

UNIVERSITY OF BAHRAIN
College of Science
Department of Mathematics



BIAS ESTIMATION AND JACKKNIFING IN REGRESSION

BY

NASEEMA MOHAMMED GHULOOM HASAN

**A Thesis Submitted in partial Fulfillment of the
Requirements for the Degree**

Of

Master of Science

**Under the Supervision of
Dr. Moh'd Akkas Ali**

January 2002



12th January 2003

The Thesis Defence Committee considers the thesis entitled:

"Bias Estimation and Jackknifing in Regression"

Submitted by *Naseema Mohammed Gloom Hassan* (I.D. No. 19829416) to the Department of Mathematics is satisfactory and acceptable for the Master of Science Degree in Mathematics.

Thesis Defence Committee:

Dr. Mohammed Akkas Ali
Supervisor

Department of Mathematics
University of Bahrain

Dr. Mariam Ahmed Al-Mannai
Internal Examiner

Department of Mathematics
University of Bahrain

Prof. Robert Michael Loynes
External Examiner

School of Mathematics
and Statistics
Sheffield, U.K.

Dr. A. Abusafiya
Chairman

Department of Mathematics
University of Bahrain

ACKNOWLEDGMENTS

By the grace of Allah , I have succeeded in completing my thesis for the Degree of Master of Science. I wish to express my gratitude and profound thanks to my supervisor Dr. Mohammed Akkas Ali, for his valuable advice during his supervision of this research. I am grateful to my family and my colleagues in the Department of Mathematics at the Directorate of Curricula, Ministry of Education, Kingdom of Bahrain for their continuous encouragement.

Finally, my heartfelt thanks go to all concerned.

Contents

Acknowledgments	ii
Introduction	v
Chapter 1: Regression	
1.1 Introduction	2
1.2 Point Estimation	6
1.2.1 Maximum Likelihood Estimators of β and σ^2	6
1.2.2 The Properties of $\hat{\beta}$ and $\hat{\sigma}^2$	7
1.3 Estimation of E (Y)	22
1.4 Analysis of Residuals	23
1.4.1 Properties of the Residuals	24
1.5 Testing a Statistical Hypothesis	27
1.5.1 The General Linear Hypothesis	28
1.5.2 Confidence Region for Regression Coefficients	33
Chapter 2: Biased Estimators	
2.1 Introduction	42
2.2 Multicollinearity	44
2.2.1 Exact – Multicollinearity	51
2.2.2 Near – Multicollinearity	52
2.3 James – Stein Estimation	59
2.4 Shrinkage Estimator	61
2.5 Ridge Regression	66
2.5.1 Properties of Ridge Estimator $\hat{\beta}_*$	68
2.5.2 Remarks	75
2.6 Principal Component Regression	77

Chapter 3: Errors of Measurement in the Independent Variables

3.1 Introduction	82
3.2 Errors in the Independent Variables of a General Linear Model	86
3.3 Simultaneous Estimation of Linear Functions of Parameters when Independent Variables are Subject to Measurement Errors	91
3.4 Estimation of Parameters in a General Linear Model when some of the Independent Variables are Missing	94
3.4.1 Methods of Estimation	96

Chapter 4: Reduction of Bias by Jackknifing

4.1 Introduction	100
4.2 Jackknifing	101
4.3 Bias and Mean Square Error of $\hat{\beta}_*$	108
4.4 Comparison of Bias	113

Concluions	118
-------------------	-----

References	120
-------------------	-----

Introduction

Multiple linear regression is one of the most widely used statistical methods. It is used by data analyst in nearly every field of science and technology as well as the social sciences, economics, finance, and many of the physical and chemical sciences. In vast majority of regression analyses the least squares approach is used to fit a linear model to the data. The coefficients of regression play an important role in fitting a linear model to the data. The accuracy of the model depends on the accuracy of the estimation of the parameters. The accuracy of an estimator can be affected by several factors, such as influential observation and errors of measurements. In this thesis an attempt is made to reduce the bias of the least squares estimator of the vector of unknown regression coefficients, when a subset of the observations is subject to measurement errors. The thesis consists of four chapters. Below a brief description of the contents of the thesis is given.

In chapter one, we discuss the estimation of unknown regression coefficients. Also we study the distribution and the properties of the estimators.

Chapter two, is devoted to the biased estimation in regression. Here we discuss the multicollinearity and four methods of estimation: James – Stein, shrinkage, ridge regression, and principal component regression.

In chapter three, we discuss the estimation of parameters of a simple linear regression model when some values of the independent variables are subject of rounding errors . We obtain bounds for the relative squared bias and mean square error of the estimator of linear combination of the regression coefficients. Impacts of errors of measurement on the simultaneous estimation of linear functions of the unknown regression coefficients are also considered.

Chapter four deals with a class of Jackknife estimators for the vector of unknown regression coefficients. Some principal properties of this class of estimators are observed.

Chapter 1

REGRESSION

1.1 Introduction:

One of the aims of science is to find, to describe, and to predict relationships among events in the world in which we live. One way that this is accomplished is by finding a formula or an equation that relates quantities in the real world. We may be interested, for example, in the relationship between temperature and pressure in a chemical process; or in the relationship between the number of apples on trees of an orchard and the amount of fertilizer the trees receive; or we may be interested in the relationship of supply, demand, and price of certain commodities; or in how a certain vaccine affects a disease; or in the relationship of rainfall, temperature, and humidity; or in the yields of various varieties of wheat.

Regression analysis enables us to ascertain and utilize a relation between a variable of interest, called the dependent or response variable, and one or more variables called independent, explanatory, or predictor variables.

Regression analysis is often used to predict the response variable from knowledge of the independent variables. At other times, regression analysis is utilized primarily for examining the nature of a relationship between the independent variables and the response variable.

In most empirical studies where interest centers on a dependent variable y , the value of y is not uniquely determined when the level of the independent variable x is specified. Relations where y is not uniquely determined from knowledge of x are called statistical relations. A statistical relation between two variables x and y is not exact; the value of y is not uniquely determined when the value of x is specified.

Regression models containing two or more independent variables are called multiple regression models. The model , for example,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon,$$

is a multiple regression model containing p independent variables , $x_1, x_2, \dots$, and x_p . $\beta_0, \beta_1, \dots, \beta_p$ are the regression coefficients. These coefficients are unknown . y is the dependent variable and ε is called the random error . $\beta_0, \beta_1, \dots, \beta_p$ are referred to as the parameters of the model.

For statistical analysis, we take n independent random observations on y corresponding to n selected levels of $x_1, x_2, \dots, x_p$. The i th observation of y can be expressed as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i, \text{ where } i = 1, 2, \dots, n.$$

The above model can be rewritten as:

$$Y = X\beta + \varepsilon, \quad (1.1)$$

where,

$$Y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix}',$$

an $n \times 1$ column vector of observations,

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix},$$

is an $n \times (p+1)$ matrix consisting of a column of ones, and p columns of observations on the p independent variables, $x_1, x_2, \dots$, and x_p .

$$\beta = \begin{pmatrix} \beta_0 & \beta_1 & \dots & \beta_p \end{pmatrix}',$$

is a $(p+1) \times 1$ vector of regression coefficients, and

$$\varepsilon = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & . & . & . & . & \varepsilon_n \end{pmatrix}',$$

is an $n \times 1$ unobservable vector of random errors .It is assumed $E(\varepsilon) = O$, where O is a null vector, and $cov(\varepsilon) = \sigma^2 V$, where V is a positive definite matrix, σ^2 is a positive scalar constant.

The nature of the model $Y = X\beta + \varepsilon$ usually depends on,

- * the distribution of ε .
- * the structure of the covariance matrix $\sigma^2 V$, and
- * the rank and structure of X .

Let us consider two cases for the distribution of ε .

Case (1): ε is distributed normally with mean vector zero and covariance matrix $\sigma^2 I$, that is, $\varepsilon \sim N(O, \sigma^2 I)$

Case (2) : ε has an unknown distribution with $E(\varepsilon) = O$, and $cov(\varepsilon) = \sigma^2 V$.

For case (1), we sometimes generalize slightly and consider the situation when ε is distributed $N(O, \sigma^2 V)$, where $\sigma^2 > 0$ and V is a known positive definite matrix. Case (1) is equivalent to saying that each component of ε is normally distributed with mean 0 and unknown variance σ^2 and that the components of ε are mutually independent. Once the distribution of Y or ε is specified, the investigator may consider the problem of estimation of parameters of the model.

In the following section we shall address the problem of estimation of the parameters $\beta_0, \beta_1, \dots, \beta_p$ and σ^2 .

1.2 Point Estimation

Let us obtain the maximum likelihood estimators (MLE) of β and σ^2 , and then we shall study some of their properties .

1.2.1 Maximum Likelihood Estimators (MLE) Of β And σ^2

Since Y is distributed $\sim N(X\beta, \sigma^2 V)$. The likelihood function is:

$$L(\beta, \sigma^2; Y, X) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\sigma^2 V|^{-\frac{n}{2}} \times \exp \left[\frac{-1}{2\sigma^2} (Y - X\beta)' V^{-1} (Y - X\beta) \right]$$

Using logarithms we get

$$\begin{aligned} \log L(\beta, \sigma^2 : Y, X) &= \frac{-n}{2} \log 2\pi - \frac{1}{2} \log |\sigma^2 V| - \frac{1}{2\sigma^2} [(Y - X\beta)' V^{-1} (Y - X\beta)] \\ &= \frac{-n}{2} \log 2\pi - \frac{1}{2} \log |\sigma^2 V| - \frac{1}{2\sigma^2} (Y' V^{-1} Y - 2\beta' X' V^{-1} Y + \beta' X' V^{-1} X \beta) \\ &= \frac{-n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |V| - \frac{1}{2\sigma^2} [(Y' V^{-1} Y - 2\beta' X' V^{-1} Y + \beta' X' V^{-1} X \beta)] \end{aligned}$$

To find the values of β and σ^2 that maximize the likelihood function, we find partial derivatives of $\log L(\beta, \sigma^2 : Y, X)$ with respect to β and σ^2 and set these derivatives equal to zero. We denote the solution of the equations for β and σ^2 when the derivatives are set equal to zero, by $\hat{\beta}$ and $\hat{\sigma}^2$. This gives

$$X' V^{-1} X \hat{\beta} = X' V^{-1} Y .$$

Since X is a matrix of full rank, and V is a positive definite matrix, therefore

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y = HY, \text{ where } H = (X'V^{-1}X)^{-1}X'V^{-1}, \quad (1.2)$$

and,

$$\hat{\sigma}^2 = \frac{1}{n} (Y - X\hat{\beta})' V^{-1} (Y - X\hat{\beta}). \quad (1.3)$$

1.2.2 The Properties Of $\hat{\beta}$ and $\hat{\sigma}^2$

Here we examine some properties of the estimators $\hat{\beta}$ and $\hat{\sigma}^2$ and determine their distributions.

Distribution of $\hat{\beta}$

It follows from (1.2) that $\hat{\beta}$ has $(p+1)$ - variate normal distribution.

To specify the distribution completely, we must find the mean vector and covariance matrix of $\hat{\beta}$. We have,

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y = HY,$$

where H is defined in (1.2)

$$E[\hat{\beta}] = E[HY] = H[E(Y)] = HX\beta = (X'V^{-1}X)^{-1}X'V^{-1}X\beta = \beta.$$

The covariance of $\hat{\beta}$ is given by

$$\begin{aligned}
cov(\hat{\beta}) &= E \left\{ \left[\hat{\beta} - E(\hat{\beta}) \right] \left[\hat{\beta} - E(\hat{\beta}) \right]' \right\} \\
&= E \left\{ [HY - E(HY)] [HY - E(HY)]' \right\} \\
&= E \left\{ [HY - HE(Y)] [HY - HE(Y)]' \right\} \\
&= E \left\{ [H(X\beta + \varepsilon) - HX\beta] [H(X\beta + \varepsilon) - HX\beta]' \right\} \\
&= E \left\{ [HX\beta + H\varepsilon - HX\beta] [HX\beta + H\varepsilon - HX\beta]' \right\} \\
&= E \left\{ [H\varepsilon] [H\varepsilon]' \right\} \\
&= E [H\varepsilon\varepsilon'H'] \\
&= H [E(\varepsilon\varepsilon')] H' \\
&= H [\sigma^2 V] H' \\
&= \sigma^2 (X'V^{-1}X)^{-1} X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-1} \\
&= \sigma^2 \left[(X'V^{-1}X)^{-1} \right].
\end{aligned}$$

Thus,

$$\hat{\beta} \sim N \left(\beta, \sigma^2 (X'V^{-1}X)^{-1} \right). \quad (1.4)$$

We now show that the $p+1$ estimators $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$, and $\hat{\sigma}^2$ are sufficient statistics for $\beta_0, \beta_1, \dots, \beta_p$ and σ^2 respectively. The probability density function (p.d.f) of Y is given by

$$f(Y : X; \beta, \sigma^2) = \left(\frac{1}{2\pi} \right)^{\frac{n}{2}} |\sigma^2 V|^{\frac{-1}{2}} \times \exp \left[\frac{-1}{2\sigma^2} (Y - X\beta)' V^{-1} (Y - X\beta) \right]. \quad (1.5)$$

