A Study of Estimators in Linear Models

Anna Nozza

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Masters of Science at
Concordia University
Montreal, Quebec, Canada

December 1999

0-612-47801-7

ABSTRACT

A Study of Estimators in Linear Models

Anna Nozza

This thesis is a study of Estimators, particularly in Linear Models. The newest technology of Bootstrap Methodology is employed in the estimation procedure. We present a survey of the Bootstrap Methodology in the beginning and move on to some serious problems in Linear Model estimation procedure. We have worked out the conditions under which the estimators of nonstandard linear models will be best linear unbiased estimators. Furthermore, we have shown that the estimators of other linear models bear a linear relationship with least-squares estimators. Finally, we have worked out the finite sample properties of two-stage least-squares using the Bootstrap Methodology.

ACKNOWLEDGMENTS

I would like to express my gratitude to all those who have assisted me towards the completion of this thesis.

A special thanks goes to my supervisor, Dr. T.D. Dwivedi, whereby without him, I would never have finished this thesis. His encouragement and contribution are behind the completion of this thesis. Many thanks also goes to Dr. R. Tiwari for many discussions during this work.

A deep affection goes to my sister, Maria Grazia Nozza, whose computer skills proved to be, without a doubt, invaluable.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION AND REVIEW

1.1 HISTORY

The accuracy and speed of computers have revolutionized statistics including every other field of human activity and science. Apart from conventional statistical methods that relies on mathematical equations, Dr. Bradley Efron, a Stanford University statistician, introduced "the bootstrap method" in 1977, a simple but powerfully effective computational device for approximating quantities that are almost impossible to compute analytically (Kolata, 1988). Through the superior mathematical abilities of computers, the bootstrap method allows statisticians to predict the reliability of data analysis, to derive relevant information from the data, to see patterns and fluctuations in data, and ultimately to discover solutions and connections in simulated samplings.

In his definitive paper, The 1977 Rietz Lecture, Dr. Efron (1979) radicalized the statistics field when he proposed that the bootstrap method was more dependable and more applicable than the Jackknife, a simple statistical tool. Unfortunately, in the following years, there has been abundant research in the jackknife, but research on the bootstrap method is far and in between. [Wu 1986]

Nonetheless, after the initial skepticism by some theoretical statisticians and the euphoria by other academicians, scientists from different disciplines (psychologists, physicists, geologists, etc) now use the bootstrap method. K. Singh (1981) provided mathematical proof for the validity of the Bootstrap methodology.

1.2 DESCRIPTION OF THE BOOTSTRAP METHOD

The essential problem in statistics is how to derive, from a sample data, conclusions applicable to a population.

"Statistical theory attempts to answer three basic questions:

1.How should I collect my data?

2.How should I analyze and summarize the data that I've collected?

3.How accurate are my data summaries?" [Efron & Tibshirani, 1993]

Statistical theory aims to answer the second and third questions. Dr. Efron's method, (dubbed 'the bootstrap method' because data "pulls itself up by its own bootstraps"), allows data to generate artificial data sets through which their reliability can be evaluated (Sim, 1989). In other words, the bootstrap method assumes that the collected data is an exact replica of the entire population; it resamples the data several times (creating "the bootstrap samples") in order to compute or estimate statistical functions (ex; mean, variance …). For example, we obtain the mean of the computed samples as an estimate of the expected value by bootstrapping several times the original sample.

A University psychology professor is testing whether a new school curriculum would improve the scores of 100 grade 6 children having difficulties in math. Using the bootstrap method, the computer would generate a new data set from the original 100 measurements. At random, it would pick a measure from among the 100, return it to the original data set, pick a second number, return it again, pick a third, return in again, and so on and so forth until 100 measurements are chosen. From the original data, innumerable new data sets can now be generated. Researchers can verify and quantify their original assumptions by comparing the results of the bootstrapped data with the results of the original data. Data resampling provides statistical inference.

To further explain the procedure of the bootstrap method, we select a random sample from a population, say **X**, with probability density function *(f)* and cumulative distribution function (F). The symbol **X** represents the population of measurements whereas $\boldsymbol{x}$ represents a random sample of size $\boldsymbol{n}$ from **X**. In other words, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, ... , $\boldsymbol{x}_n$ is a collection of measurements from **X**. We will use the bootstrap method to verify how accurately a statistic (denoted by **T**), calculated from $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, ..., $\boldsymbol{x}_n$ can estimate the corresponding population characteristic (denoted by $\boldsymbol{\theta}$).

Before we use the bootstrap method to estimate standard errors, to correct for bias and to test hypothesis, we must clarify the differences between parametric and nonparametric models. Parametric models are based on the assumption that both the mathematical model *(f)* and the constants are known; when no such mathematical model exists and conditions are unknown, the statistical analysis is nonparametric. The only assumption for non-parametric models is that the random variable $\boldsymbol{x}_j$ are independent and identically distributed (iid). Nonetheless, whether a parametric model is used or not, a nonparametric analysis can still be performed to assess the robustness of the conclusions drawn from the parametric analysis (Davidson & Hinkley, 1997).

The empirical distribution is the primary component of the nonparametric analysis. Efron & Tibshirani (1993), defined the empirical distribution as "Having observed a random sample of size n from a probability distribution F,

$$F \rightarrow (x_1, x_2, \ldots, x_n),$$

the empirical distribution function $\hat{F}$ is defined to be the discrete distribution that puts probability $\frac{1}{n}$ on each value x_i, where $i = 1, 2, \ldots, n$. In other words, $\hat{F}$ assigns to a set A in the sample space of x its empirical probability

$$\text{Prob}\{A\} = \#\{x_i \in \mathbf{A}\} / n,$$

the proportion of the observed sample $x = (x_1, x_2, \ldots, x_n)$ occurring in **A**." The empirical distribution function (EDF) $\hat{F}$, defined as the sample proportion

$$\hat{F}(x) = \frac{\#\{x_i \leq x\}}{n}$$

There are many simple statistics (mean, median, correlation, etc) that are extensions of the empirical distribution function. The EDF ($\hat{F}$) is a simple estimate of the entire distribution F.

The most obvious example is the mean of the EDF whereby

$$\bar{x} = \sum \frac{1}{n} x_i$$

Parameters and statistics determine statistical inference. A parameter is a function of the probability distribution F whereas a statistic is a function of the sample x. Statisticians sometimes write parameters directly as functions of F which can be expressed as

$$\theta = t(F)$$

where $t(\cdot)$ is a statistical function (or simply a numerical evaluation expression) for computing t from F. We then apply the 'plug-in principle' of estimating from samples. "The plug-in estimate of a parameter $\theta = t(F)$ is defined to be

$$\hat{\theta} = t(\hat{F}).$$

In other words, we estimate the function $\theta = t(F)$ of the probability distribution F by the same function of the empirical distribution $\hat{F}$, $\hat{\theta} = t(\hat{F})$" [Efron & Tibshirani, 1993]. Some examples are the mean and the variance of **X**.

The bootstrap method has been most effectively used in the estimation of standard errors, in the correction for bias, and in the testing of hypothesis. The following is a brief description of each of these concepts.

1.2.1 BOOTSTRAP ESTIMATE OF STANDARD ERROR

In 1993, Efron and Tibishirani wrote an algorithm in *An Introduction to the Bootstrap* on how to calculate the estimate of standard error using the bootstrap method.

Let $\hat{F}$ be the empirical distribution. With probability $\frac{1}{n}$ on each of the observed values x_i, where $i = 1, 2, \ldots, n$. Let $(x_1^*, x_2^*, \ldots, x_n^*)$ be a bootstrap sample of size n drawn with replacement from the population of x. The bootstrap sample

$$x^* = (x_1^*, x_2^*, \ldots, x_n^*)$$

consists of elements drawn with replacement from $x = (x_1, x_2, \ldots, x_n)$ whereby each x_i may not be unique, i.e. it may appear once, twice, or even never. We select **B** independent bootstrap samples.

To each of these bootstrap samples is a bootstrap replication of $\hat{\theta}$, whereby

$$\hat{\theta}^* = s(x^*).$$

$s(x^*)$ uses the same function $s(\cdot)$ that was applied to x. For example, if $s(x)$ is the sample mean $\bar{x}$, then $s(x^*)$ is the mean of the bootstrap data $\bar{x}^* = \sum_{i=1}^{n} \frac{x_i^*}{n}$ Our final step is to estimate the standard error $se_F(\hat{\theta})$ by using the sample standard deviation of all **B** replications. We use the plug-in estimate that uses the empirical distribution function $\hat{F}$ in place of unknown F. In other words, $se_F(\hat{\theta})$ is defined by $se_{\hat{F}}(\hat{\theta}^*)$. Consequently,

$$se_B = \left\{ \frac{\sum_{b=1}^{B} [\hat{\theta}_b^* - \hat{\theta}^*]^2}{B - 1} \right\}^{\frac{1}{2}}$$

where

$$\hat{\theta}^* = \frac{\sum_{b=1}^{B} \hat{\theta}_b^*}{B}$$

As $\mathbf{B} \rightarrow \infty$, (i.e. $\mathbf{B}$ is very large), the empirical standard deviation approaches the population standard deviation.

$$\lim_{B \to \infty} se_B = se_{\hat{F}}(\hat{\theta}^*)$$

1.2.2 BIAS CORRECTION

Bias is defined as the difference between the expectation of an estimator $\hat{\theta}$ and the quantity θ being estimated. The bootstrap algorithm described in Efron & Tibashrani [1993] can easily be adapted to give estimates of bias as well as estimates of standard errors. Although the estimator is consistent, it might still be biased. In 1999, Bergstrom described a simple procedure outlining the application of bootstrap to bias correction. He wrote that "the relation of the bootstrap sample to the original sample is the same as the relation between the original sample and the true population". Bergstrom (1999) believes that the bias is constant and does not vary with the parameter value.

Biases are harder to estimate than standard errors. Due to the high variability in bias, bias correction can be dangerous to use and hence problematic. Correcting the bias may cause a larger increase in the standard error and consequently increase the mean squared error of the estimator. This is also explored in MacKinnan & Smith [1998] and in Ferrari & Cribari-Neto [1998].

If the bias is small compared to the estimated standard error, then it is safe to use the statistic $\hat{\theta}$. However, if the bias is large compared to the estimated standard error, then this may indicate that the statistic $\hat{\theta} = s(x)$ may not be a good estimate of the parameter θ [Efron and Tibashrani (1993)]

1.2.3. Testing of Hypothesis

The primary reason for using bootstrap tests rather than asymptotic tests is due to the fact that asymptotic tests may be biased (Bergstrom, 1999). Another feature is that their empirical sizes would converge to the true sizes faster than asymptotic tests, proven in Bergstrom (1999).

