, S.F. (1981). The Theory of Linear Models and Multivariate Analysis. *Wiley Series in Probability and Statistics*, 202

Hawkins, D. M., Bradu D. and Kass G. V. (1984). Location of Several Outliers in Multiple-Regression Data Using Elemental Sets. *Technometrics*, 26, 197-208

Hawkins, D. M. (1993). The Accuracy of Elemental Set Approximations for Regression. *Journal of the American Statistical Association*, 88, 580-589

Hoerl, A. E. and Kennard, R. W. (1980). A Note on Least Squares Estimates. *Communications in Statistics: Simulation and Computation,* 9, 315-317

Jin, C. (2012). New Thoughts on Using Elemental and Non-Elemental Sets to Produce Robust Estimates of Regression Coefficients. *Unpublished Masters Thesis*, Emory University.

Mayo, M. S. and Gray, J. B. (1997). Elemental Subsets: The building Blocks of Regression. *The American Statistician,* 51, 122-129

Rao C.R. and Mitra S. K. (1971). Generalized Inverse of Matrices and its Applications. *New York: John Wiley & Sons*, 240.

Rousseeuw, P.J. (1984). Least Median Squares Regression. *Journal of the American Statistical Association*, 79, 871-880

Rubin, D. B. (1980). Composite Points in Weighted Least Squares Regressions. *Technometrics*, 22, 343-348

Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association,* 63, 1379-1389

Sheynin, O. B. (1973). R. J. Boscovich's Work on Probability. *Archive for History of Exact Sciences*, 9, 306-324

Smyth, G. K. and Hawkins, D. M. (2000). Robust Frequency Estimation Using Elemental Sets. *Journal of Computational and Graphical Statistics,* 9, 196-214

Stormberg, A. J. (1993). Computation of High Breakdown Nonlinear Regression Parameters. *Journal of the American Statistical Association,* 88, 237-244

Theil, H. (1950). A Rank-Invariant Method of Linear and Polynomial Regression Analysis. *Konlijke Nederlandsche Akademie Wetenschappen Proceedings, Ser. A*, 53, 1397-1412