Now, we consider the exponent,

$$\begin{aligned}
& (Y - X\beta)' V^{-1} (Y - X\beta) \\
&= \left[(Y - X\hat{\beta}) - X(\beta - \hat{\beta}) \right]' V^{-1} \left[(Y - X\hat{\beta}) - X(\beta - \hat{\beta}) \right] \\
&= (Y - X\hat{\beta})' V^{-1} (Y - X\hat{\beta}) - 2(\beta - \hat{\beta})' X' V^{-1} (Y - X\hat{\beta}) + (\beta - \hat{\beta})' X' V^{-1} X (\beta - \hat{\beta}) \\
&= (Y - X\hat{\beta})' V^{-1} (Y - X\hat{\beta}) - 2(\beta - \hat{\beta})' [X' V^{-1} Y - X' V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1} Y] \\
&\quad + (\beta - \hat{\beta})' X' V^{-1} X (\beta - \hat{\beta}) \\
&= (Y - X\hat{\beta})' V^{-1} (Y - X\hat{\beta}) - 2(\beta - \hat{\beta})' [X' V^{-1} Y - X' V^{-1} Y] + (\beta - \hat{\beta})' X' V^{-1} X (\beta - \hat{\beta}) \\
&= (Y - X\hat{\beta})' V^{-1} (Y - X\hat{\beta}) + (\beta - \hat{\beta})' X' V^{-1} X (\beta - \hat{\beta}) \\
&= n\hat{\sigma}^2 + (\beta - \hat{\beta})' X' V^{-1} X (\beta - \hat{\beta}) \tag{1.6}
\end{aligned}$$

If we substitute equation (1.6) into (1.5), we get

$$\begin{aligned}
f(Y : X; \beta, \sigma^2) &= \left[\left(\frac{1}{2\pi} \right)^{\frac{n}{2}} |\sigma^2 V|^{-\frac{n}{2}} \times \exp \frac{-1}{2\sigma^2} \left\{ n\hat{\sigma}^2 + (\beta - \hat{\beta})' X' V^{-1} X (\beta - \hat{\beta}) \right\} \right] \\
&= g(\hat{\beta}, \hat{\sigma}^2 : \beta, \sigma^2) \cdot h(Y),
\end{aligned}$$

where ,

$$\begin{aligned}
g(\hat{\beta}, \hat{\sigma}^2 : \beta, \sigma^2) &= \left[\left(\frac{1}{2\pi} \right)^{\frac{n}{2}} |\sigma^2 V|^{-\frac{n}{2}} \times \exp \frac{-1}{2\sigma^2} \left\{ n\hat{\sigma}^2 + (\beta - \hat{\beta})' X' V^{-1} X (\beta - \hat{\beta}) \right\} \right], \text{ and} \\
h(Y) &= 1.
\end{aligned}$$

Hence $\hat{\beta}$ and $\hat{\sigma}^2$ are jointly sufficient for β and σ^2 respectively.

Distribution of $\hat{\sigma}^2$

From (1.3) we have,

$$\begin{aligned}
 \hat{\sigma}^2 &= \frac{1}{n} (Y - X\hat{\beta})' V^{-1} (Y - X\hat{\beta}) \\
 n\hat{\sigma}^2 &= (Y - XHY)' V^{-1} (Y - XHY), \text{ where } H \text{ is given in (1.2)} \\
 &= Y'(I - XH)' V^{-1} (I - XHY) Y, \text{ let } K = (I - XH) \\
 &= Y' K' V^{-1} K Y \\
 &= (X\beta + \varepsilon)' K' V^{-1} K (X\beta + \varepsilon) \\
 &= (\beta' X' K' + \varepsilon' K') V^{-1} (KX\beta + K\varepsilon), \text{ since } KX = 0.
 \end{aligned}$$

Therefore,

$$n\hat{\sigma}^2 = \varepsilon' K' V^{-1} K \varepsilon,$$

where

$$K = (I - X(X'V^{-1}X)^{-1}X'V^{-1}) \quad (1.7)$$

The following theorem leads to the distribution of $\hat{\sigma}^2$.

Theorem 1 *Let the $n \times 1$ unobservable vector of random errors $\varepsilon = (\varepsilon_1 \varepsilon_2 \dots \varepsilon_n)'$ have a multivariate normal distribution with mean zero and dispersion matrix $\sigma^2 V$, then the quadratic form $\varepsilon' \left(\frac{K' V^{-1} K}{\sigma^2} \right) \varepsilon = \frac{n\hat{\sigma}^2}{\sigma^2}$ has a chi-square distribution with $(n - p - 1)$ degrees of freedom.*

Proof. Let $W = \varepsilon' \left(\frac{K' V^{-1} K}{\sigma^2} \right) \varepsilon = \frac{n\hat{\sigma}^2}{\sigma^2}$

The moment - generating function (*m.g.f*) of W is given by

$$\begin{aligned}
M_W(t) &= E[e^{Wt}] \\
&= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{Wt} \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\sigma^2 V|^{-\frac{1}{2}} \times \exp \left[\frac{-1}{2\sigma^2} \varepsilon' V^{-1} \varepsilon \right] d\varepsilon_1 \dots d\varepsilon_n \\
&= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\sigma^2 V|^{-\frac{1}{2}} \times \exp \left[\frac{-1}{2\sigma^2} (\varepsilon' V^{-1} \varepsilon - 2\sigma^2 Wt) \right] d\varepsilon_1 \dots d\varepsilon_n \\
&= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\sigma^2 V|^{-\frac{1}{2}} \times \exp \left[\frac{-1}{2\sigma^2} (\varepsilon' V^{-1} \varepsilon - 2\varepsilon' K' V^{-1} K \varepsilon t) \right] d\varepsilon_1 \dots d\varepsilon_n \\
&= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\sigma^2 V|^{-\frac{1}{2}} \times \exp \left[\frac{-1}{2\sigma^2} \varepsilon' (V^{-1} - 2K' V^{-1} K t) \varepsilon \right] d\varepsilon_1 \dots d\varepsilon_n \\
&= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\sigma^2 V|^{-\frac{1}{2}} \\
&\quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp \left[\frac{-1}{2\sigma^2} \varepsilon' V^{-\frac{1}{2}} \left(I - 2V^{\frac{1}{2}} K' V^{-1} K V^{\frac{1}{2}} t \right) V^{-\frac{1}{2}} \varepsilon \right] d\varepsilon_1 \dots d\varepsilon_n \\
&= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\sigma^2 V|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp \left[\frac{-1}{2\sigma^2} \varepsilon' V^{-\frac{1}{2}} Q V^{-\frac{1}{2}} \varepsilon \right] d\varepsilon_1 \dots d\varepsilon_n,
\end{aligned}$$

where $Q = (I - 2V^{\frac{1}{2}} K' V^{-1} K V^{\frac{1}{2}} t)$,

$$= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\sigma^2 V|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp \left[\frac{-1}{2} \varepsilon' \left(V^{\frac{1}{2}} Q^{-1} V^{\frac{1}{2}} \sigma^2 \right)^{-1} \varepsilon \right] d\varepsilon_1 \dots d\varepsilon_n$$

$$\begin{aligned}
&= \frac{\left| \sigma^2 V^{\frac{1}{2}} Q^{-1} V^{\frac{1}{2}} \right|^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}} \left| \sigma^2 V \right|^{\frac{1}{2}} \left| \sigma^2 V^{\frac{1}{2}} Q^{-1} V^{\frac{1}{2}} \right|^{\frac{1}{2}}} \\
&\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left[\frac{-1}{2} \varepsilon' \left(V^{\frac{1}{2}} Q^{-1} V^{\frac{1}{2}} \sigma^2 \right)^{-1} \varepsilon \right] d\varepsilon_1 \dots d\varepsilon_n.
\end{aligned}$$

Clearly,

$$\frac{1}{(2\pi)^{\frac{n}{2}} \left| \sigma^2 V^{\frac{1}{2}} Q^{-1} V^{\frac{1}{2}} \right|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left[\frac{-1}{2} \varepsilon' \left(V^{\frac{1}{2}} Q^{-1} V^{\frac{1}{2}} \sigma^2 \right)^{-1} \varepsilon \right] d\varepsilon_1 \dots d\varepsilon_n = 1$$

Thus, we have

$$\begin{aligned}
M_W(t) &= \frac{\left| \sigma^2 V^{\frac{1}{2}} Q^{-1} V^{\frac{1}{2}} \right|^{\frac{1}{2}}}{\left| \sigma^2 V \right|^{\frac{1}{2}}} \\
&= \frac{\left| \sigma^2 V \right|^{\frac{1}{2}} \left| Q^{-1} \right|^{\frac{1}{2}}}{\left| \sigma^2 V \right|^{\frac{1}{2}}} \\
&= |Q|^{\frac{-1}{2}} \\
&= \left[\left| I - 2V^{\frac{1}{2}} K' V^{-1} K V^{\frac{1}{2}} t \right| \right]^{\frac{-1}{2}}.
\end{aligned}$$

Since $K = I - X(X'V^{-1}X)^{-1}X'V^{-1} = I - XH$, $H = (X'V^{-1}X)^{-1}X'V^{-1}$

Therefore,

$$\begin{aligned}
V^{\frac{-1}{2}} K &= V^{\frac{-1}{2}} - V^{\frac{-1}{2}} X(X'V^{-1}X)^{-1}X'V^{-1} \\
&= \left[I - V^{\frac{-1}{2}} X(X'V^{-1}X)^{-1}X'V^{\frac{-1}{2}} \right] V^{\frac{-1}{2}},
\end{aligned}$$

and

$$V^{\frac{-1}{2}} K V^{\frac{1}{2}} = \left[I - V^{\frac{-1}{2}} X (X' V^{-1} X)^{-1} X' V^{\frac{-1}{2}} \right].$$

Next,

$$\begin{aligned} K' V^{\frac{-1}{2}} &= [I - V^{-1} X (X' V^{-1} X)^{-1} X'] V^{\frac{-1}{2}} \\ &= V^{\frac{-1}{2}} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{\frac{-1}{2}}, \end{aligned}$$

and

$$V^{\frac{1}{2}} K' V^{\frac{-1}{2}} = \left[I - V^{\frac{-1}{2}} X (X' V^{-1} X)^{-1} X' V^{\frac{-1}{2}} \right].$$

Thus,

$$\begin{aligned} V^{\frac{1}{2}} K' V^{\frac{-1}{2}} V^{\frac{-1}{2}} K V^{\frac{1}{2}} &= \left[I - V^{\frac{-1}{2}} X (X' V^{-1} X)^{-1} X' V^{\frac{-1}{2}} \right] \left[I - V^{\frac{-1}{2}} X (X' V^{-1} X)^{-1} X' V^{\frac{-1}{2}} \right] \\ &= I - V^{\frac{-1}{2}} X (X' V^{-1} X)^{-1} X' V^{\frac{-1}{2}} - V^{\frac{-1}{2}} X (X' V^{-1} X)^{-1} X' V^{\frac{-1}{2}} \\ &\quad + V^{\frac{-1}{2}} X (X' V^{-1} X)^{-1} X' V^{\frac{-1}{2}} V^{\frac{-1}{2}} X (X' V^{-1} X)^{-1} X' V^{\frac{-1}{2}} \\ &= I - V^{\frac{-1}{2}} X (X' V^{-1} X)^{-1} X' V^{\frac{-1}{2}}. \end{aligned}$$

Now we consider,

$$\begin{aligned} &\left(V^{\frac{1}{2}} K' V^{\frac{-1}{2}} V^{\frac{-1}{2}} K V^{\frac{1}{2}} \right) \left(V^{\frac{1}{2}} K' V^{\frac{-1}{2}} V^{\frac{-1}{2}} K V^{\frac{1}{2}} \right) \\ &= \left[V^{\frac{1}{2}} (I - X (X' V^{-1} X)^{-1} X' V^{-1})' V^{\frac{-1}{2}} V^{\frac{-1}{2}} (I - X (X' V^{-1} X)^{-1} X' V^{-1}) V^{\frac{1}{2}} \right] \\ &\quad \cdot \left[V^{\frac{1}{2}} (I - X (X' V^{-1} X)^{-1} X' V^{-1})' V^{\frac{-1}{2}} V^{\frac{-1}{2}} (I - X (X' V^{-1} X)^{-1} X' V^{-1}) V^{\frac{1}{2}} \right] \end{aligned}$$

$$\begin{aligned}
&= \left[V^{\frac{1}{2}} (I - V^{-1}X(X'V^{-1}X)^{-1}X') V^{-1} (I - X(X'V^{-1}X)^{-1}X'V^{-1}) V^{\frac{1}{2}} \right] \\
&\quad \cdot \left[V^{\frac{1}{2}} (I - V^{-1}X(X'V^{-1}X)^{-1}X') V^{-1} (I - X(X'V^{-1}X)^{-1}X'V^{-1}) V^{\frac{1}{2}} \right] \\
&= \begin{bmatrix} V^{\frac{1}{2}} (V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}) (I - X(X'V^{-1}X)^{-1}X'V^{-1}) \\ V (V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}) (I - X(X'V^{-1}X)^{-1}X'V^{-1}) V^{\frac{1}{2}} \end{bmatrix} \\
&= V^{\frac{1}{2}} \left\{ \begin{bmatrix} V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} + \\ V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \\ [(I - X(X'V^{-1}X)^{-1}X'V^{-1}) (I - X(X'V^{-1}X)^{-1}X'V^{-1})] \end{bmatrix} \right\} V^{\frac{1}{2}} \\
&= V^{\frac{1}{2}} \left\{ \begin{bmatrix} [V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}] \cdot \\ I - X(X'V^{-1}X)^{-1}X'V^{-1} - X(X'V^{-1}X)^{-1}X'V^{-1} + \\ X(X'V^{-1}X)^{-1}X'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \end{bmatrix} \right\} V^{\frac{1}{2}} \\
&= V^{\frac{1}{2}} [I - V^{-1}X(X'V^{-1}X)^{-1}X'] V^{-1} [I - X(X'V^{-1}X)^{-1}X'V^{-1}] V^{\frac{1}{2}} \\
&= V^{\frac{1}{2}} K' V^{-1} K \\
&= V^{\frac{1}{2}} K' V^{\frac{-1}{2}} V^{\frac{-1}{2}} K V^{\frac{1}{2}},
\end{aligned}$$

showing that $V^{\frac{1}{2}} K' V^{\frac{-1}{2}} V^{\frac{-1}{2}} K V^{\frac{1}{2}}$ is an idempotent matrix.