1.3 BOOTSTRAPPING REGRESSION MODELS

Bootstrapping can also be applied to regression models. According to Efron & Tibshirani [1993], regression models are among the most useful and most widely used of statistical methods. Regression models evaluate the effects of many possible explanatory variables on a response variable. Since Legendre & Gauss of the 19th century, the multiple linear regression model is given by

$$\boldsymbol{Y} = \mathbf{X} \ \beta + \varepsilon$$

In this equation, $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ (called a parameter vector or regression parameter) is a $\mathbf{p \times 1}$ vector of unknown parameters. The main goal of the regression analysis is to infer β from the observed data where $\mathbf{X}$ is an $\boldsymbol{n} \times \mathbf{p}$ matrix of full rank where $\mathbf{p} \le \boldsymbol{n}$, and $\boldsymbol{Y}$ (called the response vector) an $\boldsymbol{n} \times \mathbf{1}$ data vector. The error terms ε, an $\boldsymbol{n} \times \mathbf{1}$ vector, are assumed to be a random sample from an unknown error distribution F having expectation 0,

$$F \rightarrow (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = \varepsilon \qquad \left[\mathrm{E}_F(\varepsilon) = 0 \right]$$

Using the conventional least-square estimate to the so-called normal equations, we need to minimize the residual squared error denoted by,

$$\mathbf{RSE}_b = (Y - \mathbf{X}\beta)'(Y - \mathbf{X}\beta).$$

Minimizing $\mathbf{RSE}_b$ yields

$$\mathbf{X}'\mathbf{X}\,\beta = \mathbf{X}'Y$$

or simply

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$$

The question now is: how close is $\hat{\beta}$ to β? We will compare the bootstrap approximation with the standard asymptotics. The main assumptions are the following:

i. The matrix $\mathbf{X}$ is not random;

ii. The error components $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent, with the distribution having mean 0 and the finite variance θ^2; F and θ^2 are unknown. [Freedman, 1981]

Let θ_F^2 be the variance of the error ε_i given by

$$\theta_F^2 = var_F(\varepsilon)$$

And let the standard error of $\hat{\beta}$ be given by

$$se\ \hat{\beta} = \{\theta_F^2 (\mathbf{X}'\mathbf{X})^{-1}\}^{\frac{1}{2}}.$$

Since θ_F is estimated by $\hat{\theta}_F$, we obtain the estimated standard errors given by

$$se(\hat{\beta}) = \{\hat{\theta}_F^2 (\mathbf{X}'\mathbf{X})^{-1}\}^{\frac{1}{2}}$$

where

$$\hat{\theta}_F^2 = \frac{\varepsilon'\varepsilon}{n-p} = \frac{(Y - \mathbf{X}\hat{\beta})'(Y - \mathbf{X}\hat{\beta})}{n-p} .$$

The bootstrap method gives the same asymptotic results as the linear regression model (Freedman, 1981). Hence, we are assured that the bootstrap delivers the same results that we can then analyze mathematically. The bootstrap method is best applied to more general regression models

(i) that have no mathematical solution;

(ii) where the parameter vector β is non-linear;

(iii) and they use fitting methods other than least square.

[Efron & Tibshirani, 1993]

The probability model has basically two components; β, the parameter vector of regression coefficients and F, the probability distribution of the error terms ε. In regression, the centered residuals are resampled. β is not known, however, we use $\hat{\beta}$, the least-squares estimate of β. If β were known, then the error terms are calculated by

$$\varepsilon_i = y_i - x_i \beta \quad \text{for } i = 1, 2, \ldots, n$$

and estimate F by their empirical distribution. Since β is not known, then we use $\hat{\beta}$ to calculate the residuals.

$$\hat{\varepsilon}_i = Y_i - x_i \hat{\beta} \quad \text{for } i = 1, 2, \ldots, n$$

The estimate of F is $\hat{F}$, the empirical distribution of the $\hat{\varepsilon}_i$ so that $\hat{F}$ puts mass (or probability) $\frac{1}{n}$ at $\hat{\varepsilon}_i$ for $i = 1, 2, \ldots, n$ with expectation equal to 0. Given $\hat{\beta}$ & $\hat{F}$, let

$$Y^* = X\hat{\beta} + \varepsilon^*$$

where Y^* is generated from the data, using the regression model with $\hat{\beta}$ & $\hat{F}$ as the empirical distribution of $\hat{\varepsilon}$ and whereby each ε_i^* equals any one of the n values $\hat{\varepsilon}_j$ with mass $1/n$.

If we would give $\mathbf{X}$ & Y^* to another statistician in order to estimate β, then the estimate would be

$$\hat{\beta}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y^*.$$

Here, by applying the bootstrap principle, the distribution of $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$ approximates closely the distribution of $\sqrt{n}(\hat{\beta} - \beta)$. . [Freedman, 1981]

The normal equations will yield the bootstrap least-squares estimate $\hat{\beta}^*$ given by

$$\hat{\beta}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y^*$$

1.4 THE SELECTION MATRIX

A selection matrix is a binary matrix of order $m \times n$ that takes on the value of 0 or 1 with probability $1/n$ and that the value will be 1. The $m \times n$ matrix randomly selects m elements (with replacement) from a set of n elements. The selected element is recorded and then returned to the original set.

A one is placed in the first column of the first row and zeroes placed in the remaining columns. The same process is repeated throughout the matrix for every row.

Let $(\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_n^*)$ denote a bootstrap sample of size n from $\hat{F}$. For all $j = 1, 2, \ldots, J$, then

$$\varepsilon_j^* = (\varepsilon_{(j)}^*, \ldots, \varepsilon_{(j)n}^*).$$

Consequently,

$$y_{(j)}^* = X\beta + \varepsilon_{(j)}^*.$$

Solving for normal equations,

$$\hat{\beta}_{(j)} = (X'X)^{-1} X' y_{(j)}*$$

Substituting for $y_{(j)}*$,

$$\hat{\beta}_{(j)} = (X'X)^{-1} X' \, [X\beta + \varepsilon_{(j)}*]$$

$$= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} \, \beta_{(j)} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \, \varepsilon_{(j)}*.$$

As a result,

$$\hat{\beta}_{(j)} = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \, \varepsilon_{(j)}*.$$

We will look at some properties of $\hat{\beta}_{(j)}$ as it is applied to the selection matrix. But first, some notations and definitions must be supplied.

Let

S denote a set of n elements

$\mathbf{S}_j$ denote a $m \times n$ selection matrix corresponding to the j^{th} bootstrap replication, $j = 1, \ldots, J$.

$\mathbf{S}_{jr}$ denote the r^{th} row of $\mathbf{S}_j$ having zero everywhere except in one position–the unity position, $r = 1, \ldots, m$

$\mathbf{S}_{jri}$ denote the unity position in $\mathbf{S}_{jr}$ where i represents the integer randomly selected with replacement from $\mathbf{S}$. In other words, $\mathbf{S}_{jri}$ is a random variable which assigns 0 or 1 to each element with probability $\frac{1}{n}$ that it will be "1" and $\left(1 - \frac{1}{n}\right)$ that is will be "0".

$\mathbf{S}_j'$ denote the transpose of $\mathbf{S}_j$.

$\mathbf{S}_{jr}'$ denote the rth column of $\mathbf{S}_j'$.

$$\mathbf{S}_j \quad \begin{bmatrix} \mathbf{S}_{j1} \\ \mathbf{S}_{j2} \\ \vdots \\ \mathbf{S}_{jn} \end{bmatrix}_{m \times 1}$$

$$\mathbf{S}_j' \quad \left[\mathbf{S}_{j1}', \mathbf{S}_{j2}', \ldots, \mathbf{S}_{jm}' \right]_{1 \times m}$$

Using the preceding definitions:

$$\mathbf{S}_j' \, \mathbf{S}_j = \sum_{r=1}^{m} \left[\mathbf{S}_{jr}' \quad \mathbf{S}_{jr} \right]$$

$\mathbf{S}_{jr}' \, \mathbf{S}_{jr}$ is an $n \times n$ identity matrix

$$\mathbf{S}_j \, \mathbf{S}_j' = \sum_{r=1}^{m} \left[\mathbf{S}_{jr} \quad \mathbf{S}_{jr}' \right]$$

Resulting in the following theorems from Sim (1989):

$$\lim_{j \to \infty} \left\{ \frac{1}{J} \sum_{j=1}^{J} \mathrm{S}_j \right\} = \frac{1}{n} \, \mathbf{u}_{m \times n}$$

where $\mathbf{u}_{m \times n}$ is the unity matrix of order $m \times n$.

$$\lim_{j \to \infty} \left\{ \frac{1}{J} \sum_{j=1}^{J} \left[\mathrm{S}_{jr} \, ' \, \mathrm{S}_{jr} \right] \right\} = \frac{1}{n} \, \mathrm{I}_n$$

Summing for all $r = 1, \ldots, m$

$$\lim_{j \to \infty} \left\{ \frac{1}{J} \sum_{j=1}^{J} \sum_{r=1}^{m} \left[\mathrm{S}_{jr} \, ' \, \mathrm{S}_{jr} \right] \right\} = \frac{1}{n} \, (m) \, \mathrm{I}_n$$

and,

$$\lim_{j \to \infty} \left\{ \frac{1}{J} \sum_{j=1}^{J} \mathrm{S}_j \, \mathrm{S}_j' \right\} = \frac{1}{n} (n-1) \cdot \mathrm{I}_m + \frac{1}{n} \, \mathbf{u}_{m \times m}$$

If we let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ and its sample mean as $\bar{\varepsilon} = \sum_{i=1}^{n} \frac{\varepsilon_i}{n}$ and replace into

$$\lim_{j\to\infty} \left\{ \frac{1}{J} \sum_{j=1}^{J} S_j \right\} = \frac{1}{n} \; \mathbf{u}_{m\times n}$$

we get

$$\lim_{J\to\infty} \left\{ \frac{1}{J} \sum_{j=1}^{J} S_j \, \varepsilon \right\} = \varepsilon \; \mathbf{u}_{m\times 1},$$

since

$$\mathbf{u}_{m\times n} \, \varepsilon = n \, \bar{\varepsilon} \, \mathbf{u}_{m\times 1}.$$

Similarly,

$$\lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} \mathbf{X}' S_j \, \varepsilon = \varepsilon \; \mathbf{X}' \; \mathbf{u}_{m\times l} .$$

Let $\mathbf{S}_n^2 = \frac{\varepsilon' \varepsilon}{n}$, then

$$\lim_{J\to\infty} \left[\frac{1}{J} \sum_{j=1}^{J} (\mathbf{S}_j \, \varepsilon)' (\mathbf{S}_j \, \varepsilon) \right] = \lim_{J\to\infty} \left[\frac{1}{J} \sum_{j=1}^{J} \varepsilon' \mathbf{S}_j{}' \mathbf{S}_j \, \varepsilon \right]$$

$$= \varepsilon' \left[\lim_{j\to\infty} \frac{1}{J} \sum_{j=1}^{J} \mathbf{S}_j{}' \mathbf{S}_j \right] \varepsilon = \varepsilon' \frac{1}{n} m \; \mathrm{I}_m \; \varepsilon \qquad m \frac{\varepsilon' \varepsilon}{n} = m \, s_n^2$$

or

$$m\sigma^2 \text{ as } n \to \infty$$

CHAPTER 2

GENERALIZED LEAST SQUARES ESTIMATORS AND ITS RELATIONSHIPS WITH OTHER ESTIMATORS

2.1 Introduction

In the linear model $\mathbf{Y}=\mathbf{X}\beta + \varepsilon$, where $\mathbf{X}$ is $\mathbf{n}\times\mathbf{p}$ matrix of rank $\mathbf{p}$, the Gauss-Markoff theorem gives all the essential properties of the estimate $\hat{\beta}$ of β. However in many practical situations when the rank of the design matrix $\mathbf{X}$ is $\mathbf{r}$ such that $\mathbf{r}\leq\mathbf{p}<\mathbf{n}$, then a practitioner faces all kinds of problems, more so if $\varepsilon \simeq \mathbf{N}(\mathbf{0},\mathbf{V})$ where $\mathbf{V}$ is the variance-covariance matrix. In this chapter, we show that if $\mathbf{X}$ is in the column space of exactly $\mathbf{r}$ eigenvectors of $\mathbf{V}$ each of which is associated with a positive eigenvalue, then the set of weighted least square (LS) estimates of β is identical to the set of **LS** estimates of β. Subsequently the general form of a BLUE in terms of weighted least squares is developed. Also the result is refined when $\mathbf{0}$ is in the sample space. We have utilised Moore-Penrose generalised inverse in order to find estimates. For completeness, the definition and an easiest method of computation is also given. A method called Dwivedi's method of rank factorisation, is also included to facilitate the computations of generalised inverses in practical problems.