Hence,

$$\begin{aligned}
\text{Rank}(V^{\frac{1}{2}}K'V^{\frac{-1}{2}}V^{\frac{-1}{2}}KV^{\frac{1}{2}}) &= \text{Tr}(V^{\frac{1}{2}}K'V^{\frac{-1}{2}}V^{\frac{-1}{2}}KV^{\frac{1}{2}}) \\
&= \text{Tr} \left[I - V^{\frac{-1}{2}}X(X'V^{-1}X)^{-1}X'V^{\frac{-1}{2}} \right] \\
&= \text{Tr} \left[I - X'V^{\frac{-1}{2}}V^{\frac{-1}{2}}X(X'V^{-1}X)^{-1} \right] \\
&= \text{Tr} [I_{(n)} - I_{(p+1)}] \\
&= n - (p + 1) \\
&= n - p - 1.
\end{aligned}$$

Thus,

$$M_{\frac{n\hat{\sigma}^2}{\sigma^2}}(t) = (1 - 2t)^{\frac{-1}{2}(n-p-1)},$$

and therefore $\varepsilon' \left(\frac{K'V^{-1}K}{\sigma^2} \right) \varepsilon = \frac{n\hat{\sigma}^2}{\sigma^2}$ has a chi - square distribution with $(n - p - 1)$ degrees of freedom. ■

It follows from Theorem 1 that

$$M_{\frac{n\hat{\sigma}^2}{\sigma^2}}(t) = (1 - 2t)^{\frac{-1}{2}(n-p-1)}$$

Hence,

$$\begin{aligned}
M'_{\frac{n\hat{\sigma}^2}{\sigma^2}}(t) &= \frac{-1}{2}(n - p - 1)(1 - 2t)^{\frac{-1}{2}(n-p+1)}(-2) \\
&= (n - p - 1)(1 - 2t)^{\frac{-1}{2}(n-p+1)},
\end{aligned}$$

and

$$E \left[\frac{n\hat{\sigma}^2}{\sigma^2} \right] = M'_{\frac{n\hat{\sigma}^2}{\sigma^2}}(0) = (n - p - 1).$$

Showing that $\frac{n\hat{\sigma}^2}{(n - p - 1)}$ is an unbiased estimator of σ^2 .

Next,

$$\begin{aligned} M''_{\frac{n\hat{\sigma}^2}{\sigma^2}}(t) &= \frac{-1}{2}(n - p + 1)(n - p - 1)(1 - 2t)^{\frac{-1}{2}(n - p + 1)}(-2) \\ M''_{\frac{n\hat{\sigma}^2}{\sigma^2}}(0) &= (n - p + 1)(n - p - 1) \\ M''_{\frac{n\hat{\sigma}^2}{\sigma^2}}(0) &= n^2 - 2np + p^2 - 1. \end{aligned}$$

Hence,

$$\begin{aligned} Var\left(\frac{n\hat{\sigma}^2}{\sigma^2}\right) &= (n^2 - 2np + p^2 - 1) - (n - p - 1)^2 \\ &= (n^2 - 2np + p^2 - 1) - (n^2 - 2np - 2n + 2p + p^2 + 1) \\ &= 2n - 2p - 2 \\ &= 2(n - p - 1), \\ Var(n\hat{\sigma}^2) &= 2\sigma^4(n - p - 1) \end{aligned}$$

$$\begin{aligned} Var\left(\frac{n\hat{\sigma}^2}{n - p - 1}\right) &= \frac{2\sigma^4(n - p - 1)}{(n - p - 1)^2} \\ &= \frac{2\sigma^4}{(n - p - 1)}. \end{aligned}$$

Now we show that the random variables $\hat{\beta} = HY$ and $\hat{\sigma}^2 = \frac{Y'(K'V^{-1}K)Y}{n}$ are independent.

Theorem 2 Let $\hat{\beta} = HY$ and $\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})' V^{-1} (Y - X\hat{\beta})}{n - p - 1}$,

where $H = (X'V^{-1}X)^{-1} X'V^{-1}$. Then $\hat{\beta}$ and $\hat{\sigma}^2$ are independent.

Proof. We observe that $\hat{\beta}$ is linear in Y and,

$$\begin{aligned}\hat{\sigma}^2 &= \frac{Y'(I - XH)' V^{-1} (I - XH)Y}{n} \\ &= \frac{Y'K'V^{-1}KY}{n}, \text{ where } K = I - XH\end{aligned}$$

is quadratic in Y . It follows from Rawlings, Patula and Dickey (1998, p – 116) that $\hat{\beta}$ and $\hat{\sigma}^2$ are independent if $HV(K'V^{-1}K) = 0$. To show that here

$$HV(K'V^{-1}K) = 0,$$

we consider,

$$\begin{aligned}HV(K'V^{-1}K) &= HVK'(V^{-1}K) \\ &= (X'V^{-1}X)^{-1}X'V^{-1}V \left[(I - X(X'V^{-1}X)^{-1}X'V^{-1})' \right] (V^{-1}K) \\ &= (X'V^{-1}X)^{-1}X' \left[(I - V^{-1}X(X'V^{-1}X)^{-1}X') \right] (V^{-1}K) \\ &= \left[((X'V^{-1}X)^{-1}X' - (X'V^{-1}X)^{-1}X'V^{-1}X(X'V^{-1}X)^{-1}X') \right] (V^{-1}K) \\ &= \left[((X'V^{-1}X)^{-1}X' - (X'V^{-1}X)^{-1}X') \right] (V^{-1}K) \\ &= 0\end{aligned}$$

Hence $\hat{\beta}$ and $\hat{\sigma}^2$ are independent . ■

Theorem 3 Let $H = (X'V^{-1}X)^{-1}X'V^{-1}$ and $\hat{\beta} = HY$. Further, let $\tilde{\beta} = AY$ be any other linear estimator of β , where A is a $(p+1) \times n$ matrix of known constants. If $AX = I$, then

$$E[\tilde{\beta}] = \beta, \text{ and } cov[(\tilde{\beta} - \hat{\beta}), \hat{\beta}] = 0$$

Proof. Clearly $E[\tilde{\beta}] = AE[Y] = AX\beta = \beta$, provided $AX=I$.

Thus, if $AX=I$, then $\tilde{\beta}$ is unbiased estimator of β .

Next,

$$\tilde{\beta} = (A - H)Y + HY = (A - H)Y + \hat{\beta}$$

$$\tilde{\beta} - \hat{\beta} = (A - H)Y.$$

Now, by definition,

$$\begin{aligned} cov[(\tilde{\beta} - \hat{\beta}), \hat{\beta}] &= E\left[\left\{(\tilde{\beta} - \hat{\beta}) - E(\tilde{\beta} - \hat{\beta})\right\}\left\{\hat{\beta} - E(\hat{\beta})\right\}'\right] \\ &= E\left[\{(A - H)Y - E(A - H)Y\}\left\{\hat{\beta} - E(\hat{\beta})\right\}'\right] \end{aligned}$$

Since, $E(\hat{\beta}) = \beta$ and $E(Y) = E(X\beta + \varepsilon) = X\beta$, we have

$$\begin{aligned} cov[(\tilde{\beta} - \hat{\beta}), \hat{\beta}] &= E\left[\{(A - H)Y - (A - H)X\beta\}\{\beta' - \beta'\}\right] \\ &= E\left[\{(AY - HY) - (AX\beta - HX\beta)\}\{Y'H' - \beta'\}\right] \\ &= E\left[\{A(X\beta + \varepsilon) - H(X\beta + \varepsilon) - AX\beta + \beta\}\{(\beta'X' + \varepsilon')H' - \beta'\}\right] \\ &= E\left[\{AX\beta + A\varepsilon - HX\beta - H\varepsilon - AX\beta + \beta\}\{\beta'X'H' + \varepsilon'H' - \beta'\}\right] \\ &= E\left[\{A\varepsilon - HX\beta - H\varepsilon + \beta\}\{\beta'X'H' + \varepsilon'H' - \beta'\}\right]. \end{aligned}$$

Since, $HX = I$, the identity matrix, we have

$$\begin{aligned}
 cov \left[\left(\tilde{\beta} - \hat{\beta} \right), \hat{\beta} \right] &= E \left[\{A\varepsilon - \beta - H\varepsilon + \beta\} \{\beta' + \varepsilon' H' - \beta'\} \right] \\
 &= E \left[\{A\varepsilon - H\varepsilon\} \{\varepsilon' H'\} \right] \\
 &= E \left[A\varepsilon \varepsilon' H' - H\varepsilon \varepsilon' H' \right] \\
 &= E \left[A\varepsilon \varepsilon' H' \right] - E \left[H\varepsilon \varepsilon' H' \right] \\
 &= A\sigma^2 V H' - H\sigma^2 V H' \\
 &= \sigma^2 (A - H) V H' \\
 &= \sigma^2 (A - H) V V^{-1} X (X' V^{-1} X)^{-1} \\
 &= \sigma^2 (A - H) X (X' V^{-1} X)^{-1} \\
 &= \sigma^2 (AX - HX) (X' V^{-1} X)^{-1}.
 \end{aligned}$$

Note that $AX = HX$,

hence,

$$cov \left[\left(\tilde{\beta} - \hat{\beta} \right), \hat{\beta} \right] = \sigma^2 (AX - HX) (X' V^{-1} X)^{-1} = 0$$

This is the complete proof of the theorem. ■

Now, we show that $\hat{\beta} = (X' V^{-1} X)^{-1} X' V^{-1} Y = HY$ is the best linear unbiased estimator (BLUE) of β .

Let $\tilde{\beta} = AY$, be a linear estimator of β , where A is a $(p+1) \times n$ arbitrary matrix satisfying the condition $AX = I$.

Let

$$\begin{aligned}\tilde{\beta} &= AY = (A - H)Y + HY \\ &= (A - H)Y + \hat{\beta},\end{aligned}$$

where $\hat{\beta} = HY$, is the (MLE) of β and H is given in (1.2).

Note that

$$\begin{aligned}E(\tilde{\beta}) &= (A - H)E(Y) + E(\hat{\beta}) \\ &= (A - H)X\beta + \beta = \beta,\end{aligned}$$

and

$$(A - H)X = AX - HX = 0$$

Therefore,

$$\begin{aligned}\text{cov}(\tilde{\beta}) &= E\left[\left\{\tilde{\beta} - E(\tilde{\beta})\right\}\left\{\tilde{\beta} - E(\tilde{\beta})\right\}'\right] \\ &= E\left[\left\{(A - H)Y + (\hat{\beta} - \beta)\right\}\left\{(A - H)Y + (\hat{\beta} - \beta)\right\}'\right] \\ &= E\left[\left\{(A - H)(X\beta + \varepsilon) + (\hat{\beta} - \beta)\right\}\left\{(A - H)(X\beta + \varepsilon) + (\hat{\beta} - \beta)\right\}'\right] \\ &= E\left[\left\{(A - H)\varepsilon + (\hat{\beta} - \beta)\right\}\left\{(A - H)\varepsilon + (\hat{\beta} - \beta)\right\}'\right] \\ &= E\left[(A - H)\varepsilon\varepsilon'(A - H)'\right] + E\left[(\hat{\beta} - \beta)\varepsilon'(A - H)'\right] \\ &\quad + E\left[(A - H)\varepsilon(\hat{\beta} - \beta)'\right] + E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right],\end{aligned}$$

since

$$E[\varepsilon\varepsilon'] = \sigma^2 V,$$

and

$$\begin{aligned}
E \left[\varepsilon \left(\hat{\beta} - \beta \right)' \right] &= E \left[\varepsilon (HY - \beta)' \right] \\
&= E \left[\varepsilon \{ H(X\beta + \varepsilon) - \beta \}' \right] \\
&= E \left[\varepsilon \{ HX\beta + H\varepsilon - \beta \}' \right] \\
&= E \left[\varepsilon \{ \beta + H\varepsilon - \beta \}' \right] \\
&= E \left[\varepsilon \{ H\varepsilon \}' \right] \\
&= E \left[\varepsilon \varepsilon' \right] H \\
&= \sigma^2 V H,
\end{aligned}$$

hence we have

$$\begin{aligned}
\text{cov} \left(\check{\beta} \right) &= \sigma^2 (A - H) V (A - H)' + \sigma^2 H V (A - H)' + \\
&\quad \sigma^2 (A - H) V H' + \text{cov} \left(\hat{\beta} \right).
\end{aligned}$$

Now,

$$\begin{aligned}
V H' &= V \left[(X' V^{-1} X)^{-1} X' V^{-1} \right]' \\
&= V V^{-1} X (X' V^{-1} X)^{-1} \\
&= X (X' V^{-1} X)^{-1},
\end{aligned}$$

and,

$$H V = (V H')' = (X' V^{-1} X)^{-1} X'.$$

Therefore,

$$H V (A - H)' = (X' V^{-1} X)^{-1} X' (A - H)'$$

$$\begin{aligned}
&= (X'V^{-1}X)^{-1}(AX - HX)' \\
&= 0.
\end{aligned}$$

Similarly,

$$(A - H)VH' = 0.$$

This gives,

$$\text{cov}(\tilde{\beta}) = \sigma^2 (A - H)V(A - H)' + \text{cov}(\hat{\beta}).$$

Since, $(A - H)V(A - H)'$ is positive semi - definite matrix, thus

$$\text{cov}(\tilde{\beta}) \geq \text{cov}(\hat{\beta}).$$

Hence $\hat{\beta}$ is the (BLUE) of β , in the sense that

$$E(\hat{\beta}) = \beta,$$

and

$$\text{cov}(\hat{\beta}) \leq \text{cov}(\tilde{\beta}).$$

1.3 Estimation of E (Y)

An estimator of the mean vector of Y is

$$\begin{aligned}
\hat{Y} &= X\hat{\beta} = X(X'V^{-1}X)^{-1}X'V^{-1}Y \\
&= XHY,
\end{aligned}$$

where,

$$H = (X'V^{-1}X)^{-1}X'V^{-1}.$$

Note that , $\hat{Y}$ is a linear function of Y , with

$$\begin{aligned} E(\hat{Y}) &= XHE(Y) \\ &= X(X'V^{-1}X)^{-1}X'V^{-1}X\beta \\ &= X\beta. \end{aligned}$$

Hence $\hat{Y}$ is an unbiased estimator of the mean of Y .

$$\begin{aligned} Var(\hat{Y}) &= Var(X\hat{\beta}) \\ &= X Var(\hat{\beta})X' \\ &= X\sigma^2 [(X'V^{-1}X)^{-1}] X' \\ &= \sigma^2 X(X'V^{-1}X)^{-1}X'. \end{aligned}$$

1.4 Analysis of Residuals

The residuals are defined by $e_i = Y_i - \hat{Y}_i$, $i = 1, 2, \dots, n$, where Y_i is an observation and $\hat{Y}_i$ is the corresponding fitted value obtained by use of the fitted regression equation,

$$\hat{Y} = X\hat{\beta}, \text{ with } \hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y.$$

We can see from this definition that the residuals e_i are the differences between what is actually observed, and what is predicted by the regression equation; that is the amount which the regression equation has not been able to explain. Thus we can think of the e_i as the observed errors if the model is correct. The e_i 's are also called observed residuals.

residuals. The observed residuals, however, are not independent and do not have common variance.

Let,

$$\begin{aligned}
 e &= Y - X\hat{\beta} \\
 &= Y - X(X'V^{-1}X)^{-1}X'V^{-1}Y \\
 &= (I - X(X'V^{-1}X)^{-1}X'V^{-1})Y \\
 &= KY
 \end{aligned}$$

where K is defined in (1.7). The residuals vector e reflects the lack of agreement between the observed Y and estimated $\hat{Y}$.

1.4.1 Properties of the Residuals

Now we study some properties of the residuals. Obviously,

$$\begin{aligned}
 E[e] &= E[KY] \\
 &= KE[Y] \\
 &= KX\beta,
 \end{aligned}$$

where K is given in (1.7).

Since

$$KX = 0,$$

hence

$$E[e] = O,$$

where O is an $n \times 1$ vector of zeros.

$$\begin{aligned} \text{Var}(e) &= \text{Var}(KY) \\ &= K' \text{Var}(Y) K \\ &= K' \sigma^2 V K \\ &= \sigma^2 K' V K, \end{aligned}$$

if $Y \sim N(X\beta, \sigma^2 V)$, then $e = KY$ is distributed $\sim N(O, \sigma^2 K' V K)$.

We have

$$\begin{aligned} e &= KY \\ &= K(X\beta + \varepsilon), \end{aligned}$$

where

$$Y = X\beta + \varepsilon, \text{ and } K = I - XH.$$

So,

$$e = KX\beta + K\varepsilon, \text{ but } KX = 0.$$

Hence,

$$e = K\varepsilon.$$

That is, the elements of e are linear functions of elements of ε .

Next, we show that if the assumption of normality holds, then e and $\hat{\beta}$ are independent. We have $e = K\varepsilon \sim N(O, \sigma^2 K'VK)$ and $\hat{\beta} = HY \sim N(\beta, \sigma^2(X'V^{-1}X)^{-1})$.

Thus,

$$\begin{aligned}
cov(e, \hat{\beta}) &= E \left[\{e - E(e)\} \{\hat{\beta} - E(\hat{\beta})\}' \right] \\
&= E \left[(e - O) (\hat{\beta} - \beta)' \right] \\
&= E \left[(e) (\hat{\beta}' - \beta') \right] \\
&= E \left[(K\varepsilon)(Y'H' - \beta') \right] \\
&= E \left[(K\varepsilon Y'H' - K\varepsilon \beta') \right] \\
&= E \left[(K\varepsilon(X\beta + \varepsilon)'H' - K\varepsilon \beta') \right] \\
&= E \left[(K\varepsilon \beta' X'H' + K\varepsilon \varepsilon' H' - K\varepsilon \beta') \right] \\
&= E \left[(K\varepsilon \beta' + K\varepsilon \varepsilon' H' - K\varepsilon \beta') \right], \text{ since } X'H' = I \text{ and } E(\varepsilon) = 0 \\
&= E \left[(K\varepsilon \varepsilon' H') \right] \\
&= KE \left[\varepsilon \varepsilon' \right] H' \\
&= K\sigma^2 V H' \\
&= K\sigma^2 V V^{-1} X (X'V^{-1}X)^{-1} \\
&= \sigma^2 K X (X'V^{-1}X)^{-1} \\
&= 0, \text{ since } KX = 0.
\end{aligned}$$

Therefore, the normal vectors e and $\hat{\beta}$ are independent.

1.5 Testing a Statistical Hypothesis

The testing of a statistical hypothesis is the application of an explicit set of rules for deciding whether to accept the null hypothesis or to reject it in favor of the alternative hypothesis. Suppose, for example, that a statistician wants to test the null hypothesis $\theta = \theta_0$ against the alternative hypothesis $\theta = \theta_1$. In order to make a choice, he will generate a sample data by conducting an experiment and then compute the value of a test statistic, which will tell him for each possible outcome of the sample space what action to take. The test procedure, therefore, partitions the possible values of the test statistic into two subsets: an acceptance region for H_0 and a rejection region for H_0 .

The procedure just described can lead to two kinds of error. For instance, if the true value of the parameter θ is θ_0 and the statistician incorrectly concludes that $\theta = \theta_1$, he is committing an error referred to as a type I error. On the other hand, if the true value of the parameter θ is θ_1 and the statistician incorrectly concludes that $\theta = \theta_0$, he is committing a second kind of error referred to as a type II error.

Definition 1.1

1- Rejection of the null hypothesis when it is true is called a type I error; the probability of committing a type I error is denoted by α .

2- Acceptance of the null hypothesis when it is false is called a

type II error; the probability of committing a type II error is denoted by β .

It is customary to refer to the rejection region for H_0 as the critical region of the test, and to the probability of obtaining a value of the statistic inside the critical region when H_0 is true as the size of the critical region. Thus, the size of a critical region is just the probability α of committing a type I error. This probability is also called the level of significance of the test.

1.5.1 The General Linear Hypothesis

We consider the model

$$Y = X\beta + \varepsilon,$$

where $\varepsilon \sim (O, \sigma^2 V)$, Y, X and β are defined in (1.1).

A general linear hypothesis for β can be stated as,

$$H_0 : C\beta - \gamma = O,$$

where C is an $m \times (p+1)$ matrix of known values with $\text{Rank}(C) = m < (p+1)$, β is a $(p+1)$ dimensional vector of regression coefficients, γ is a $m \times 1$ vector of known constants and O denotes a null vector of appropriate dimension.

Let $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$ be the least square estimator for β . It is known from (1.4) that

$$\hat{\beta} \sim N(\beta, \sigma^2(X'V^{-1}X)^{-1}).$$

$$\text{Let } D = \frac{1}{\sigma} [C(X'V^{-1}X)^{-1}C']^{\frac{-1}{2}}.$$

Then, under H_o

$$[D(C\hat{\beta} - \gamma)] \sim N(O, \sigma^2 DC(X'V^{-1}X)^{-1}C'D').$$

Next,

$$\begin{aligned} \sigma^2 DC(X'V^{-1}X)^{-1}C'D' &= \sigma^2 \left\{ \frac{1}{\sigma} [C(X'V^{-1}X)^{-1}C']^{\frac{-1}{2}} [C(X'V^{-1}X)^{-1}C'] \right\} \\ &= I. \end{aligned}$$

Thus, under H_o

$$[D(C\hat{\beta} - \gamma)] \sim N(O, I).$$

Therefore,

$$W^* = (C\hat{\beta} - \gamma)' D' D (C\hat{\beta} - \gamma) \sim \chi_{(m)}^2.$$

Further, from Theorem 1 (page - 10) we have

$$W = \varepsilon' \left(\frac{K'V^{-1}K}{\sigma^2} \right) \varepsilon = \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-p-1)}^2.$$

Now, we show that the random variables W and W^* are independently distributed.

We replace γ in w^* by $C\beta$. This gives

$$\begin{aligned}
w^* &= \frac{1}{\sigma^2} (C\hat{\beta} - C\beta)' [C(X'V^{-1}X)^{-1}C']^{-1} (C\hat{\beta} - C\beta) \\
&= \frac{1}{\sigma^2} (\hat{\beta} - \beta)' C' [C(X'V^{-1}X)^{-1}C']^{-1} C(\hat{\beta} - \beta) \\
&= \frac{1}{\sigma^2} \left\{ \begin{array}{c} [(X'V^{-1}X)^{-1}X'V^{-1}(X\beta + \varepsilon) - \beta]' C' [C(X'V^{-1}X)^{-1}C']^{-1} C \\ [(X'V^{-1}X)^{-1}X'V^{-1}(X\beta + \varepsilon) - \beta] \end{array} \right\} \\
&= \frac{1}{\sigma^2} \left\{ \begin{array}{c} [(X'V^{-1}X)^{-1}X'V^{-1}X\beta + (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon - \beta]' C' [C(X'V^{-1}X)^{-1}C']^{-1} C \\ [(X'V^{-1}X)^{-1}X'V^{-1}X\beta + (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon - \beta] \end{array} \right\} \\
&= \frac{1}{\sigma^2} \left\{ \begin{array}{c} [\beta + (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon - \beta]' C' [C(X'V^{-1}X)^{-1}C']^{-1} C \\ [\beta + (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon - \beta] \end{array} \right\} \\
&= \frac{1}{\sigma^2} \left\{ [(X'V^{-1}X)^{-1}X'V^{-1}\varepsilon]' C' [C(X'V^{-1}X)^{-1}C']^{-1} C [(X'V^{-1}X)^{-1}X'V^{-1}\varepsilon] \right\} \\
&= \frac{1}{\sigma^2} \left\{ [H\varepsilon]' C' [C(X'V^{-1}X)^{-1}C']^{-1} C [H\varepsilon] \right\}, \text{ where } H = (X'V^{-1}X)^{-1}X'V^{-1}
\end{aligned}$$

Thus,

$$w^* = \frac{1}{\sigma^2} \varepsilon' \left\{ H' C' [C(X'V^{-1}X)^{-1}C']^{-1} C H \right\} \varepsilon.$$

Since w and w^* are quadratic forms , in ε , it is well - known that w and w^* will be independently distributed if

$$[K'V^{-1}K] [V] \left[H'C' [C(X'V^{-1}X)^{-1}C']^{-1} CH \right] = 0,$$

where $K = I - XH$, and $H = (X'V^{-1}X)^{-1}X'V^{-1}$.

Now we consider the matrix

$$[K'V^{-1}K] [V] = [K'V^{-1}KV] .$$

Note that $VH' = X(X'V^{-1}X)^{-1}$, $HX = I$ and

$$\begin{aligned} KVH' &= (I - XH) X(X'V^{-1}X)^{-1} \\ &= (X - XHX)(X'V^{-1}X)^{-1} \\ &= (X - X)(X'V^{-1}X)^{-1} \\ &= 0. \end{aligned}$$

Hence $K'V^{-1}KVH' = 0$, and therefore

$$[K'V^{-1}K] [V] \left[H'C' [C(X'V^{-1}X)^{-1}C']^{-1} CH \right] = 0.$$

Thus, w and w^* are independent, $w \sim \chi^2_{(n-p-1)}$ and $w^* \sim \chi^2_{(m)}$.

Therefore the statistic,

$$\begin{aligned}
 F &= \frac{\frac{W^*}{m}}{\frac{W}{(n-p-1)}} \\
 &= \frac{n-p-1}{m} \cdot \frac{W^*}{W} \\
 &= \frac{n-p-1}{m} \cdot \frac{(C\hat{\beta} - \gamma)' [C(X'V^{-1}X)^{-1}C']^{-1} (C\hat{\beta} - \gamma)}{n\hat{\sigma}^2}, \quad (1.8)
 \end{aligned}$$

has the F - distribution with $\nu_1 = m$ and $\nu_2 = n - p - 1$ degrees of freedom.

The statistic F can be used to test the hypothesis :

$$H_0 : C\beta - \gamma = O.$$

We reject $H_o : C\beta - \gamma = O$ if $F \geq F_\alpha(\nu_1, \nu_2)$, where $F_\alpha(\nu_1, \nu_2)$ is such that

$$P[F \geq F_\alpha(\nu_1, \nu_2)] = \alpha.$$

Otherwise , we accept H_o .

1.5.2 Confidence Region for Regression Coefficients

If we replace γ in (1.8) by $C\beta$, we get

$$F = \frac{n-p-1}{m} \cdot \frac{(C\hat{\beta} - C\beta)' [C(X'V^{-1}X)^{-1}C']^{-1} (C\hat{\beta} - C\beta)}{n\hat{\sigma}^2}.$$

This statistic has the F -distribution with $\nu_1 = m$ and $\nu_2 = n-p-1$ degrees of freedom.

Therefore,

$$P[F < F_\alpha(\nu_1, \nu_2)] = 1 - \alpha,$$

Hence, the region

$$\frac{n-p-1}{m} \cdot \frac{(C\hat{\beta} - C\beta)' [C(X'V^{-1}X)^{-1}C']^{-1} (C\hat{\beta} - C\beta)}{n\hat{\sigma}^2} < F_\alpha(\nu_1, \nu_2),$$

constitutes a $(1 - \alpha)$ 100% confidence region for $C\beta$.