Furthermore it is shown that all other estimators of linear models bear a linear relationship with LS estimators.

2.2 Definitions of Generalized Inverses

Generalized Inverse: Let $\mathbf{A}$ be an $\mathbf{m}\times\mathbf{n}$ matrix. Let $\mathbf{A}^{+}$ be the generalized inverse (g-inverse) of $\mathbf{A}$ if and only if

(i) $\mathbf{A}\mathbf{A}^{+}$ is symmetric

(ii) $\mathbf{A}^{+}\mathbf{A}$ is symmetric

(iii) $\mathbf{A}\,\mathbf{A}^{+}\mathbf{A} = \mathbf{A}$

(iv) $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+}=\mathbf{A}^{+}$.

Rao g-inverse: Let $\mathbf{A}$ be an $\mathbf{m}\times\mathbf{n}$ matrix. There exists $\mathbf{A}^{g}$ such that

$$\mathbf{A}\,\mathbf{A}^{g}\mathbf{A} = \mathbf{A}.$$

Remark: $\mathbf{A}^{+}$ is called Moore-Penrose generalized inverse and is unique. However, the Rao g-inverse $\mathbf{A}^{g}$ is not unique. The generalized inverse which we have used in our work is $\mathbf{A}^{+}$.

2.3 Method for Computing Generalized Inverses

It is a well-known fact in the mathematical literature that if $\mathbf{A}$ is an $\mathbf{m}\times\mathbf{n}$ matrix of rank $\mathbf{r}$ then $\mathbf{A}_{m\times n} = \mathbf{B}_{m\times r}\mathbf{C}_{r\times m}$ where $\mathbf{r}$ is the rank of $\mathbf{B}$ & $\mathbf{C}$. The purpose of this section is to propose a direct method to factorize $\mathbf{A}$ in the above form known in the literature as Dwivedi's method.

Let the given matrix be denoted by

$$\mathbf{A} = (\mathbf{a}_{ij})_{m\times n}$$

Now select any non-zero element of $\mathbf{A}$ say $\mathbf{a}_{hk}$ and subtract the $\mathbf{m}\mathbf{x}\mathbf{n}$ matrix $\mathbf{u}_1\mathbf{v}_1$ from $\mathbf{A}$, where $\mathbf{u}_1' = \mathbf{a}_{hk}^{-1}\,[\mathbf{a}_{1k}, \mathbf{a}_{2k}, \ldots, \mathbf{a}_{mk}]$ and $\mathbf{v}_1 = [\mathbf{a}_{h1}, \mathbf{a}_{h2}, \ldots, \mathbf{a}_{hn}]$. The difference will be denoted by $\mathbf{A}_1$. It is easily seen that all elements of the h^{th} row and the k^{th} column of $\mathbf{A}_2 = \mathbf{A}\text{-}\mathbf{A}_1$ are zero. If $\mathbf{A}_1=\mathbf{0}$, $\mathbf{A}=\mathbf{u}_1\mathbf{v}_1$ is a factorization of the required type. If $\mathbf{A}_2\neq\mathbf{0}$ the above process is applied to $\mathbf{A}_2$. This yields a matrix $\mathbf{A}_3$ with at least two null columns and two null rows. Since each step increases the number of null rows and columns by at least one, it is clear that the above process will terminate after a finite number of steps.

Suppose,

$$\mathbf{A} - \mathbf{u}_1\mathbf{v}_1 = \mathbf{A}_1$$

$$\mathbf{A}_1 - \mathbf{u}_2\mathbf{v}_2 = \mathbf{A}_2$$

$$\ldots$$

$$\mathbf{A}_{r-2} - \mathbf{u}_{r-1}\,\mathbf{v}_{r-1} = \mathbf{A}_{r-1} \neq \mathbf{0}$$

$$\mathbf{A}_{r-1} - \mathbf{u}_r\mathbf{v}_r = \mathbf{0}$$

Then

$$\mathbf{A} = \mathbf{u}_1\mathbf{v}_1 + \mathbf{u}_2\mathbf{v}_2 + \ldots + \mathbf{u}_r\mathbf{v}_r = \mathbf{BC},$$

where

$$\mathbf{B} = [\mathbf{u}_1\mathbf{u}_2 \ldots \mathbf{u}_r]1 \quad \mathbf{C}' = [\mathbf{v}_1\mathbf{v}_2 \ldots \mathbf{v}_r].$$

It is an elementary exercise to verify that the columns of $\mathbf{B}_{m \times r}$ and those of $\mathbf{C}_{n \times r}'$ are linearly independent. This establishes the fact that $\mathbf{B}_{m \times r}$ and $\mathbf{C}_{r \times n}$ as obtained above are indeed rank factors of $\mathbf{A}$ with rank $\mathbf{r}$.

Remarks

If $\mathbf{A}_{m \times n} = \mathbf{B}_{m \times r}\mathbf{C}_{r \times n}$ is a rank factorization of a complex matrix $\mathbf{A}$, the Moore-Penrose inverse of $\mathbf{A}$ is given by

$$\mathbf{A}^{+} = \mathbf{C}^{*} (\mathbf{C}\mathbf{C}^{*})^{-1} (\mathbf{B}^{*}\mathbf{B})^{-1} \mathbf{B}^{*},$$

where * indicates complex conjugate transpose (Greville, 1960).

The above method makes it much easier to compute the generalized inverse in practical situations.

2.4 Best Linear Unbiased Estimators (BLUE) in terms of Weighted Least Squares Estimators

There have appeared articles on the equivalence of simple least squares estimators, (SLSE)s, and best linear unbiased estimators (BLUE)s of estimable functions of β in the linear model $\mathbf{Y} = \mathbf{X}\beta+\varepsilon$. Zyskind (1967) has given a bibliography. One of the main points proven is that, for any estimable function of β, a BLUE is a SLSE if and only if the $\mathbf{n}\times\mathbf{p}$ matrix $\mathbf{X}$ of rank $\mathbf{r}$ is a linear combination of exactly $\mathbf{r}$ eigenvectors of $\mathbf{V}$, the covariance matrix (Zyskind, 1969). The property of estimability is used in the proof. Our purpose is to show the set of weighted least squares estimators (WLSE)s of β is the same as the set of simple least squares estimator (SLSE)s of β under the same condition on $\mathbf{X}$ and the eigenvectors of $\mathbf{V}$, provided the eigenvectors are associated with positive eigenvalues, but with no reference to an estimable function. Subsequently, the development of the general form of a BLUE will show $\mathbf{X}$ is in the column space of the eigenvectors of $\mathbf{V}$ which are associated with positive eigenvectors when $\mathbf{0}$ is in the sample space of the linear function $\mathbf{m}'\mathbf{y}$. The converse of the original theorem will be developed using Zyskind's result.

The implication of statistical theory and practice is that we can easily develop more applications oriented examples of when we would wish to estimate $\lambda'\beta$, even if it is not estimable, than generally appear in the Mathematical literature [Zyskind (1969), Kruskal (1968). In the case of non-estimability λ will be equal $\mathbf{X}'\mathbf{a}+\mathbf{b}$, with $\mathbf{b}$ in $\mathbf{C}(\perp\mathbf{X}')$, (Graybill, 1961) where $\mathbf{C}(\mathbf{M})$ is the column space of $\mathbf{M}$ and $\mathbf{C}(\perp\mathbf{X}')$ is the orthogonal complement of the column space of $\mathbf{X}'$. If $\mathbf{b}'\mathbf{b}$ is small enough we still may wish to estimate $\lambda'\beta$ with $\lambda'\hat{\beta}$ for an appropriate estimator $\hat{\beta}$. Fortunately $\mathbf{b}'\mathbf{b}$ is easy to obtain. Solving $\mathbf{X}\lambda = \mathbf{X}\mathbf{X}'\mathbf{a}$ for $\mathbf{a}$ and substituting we get $\mathbf{b}'\mathbf{b} = \lambda'(\mathbf{I}-\mathbf{X}\mathbf{X}^{+})\lambda$. This is akin to a concept of Goldman and Zelen (1964). Their corollary 1.3 is easily adapted to show that if $\hat{\beta}$ is the least squares estimator, in the case $\mathbf{V} = \sigma^2\mathbf{I}$, then $\lambda'\hat{\beta}$ is the BLUE of $\mathbf{X}'\mathbf{a}$.

In the subsequent portion of this thesis we identify the (SLSE) as a solution to $(\mathbf{X}'\mathbf{X})\hat{\beta}(\mathbf{s}) = \mathbf{X}'\mathbf{y}$ and the (WLSE) as the solution of $(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})\hat{\beta}(\mathbf{w}) = \mathbf{X}'\mathbf{V}^{+}\mathbf{y}$. In this thesis $\mathbf{V}^{+}$ is the Moore-Penrose generalized inverse of the matrix $\mathbf{V}$, and our representation of the (WLSE) is a simple extension of a well-known theorem [Graybill (1969), pg. 159].

Theorem: If the coefficient matrix $\mathbf{X}$ is $\mathbf{n \times p}$ of rank $\mathbf{r \le p < n}$, the covariance matrix $\mathbf{V}$ is $\mathbf{n \times n}$ of rank $\ge \mathbf{r}$, $(\mathbf{P}_1, \mathbf{P}_2)$ is an $\mathbf{n \times n}$ set of orthogonal vectors of $\mathbf{V}$ such that $\mathbf{P}_1$ is $\mathbf{n \times r}$ and is associated with positive eigenvalues of $\mathbf{V}$, and $\mathbf{X} = \mathbf{P}_1\mathbf{A}$ for some $\mathbf{A}$; then $\{\hat{\mathbf{B}}(\mathbf{w})\} = \{\hat{\mathbf{B}}(s)\}$.

Note: If $\mathbf{V}$ is positive definite then the conditions on the rank of $\mathbf{V}$ are met as is the association of $\mathbf{P}_1$ with positive eigenvalues of $\mathbf{V}$.