Example 1

The following are the data on the years of experience (x_1), incomes (y) of five engineers working for a company and the number of years (x_2) they studied at the university,

$$\begin{array}{lcl} x_1 : & 2 & 3 \quad 5 \quad 10 \quad 2 \\ x_2 : & 4 & 2 \quad 5 \quad 2 \quad 4 \\ y : & 6 & 5 \quad 8 \quad 5 \quad 4 \end{array}.$$

Suppose that the data satisfy the model,

$$Y = X\beta + \varepsilon, \text{ where } \varepsilon \sim N(O, \sigma^2 I),$$

where ,

$$X = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 2 \\ 1 & 5 & 5 \\ 1 & 10 & 2 \\ 1 & 2 & 4 \end{bmatrix}, Y = \begin{bmatrix} 6 \\ 5 \\ 8 \\ 5 \\ 4 \end{bmatrix}, \text{ and } \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$

Now,

$$X'X = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 2 \\ 1 & 5 & 5 \\ 1 & 10 & 2 \\ 1 & 2 & 4 \end{bmatrix}' \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 2 \\ 1 & 5 & 5 \\ 1 & 10 & 2 \\ 1 & 2 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 22 & 17 \\ 22 & 142 & 67 \\ 17 & 67 & 65 \end{bmatrix}.$$

Thus,

$$(X'X)^{-1} = \frac{1}{1323} \begin{bmatrix} 4741 & -291 & -940 \\ -291 & 36 & 39 \\ -940 & 39 & 226 \end{bmatrix}$$

Next,

$$X'Y = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 2 \\ 1 & 5 & 5 \\ 1 & 10 & 2 \\ 1 & 2 & 4 \end{bmatrix}' \begin{bmatrix} 6 \\ 5 \\ 8 \\ 5 \\ 4 \end{bmatrix}$$

$$= \begin{bmatrix} 28 \\ 125 \\ 100 \end{bmatrix}.$$

It follows from (1.2) that,

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

$$\hat{\beta} = \frac{1}{1323} \begin{bmatrix} 4741 & -291 & -940 \\ -291 & 36 & 39 \\ -940 & 39 & 226 \end{bmatrix} \begin{bmatrix} 28 \\ 125 \\ 100 \end{bmatrix}$$

$$= \frac{1}{1323} \begin{bmatrix} 2373 \\ 252 \\ 1155 \end{bmatrix}$$

$$= \begin{bmatrix} 1.7937 \\ 0.1905 \\ 0.8730 \end{bmatrix}$$

Further,

$$\text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

$$= \frac{\sigma^2}{1323} \begin{bmatrix} 4741 & -291 & -940 \\ -291 & 36 & 39 \\ -940 & 39 & 226 \end{bmatrix}$$

$$\hat{Y} = X\hat{\beta}$$

$$= \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 2 \\ 1 & 5 & 5 \\ 1 & 10 & 2 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1.7937 \\ 0.1905 \\ 0.8730 \end{bmatrix}$$

$$= \begin{bmatrix} 5.6667 \\ 4.1112 \\ 7.1112 \\ 5.4447 \\ 5.6667 \end{bmatrix},$$

and

$$\begin{aligned} e &= Y - X\hat{\beta} \\ &= \begin{bmatrix} 6 \\ 5 \\ 8 \\ 5 \\ 4 \end{bmatrix} - \begin{bmatrix} 5.6667 \\ 4.1112 \\ 7.1112 \\ 5.4447 \\ 5.6667 \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} 0.3333 \\ 0.8888 \\ 0.8888 \\ -0.4447 \\ -1.6667 \end{bmatrix}.$$

Now, from (1.3) we have

$$\hat{\sigma}^2 = \frac{1}{n} (Y - X\hat{\beta})' V^{-1} (Y - X\hat{\beta}), \text{ here } V = I \text{ and } n = 5.$$

Thus,

$$\begin{aligned}
 \hat{\sigma}^2 &= \frac{1}{n} (Y - X\hat{\beta})' (Y - X\hat{\beta}) \\
 &= \frac{1}{5} \begin{bmatrix} 0.3333 \\ 0.8888 \\ 0.8888 \\ -0.4447 \\ -1.6667 \end{bmatrix}' \begin{bmatrix} 0.3333 \\ 0.8888 \\ 0.8888 \\ -0.4447 \\ -1.6667 \end{bmatrix} \\
 &= \frac{1}{5} (4.6668) \\
 &= 0.9334
 \end{aligned}$$

Let $H_0 : C\beta = \gamma$, where

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix}, \text{ and } \gamma = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ be the hypothesis of our interest.}$$

Then we have

$$C\hat{\beta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1.7937 \\ 0.1905 \\ 0.8730 \end{bmatrix}$$

$$= \begin{bmatrix} 1.7937 \\ -0.6825 \end{bmatrix},$$

and

$$C(X'X)^{-1}C' = \frac{1}{1323} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 4741 & -291 & -940 \\ -291 & 36 & 39 \\ -940 & 39 & 226 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \frac{1}{1323} \begin{bmatrix} 4741 & -291 & -940 \\ 649 & -3 & -187 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \frac{1}{1323} \begin{bmatrix} 4741 & 649 \\ 649 & 184 \end{bmatrix}$$

$$= \begin{bmatrix} 3.5835 & 0.4906 \\ 0.4906 & 0.13908 \end{bmatrix}.$$

Hence,

$$\left[C(X'X)^{-1}C' \right]^{-1} = \frac{1}{0.2577} \begin{bmatrix} 0.13908 & -0.4906 \\ -0.4906 & 3.5835 \end{bmatrix}$$

Therefore, from (1.8) we have

$$F = \frac{n-p-1}{m} \cdot \frac{(C\hat{\beta} - \gamma)' [C(X'V^{-1}X)^{-1}C']^{-1} (C\hat{\beta} - \gamma)}{n\hat{\sigma}^2}.$$

Setting $V = I$ and $\gamma = O$ in the above expression we get,

$$\begin{aligned}
F &= \frac{\begin{bmatrix} 1.7937 \\ -0.6825 \end{bmatrix}' \begin{bmatrix} 0.13908 & -0.4906 \\ -0.4906 & 3.5835 \end{bmatrix} \begin{bmatrix} 1.7937 \\ -0.6825 \end{bmatrix}}{(0.2577)(4.6668)} \\
&= \frac{3.3179}{1.2026} \\
&= 2.759.
\end{aligned}$$

Note that here $\nu_1 = m = \text{Rank}(C) = 2$ and $\nu_2 = n - p - 1 = 5 - 2 - 1 = 2$.

If we consider the level of significance $\alpha = 0.05$, then we have from the table

$$F_{\alpha}(\nu_1, \nu_2) = F_{0.05}(2, 2) = 19.0$$

Since the observed value of F is less than $F_{\alpha}(\nu_1, \nu_2)$, the null hypothesis H_0 cannot be rejected.

Finally, to find the $(1 - \alpha) 100\%$ confidence region for $C\beta$ we note that

$$P[F < F_{\alpha}(\nu_1, \nu_2)] = 1 - \alpha.$$

Considering $\alpha = 0.05$, we get

$$P\left[\frac{n-p-1}{m} \cdot \frac{(C\hat{\beta} - C\beta)' [C(X'V^{-1}X)^{-1}C']^{-1} (C\hat{\beta} - C\beta)}{n\hat{\sigma}^2} < 19.0\right] = 0.95.$$

Now,

$$\begin{aligned} C\hat{\beta} - C\beta &= \begin{bmatrix} 1.7937 \\ -0.6825 \end{bmatrix} - \begin{bmatrix} \beta_0 \\ \beta_1 - \beta_2 \end{bmatrix} \\ &= \begin{bmatrix} 1.7937 - \beta_0 \\ -0.6825 - (\beta_1 - \beta_2) \end{bmatrix}. \end{aligned}$$

Thus,

$$\begin{aligned} &\frac{(C\hat{\beta} - C\beta)' [C(X'V^{-1}X)^{-1}C']^{-1} (C\hat{\beta} - C\beta)}{n\hat{\sigma}^2} \\ &= \frac{\begin{bmatrix} 1.7937 - \beta_0 \\ -0.6825 - (\beta_1 - \beta_2) \end{bmatrix}' \begin{bmatrix} 0.13908 & -0.4906 \\ -0.4906 & 3.5835 \end{bmatrix} \begin{bmatrix} 1.7937 - \beta_0 \\ -0.6825 - (\beta_1 - \beta_2) \end{bmatrix}}{(0.2577)(4.6668)} \end{aligned}$$

Hence,

$$P \left[\begin{aligned} &0.13908 (1.7937 - \beta_0)^2 - 0.9812 \{-0.6825 - (\beta_1 - \beta_2)\} (1.7937 - \beta_0) \\ &+ 3.5835 \{-0.6825 - (\beta_1 - \beta_2)\}^2 < 22.85 \end{aligned} \right] = 0.95.$$

The values of β_0, β_1 and β_2 satisfying ,

$$\begin{aligned} &0.13908 (1.7937 - \beta_0)^2 - 0.9812 \{-0.6825 - (\beta_1 - \beta_2)\} (1.7937 - \beta_0) \\ &+ 3.5835 \{-0.6825 - (\beta_1 - \beta_2)\}^2 < 22.85 \end{aligned}$$

constitute a 95% confidence region for $C\beta$.

Chapter 2

BIASED ESTIMATORS

2.1 Introduction

It is well known that the least squares estimators of the regression coefficients are the best linear unbiased estimators. That is, of all possible estimators that are both linear functions of the data and unbiased for the parameters being estimated, the least squares estimators have the smallest variance. In the presence of collinearity, however, this minimum variance may be unacceptably large. Relaxing the least squares condition that estimators be unbiased opens for consideration a much larger set of possible estimators from which one with better properties in the presence of collinearity might be found. Biased regression refers to this class of regression methods in which unbiasedness is no longer required. Several biased regression methods have been proposed as solution to the collinearity problem; for example Stein's (1960) shrinkage,

Hoerl and Kennard's (1970a, 1970b), ridge regression and principal component regression see (Lott, 1973; Hawkins, 1973; Hocking, Speed, and Lynn, 1976; Marquardt, 1970; Webster, Gunst, and Mason, 1974). Although ridge regression has received the greatest acceptance, all have been used with apparent success in various problems. Nevertheless, biased regression methods have not been universally accepted and should be used with caution. The mean squared error (MSE) justification for biased regression methods makes it clear that such methods can provide better estimates of the parameters in the sense of mean squared. It does not necessarily follow that a biased regression solution is acceptable or even "better" than the least squares solution for purposes other than estimation of parameters.

The biased regression methods do not seem to have much to offer when the objective is to assign some measure of "relative importance" to the independent variables involved in collinearity. In essence, the biased estimators of the regression coefficients for the variables involved in the collinearity are weighted averages of the least squares regression coefficients for those variables. The best recourse to the collinearity problem when the objective is to assign relative importance is to recognize that the data are inadequate for the purpose .

Ridge regression and principal component regression are two commonly used biased regression methods. The biased regression meth-

ods attack the collinearity problem by computationally suppressing the effects of the collinearity. Ridge regression does this by reducing the apparent magnitude of the correlations. Principal component regression attacks the problem by regressing Y on the important principal components and then parceling out the effect of the principal component variables to the original variables.

2.2 Multicollinearity

The collinearity problem in regression arises when at least one linear function of the independent variables is nearly equal to zero. Technically, a set of vectors is collinear when a linear function is exactly equal to zero. In general discussions of the collinearity problem, the term “collinear ” is often used to apply to linear functions that are only approximately zero. This near - singularity may arise in several ways.

1- A mathematical constraint on variables that forces them to add to a constant will generate a collinearity. For example, frequencies of alleles at a locus will add to one if the frequencies of all alleles are recorded, or nearly to one if a rare allele is not scored. Generating new variables as transformations of other variables can produce a collinearity among the set of variables involved. Ratios of variables or powers of variables

frequently will be nearly collinear with the original variables.

2- Component variables of a system may show near - linear dependencies because of the biological, physical, or chemical constraints of the system. Various measures of size of an organism will show dependencies as will amounts of chemicals in the same biological pathway, or measures of rainfall, temperature, and elevation in an environmental system. Such correlational structures are properties of the system and can be expected to be present in all observational data obtained from the system.

3- Inadequate sampling may generate data in which the near - linear dependencies are an artifact of the data collection process. Unusual circumstances also can cause unlikely correlations to exist in the data, correlations that may not be present in later samplings or samplings from other similar population.

4- A bad experimental design may cause some model effects to be nearly completely confounded with others. This is the result of choosing levels of the experimental factors in such a way that there are near linear dependencies among the columns of X representing the different factors. Usually, experimental designs are constructed so as to ensure that the

different treatment factors are othogonal, or very nearly orthogonal, to each other.

One may not always be able to clearly identify the origin of the collinearity problem but it is important to understand its nature as much as possible. Knowing the nature of the collinearity problem will often suggest to the astute researcher its origin and, in turn, appropriate ways of handling the problem and of interpreting the regression results.

Consider the standard model for multiple linear regression,

$$Y = X\beta + \varepsilon, \text{ where } \varepsilon \sim N(O, \sigma^2 I), \quad (2.1)$$

Y and ε are $n \times 1$ vectors of observations and random errors respectively, X is an $n \times p$ matrix of input values, β is a $p \times 1$ vector of unknown regression coefficients, σ^2 is an unknown scalar constant, I denotes the identity matrix of dimension n and O stands for the null vector of appropriate dimension. Further, let

$$\tilde{\beta}_* = (X'X)^{-1}X'Y, \quad (2.2)$$

be the least squares estimator of β , where $\tilde{\beta}_*$ is the best linear unbiased estimator (BLUE) of β .

Note that,

$$\begin{aligned}
 \tilde{\beta}'_* \tilde{\beta}_* &= \left[(X'X)^{-1} X'Y \right]' \left[(X'X)^{-1} X'Y \right] \\
 &= Y'X(X'X)^{-1}(X'X)^{-1}X'Y \\
 &= Y'X(X'X)^{-2}X'Y.
 \end{aligned}$$

Thus,

$$\begin{aligned}
 E \left[\tilde{\beta}'_* \tilde{\beta}_* \right] &= E \left[Y'X(X'X)^{-2}X'Y \right] \\
 &= E \left[\text{Tr} \left\{ Y'X(X'X)^{-2}X'Y \right\} \right] \\
 &= E \left[\text{Tr} \left\{ X(X'X)^{-2}X'YY' \right\} \right] \\
 &= \text{Tr} \left\{ X(X'X)^{-2}X'E(YY') \right\}.
 \end{aligned}$$

Now,

$$\begin{aligned}
 E(YY') &= E \left[(X\beta + \varepsilon)(X\beta + \varepsilon)' \right] \\
 &= E \left[X\beta\beta'X' + X\beta\varepsilon' + \varepsilon\beta'X' + \varepsilon\varepsilon' \right] \\
 &= X\beta\beta'X' + \sigma^2 I, \text{ where } E(\varepsilon) = O.
 \end{aligned}$$

Hence,

$$\begin{aligned}
 E \left[\tilde{\beta}'_* \tilde{\beta}_* \right] &= \text{Tr} \left[X(X'X)^{-2}X' \{ X\beta\beta'X' + \sigma^2 I \} \right] \\
 &= \beta'X'X(X'X)^{-2}X'X\beta + \sigma^2 \text{Tr} \left\{ X(X'X)^{-2}X' \right\} \\
 &= \beta'\beta + \sigma^2 \text{Tr} \left\{ (X'X)^{-2}X'X \right\} \\
 &= \beta'\beta + \sigma^2 \text{Tr}(X'X)^{-1}.
 \end{aligned}$$

Let,

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ be the eigenvalues of $(X'X)$.

Therefore,

$\frac{1}{\lambda_1} \leq \frac{1}{\lambda_2} \leq \dots \leq \frac{1}{\lambda_p}$ are the eigenvalues of $(X'X)^{-1}$.

And,

$$Tr(X'X)^{-1} = \sum_{i=1}^p \frac{1}{\lambda_i}.$$

Moreover,

$$E [\tilde{\beta}'_* \tilde{\beta}_*] = \beta' \beta + \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i},$$

where λ_i 's are the eigenvalues of $(X'X)$. When near multicollinearity exists $|X'X| \approx 0$ and $\sum_{i=1}^p \frac{1}{\lambda_i}$ will be very large.

Therefore if multicollinearity exists then the expected length of $\tilde{\beta}'_* \tilde{\beta}_*$ is much larger than the length of $\beta' \beta$. But we want $E [\tilde{\beta}'_* \tilde{\beta}_*]$ to be closed to $\beta' \beta$.

Hoerl and Kennard (1970) suggested $\hat{\beta}_*$ such that $\hat{\beta}'_* \hat{\beta}_*$ has length smaller than $\tilde{\beta}'_* \tilde{\beta}_*$.

Next,

$$\begin{aligned}
 cov(\tilde{\beta}_*) &= \sigma^2 (X'X)^{-1} \\
 MSE(\tilde{\beta}_*) &= E \left[(\tilde{\beta}_* - \beta)' (\tilde{\beta}_* - \beta) \right] \\
 &= Tr \, cov(\tilde{\beta}_*) \\
 &= \sigma^2 Tr \, (X'X)^{-1} \\
 &= \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}.
 \end{aligned}$$

If some of the λ 's are very small $Tr \, cov(\tilde{\beta}_*)$ will be very large.

Suppose we have

$$X'X = \begin{bmatrix} 100.00 & 9.99 \\ 9.99 & 1.00 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{0.2} \begin{bmatrix} 1.00 & -9.99 \\ -9.99 & 100.00 \end{bmatrix}, \text{ where } |X'X| \approx 0.2$$

$$= \begin{bmatrix} 5 & -49.95 \\ -49.95 & 500 \end{bmatrix}.$$

Hence ,

$$Var(\tilde{\beta}_*) = \sigma^2 \begin{bmatrix} 5 & -49.95 \\ -49.95 & 500 \end{bmatrix}, \quad \tilde{\beta}_* = \begin{bmatrix} \tilde{\beta}_{*0} \\ \tilde{\beta}_{*1} \end{bmatrix}.$$

2.2.1 Exact - Multicollinearity

Now suppose that $\text{Rank}(X) = r < p+1$ (exact collinearity). Then we can partition X as follows

$$X = \begin{bmatrix} X_1 & \vdots & X_2 \end{bmatrix},$$

such that $\text{Rank}(X_1) = r$, where X_1 is $n \times r$ matrix and X_2 is $n \times (p+1-r)$ matrix. So that

$$X_2 = X_1 \omega, \text{ where } \omega \text{ is an arbitrary positive constant}$$

Next,

$$\begin{aligned} Y &= X\beta + \varepsilon \\ &= X_1\beta_1 + X_2\beta_2 + \varepsilon \\ &= X_1\beta_1 + (X_1\omega)\beta_2 + \varepsilon \\ &= X_1(\beta_1 + \omega\beta_2) + \varepsilon \\ &= X_1\delta + \varepsilon, \text{ where } \delta = (\beta_1 + \omega\beta_2). \end{aligned}$$

The new model is called the reduced model. The least squares estimate of δ is

$$\hat{\delta} = (X_1'X_1)^{-1}X_1'Y,$$

and

$$\text{Var}(\hat{\delta}) = \sigma^2 (X_1'X_1)^{-1}.$$

If $\text{Rank}(X) = r < p + 1$, then all the regression coefficients can not be estimated individually but some linear combinations of the parameter can be estimated. The predicted value of Y is given by

$$\tilde{Y} = X_1 \hat{\delta}.$$

2.2.2 Near - Multicollinearity

Consider a model of the form

$$Y = \beta_0 J + X_* \beta_{**} + \varepsilon, \quad (2.3)$$

$$= \begin{bmatrix} J & X_* \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_{**} \end{bmatrix} + \varepsilon$$

$$Y = X\beta + \varepsilon,$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad J = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

β_0 is the unknown intercept, and

$$\beta_{**} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}.$$

And,

$$X_* = \begin{bmatrix} x_{11}^* & x_{12}^* & \dots & x_{1p}^* \\ x_{21}^* & x_{22}^* & \dots & x_{2p}^* \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1}^* & x_{n2}^* & \dots & x_{np}^* \end{bmatrix}$$

where X_* is an $n \times p$ matrix of known values , $\text{Rank} (X_*) = p$, X_* is a matrix of full rank. The $(j,i)th$ element of X_* is $x_{ji}^* = \frac{(x_{ji} - \bar{x}_i)}{S_i}$,

where $\bar{x}_i = \frac{\sum_{j=1}^n x_{ji}}{n}$, $S_i^2 = \frac{\sum_{j=1}^n (x_{ji} - \bar{x}_i)^2}{(n-1)}$, $i = 1, 2, \dots, p$.

So our model ,

$$y = \beta_0 + \beta_1 \frac{(x_{11} - \bar{x}_1)}{S_1} + \beta_2 \frac{(x_{12} - \bar{x}_2)}{S_2} + \dots + \beta_p \frac{(x_{1p} - \bar{x}_p)}{S_p} + \epsilon.$$

For example ,

$$y_1 = \beta_0 + \beta_1 x_{11}^* + \beta_2 x_{12}^* + \cdots + \beta_p x_{1p}^* + \varepsilon_1.$$

The Properties Of $\hat{\beta}_0$ and $\hat{\beta}_*$

Now to find the least squares estimates of β_0 and β_{**} multiply

(2.3) by J' ,

$$\begin{aligned} J'Y &= \beta_0 J'J + (J' X_*) \beta_{**} + J'\varepsilon \\ \sum_{i=1}^n Y_i &= n\beta_0 + \left(\sum_{j=1}^n x_{j1}, \sum_{j=1}^n x_{j2}, \dots, \sum_{j=1}^n x_{jp} \right) \beta_{**} + \sum_{i=1}^n \varepsilon_j, \end{aligned}$$

where

$$\begin{aligned} x_{11}^* &= \frac{(x_{11} - \bar{x}_1)}{S_1}, \\ x_{21}^* &= \frac{(x_{21} - \bar{x}_1)}{S_1}, \\ &\vdots \\ x_{n1}^* &= \frac{(x_{n1} - \bar{x}_1)}{S_1} \\ \sum_{j=1}^n x_{j1}^* &= \frac{\left(\sum_{j=1}^n x_{j1} - n\bar{x}_1 \right)}{S_1} \\ &= \frac{(n\bar{x}_1 - n\bar{x}_1)}{S_1} \\ &= 0, \end{aligned}$$

similarly for other.

Therefore,

$$\sum_{i=1}^n Y_i = n\beta_0 + \sum_{i=1}^n \varepsilon_i.$$

Thus,

$$\bar{Y} = \beta_0 + \bar{\varepsilon}. \quad (2.4)$$

Next,

$$\begin{aligned} X'X &= \begin{bmatrix} J' \\ X'_* \end{bmatrix} \begin{bmatrix} J & X_* \end{bmatrix} \\ &= \begin{bmatrix} J'J & J'X_* \\ X'_*J & X'_*X_* \end{bmatrix} \\ &= \begin{bmatrix} n & 0 \\ 0 & X'_*X_* \end{bmatrix}. \end{aligned}$$

Hence,

$$\begin{aligned} (X'X)^{-1} &= \begin{bmatrix} n & 0 \\ 0 & X'_*X_* \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & (X'_*X_*)^{-1} \end{bmatrix} \quad (2.5) \end{aligned}$$

Further,

$$X'Y = \begin{bmatrix} J' \\ X'_* \end{bmatrix} Y$$

$$= \begin{bmatrix} J'Y \\ X_*'Y \end{bmatrix}.$$

Since ,

$$\tilde{\beta}_* = (X'X)^{-1}X'Y, \text{ as defined in (2.2)}$$

$$= \begin{bmatrix} n & 0 \\ 0 & X_*'X_* \end{bmatrix}^{-1} \begin{bmatrix} J'Y \\ X_*'Y \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & (X_*'X_*)^{-1} \end{bmatrix} \begin{bmatrix} J'Y \\ X_*'Y \end{bmatrix}$$

$$= \begin{bmatrix} \bar{Y} \\ (X_*'X_*)^{-1} X_*'Y \end{bmatrix}.$$

Note that,

$$\tilde{\beta}_* = \begin{bmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_{**} \end{bmatrix}.$$

Hence,

$$\tilde{\beta}_0 = \bar{Y}, \text{ and } \tilde{\beta}_{**} = (X_*'X_*)^{-1} X_*'Y$$

Now, we have

$$\tilde{\beta}_* = (X'X)^{-1}X'Y, \text{ where } \tilde{\beta}_* \text{ is given in (2.2),}$$

So,

$$E(\tilde{\beta}_*) = \beta.$$

And,

$$\begin{aligned} \text{Var}(\tilde{\beta}_*) &= \sigma^2 (X'X)^{-1} \\ &= \sigma^2 \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & (X_*'X_*)^{-1} \end{bmatrix}, \text{ where } (X'X)^{-1} \text{ is given in (2.5)} \\ &= \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \sigma^2 (X_*'X_*)^{-1} \end{bmatrix} \end{aligned} \quad (2.6)$$

Thus,

$$\begin{aligned} E(\tilde{\beta}_0) &= E(\bar{Y}) \\ &= E(\beta_0 + \bar{\epsilon}), \text{ where } \bar{Y} \text{ is defined in (2.4)} \\ &= \beta_0. \end{aligned}$$

And,

$$\begin{aligned} E(\tilde{\beta}_{**}) &= E((X_*'X_*)^{-1} X_*'Y) \\ &= (X_*'X_*)^{-1} X_*'E(Y) \\ &= (X_*'X_*)^{-1} X_*'E(Y) \\ &= (X_*'X_*)^{-1} X_*'X_*\beta_{**} \\ &= \beta_{**}. \end{aligned}$$

Next from (2.6) we note that,

$$\text{Var}(\tilde{\beta}_0) = \frac{\sigma^2}{n}.$$

And,

$$\text{Var}(\tilde{\beta}_{**}) = \sigma^2 (X'_* X_*)^{-1}.$$

Finally,

$$\begin{aligned} \text{cov}(\tilde{\beta}_0, \tilde{\beta}_{**}) &= E \left[\left\{ \tilde{\beta}_0 - E(\tilde{\beta}_0) \right\} \left\{ \tilde{\beta}_{**} - E(\tilde{\beta}_{**}) \right\} \right] \\ &= E \left[\left\{ \bar{Y} - \beta_0 \right\} \left\{ (X'_* X_*)^{-1} X'_* Y - \beta_{**} \right\} \right] \\ &= E \left[\left\{ \beta_0 + \bar{\varepsilon} - \beta_0 \right\} \left\{ (X'_* X_*)^{-1} X'_* (\beta_0 J + X_* \beta_{**} + \varepsilon) - \beta_{**} \right\} \right], \end{aligned}$$

where Y and $\bar{Y}$ are given in (2.3) and (2.4) respectively. Thus,

$$\begin{aligned} \text{cov}(\tilde{\beta}_0, \tilde{\beta}_{**}) &= E \left[\left\{ \bar{\varepsilon} \right\} \left\{ \begin{aligned} &(X'_* X_*)^{-1} X'_* J \beta_0 + (X'_* X_*)^{-1} X'_* X_* \beta_{**} \\ &+ (X'_* X_*)^{-1} X'_* \varepsilon - \beta_{**} \end{aligned} \right\} \right] \\ &= E \left[\left\{ \frac{J' \varepsilon}{n} \right\} \left\{ 0 + \beta_{**} + (X'_* X_*)^{-1} X'_* \varepsilon - \beta_{**} \right\} \right] \\ &= E \left[\frac{1}{n} (X'_* X_*)^{-1} X'_* \varepsilon \varepsilon' J \right] \\ &= \left[\frac{1}{n} (X'_* X_*)^{-1} X'_* J \right] E[\varepsilon \varepsilon'] \\ &= 0. \end{aligned}$$

Hence $\tilde{\beta}_0$ and $\tilde{\beta}_{**}$ are independent.