Proof: The statement on the rank of $\mathbf{V}$ and $\mathbf{P}_1$ is associated with positive eigenvalues of $\mathbf{V}$ permits us to write

2.4.1 $$\mathbf{X}'\mathbf{V}^{+}\mathbf{X} = \mathbf{A}'\ \mathbf{P}_1'\ (\mathbf{P}_1, \mathbf{P}_2)\begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix} \begin{pmatrix} \mathbf{P}_1' \\ \mathbf{P}_2' \end{pmatrix} \mathbf{P}_1\ \mathbf{A}$$

where $\mathbf{D}_1$ is an $\mathbf{r \times r}$ diagonal matrix of positive elements and $\mathbf{D}_2$ is diagonal. Hence,

2.4.2 $$(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{+} = \mathbf{A}^{+}\ (\mathbf{A}'\mathbf{D}_1)^{+}$$

because $(\mathbf{A}'\mathbf{D}_1)$ is $\mathbf{p \times r}$ of rank $\mathbf{r}$ [Graybill (1969), pg.102]. We know

(2.4.3) $\hat{\beta}(\mathbf{w}) = (\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{+}\mathbf{X}'\mathbf{V}^{+}\mathbf{y} + [\mathbf{I} - (\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{+}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})]\mathbf{h}$

for any $\mathbf{p}\times\mathbf{1}$ vector $\mathbf{h}$ [Graybill (1969), pg 104]. Equivalently

2.4.4 $\hat{\beta}(\mathbf{w}) = \mathbf{A}^{+}(\mathbf{A}'\mathbf{D}_1)^{+}\mathbf{A}_1\mathbf{D}_1\mathbf{P}_1'\mathbf{y} + [\mathbf{I} - \mathbf{A}^{+}(\mathbf{A}'\mathbf{D}_1)^{+}\mathbf{A}'\mathbf{D}_1\mathbf{A}]\ \mathbf{h}.$

It is well known, however, for any matrix $\mathbf{K}$ that $(\mathbf{K}'\mathbf{K})^{+}\mathbf{K}'\mathbf{K} = \mathbf{K}^{+}\mathbf{K}$. Hence, on identifying $(\mathbf{A}'\mathbf{D}_1)$ in (2.4.4) as $\mathbf{K}$, and recognizing $\mathbf{AA}'$ is $\mathbf{r}\times\mathbf{r}$ of rank $\mathbf{r}$, we obtain

2.4.5 $\hat{\beta}(\mathbf{w}) = \mathbf{A}^{+}\mathbf{P}_1'\mathbf{y} + [\mathbf{I} - (\mathbf{A}^{+}\mathbf{A})]\ \mathbf{h}.$

On the other hand,

2.4.6 $\hat{\beta}(\mathbf{s}) = (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{y} + [\mathbf{I} - (\mathbf{X}'\mathbf{X})^{+}(\mathbf{X}'\mathbf{X})]\mathbf{h}$

$$= (\mathbf{A}'\mathbf{A})^{+}\mathbf{A}'\mathbf{P}_1'\mathbf{y} + [\mathbf{I} - (\mathbf{A}'\mathbf{A})^{+}\mathbf{A}'\mathbf{A}]\ \mathbf{h}$$

$$= \mathbf{A}^{+}\mathbf{P}_1'\mathbf{y} + [\mathbf{I} - (\mathbf{A}^{+}\mathbf{A}]\ \mathbf{h}.$$

Before starting the next part of the development we will prove a lemma which will be needed subsequently.

Lemma: $\mathbf{m}'\mathbf{y} = (\mathbf{q}+\mathbf{a})'\mathbf{y} = \mathbf{q}'\mathbf{y} + \mathbf{r}$, where $\mathbf{r}$ is a constant, is a BLUE of $\mathbf{E}\,[\mathbf{m}'\mathbf{y}]$ if and only if $\mathbf{q}'\mathbf{y}$ is a BLUE of $\mathbf{E}\,[\mathbf{q}'\mathbf{y}]$.

Proof: Certainly $\mathbf{q}'\mathbf{y}$ is a BLUE of $\mathbf{E}\,[\mathbf{q}'\mathbf{y}]$. Thus there exists a vector $\mathbf{k}$ such that $\mathbf{k}'\mathbf{y}$ is a BLUE of $\mathbf{E}[\mathbf{q}'\mathbf{y}]$. Then $(\mathbf{k}+\mathbf{a})'\mathbf{y} = \mathbf{k}'\mathbf{y} + \mathbf{r}$ is also a BLUE of $\mathbf{E}\,[\mathbf{m}'\mathbf{y}]$. Since the variance of $\mathbf{m}'\mathbf{y}$ is identical to the variance of $\mathbf{q}'\mathbf{y}$, which is at least as large as that of $\mathbf{k}'\mathbf{y}$, then $\mathbf{m}'\mathbf{y}$ is a BLUE of $\mathbf{E}\,[\mathbf{m}'\mathbf{y}]$ if and only if $\mathbf{k}=\mathbf{q}$.

The preceeding theorem only showed that $\{\beta(\mathbf{s})\}$ is identical to $\{\beta(\mathbf{w})\}$ under the stated conditions and does not provide a Gauss-Markov type theorem on (BLUE)s. To develop this further we start, for the sake of clarity of exposition, with some well-known results. Consider first non-singular $\mathbf{V}$ and that $\mathbf{m}'\mathbf{y}$ is a BLUE of its expectation, $\mathbf{m}'\mathbf{X}\beta$. Then, since a non-singular $\mathbf{L}$ exists with $\mathbf{LVL}' = \sigma^2\mathbf{I}$, we have $\mathbf{m}'\mathbf{y} = \mathbf{m}'\mathbf{L}^{-1}\mathbf{Ly} = (\mathbf{L}'^{-1}\mathbf{m})'\mathbf{z}$ is a BLUE of its expectation. Now $\mathbf{z}$ has mean $\mathbf{L}\times\beta$ and covariance matrix $\sigma^2\mathbf{I}$. For such a covariance matrix a BLUE is a SLSE, as is well known. Hence, $\mathbf{m}'\mathbf{y}$ is a BLUE of $\mathbf{m}'\mathbf{X}\beta$ if and only if

2.4.7 $$\mathbf{m}'\mathbf{y} = (\mathbf{L}'^{-1}\mathbf{m})'\mathbf{z} = \mathbf{m}'\mathbf{X}\,\{(\mathbf{X}'\mathbf{L}'\mathbf{LX})^{+}\mathbf{X}'\mathbf{L}'\mathbf{z} + [\mathbf{I} - (\mathbf{X}'\mathbf{L}'\mathbf{LX})^{+}(\mathbf{X}'\mathbf{L}'\mathbf{LX})]\,\mathbf{h}\}$$

for any $\mathbf{p}\times\mathbf{1}$ vector $\mathbf{h}$. One notes that this reduces to

2.4.8 $$\mathbf{m}'\mathbf{y} = \mathbf{m}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{+}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

on observing that

2.4.9 $$\mathbf{X}\ [\mathbf{I} - (\mathbf{X}'\mathbf{L}'\mathbf{L}\mathbf{X})^{+}(\mathbf{X}'\mathbf{L}'\mathbf{L}\mathbf{X})].$$

$$= \mathbf{L}^{-1}\mathbf{L}\mathbf{X}[\mathbf{I}-(\mathbf{L}\mathbf{X})^{+}\mathbf{L}\mathbf{X}] = \mathbf{0}$$ [Graybill (1969), pg. 111]

Also, since $\mathbf{Z}$ has variance $\sigma^2\mathbf{I}$, $[(\mathbf{L}')^{-1}\mathbf{m}]'\mathbf{z}$ is a BLUE of its expectation if and only if $(\mathbf{L}')^{-1}\mathbf{m}\ \varepsilon\ \mathbf{C}(\mathbf{L}\mathbf{X})$. Equivalently, it is a BLUE if and only if $\mathbf{Vm}=\mathbf{Xb}$ for some vector $\mathbf{b}$.

However, if V is of rank $\mathbf{q} < \mathbf{n}$ an orthogonal matrix $\mathbf{P} = (\mathbf{P}_1, \mathbf{P}_2)$, where $\mathbf{P}_1$ is $\mathbf{n}\times\mathbf{q}$ and $\mathbf{P}_2$ is $\mathbf{n}\times(\mathbf{n}-\mathbf{q})$, exists such that

2.4.10 $$\mathbf{P}'\mathbf{V}\mathbf{P} = \begin{pmatrix} \mathbf{D}_{q\times q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

We have then the BLUE of $\mathbf{m}'\mathbf{X}\beta$ is

2.4.11 $$\mathbf{m}'\mathbf{y} = \mathbf{m}'\mathbf{P}\mathbf{P}'\mathbf{y} = (\mathbf{m}'\mathbf{P}_1,\ \mathbf{m}'\mathbf{P}_2)\begin{pmatrix} \mathbf{P}_1'\mathbf{y} \\ \mathbf{P}_2'\mathbf{y} \end{pmatrix}$$

$$= \mathbf{m}'\mathbf{P}_1\mathbf{z}_1 + \mathbf{m}'\mathbf{P}_2\mathbf{z}_2.$$

But $\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix}$ has covariance matrix $\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ so that $\mathbf{m}'\mathbf{P}_2\mathbf{z}_2$ is a constant. Thus $\mathbf{m}'\mathbf{y}$ a BLUE of its expectation is identical to $\mathbf{m}'\mathbf{P}_1\mathbf{z}_1$ being a BLUE of its expectation, $(\mathbf{m}'\mathbf{P}_1)(\mathbf{P}_1'\mathbf{X}\boldsymbol{\beta})$. We have then by (2.4.7) and (2.4.8) that $\mathbf{m}'\mathbf{y}$ is a BLUE of its expecation if and only if

2.4.12 $$\mathbf{m}'\mathbf{y} = \mathbf{m}'\mathbf{P}_1\mathbf{P}_1'\mathbf{X}\left\{ (\mathbf{X}'\mathbf{P}_1\mathbf{D}^{-1}\mathbf{P}_1'\mathbf{X})^{+}\mathbf{X}'\mathbf{P}_1\mathbf{D}^{-1}\mathbf{z}_1 \right\} + \mathbf{m}'\mathbf{P}_2\mathbf{P}_2'\mathbf{X}\boldsymbol{\beta}.$$

The second term of (2.4.12) follows since $\mathbf{m}'\mathbf{P}_2\mathbf{z}_2$ is a constant which is identically equal to $\mathbf{m}'\mathbf{P}_2\mathbf{P}_2'\mathbf{y}$, which as noted by Goldman and Zelen (1964), must equal its expected value.