2.3 James - Stein Estimation

Consider the standard model for multiple linear regression,

$$Y = X\beta + \varepsilon, \text{ where } \varepsilon \sim N(0, \sigma^2 I).$$

The Gauss - Markov Theorem assures us that we cannot find linear unbiased estimators of regression coefficients β which have smaller variances than the least squares estimators $\tilde{\beta}_*$. Thus it is natural to look at biased estimators for an improvement over least squares estimators. If we shrink the least squares estimators toward 0 (by looking at estimators $k\tilde{\beta}_*$, $k = \text{diag}(k_{11}, k_{22}, \dots, k_{pp})$ and $k_{ii} \in [0, 1]$), the resulting estimators will be biased, but have smaller variances than $\tilde{\beta}_*$. Hoerl and Kennard (1970) have shown that such a k always exists (see Section 5).

Unfortunately, except for the James - Stein estimator, the optimum k depends on the unknown parameters, which must be estimated from the data. The distribution of any estimator $\hat{k}$ that depends on the data (including the James - Stein) has apparently not been successfully approached from a classical point of view.

Stein (1956) showed that, given Y a p - dimensional $\sim N(\theta, \sigma^2 I)$, with $p \geq 3$, y_i the usual unbiased estimator for θ_i is inadmissible given a squared loss function. Also James and Stein (1961) have shown that an estimator of θ the form $\left(1 - \frac{b^*}{Y'Y}\right)Y$ has a lower loss for all values

of θ , if $b^* \in (0, 2(p-2))$, and is minimized for $b^* = p - 2$. An interesting point they make is that the James - Stein estimator is not itself admissible. Although the notation conceals this, we can pick any origin we want by a translation. Thus the James - Stein estimator is not location invariant. Lindley (1962) suggested letting the data pick the origin, so that the mean $\bar{Y}$ is the new origin. His estimator is $\bar{Y} + \left\{ 1 - \frac{(p-3)}{(Y - \bar{Y})' (Y - \bar{Y})} \right\} (Y - \bar{Y})$, where $\bar{Y}$ is a vector of $\bar{Y}$'s and the estimator is written in terms of the original co-ordinate system. Notice the change in the coefficient of the fraction due to the loss of one degree of freedom in estimating $\bar{Y}$. Lindley's estimator shrinks toward the group mean, which seems more intuitively satisfying than shrinking towards what may be an arbitrary origin.

An easy extension of the James Stein estimator occurs when $Y \sim N(\theta, \sigma^2 I)$, σ^2 known; this is

$$Y_{js}^* = \left(1 - \frac{(p-2)\sigma^2}{Y'Y} \right) Y.$$

If σ^2 is unknown, we could substitute our usual estimate of the variance, $\hat{\sigma}_*^2$, to obtain

$$Y_{js} = \left(1 - \frac{(p-2)\nu\hat{\sigma}_*^2}{(\nu+2)Y'Y} \right) Y,$$

where $\hat{\sigma}_*^2$ has ν degrees of freedom and is independent of Y . Finally, James and Stein (1961) treat the case where an estimate S of the variance - covariance matrix, where S being distributed as a Wishart

distribution with ν degrees of freedom . Then we use

$$\left(1 - \frac{(p-2)}{(\nu-p+3) Y' S^{-1} Y}\right) Y.$$

2.4 Shrinkage Estimator

Another class of biased estimators introduced by Stein (1960), Hoerl (1962) and others, in their papers, they introduced a new class of estimators to reduce the length of $\tilde{\beta}_*$, this class is known as shrinkage estimators. While the original thrust of ridge regression was to alleviate effects of multicollinearity that of the shrinkage estimator was aimed only at reducing the mean square error. Consider the linear regress on model,

$$Y = X\beta + \varepsilon, \text{ where } \varepsilon \sim N(O, \sigma^2 I).$$

It is well known that $\tilde{\beta}_*$ is the least squares estimator of β with variance

$$Var(\tilde{\beta}_*) = \sigma^2 (X'X)^{-1},$$

and the expected squared length of $\tilde{\beta}_*$ is

$$E(\tilde{\beta}_*' \tilde{\beta}_*) = \beta' \beta + \sigma^2 Tr[(X'X)^{-1}].$$

The Stein - rule estimator can be written (see, e.g. Judge et al.1980, p.67 - 68) as

$$\tilde{\beta}_s = c \tilde{\beta}_* , \tag{2.7}$$

where $c = 1 - \frac{kS_*^2}{\tilde{\beta}_*'X'X\tilde{\beta}_*}$, k is an arbitrary positive constant ,

$$S_*^2 = (Y - X\tilde{\beta}_*)'(Y - X\tilde{\beta}_*) .$$

The estimator $\tilde{\beta}_s$ is biased. Thus, in order to compare it with the least squares estimator $\tilde{\beta}_*$, we need to consider the mean squared error matrix,

$$MSE(\tilde{\beta}_s) = E \left[(\tilde{\beta}_s - \beta) (\tilde{\beta}_s - \beta)' \right] ,$$

and compare it with $E \left[(\tilde{\beta}_* - \beta) (\tilde{\beta}_* - \beta)' \right]$, the covariance matrix of $\tilde{\beta}_*$.

It is well known that

$$E \left[(\tilde{\beta}_* - \beta) (\tilde{\beta}_* - \beta)' \right] = \sigma^2(X'X)^{-1} .$$

However, it is more convenient to compare the scalar quantities

$$Tr \left[(X'X)MSE(\tilde{\beta}_s) \right] \text{ and } Tr \left[(X'X)MSE(\tilde{\beta}_*) \right] .$$

Note that

$$Tr \left[(X'X)MSE(\tilde{\beta}_s) \right] = E \left[(\tilde{\beta}_s - \beta)' (X'X) (\tilde{\beta}_s - \beta) \right] ,$$

and

$$\begin{aligned} Tr \left[(X'X)MSE(\tilde{\beta}_*) \right] &= Tr \left[(X'X)E \left[(\tilde{\beta}_* - \beta) (\tilde{\beta}_* - \beta)' \right] \right] \\ &= E \left[Tr \left\{ (X'X) (\tilde{\beta}_* - \beta) (\tilde{\beta}_* - \beta)' \right\} \right] \\ &= E \left[(\tilde{\beta}_* - \beta)' (X'X) (\tilde{\beta}_* - \beta) \right] \\ &= \sigma^2 Tr \left[(X'X)^{-1} (X'X) \right] \\ &= \sigma^2 p . \end{aligned}$$

It has been shown by Stein (1961) that $0 < k < 2(p-2)(n+P+2)^{-1}$ and $p \geq 3$, the former quantity - i.e $E \left[(\tilde{\beta}_s - \beta)' (X'X) (\tilde{\beta}_s - \beta) \right]$ is always less than $p\sigma^2$ and attains its minimum value at $k = \frac{(p-2)}{(n-P+2)}$. It may be worth to mention that this value of k does not depend on any parameter.

It is well known that if the regression data is multicollinear, then the expected squared length of $\tilde{\beta}_*$ can be much greater than the squared length of the true parameter β (see, e.g., Judge et al. 1980, p.472), we note that

$$\begin{aligned} \tilde{\beta}_s' \tilde{\beta}_s &= c^2 \tilde{\beta}_* \tilde{\beta}_*, \text{ where } \tilde{\beta}_* \text{ is given in (2.7)} \\ &\leq \tilde{\beta}_* \tilde{\beta}_*. \end{aligned}$$

Although, the Stein's estimator has been generalized before (see, e.g., Judge et al, 1980, p.67 - 68), the extension here is obtained by minimizing the MSE of a linear estimator

$$\tilde{\beta}_{*s} = C_* Y,$$

where C_* is the arbitrary $(p \times n)$ matrix . The matrix C_* is defined by

$$C_* = \beta \beta' X' (X \beta \beta' X' + \sigma^2 I)^{-1}.$$

This estimator was derived by Thiel (1971, p.125). Hence the non - operational minimum MSE estimator of β is

$$\tilde{\beta}_{*s} = \frac{\beta \beta' X' Y}{\sigma^2 + \beta' X' X \beta}$$

and its MSE is

$$E \left[\left(\tilde{\beta}_{*s} - \beta \right) \left(\tilde{\beta}_{*s} - \beta \right)' \right] = \frac{\sigma^2 \beta \beta'}{\sigma^2 + \beta' X' X \beta}.$$

The properties of $\tilde{\beta}_s$ were studied by Swamy and Mehta (1976 and 1977).

Hosmane (1988) suggested a class of generalized Stein estimators of regression coefficients. Ali (1989) proposed another class of estimators. This class is better than the estimators developed by Hosmane (1988).

Ali (1989) defined a linear estimator

$$b = L'Y,$$

where L' is an arbitrary $(p \times n)$ matrix, of β . The squared bias and the trace of the dispersion matrix of b are given by

$$\begin{aligned} \text{Bias}(b) &= [E(b) - \beta]' [E(b) - \beta] \\ &= \beta' (L'X - I)' (L'X - I) \beta, \end{aligned}$$

and

$$\begin{aligned} \text{Var}(b) &= E [\{b - E(b)\}' \{b - E(b)\}] \\ &= \sigma^2 \text{Tr}(L'L), \end{aligned}$$

respectively. He obtained a generalized Stein estimator of β by minimizing

$$F_* = \beta' (L'X - I)' (L'X - I) \beta + m\sigma^2 \text{Tr}(L'L) \quad (2.8)$$

where m is an arbitrary non- negative scalar constant. We observe that (2.8) consists of two parts, one representing the bias and the other variance. Here the choice of m in F_* represents the weight we attach to the variance. m may be suitably chosen depending on the importance of variance in the estimator. Note that if $m = 0$, then $F_* = \text{Bias}(b)$ and if $m = 1$, then $F_* = \text{Bias}(b) + \text{Var}(b)$. The case $m = 1$ was considered by Hosmane (1988) . Ali (1989) showed that $m = 1$ is not the best choice of m .

The matrix L that minimizes F_* is given by

$$L_* = \left[X\beta\beta'X' + (m\sigma^2)I \right]^{-1} X\beta\beta'.$$

If $m > 0$, then we can rewrite L_* as

$$L_* = (m\sigma^2)^{-1} \left[1 - \frac{X\beta\beta'X'}{m\sigma^2 + \beta'X'X\beta} \right] X\beta\beta'. \quad (2.9)$$

See, Rao (1973, p.33). Substituting L_* given by (2.9) for L in $b = L'Y$ we obtain

$$b_* = [m\sigma^2 + \beta'X'X\beta]^{-1} (\beta'X'X\tilde{\beta}_*)\beta, \quad (2.10)$$

where $\tilde{\beta}_*$ is the least square estimator of β . The estimator b_* is not operational. To make it operational Ali (1989) replaced β and σ^2 in (2.10) by their unbiased estimators $\tilde{\beta}_*$ and

$$\hat{\sigma}_*^2 = (n - p)^{-1} (Y - X\tilde{\beta}_*)'(Y - X\tilde{\beta}_*),$$

respectively. This gives

$$b_+ = \left[I - m\hat{\sigma}_*^2(m\hat{\sigma}_*^2 + \tilde{\beta}_*'X'X\tilde{\beta}_*)^{-1} \right] \tilde{\beta}_*.$$

He also examined the properties of b_+ . He found that b_+ , in the sense of mean squared error, dominates the least square estimator as well as the generalized Stein estimator developed by Hosmane (1988).

2.5 Ridge Regression

This approach was proposed by Hoerl and Kennard (1970). It is used to stabilize the least squares estimator of the regression coefficients when near multicollinearity exists. Hoerl and Kennard(1970) suggested to estimate β of the model (2.1),

$$Y = X\beta + \varepsilon, \text{ where } \varepsilon \sim N(O, \sigma^2 I).$$

One such set of estimators is given by

$$\hat{\beta}_* = (X'X + kI)^{-1}X'Y,$$

where $k > 0$ and I is the identity matrix. Let,

$$\tilde{\beta}_* = (X'X)^{-1}X'Y,$$

be the least squares estimator of β .

If,

$$k = 0, \text{ then } \hat{\beta}_* = \tilde{\beta}_*.$$

Suppose that $\hat{\alpha}$ is another arbitrary estimator of β . Then the residual sum of squares is given by

$$\begin{aligned}\hat{\varepsilon}'\hat{\varepsilon} &= (Y - X\hat{\alpha})'(Y - X\hat{\alpha}) \\ &= \left[(Y - X\tilde{\beta}_*) + X(\tilde{\beta}_* - \hat{\alpha}) \right]' \left[(Y - X\tilde{\beta}_*) + X(\tilde{\beta}_* - \hat{\alpha}) \right] \\ &= (Y - X\tilde{\beta}_*)'(Y - X\tilde{\beta}_*) + (Y - X\tilde{\beta}_*)'X(\tilde{\beta}_* - \hat{\alpha}) \\ &\quad + (\tilde{\beta}_* - \hat{\alpha})'X'(Y - X\tilde{\beta}_*) + (\tilde{\beta}_* - \hat{\alpha})'X'X(\tilde{\beta}_* - \hat{\alpha}).\end{aligned}$$

Consider,

$$\begin{aligned}X'(Y - X\tilde{\beta}_*) &= X'Y - X'\tilde{\beta}_* \\ &= X'Y - X'X(X'X)^{-1}X'Y \\ &= X'Y - X'Y \\ &= 0.\end{aligned}$$

Thus,

$$\begin{aligned}\hat{\varepsilon}'\hat{\varepsilon} &= (Y - X\tilde{\beta}_*)'(Y - X\tilde{\beta}_*) + (\tilde{\beta}_* - \hat{\alpha})'X'X(\tilde{\beta}_* - \hat{\alpha}) \\ &= \tilde{\varepsilon}'\tilde{\varepsilon} + (\tilde{\beta}_* - \hat{\alpha})'X'X(\tilde{\beta}_* - \hat{\alpha}).\end{aligned}$$

But $(\tilde{\beta}_* - \hat{\alpha})'X'X(\tilde{\beta}_* - \hat{\alpha}) > 0$ because $X'X$ is a positive definite matrix.