By our earlier comments a necessary and sufficient condition for $\mathbf{m}'\mathbf{P}_1\mathbf{z}_1$ to be a BLUE of $\mathbf{m}'\mathbf{P}_1\mathbf{P}_1'\mathbf{X}\boldsymbol{\beta}$ is that $\mathbf{D}\mathbf{P}_1'\mathbf{m} = \mathbf{P}_1'\mathbf{X}\mathbf{b}$ for some vector $\mathbf{b}$. Equivalently $\mathbf{m}'\mathbf{y}$ is a BLUE of its mean if and only if

2.4.13 $$\mathbf{V}\mathbf{m} = \mathbf{P}_1\mathbf{D}\mathbf{P}_1'\mathbf{m} = \mathbf{P}_1\mathbf{P}_1'\mathbf{X}\mathbf{b}.$$

(Actually for $\mathbf{m}'\mathbf{y}$ to be a BLUE it is known that $\mathbf{V}\mathbf{m} \in \mathbf{C}(\mathbf{X})$ (Zyskind, 1969), but for present purposes there is some merit in having our conclusion in the form of (2.4.13). From (2.4.13) it follows that we must have, for some $\mathbf{h}$

2.4.14 $$\mathbf{m} = \mathbf{V}^{+}\mathbf{P}_1\mathbf{P}_1{}'\mathbf{X}\mathbf{b} + (\mathbf{I}-\mathbf{V}^{+}\mathbf{V})\,\mathbf{h}.$$

Now the columns of $\mathbf{P}$ span $\mathbf{n}$ space, so we can write $\mathbf{X} = \mathbf{P}_1\mathbf{A}_1 + \mathbf{P}_2\mathbf{A}_2$ for some matrices $\mathbf{A}_1$ and $\mathbf{A}_2$. Also from (2.4.10) it follows that $\mathbf{V} = \mathbf{P}_1\mathbf{D}\mathbf{P}_1{}'$ and $\mathbf{V}^{+} = \mathbf{P}_1\mathbf{D}^{-1}\mathbf{P}_1{}'$. Thus (2.4.14) reduces to

2.4.15 $$\mathbf{m} = \mathbf{V}^{+}\mathbf{X}\mathbf{b} + \mathbf{P}_2\mathbf{P}_2{}'\mathbf{h}.$$

Subsituting (2.4.15) into the second term on the right of (2.4.12) and recalling the form of $\mathbf{V}^{+}$ we have $\mathbf{m}'\mathbf{y}$ is a BLUE of $\mathbf{m}'\mathbf{X}\beta$ if and only if

2.4.16 $$\mathbf{m}'\mathbf{y} = \mathbf{m}'\mathbf{P}_1\mathbf{P}_1{}'\mathbf{X}\ \{\ (\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{+}\mathbf{X}'\mathbf{V}^{+}\mathbf{y}\ \} + \mathbf{h}'\mathbf{P}_2\mathbf{A}_2\beta.$$

Now consider the ms for which $\mathbf{m}'\mathbf{y}$ is a BLUE of $\mathbf{m}'\mathbf{X}\beta$. We have shown the totality of all such ms is the set that satisfies (2.4.15) for some $\mathbf{b}$ and $\mathbf{h}$. Hence, if $\mathbf{0}$ is in the sample space of $\mathbf{m}'\mathbf{y}$ for all $\mathbf{m}$ (for example, if $\mathbf{y}$ has a multivariate normal distribution) then (2.4.16) shows that $\mathbf{h}'\mathbf{P}_2\mathbf{A}_2\beta = \mathbf{0}$ for any $\mathbf{h}$ and β. Thus $\mathbf{P}_2\mathbf{A}_2 = \mathbf{0}$, $\mathbf{X} = \mathbf{P}_1\mathbf{A}_1$, and (2.4.13) shows $\mathbf{V}\mathbf{m}\ \varepsilon\ \mathbf{C}(\mathbf{x})$, and (2.4.16) becomes

2.4.17 $$\mathbf{m}'\mathbf{y} = \mathbf{m}'\mathbf{X}\,(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{+}\mathbf{X}'\mathbf{V}^{+}\mathbf{y}.$$

The converse requires no comment on the sample space of $\mathbf{y}$. If $\mathbf{X} = \mathbf{P}_1\mathbf{A}_1$, where $\mathbf{P}_1$ is associated with positive eigenvalues of $\mathbf{V}$, then, using the general form of a BLUE in (2.4.12) we note (2.4.17) is the BLUE of its mean. We have then a somewhat more general result than corollary 1.1 of (Zyskind, 1969) in the sense that we have proved the following theorem.

<u>Theorem</u>: If $\mathbf{0}$ is in the sample space of $\mathbf{m}'\mathbf{y}$ for all m then $\mathbf{m}'\mathbf{y}$ is a BlUE of its expectation only if $\mathbf{X}$ is in the column space of the eigenvectors of $\mathbf{V}$ associated with positive eigenvalues. The vector $\mathbf{m}$ is of the form in (2.4.15) for some $\mathbf{h}$ and $\mathbf{b}$. The general form of a BLUE is given in (2.4.16), and for $\mathbf{0}$ in the sample space, by (2.4.17). Conversely, if $\mathbf{X}$ is in the column space of the positive eigenvectors then a BLUE is given by (2.4.17) and $\mathbf{m}$ has form (2.4.15).

Perhaps it should be pointed out that the restriction on $\mathbf{0}$ being in the sample space of $\mathbf{m}'\mathbf{y}$ is a minimal one. If $\mathbf{m}'\mathbf{y}$ is to be a BLUE of $\mathbf{m}'\mathbf{X}\beta$ for all β then it must be one when $\beta = \mathbf{0}$. Thus for a continuous sample space it is most reasonable that $\mathbf{0}$ be in it.

We are almost in position to prove a converse to the first theorem. Since the columns of $\mathbf{P}$ span $\mathbf{n}$ space we know that for some set of columns of $\mathbf{P}$ which are assembled in the matrix $\mathbf{P}_I$, where $\mathbf{P}_I$ is $\mathbf{n} \times \mathbf{k}$ and $\mathbf{k} \leq \mathbf{n}$, $\mathbf{X}_I = \mathbf{P}_I\mathbf{A}$, if $\mathbf{X}_I$ is a matrix of the $\mathbf{r}$ independent columns of $\mathbf{X}$. Let $\mathbf{q} = \boldsymbol{min} \left\{ \mathbf{k} \mid \mathbf{X}_I = \mathbf{P}_I\mathbf{A} \right\}$.

<u>Corollary</u>: Let $\mathbf{X}$ be $\mathbf{n} \times \mathbf{p}$ of rank $\mathbf{r} \leq \mathbf{p}$ and $\mathbf{X}_I$ a set of $\mathbf{r}$ independent columns of $\mathbf{X}$. Let the related $\mathbf{n} \times \mathbf{q}$ matrix (as described in the preceeding paragraph) be $\mathbf{P}_I$ which is made up of orthogonal eigenvectors of $\mathbf{V}$ and each of its columns be associated with positive eigenvalues of $\mathbf{V}$. Then if $\left\{ \hat{\beta}(\mathbf{w}) \right\} = \left\{ \hat{\beta}(\mathbf{s}) \right\}$ we have $\mathbf{r} = \mathbf{q}$.

<u>Proof</u>: By the preceeding theorem, if $\lambda'\mathbf{b}$ is estimable then $\lambda = \mathbf{X}'\mathbf{m}$ for some $\mathbf{m}$ and its BLUE is (2.4.17). However, by hypothesis the BLUE is identical to the SLSE. The result follows from Zyskind's conclusions (Zyskind, 1969).

2.5 Relationships Among Estimators in Linear Models

2.5.1 Introduction

In the linear models, in simultaneous equations systems, it has been observed that the traditional method of OLS (ordinary least squares) estimators fails to estimate the parameters consistently. Thus, the number of estimation procedures has been developed for estimating the parameters. They are classified in two (2) groups such as limited information methods and systems method. The families of limited information methods maximum are double K-class, K-class, two-stage least squares, limited information maximum likelihood. Following Dwivedi (1992), a linear relation is developed.

The families of systems method are full information maximum likelihood, 3SLS, double h-vector class, double K-matrix class estimators, and etc. For detail study regarding the relationship among these estimators see Srivastava & Tiwari (1990), Tiwari (1986), Srivastava & Tiwari (1986), and Srivastava & Tiwari (1977).

2.5.2 Model Specification & Estimators

Suppose the equation to be estimated is

$$\begin{aligned} \mathbf{y} &= \mathbf{Y}\gamma + \mathbf{X}_1\beta + \mathbf{u} \\ &= \mathbf{Z}\delta + \mathbf{u} \\ \mathbf{Z} &= [\,\mathbf{Y}\mathbf{X}_1\,] \quad \delta = \begin{pmatrix} \gamma \\ \beta \end{pmatrix} \end{aligned} \tag{2.5.1}$$

where $\mathbf{y}$ is a $\mathbf{T}\times\mathbf{1}$ column vector of observations on the jointly dependent variables, $\mathbf{Y}$ is a $\mathbf{T}\times\mathbf{m}$ matrix of observations on jointly dependent variables, $\mathbf{X}_1$ is a $\mathbf{T}\times\mathbf{L}$ matrix of predetermined variables, γ, β are associated coefficient vectors and $\mathbf{u}$ is a column vector of $\mathbf{T}$ unobserved structural disturbances.

Assuming the equation (2.5.1) to be identifiable the double k-class estimator as proposed by Nagar (1962) of δ is given by

$$\hat{\delta}_{Dkc} = [\,\mathbf{Z}'(\mathbf{I}-\mathbf{k}_1\mathbf{M})\,\mathbf{Z}\,]^{-1}\mathbf{Z}'(\mathbf{I}-\mathbf{k}_2\mathbf{M})\,\mathbf{y} \tag{2.5.2}$$

where $\mathbf{k}_1$ and $\mathbf{k}_2$ are the characterizing scalers and

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{2.5.3}$$

$\mathbf{X}$ being $\mathbf{T}\mathbf{x}\Lambda$ matrix assumed to be of full column rank, of observations on all the Λ predetermined variables in the model.

Notice that if $\mathbf{k}_1 = \mathbf{k}_2$, we get k-class estimators:

$$\hat{\delta}_{kc} = [\,\mathbf{Z}'(\mathbf{I} - \mathbf{k}_1\mathbf{M})\,\mathbf{Z}\,]^{-1}\mathbf{Z}' \; [\,\mathbf{I} - \mathbf{k}_1\,\mathbf{M}\,]\;\mathbf{y} \qquad (2.5.4)$$

and if we $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{1}$ we get the two stage least squares estimators:

$$\hat{\delta}_{2SLS} = [\,\mathbf{Z}'\mathbf{M}^*\mathbf{Z}\,]^{-1}\mathbf{Z}'\mathbf{M}^*\mathbf{y} \qquad (2.5.5)$$

where

$$\mathbf{M}^* = \mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'. \qquad (2.5.6)$$

2.5.3. Relationships among estimators

Consider equation (2.5.2)

$$\hat{\delta}_{Dkc} = [\,\mathbf{Z}'(\mathbf{I} - \mathbf{k}_1\mathbf{M})\,\mathbf{Z}\,]^{-1}\mathbf{Z}'\,[\,\mathbf{I} - \mathbf{k}_2\mathbf{M}\,]\;\mathbf{y}$$

$$= \hat{\delta}_{kc} + (\mathbf{k}_1 - \mathbf{k}_2)\;[\;\mathbf{Z}'(\mathbf{I} - \mathbf{k}_1\mathbf{M})\;\mathbf{Z}\,]^{-1}\mathbf{Z}'\mathbf{M}\;\mathbf{y} \qquad (2.5.7)$$

To proceed with our discussion we require the following useful result which we cite without proof. As a matter of fact the proof is quite straight forward.

Lemma: For any two matrices $\mathbf{Q}_1$ & $\mathbf{Q}_2$

$$(\mathbf{Q}_1+\mathbf{Q}_2)^{-1} = [\mathbf{I} - (\mathbf{Q}_1+\mathbf{Q}_2)^{-1}\mathbf{Q}_2]\ \mathbf{Q}_1^{-1} \tag{2.5.8}$$

provided $\mathbf{Q}_1^{-1}$ exists.

Now utilizing the above Lemma we can show that

$$[\mathbf{Z}'(\mathbf{I} - \mathbf{k}_1\mathbf{M})\mathbf{Z}]^{-1} = [\mathbf{k}_1\mathbf{Z}'\mathbf{M}^*\mathbf{Z} + (1-\mathbf{k}_1)\ \mathbf{Z}'\mathbf{Z}]^{-1}$$

$$= \frac{1}{\mathbf{k}_1}[\mathbf{I} - (1 - \mathbf{k}_1)\ [\mathbf{Z}'\ (\mathbf{I} - \mathbf{k}_1\mathbf{M})\ \mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{Z}]\quad (\mathbf{Z}'\mathbf{M}^*\mathbf{Z})^{-1} \tag{2.5.9}$$

Therefore

$$[\mathbf{Z}'(\mathbf{I} - \mathbf{k}_1\mathbf{M})\ \mathbf{Z}]^{-1}\mathbf{Z}'\ \mathbf{M}\ \mathbf{y}$$

$$= [\mathbf{Z}'(\mathbf{I} - \mathbf{k}_1\mathbf{M})\ \mathbf{Z}]^{-1}\mathbf{Z}'(\mathbf{I} - \mathbf{M}^*)\ \mathbf{y}$$

$$= [\mathbf{Z}'(\mathbf{I} - \mathbf{k}_1\mathbf{M})\ \mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{y} - [\mathbf{Z}'(\mathbf{I} - \mathbf{k}_1\mathbf{M})\mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{M}^*\mathbf{y}.$$

Now substituting from (2.5.9) to the 2nd term on the RHS of the above we obtain

$$[Z'(I - k_1 M) Z]^{-1} Z'My$$

$$= [Z'(I-k_1 M) Z]^{-1} Z'y - \frac{1}{k_1}[I - (1-k_1) [Z'(I-k_1 M) Z]^{-1} Z'Z] \hat{\delta}_{2SLS} \tag{2.5.10}$$

Again using the Lemma we can write (2.5.10) in the following form:

$$1/1-k_1 \; [I - [Z'(I - k_1 M) Z]^{-1} k_1 (Z'M^*Z)] \hat{\delta}_{OLS} - 1/k_1 \; [I - (1-k_1) [Z'(I - k_1 M) Z]^{-1} Z'Z] \hat{\delta}_{2SLS} \tag{2.5.11}$$

Now using (2.5.7) and (2.5.11) we obtain

$$\hat{\delta}_{Dkc} = \hat{\delta}_{kc} + \frac{k_1-k_2}{1-k_1}\{ I - [Z'(I - k_1 M) Z]^{-1} (k_1 Z'M^*Z) \} \hat{\delta}_{OLS}$$

$$- \frac{k_1-k_2}{k_1}\{ I - (1 - k_1) [Z'(1 - k_1 M) Z]^{-1} Z'Z \} \hat{\delta}_{2SLS}.$$

We have therefore established the following:

Theorem: Consider the equation (2.5.1); then the double k-class estimator of the parameter, denoted by $\hat{\delta}_{Dkc}$, k-class estimator, two stage least squares estimator and ordinary least squares estimator

of the parameter denoted as $\hat{\delta}_{kc}$, $\hat{\delta}_{2SLS}$ and $\hat{\delta}_{OLS}$ respectively are connected by the following identity:

$$\hat{\delta}_{Dkc} = \hat{\delta}_{kc} + \frac{k_1 - k_2}{1 - k_1}\left\{ I - k_1 \left[Z' (I - k_1 M) Z \right]^{-1} (Z' M^* Z) \right\} \hat{\delta}_{OLS}$$

$$- \frac{k_1 - k_2}{k_2}\left\{ I - (1 - k_1) \left[Z' (Z - k_1 M) Z \right]^{-1} Z' Z \right\} \hat{\delta}_{2SLS}.$$

Chapter 3

Finite Sample Properties of Bootstrap Estimate of Two-Stage Least-Squares (2SLS): An Analytical Approach

3.1 Introduction

Several methods of estimation have been proposed for the parameters of a system of simultaneous equations in econometrics. The method of two-stage least square has been very widely used and investigated. This method, like many others in this context, suffers from the drawback that it produces a biased estimator. It still remains a popular method because of its computerized simplicity and consistency in large samples.

In general, the moments of 2SLS estimator do not exist (Mariano, 1972) but under certain regularity conditions, they have been explicitly obtained (Nagar, 1959), Chaubey et al. 1984]. The expressions obtained in these papers are not easily applicable for computation of bias and standard errors in practice. Freedman (1984) has investigated the asymptotic properties of the bootstrap estimator of 2SLS. However, Hsu et al (1986) pointed out that Freedman's "theoretical results do not apply to the non-large sample case". For this purpose, they applied the bootstrap method to reduce the bias of the 2SLS estimator and investigated it empirically.

The purpose of the present thesis is to investigate the finite sample properties of the bootstrap estimator of 2SLS. The approach is analytical and offers simple estimate of bias in finite samples. In this paper, we concentrate on studying the bias properties only. However, the method is general enough to investigate the properties of arbitrary moments; and this exercise is left for further work.

The plan of this section is as follows. In section 3.2, the simultaneous equations model is described in detail for the sake of completeness; and the method of 2SLS is also presented in this section. Section 3.3 considers the bootstrap method when the disturbances are sampled from a known distribution, and section 3.4 is devoted to the general case. The latter case is, of course, more practical, but the former one is used because it facilitates some derivations in the latter case. The introduction of bootstrap is done through what is known as "selection matrix", which simplifies the derivations. It may be mentioned that the use of selection matrix was also made by Sim (1989), Sim & Fisher (1991), and Nebebe & Sim (1991).

3.2 Simultaneous Equation Model and Two Stage Least Squares

3.2.1 Description of Simultaneous Estimation Model

In general, a system of **M** linear structural equations in **M** jointly dependent and **k** predetermined variables may be expressed in algebraic form as follows:

$$\gamma_{11} y_1(t) + \dots + \gamma_{M1} y_M(t) + \beta_{11} x_1(t) + \dots + \beta_{K1} x_k(t) = u_1(t)$$
$$\gamma_{1M} y_1(t) + \dots + \gamma_{MM} y_M(t) + \beta_{1M} x_1(t) + \dots + \beta_{KM} x_k(t) = u_M(t) \qquad (3.2.1.1)$$

for t = 1, 2, ..., T. Here y's are jointly dependent and x's are predetermined variables. We have assumed that T observations are available on each of these variables. Structural disturbances in successive equations are represented by $u_1(t)$, $u_2(t)$, ..., $u_M(t)$ respectively and γ's & β's are structural coefficients. We assume that the number of jointly dependent variables are the same as the number of equations. That is, we have a complete system.

In matrix notation, the system of equations can be written as:

$$Y\Gamma + X\beta = U \qquad (3.2.1.2)$$

where Y is a $\mathbf{T}\times\mathbf{M}$ matrix of jointly dependent variables, Γ is a $\mathbf{M}\times\mathbf{M}$ matrix of parameters, X is a $\mathbf{T}\times\mathbf{K}$ matrix of predetermined variables, β is a $\mathbf{K}\times\mathbf{M}$ matrix of parameters and $\mathbf{U}$ is a $\mathbf{T}\times\mathbf{M}$ matrix of disturbances. Assuming that Γ is non singular square matrix of order $\mathbf{M}$, we can write (3.2.1.2) as follows:

$$Y_{T\times M} = -X\beta\Gamma^{-1} + U\Gamma^{-1} \quad (3.2.1.3)$$

$$Y_{T\times M} = X_{T\times K}\,\Pi_{K\times M} + V_{T\times M} \quad (3.2.1.4$$

where

$$\Pi = -\beta\Gamma^{-1} \quad and \quad V = U\Gamma^{-1}.$$

Further we make the following standard assumptions (Theil, 1971):

(i) The elements of X are non stochastic and fixed in repeated samples.

(ii) Rank $X = \mathbf{K} < \mathbf{T}$

(iii) $\underset{T\to\infty}{\rho lim} \dfrac{X'X}{T} = \sum XX$ is a positive definite matrix.

(iv) $\underset{T}{\rho lim} \dfrac{X'U}{T} = 0$

We make further assumptions on the structural disturbances:

(a) $E[\ u_i(t)\ ] = 0$ for all $i = 1, 2, \ldots, M$, $t = 1, 2, \ldots, T$.

and

(b) $E[\ u_i(t)\ u_j(t')\ ] = \sigma_{ij}$ if $t = t'$

$= 0$ if $t \neq t'$.

The M-dimensional row vectors of U are independently and identically distributed.

Furthermore, the T rows of V are also independently and identically distributed such that

$$E(V) = 0, \frac{1}{T} E\ [V'\ V] = \Omega = \Gamma^{-1} \Sigma \Gamma^{-1} \qquad (3.2.1.5)$$

where

$$\Sigma = (\sigma_{ij}) .$$

Suppose that because of a priori restrictions, $\mathbf{m + 1 < M}$ jointly dependent and $\mathbf{K_1 < K}$ predetermined variables enter the equation with non-zero coefficients. Further, the structural coefficients have been normalized by dividing the entire equation by the coefficient of one of the jointly dependent variables. Then the structural equation, in terms of normalized coefficients, may be expressed as

$$y(t) = \sum_{i=1}^{m} \gamma_i y_i(t) + \sum_{j=1}^{K_1} \beta_j x_j(t) + u(t) \qquad (3.2.1.6)$$

where the jointly dependent variable with "unit" coefficient is placed on the left-hand side and all other jointly dependent variables on the right-hand side. The coefficient γ's and β's are ratios of the original structural coefficients, which are parameters of interest (Thiel, 1971).

We can write (3.2.1.6) in matrix format as follows:

$$y = Y_1 \gamma + X_1 \beta + u \qquad (3.2.1.7)$$

where y is a $T \times 1$ column vector, Y_1 and X_1 are matrices of the order $T \times m$ and $T \times K_1$ respectively. The coefficient vectors γ and β are $m \times 1$ and $K \times 1$ respectively and u is a $T \times 1$ vector of structural disturbances.

One should note that y and Y_1 are submatrices of Y and X_1 is a submatrix of X. Supposing that the columns of Y are rearranged so that the columns of y and Y_1 occur first in Y and the columns of X_1 occur first in X. Then

$$Y = [\, y \;\vdots\; Y_1 \;\vdots\; Y_2 \,] \quad \textit{and} \quad X = [\, X_1 \;\vdots\; X_2 \,]$$

where Y_2 is the $\mathbf{T \times (M\text{-}m\text{-}1)}$ matrix of observations on those jointly dependent variables which are excluded from (3.2.1.7) and X_2 is $\mathbf{T \times K_2}$ matrix of observations on $\mathbf{K - K_1 = K_2}$ excluded predetermined variables.

Accordingly, the reduced form (3.2.1.4) can be written as

$$y = X_1\pi^* + X_2\pi_1 + v \tag{3.2.1.8}$$

$$Y_1 = X_1\Pi^* + X_2\Pi_1 + V_1 \tag{3.2.1.9}$$

$$Y_2 = X_1\Pi_2^* + X_2\Pi_2 + V_2. \tag{3.2.1.10}$$

so that (3.2.1.8) is that part of the complete reduced form which corresponds to the jointly dependent variable on the left side in (3.2.1.7); similarly, (3.2.1.9) corresponds to the right hand side jointly dependent variables of (3.2.1.7) and (3.2.1.10) corresponds to the jointly dependent variables excluded from (3.2.1.7). We should note that

$$\pi^* \textit{ is } K_1 \times 1,\ \pi_1 \textit{ is } K_2 \times 1,$$

$$\Pi^* \textit{ is } K_1 \times m,\ \Pi_1 \textit{ is } K_2 \times m,$$

$$\Pi_2^* \textit{ is } K_1 \times M\text{-}m\text{-}1,\ \Pi_2 \textit{ is } K_2 \times M\text{-}m\text{-}1 \tag{3.2.1.11}$$

and

$$\boldsymbol{v} \text{ is } T\times 1,\ \boldsymbol{V}_1 \text{ is } T\times m \text{ and } \boldsymbol{V}_2 \text{ is } T \times M\text{-}m\text{-}1. \qquad (3.2.1.12)$$

Now we may write

$$\Pi = \begin{pmatrix} \pi^* & \Pi^* & \Pi_2^* \\ \pi_1 & \Pi_1 & \Pi_2 \end{pmatrix}, V = (\, v \ V_1 \ V_2 \,)$$

Without any loss of generality, let us assume that (3.2.1.7) happens to be the first equation of the complete structural equation system (3.2.1.4). Then

$\begin{bmatrix} 1 \\ -\gamma \\ O_{M-m-1} \end{bmatrix}$ is the first column of Γ and $\begin{bmatrix} -\beta \\ O_{K_2} \end{bmatrix}$ is the first column of β where O_{M-m-1} and O_{K_2} are column vectors of M-m-1 and K_2 "zero" elements respectively, also $K_2 = K - K_1$ as defined earlier. Since the structural parameters are related with the reduced form parameters as

$$\Pi = -\mathbf{B}\Gamma^{-1} \text{ and } V = \mathrm{U}\Gamma^{-1}$$

or

$$\Pi\Gamma = -\mathbf{B} \text{ and } V\Gamma = U,$$

we have

$$\Pi \begin{bmatrix} 1 \\ -\gamma \\ O_{M-m-1} \end{bmatrix} = - \begin{bmatrix} -\beta \\ O \end{bmatrix}$$

and

$$V \begin{bmatrix} 1 \\ -\gamma \\ O_{M-m-1} \end{bmatrix} = u.$$

The important thing to note here is the following:

$$(v \quad V_1 \; V_2) \begin{bmatrix} 1 \\ -\gamma \\ O_{M-m-1} \end{bmatrix} = u.$$

$$v - V_1 \gamma = u$$

(3.2.1.12)

Hence,

$$\frac{1}{T} E \; [v - V_1 \gamma] \; ' \; [v - V_1 \gamma] = \frac{1}{T} E \; (u'u) = \sigma^2_{(say)}. \qquad (3.2.1.13)$$

3.2.2 Two-Stage Least Squares Estimation

To obtain the 2SLS estimator (3.2.1.7), we proceed as follows: First, note that we can write

$$Y_1 = X_1 \Pi^* + X_2 \Pi_1 + V_1$$

as

$$Y_1 = (X_1 \quad X_2) \begin{bmatrix} \Pi^* \\ \Pi_1 \end{bmatrix} + V_1 \qquad (3.2.2.1)$$

or

$$Y_1 = X\pi + V_1 \qquad (3.2.2.2)$$

where X is a $\mathbf{T \times K}$ matrix and π is a $\mathbf{K \times m}$ matrix of parameters and $\mathbf{V_1}$ is a $\mathbf{T \times m}$ matrix of the reduced form disturbances.

First, apply ordinary LS to each one of the reduced form equation in (3.2.2.2). Note that we can represent the estimator as follows:

$$\hat{\pi} = (X'X)^{-1} X' Y_1 , \qquad (3.2.2.3)$$

So

$$\hat{Y}_1 = X \hat{\pi} = X (X'X)^{-1} X' Y_1 . \qquad (3.2.2.4)$$

Now replacing Y_1 by $\hat{Y}_1$ in (3.2.1.7), we obtain

$$y = \hat{Y}_1 \gamma + X_1 \beta + u . \qquad (3.2.2.5)$$

$$= (\hat{Y}_1 \; X_1) \begin{pmatrix} \gamma \\ \beta \end{pmatrix} + u . \qquad (3.2.2.6)$$

Now applying LS again to (3.2.2.6), we obtain:

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix} = \begin{bmatrix} \hat{Y}_1' \hat{Y}_1 & \hat{Y}_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_1' y_1 \\ X_1' y_1 \end{bmatrix} \qquad (3.2.2.7)$$

We can represent this in the following form:

$$\textit{Let } Z = [\hat{Y} \; X_1] \; , \quad \hat{\delta} = \begin{bmatrix} \hat{\gamma} \\ \hat{\beta} \end{bmatrix} \quad \textit{and} \quad \delta = \begin{bmatrix} \gamma \\ \beta \end{bmatrix}$$

Then (3.2.2.7) can be written as

$$\hat{\delta} \;\; (Z'Z)^{-1} \; Z'y \qquad (3.2.2.8)$$

and (3.2.2.6) as $y = Z \delta + u.$

$$\therefore \hat{\delta} = (Z'Z)^{-1} \; Z'(Z\,\delta + u)$$
$$= \delta + (Z'Z)^{-1} \; Z'u.$$

Further, the estimator of σ^2 in (3.2.1.3) is given by

$$\hat{\sigma}^2 = \frac{1}{T}\, \hat{u}'\hat{u} \tag{3.2.2.9}$$

where $\hat{u} = y - Z\,\hat{\delta}$ denotes residual vector of the second stage least squares, which can be shown to be a consistent estimator.

3.3 THE BOOTSTRAP ESTIMATOR

The bootstrap methodology was first introduced by Efron (1979) and was adapted by Freedman (1984) who further studied the large sample properties of the bootstrap estimator of 2SLS. For simplicity, we first describe it as follows for the ease of known disturbances.

Let the components of the disturbance vector u be considered as a known random sample from some distribution $\mathbf{F}$. The bootstrap method considers drawing repeated random samples from the distribution $\mathbf{F}_T$ which puts equal probability mass $\frac{1}{T}$ at each of the components u_1, u_2, ..., u_T. Let u_j^* denote the j^{th} random sample (called the Bootstrap sample) in $\mathbf{J}$ repeated ramdom samples, $\mathbf{j}= 1,2,$ ... , J from $\mathbf{F}_T$. Then the boostrap consists of constructing the vectors $\mathbf{y}_j^*$ (for a known δ and $\mathbf{Z}$) as:

$$\mathbf{y}_j^* = \mathbf{Z}\ \delta + \mathbf{u}_j^* \qquad (3.3.1)$$

Note that this is not a realistic case, but is presented for the sake of clarity. Then we use the least squares method on (3.3.1) to get for the j^{th} bootstrap sample;

$$\delta_{(j)}^* = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_{(j)}^*$$
$$\hat{\mathbf{u}}_{(j)}^* = \mathbf{y}_j^* - \mathbf{Z}\ \delta_{(j)}^*$$
$$s_{(j)}^{2*} = T^{-1}\ \hat{\mathbf{u}}_{(j)}^{*\prime}\ \hat{\mathbf{u}}_{(j)}^*.$$

For a large $\mathbf{J}$, the bootstrap estimator of δ is given by

$$\delta^* = \frac{1}{J}\sum \delta_{(j)}^*. \qquad (3.3.2)$$

We now introduce an alternative representation of the bootstrap sample $\mathbf{u}_{(j)}^*$ through "selection matrix". Let $\mathbf{S}_{(j)}$ be a $\mathbf{T}\times\mathbf{T}$ matrix of rows consisting of zeroes, except at the index at which the camponent of $\mathbf{u}$ is selected where there is a unity. This will be called the "selection matrix" corresponding to the j^{th}

bootstrap sample. For example, let $\mathbf{T}=4$ and $\boldsymbol{u}^{*\prime}_{(j)} = (\boldsymbol{u}_1, \boldsymbol{u}_4, \boldsymbol{u}_2, \boldsymbol{u}_1)$ then

$$\boldsymbol{S}_{(j)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Note that each element of any row of $\boldsymbol{S}_{(j)}$ is distributed as a Bernouilli random variable with probability of success = $\frac{1}{T}$.

First we mention a few simple lemmas which will be used in proving the theorems.

<u>**Lemma 2.1**</u> (Sim & Fisher, 1991)

Let $\boldsymbol{T}$ be finite and $\mathbf{J}$ be the number of bootstrap replications, then

$$\left\{ \frac{1}{J} \sum_{j=1}^{J} \boldsymbol{S}_{(j)} \right\} \xrightarrow[as]{J} \boldsymbol{T}^{-1}\, \mathbf{E}_{T} \qquad (3.3.4)$$

where $\mathbf{E}_T$ is a $\mathbf{T}\times\mathbf{T}$ matrix of unities.

Proof:

The lemma is easily established by the use of a strong law of large numbers for the case of Bernoulli random variables.

Lemma 2.2

Let T be finite and $\mathbf{J}$ be the number of bootstrap replications, then

$$\left\{\frac{1}{J}\sum_{j=1}^{J} S_{(j)}{}'S_{(j)}\right\}\xrightarrow[as]{J} I_T \qquad (3.3.5)$$

Proof:

Let $S_{(j)} = (\mathbf{a}_{rs}^{(j)})$ where $\mathbf{a}_{rs}^{(j)}$ denotes the $(\mathbf{r},\mathbf{s})$ element of $S_{(j)}$. Note that the $(\mathbf{r},\mathbf{s})$ element of $S_{(j)}{}'S_{(j)}$ is zero when $r \neq s$ and for $\mathbf{r}=\mathbf{s}$ it equals $\mathbf{a}_r^{(j)}$ where $\mathbf{a}_r^{(j)} \simeq Bin\,(\mathbf{T}, 1/\mathbf{T})$. Hence the results again follow from the application of strong law of large numbers.

Now we can write (3.3.2) as follows;

$$\delta^* = \frac{1}{J}\sum_{j=1}^{J} (Z'Z)^{-1} Z' \, [Z\delta + s_{(j)}\, u]$$

$$= \delta + \frac{1}{J}\sum (Z'Z)^{-1} Z' \, s_{(j)}\, u$$

$$= \delta + (Z'Z)^{-1} Z' \left[\frac{1}{J} \sum s_{(j)} \right] u. \qquad (3.3.6)$$

Thus by lemma 2.1, we note that

$$\delta^* \xrightarrow[a.s.]{J} \delta + (Z'Z)^{-1} Z' \; T^{-1} \mathbf{E}_T \; u$$

which equals

$$\delta + (Z'Z)^{-1} Z' \; \bar{u} \; \mathbf{1}_T$$

where $\mathbf{1}_T$ denotes $\mathbf{T} \times \mathbf{1}$ vector of unities and $\bar{u} = T \sum_{i=1}^{T} u_i$. We state this as the following theorem.

Theorem 3.1

The bias of δ^* is zero when $\bar{u} = 0$ in a finite sample, otherwise it is given by

$$(Z'Z)^{-1} Z' \; \bar{u} \; \mathbf{1}_T \; .$$

By a similar use of lemma 2.2, we obtain the following.

Theorem 3.2

Let $s^2 = u'u / T$ and let T be finite, then

$$\frac{1}{J}\sum_{j=1}^{J} u_{(j)}^{*\prime} u_{(j)}^{*}$$

$$\left\{\frac{1}{J}\sum \; (s_{(j)}u)'(s_{(j)}u)\right\} \xrightarrow[a.s.]{J} Ts^2. \qquad (3.3.7)$$

Now we consider the case of unknown disturbances.

3.4. Bootstrap Estimator when the Disturbances are unknown:

The bootstrap algorithm in this case is as follows:

Step 1 Obtain $\hat{\delta}$, $\hat{u}$ according to the specification in (3.2.2.5).

Step 2 Construct $s_{(j)}$.

Step 3 Reconstruct the equation $y_{(j)}^* = Z\hat{\delta} + s_{(j)}\hat{u}$.

Step 4 Compute $\delta_{(j)}^* = (Z'Z)^{-1} Z' y_{(j)}^*$.

Step 5 Repeat it for $j = 1, 2, \ldots, J$.

Now we focus our attention to the bootstrap estimates obtained as above and obtain the following results.

Theorem 4.1

Let $\hat{u}$ be the residual vector at the second (2nd) stage of 2SLS and $\bar{\hat{u}} = \frac{1}{T}\sum_{i=1}^{T} \hat{u}_i$. Then for the finite **T**:

$$\hat{\delta}^* = \frac{1}{J}\sum_{j=1}^{J} \hat{\delta}_{(j)}^* \overset{a.s.}{\underset{J}{\rightarrow}} \hat{\delta} + (Z'Z)^{-1}Z' \; [\bar{\hat{u}}] \; 1_T. \qquad (3.4.1)$$

The proof of the above theorem follows along the same lines as theorem (3.1). We also note that the conditional bias of the bootstrap is zero if the average of the 2nd stage residual is zero. The term $(Z'Z)^{-1}Z' \; [\bar{\hat{u}}] \; 1_T$ may be regarded as the estimate of the bias of $\hat{\delta}$ due to the following theorem (see also Freedman, 1984).

Theorem 4.2

Let $\hat{\delta}^*$ be the boostrap estimator of δ based on a large sample and $\hat{\delta}$ be the two stage least square estimate of δ; then

$$\rho \lim_{T\to\infty} [(\hat{\delta}^* - \hat{\delta}) - (\hat{\delta} - \delta)] = 0$$

Proof:

Note that from (3.4.1)

$$\rho \lim_{T} (\hat{\delta}^* - \hat{\delta}) = \rho \lim_{T} [(Z'Z)^{-1} Z' \bar{\hat{u}} 1_T]$$

$$= \rho \lim_{T} \left[T (Z'Z)^{-1} \frac{1}{T} Z' E_T Q_z u \right]$$

where $Q_z = [I - Z(Z'Z)^{-1} Z']$. Hence

$$= \rho \lim_{T} (\hat{\delta}^* - \hat{\delta})$$

$$= \rho \lim_{T} [(Z'Z)^{-1} Z' E_T u - (Z'Z)^{-1} Z' E_T Z (Z'Z)^{-1} Z' u]$$

$$= \rho \lim \left[T(Z'Z)^{-1} \frac{1}{T} Z' E_T u - T(Z'Z)^{-1} \frac{1}{T} Z' E_T Z T (Z'Z)^{-1} \frac{1}{T} Z' u \right]$$

$$\rho \lim [T(z'z)^{-1}] \rho \lim \frac{1}{T} z' E_T u - \rho \lim T(z'z)^{-1} \rho \lim \frac{1}{T} z' E_T z \rho \lim T(z'z)^{-1} \rho \lim \frac{z'u}{T}$$

Note that all the $\rho \lim$'s are finite except $\rho \lim \frac{z' E_T u}{T}$ and $\rho \lim \frac{z'u}{T}$. These $\rho \lim$'s can be shown to be equal to zero as follows:

Since

$$z = [X\hat{\pi} \;\vdots\; X_1]$$

$$= \left[X\hat{\pi} \;\vdots\; (X_1 \;\vdots\; X_2)\begin{bmatrix} I \\ \cdots \\ O \end{bmatrix}\right]$$

$$= \quad X \left[\hat{\pi} \;\vdots \begin{bmatrix} I \\ \cdots \\ O \end{bmatrix}\right]$$

we have

$$\frac{1}{T} z'u \;=\; \left[\hat{\pi}' \;\vdots \begin{bmatrix} I \\ \cdots \\ O \end{bmatrix}\right]' \; \frac{X'u}{T}.$$

And by assumption (4) and the fact that $\rho\, lim\, \hat{\pi} = \Pi$; it follows that,

$$\rho \lim_{T} \frac{z'u}{T} \;=\; 0.$$

Similarly, $\rho\, lim\, \frac{1}{T}\, z'E_T\, u \;=\; 0$ also.

Further, since $\hat{\delta}$ is 2SLS estimator, it is well known that

$$\rho \lim_{T} (\hat{\delta}* - \delta) = 0$$

and therefore the theorem follows.

BIBLIOGRAPHY

Beran, R. "Jacknife Approximation to Bootstrap Estimate." Annals of Statistics (12), 1984: 101-118.

Bergstrom, M. Bootstrap Methods and Applications in Econometrics - A Brief Survey. Sweden: Department of Economics, 1999.

Bickel, P.J. & Freedman, D. A. "Some Asymptotic Theory for the Bootstrap." Annals of Statistics (9), 1981: 1196-1217.

Chaubey, Y.P. ; Dwivedi, T.D. ; Nagar, A. L. & Singh, M. "On the Exact Moments of Two Stage Least Squares Estimator of Structure Parameters in Simultaneous Linear Equation Models." Statistics Applications and New Directions, 1984.

Davidson, A. C. & Hinkley, D.V. Bootstrap Methods and their Application. Cambridge: Cambridge University Press, 1997.

Dhrymes, P. J. "An Identity Between Double K-Class and Two Stage Least Squares." International Economic Review (Vol. 10, No. 1), 1969: 114-117.

Dwivedi, T. D. "A Method of Rank Factorization." Sankhya (Sr. B), 1975.

Efron, B. & Tibshirani, R. J. An Introduction to the Bootstrap. New York: Chapman & Hall, 1993.

Efron, B. "Bootstrap Confidence Intervals for a Class of Parametric Problems." Biometrika [72(1)], 1985: 45-58.

Efron, B. "Bootstrap Methods: Another Look at the Jacknife." Annals of Statistics (7), 1979: 1-26. 26.

Efron, B. The Jacknife, The Bootstrap and Other Resampling Plans. Philadelphia: Society for Industrial and Applied Mathematics, 1982.

Ferrari, S. & Cribari-Neto, F. "On Bootstrap an Analytical Bias Correction." Economic Letters (58), 1998: 7-15.

Freedman, D.A. & Peters, S.C. "Bootstrapping a Regression Equation: Some Empirical Results". Journal of the American Statistical Association (79), 1984a: 97-106.

Freedman, D.A. "Bootstrapping Regression Models." Annals of Statistics (9), 1981: 1218-1228.

Freedman, D.A. "On Bootstrapping Two-Stage Least-Squares Estimates in Stationary Linear Models." Annals of Statistics [12 (3)], 1984: 827-842.

Geville, T.N.E. "Some Applications of the Pseudo-Inverse of a Matrix." SIAM [2] 1960: 15-22

Goldman, A. & Zelen, M. "Weak Generalized Inverses and Minimum Variance Linear Unbiased Estimation", Journal of Research of the National Bureau of Standards (68B), 1964: 151-172.

Graybill, F. An Introduction to Linear Statistical Models, Vol. 1. New York: McGraw-Hill Book Co. Incorporated, 1961.

Graybill F. Introduction to Matrices with Applications in Statistics. Belmont, California: Wadsworth Publishing Company, 1969.

Hsu, Y.; Lau, K.; Fung,H. & Ulveling, E.F. "Monte Carlo Studies on the Effectiveness of the Bootstrap Bias Reduction Method on 2SLS Estimates." Economic Letters (20): 233-239

Kolata, Gina. "Theorist Applies Computer Power to Uncertainty in Statistics." New York Times Nov 8, 1988.

Kruskal, W. "Parametric Augmentations and Error Structures Under Which Certain Simple Least Squares and Analysis of Variance Procedures are also Best." Journal of the American Statistical Association (64), 1969: 1353-1368.

Kruskal, W. "When are Gauss-Markov and Least Square Estimatimators Identical? A Coordinator-Free Approach." Annals of Mathematical Statistics (39), 1968: 70-75.

Lothgren, Mickael. How to Bootstrap DEA Estimators: A Monte Carlo Comparison. Stockholm: Department of Economic Statistics, Stockholm School of Economics, 1998.

MacKinnon J.G. & Smith Jr., A.A. "Approximate bias correction in econometrics." Journal of Econometrics (85), 1998: 205-230.

Maeshiro, A. "A Simple Mathematical Identity Among k-class Estimators." Journal of the American Statistical Association (Vol. LXI), 1966: pp. 368-374.

Mariano, R.S. "The Existence of Moments of the Ordinary Least Squares and Two-Stage Least Squares Estimators." Econometrica (40), 1972: 643-651

Nagar, A.L. "Double k-class Estimators of Parameters in simulataneous and then Small Sample Properties." International Economic Review (3), 1962: 168-188.

Nagar, A.L. "The Bias and the Moment Matrix of the General *k* Class Estimators of the Parameters in Simultaneous Equations." Econometrica (27), 1959.

Nebebe, F. & Sim, A.B. "The Relative Performances of Improved Ridge Estimators and an Empirical Bayes estimator: Some Monte Carlo Results. Commun. Statist-Theory Meth. 19 (9), 1990: 3469-3495.

Oi, Walter. "On the Relationship Among Different Members of the k-class." International Economic Review (Vol. 10, No. 1), 1969: 36-45.

Sim, A.B. & Fisher, G. Bootstrap Regression Estimates and Confidence Intervals: Some Finite Sample and Monte Carlo Results. Montreal: Concordia University, Preprint, Dept. of Economics, 1991.

Sim, A.B. Bootstrapping Single Equation Regression Models: Some Finite Sample Results. Montreal: Concordia University, PhD thesis in the Department of Economics, 1989.

Singh, K. "On Asymptotic Accuracy of Efron's Bootstrap". Annals of Statistics (9), 1981: 1187-1195.

Srivastava, V.K. & Tiwari, Ramji. "A Simple Derivation of the Identity Connecting Double K-Class and Two-Stage Least Squares Estimators." Sankhya (C) (vol. 39), 1977: 2.

Srivastava, V.K. & Tiwari, Ramji. "An Identity Between Double K-Matrix Class and Three-Stage Least Squares Estimators." Journal Quantitative Economics (Vol. 6, No. 1), 1990.

Srivastava, V.K. & Tiwari, Ramji. K-Matrix Class as a weighted average of the Modified Three-Stage Least Squares and the Modified Estimated Generalised Least Squares Estimators. Presented in the UGC sponsored seminar on Migration Research in Context of Development, held at Bananas Hindu University, Varanasi, 1986.

Theil, H. Principles of Econometrics. New York: John Wiley and Sons, 1971.

Tiwari, Ramji. A Simple Derivation of the Identity Between Double h-vector class and Three-Stage Least Square Estimators. Presented at the seminar on role of applied sciences in Mineral Industry, held at Indian Schools of Mines, Dhanbad, 1986.

Wu, C.F.J. "Jacknife, Bootstrap and Other Resampling Methods in Regression Analysis". Annals of Statistics (14), 1986: 1261-1295.

Zyskind, G. "On Best Linear Estimation and a General Gauss-Markov Theorem in Linear Models with Arbitrary Non-Negative Covariance Structure." SIAM Journal of Applied Mathematics (17), 1969: 1190-1202.

Zyskind, G. "On Canonical Forms, Non-Negative Covariance Matrices and Best and Simply Least Squares Linear Estimators in Linear Model." Annals of Mathematical Statistics (38), 1967: 1092-1109.