Hence,

$$\hat{\varepsilon}'\hat{\varepsilon} > \tilde{\varepsilon}'\tilde{\varepsilon}.$$

Now let $\hat{\alpha}'\hat{\alpha} = \Phi$, where Φ is given fixed number. Our problem is to minimize $(\tilde{\beta}_* - \hat{\alpha})'X'X(\tilde{\beta}_* - \hat{\alpha})$ subject to the condition $\hat{\alpha}'\hat{\alpha} = \Phi$. Let

$$\Psi = (\tilde{\beta}_* - \hat{\alpha})'X'X(\tilde{\beta}_* - \hat{\alpha}) + k(\hat{\alpha}'\hat{\alpha} - \Phi),$$

where k is the lagrange multiplier .

Differentiate Ψ with respect to $\hat{\alpha}$ we get,

$$\frac{\partial \Psi}{\partial \hat{\alpha}} = -2X'X (\tilde{\beta}_* - \hat{\alpha}) + 2k\hat{\alpha},$$

setting the derivative equal to zero and solving for $\hat{\alpha}$ we get

$$-X'X (\tilde{\beta}_* - \hat{\alpha}) + k\hat{\alpha} = 0$$

$$(X'X + kI) \hat{\alpha} = X'X \tilde{\beta}_*.$$

$$\hat{\beta}_* = \hat{\alpha} = (X'X + kI)^{-1} X'X \tilde{\beta}_*.$$

$$= (X'X + kI)^{-1} X'X (X'X)^{-1} X'Y, \text{ where } \tilde{\beta}_* \text{ is given in (2.1)}$$

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

This class of estimators is called ridge regression estimators proposed by Hoerl and Kennard (1970) .

2.5.1 Properties of Ridge Estimator $\hat{\beta}_*$

Consider the model

$$Y = X\beta + \varepsilon, \text{ where } \varepsilon \sim N(0, \sigma^2 I).$$

We have,

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

$$\begin{aligned}
&= (X'X + kI)^{-1} X' (X\beta + \varepsilon) \\
&= (X'X + kI)^{-1} (X'X\beta + X'\varepsilon) \\
&= (X'X + kI)^{-1} X'X\beta + (X'X + kI)^{-1} X'\varepsilon \\
&= (X'X + kI)^{-1} (X'X + kI - kI) \beta + (X'X + kI)^{-1} X'\varepsilon \\
&= \beta - k (X'X + kI)^{-1} \beta + (X'X + kI)^{-1} X'\varepsilon \\
\hat{\beta}_* - \beta &= (X'X + kI)^{-1} (-k\beta + X'\varepsilon).
\end{aligned}$$

Now,

$$E(\hat{\beta}_*) = \beta - k (X'X + kI)^{-1} \beta.$$

And,

$$E(\hat{\beta}_* - \beta) = -k (X'X + kI)^{-1} \beta.$$

The bias for $\hat{\beta}_*$ is

$$\text{Bias}(\hat{\beta}_*) = -k (X'X + kI)^{-1} \beta, \text{ where } \hat{\beta}_* \text{ is a biased estimator for } \beta.$$

The squared of the bias of $\hat{\beta}_*$ is given by

$$\begin{aligned}
\{\text{Bias}(\hat{\beta}_*)\}^2 &= [E(\hat{\beta}_*) - \beta]' [E(\hat{\beta}_*) - \beta] \\
&= k^2 \beta' (X'X + kI)^{-2} \beta.
\end{aligned}$$

Next,

$$\begin{aligned}
\hat{\beta}_* - E(\hat{\beta}_*) &= \beta + (X'X + kI)^{-1} (-k\beta + X'\varepsilon) - [\beta - k (X'X + kI)^{-1} \beta] \\
&= \beta - k (X'X + kI)^{-1} \beta + (X'X + kI)^{-1} X'\varepsilon - [\beta - k (X'X + kI)^{-1} \beta], \\
&= \beta - k (X'X + kI)^{-1} \beta + (X'X + kI)^{-1} X'\varepsilon - \beta + k (X'X + kI)^{-1} \beta \\
&= (X'X + kI)^{-1} X'\varepsilon.
\end{aligned}$$

Thus,

$$\begin{aligned}
 Var [\hat{\beta}_*] &= E \left[\left(\hat{\beta}_* - E(\hat{\beta}_*) \right)' \left(\hat{\beta}_* - E(\hat{\beta}_*) \right) \right] \\
 &= E \left[\left((X'X + kI)^{-1} X' \varepsilon \right)' \left((X'X + kI)^{-1} X' \varepsilon \right) \right] \\
 &= E \left[\varepsilon' X (X'X + kI)^{-1} (X'X + kI)^{-1} X' \varepsilon \right] \\
 &= Tr \left[X (X'X + kI)^{-2} X' E(\varepsilon' \varepsilon) \right] \\
 &= \sigma^2 Tr \left[(X'X + kI)^{-2} X'X \right].
 \end{aligned}$$

The mean squared error (MSE) for $\hat{\beta}_*$ is

$$\begin{aligned}
 MSE(\hat{\beta}_*) &= Var(\hat{\beta}_*) + Bias^2(\hat{\beta}_*) \\
 &= \sigma^2 Tr \left[(X'X + kI)^{-2} X'X \right] + k^2 \beta' (X'X + kI)^{-2} \beta,
 \end{aligned}$$

setting $k = 0$ we get,

$$\begin{aligned}
 MSE(\hat{\beta}_.) &= \sigma^2 Tr \left[(X'X)^{-2} X'X \right] \\
 &= \sigma^2 Tr \left[(X'X)^{-1} \right] \\
 &= MSE(\tilde{\beta}_.), \text{ because when } k = 0 \implies \hat{\beta}_. = \tilde{\beta}_.
 \end{aligned}$$

Now we define a function of k by

$$\begin{aligned}
 V_1(k) &= Bias^2(\hat{\beta}_.) \\
 &= k^2 \beta' (X'X + kI)^{-2} \beta,
 \end{aligned} \tag{2.11}$$

and let,

$$\begin{aligned}
 V_2(k) &= Var(\hat{\beta}_.) \\
 &= \sigma^2 Tr \left[(X'X + kI)^{-2} X'X \right]
 \end{aligned} \tag{2.12}$$

In the following theorem, the behavior of $V_2(k)$ is examined.

Theorem 4 *The total variance $V_2(k)$ is a continuous monotonically decreasing function of k .*

Proof. Let $Z = (X'X + kI)^{-1}$. Therefore,

$$\begin{aligned} Z^2 X'X &= (X'X + kI)^{-2} (X'X + kI - kI) \\ &= Z - kZ^2. \end{aligned}$$

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ be the eigenvalues of $(X'X)$. Thus,

$$\left(\frac{1}{\lambda_1 + k} \right) \leq \left(\frac{1}{\lambda_2 + k} \right) \leq \dots \leq \left(\frac{1}{\lambda_p + k} \right),$$

are the eigenvalues of Z ,

$$\left(\frac{1}{\lambda_1 + k} \right)^2 \leq \left(\frac{1}{\lambda_2 + k} \right)^2 \leq \dots \leq \left(\frac{1}{\lambda_p + k} \right)^2,$$

are the eigenvalues of Z^2 . Hence,

$$\begin{aligned} \text{Tr}(Z^2 X'X) &= \text{Tr}(Z) - k\text{Tr}(Z^2) \\ &= \sum_{i=1}^p \frac{1}{\lambda_i + k} - k \sum_{i=1}^p \frac{1}{(\lambda_i + k)^2} \\ &= \sum_{i=1}^p \frac{1}{\lambda_i + k} \left(1 - \frac{k}{\lambda_i + k} \right) \\ &= \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2}. \end{aligned}$$

Therefore,

$$V_2(k) = \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2}. \quad (2.13)$$

Further $V_2(k)$ define on $(0, \infty)$ and $\lim_{k \rightarrow a} V_2(k)$ exists at any $a \in (0, \infty)$.

Hence $V_2(k)$ is continuous, and it is clear that each term in $V_2(k)$ decreases when k increases. This completes the proof. ■

Next,

$$\frac{\partial (V_2(k))}{\partial k} = -2\sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3}.$$

Note that $\frac{\partial (V_2(k))}{\partial k} \rightarrow -\infty$, as $k \rightarrow 0^+$ and $\lambda_p \rightarrow 0$.

Theorem 5 *The squared bias $V_1(k)$ is a continuous monotonically increasing function of k .*

Proof. Since $X'X$ is a positive definite matrix, there exists an orthogonal matrix T such that

$$T' (X'X) T = \Delta = \text{digonal elements } (\lambda_1, \lambda_2, \dots, \lambda_p).$$

Hence,

$$T' (X'X + kI) T = \Delta + kI.$$

Therefore,

$$\begin{aligned}
 V_1(k) &= k^2 \beta' T T' (X' X + kI)^{-2} T T' \beta \\
 &= k^2 \beta' T T' (X' X + kI)^{-1} T T' (X' X + kI)^{-1} T T' \beta \\
 &= k^2 \beta' T (\Delta + kI)^{-1} (\Delta + kI)^{-1} T' \beta \\
 &= k^2 \xi' (\Delta + kI)^{-2} \xi \\
 &= k^2 \sum_{i=1}^p \frac{\xi^2}{(\lambda_i + k)^2}, \tag{2.14}
 \end{aligned}$$

where $\xi' = \beta' T = (\xi_1, \xi_2, \dots, \xi_p)$ an $(1 \times p)$ vector . Since $\lambda_i > 0$, there is no singularity in (2.14). Note, also that $V_1(0) = 0$. Therefore, $V_1(k)$ is a continuous function for $k \geq 0$, hence (2.14) can be weitten as

$$V_1(k) = \sum_{i=1}^p \frac{\xi^2}{\left(1 + \frac{\lambda_i}{k}\right)^2}.$$

The function $\left(1 + \frac{\lambda_i}{k}\right)$, $i=1,2,\dots,p$, are monotone decreasing for increasing k . Each term of $V_1(k)$ is monotone increasing . This completes the proof. ■

Differentiating (2.14) with respect to k , we obtain

$$\frac{\partial V_1(k)}{\partial k} = 2k \sum_{i=1}^p \frac{\xi^2}{(\lambda_i + k)^2} - 2k^2 \sum_{i=1}^p \frac{\xi^2}{(\lambda_i + k)^3}$$

$$\begin{aligned}
&= 2k \sum_{i=1}^p \frac{\xi^2}{(\lambda_i + k)^2} \left\{ I - \frac{k}{\lambda_i + k} \right\} \\
&= 2k \sum_{i=1}^p \frac{\xi^2 \lambda_i}{(\lambda_i + k)^3}.
\end{aligned}$$

Clearly, $\frac{\partial V_1(k)}{\partial k} \rightarrow 0$ as $k \rightarrow 0$.

Theorem 6 *There always exists a $k > 0$ such that*

$$\frac{\partial MSE(\hat{\beta}_*)}{\partial k} < 0. \quad (2.15)$$

Proof. From (2.11), (2.12), (2.13) and (2.14) we have

$$MSE(\hat{\beta}_*) = V_1(k) + V_2(k)$$

$$MSE(\hat{\beta}_*) = k^2 \sum_{i=1}^p \frac{\xi_i^2}{(\lambda_i + k)^2} + \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2}.$$

Thus,

$$\begin{aligned}
\frac{\partial MSE(\hat{\beta}_*)}{\partial k} &= 2k \sum_{i=1}^p \frac{\xi_i^2 \lambda_i}{(\lambda_i + k)^3} - 2\sigma^2 \sum_{i=1}^{p+1} \frac{\lambda_i}{(\lambda_i + k)^3} \\
&= 2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3} (\xi_i^2 k - \sigma^2).
\end{aligned}$$

So, if $k < \frac{\sigma^2}{\xi_{\max}^2}$, then (2.15) is hold. This completes the proof. ■

One particular choice for k is

$$k = \frac{\sigma^2}{\xi'\xi} = \frac{\sigma^2}{\beta'TT'\beta},$$

which is not operational. To make it operational Hoerl and Kennard (1970) suggested to use

$$\tilde{k} = \frac{\tilde{\sigma}_*^2}{\tilde{\beta}'_*\tilde{\beta}_*},$$

where $\tilde{\beta}_*$ is given in (2.2). This k also is not good because we are not sure whether $\frac{\tilde{\sigma}_*^2}{\tilde{\beta}'_*\tilde{\beta}_*}$ is close to $\frac{\sigma^2}{\beta'\beta}$ or not. In other words we are not sure that if $0 < k < \frac{\tilde{\sigma}_*^2}{\tilde{\beta}'_*\tilde{\beta}_*}$ is a part of $0 < k < \frac{\sigma^2}{\beta'\beta}$ or not.

2.5.2 Remarks

If ridge regression is used as a general mechanical technique, is it any good? That is, is it worthwhile in the sense of reducing mean squared error enough to compensate for the extra trouble involved. Casella (1977) showed that, if the estimator produced by using the matrix $(X'X + kI)$ in place of $X'X$ in the normal equation is made to satisfy minimax (MSE) conditions, the number of conditions (that is, the ratio of the largest eigenvalue of $(X'X + kI)$ to the smallest) will tend to be larger, that is, worse, than that of $X'X$. Thus the “stability” in the sense of Hoerl and Kennard (1970) is reduced. This contradicts

the intuitive justification of ridge regression which supposedly improves “stability” by adding to the diagonal of $X'X$. Similarly, Thisted (1978) shows that, if minimaxity of (MSE) is required, one will tend not to get large (MSE) improvements for $X'X$ with large condition numbers. This is discouraging for ridge regression because highly multicollinear data for which ridge regression is supposed to be effective have high condition numbers.

A number of papers containing simulations of regression situations claim to show that ridge regression estimates are better than least squares estimates when judged via mean square error. Such claims must be viewed with caution. Careful study typically reveals that the simulation has been done with effective restrictions on the parameter values - precisely the situations where ridge regression is the appropriate technique theoretically. Furthermore, simulations often use at least some small values of $\frac{\beta\beta'}{\sigma^2}$, that is, impose some very severe constraints, which increases the bias of the simulations in favor of the ridge technique. Also, these small $\frac{\beta\beta'}{\sigma^2}$ values typically lead to nonsignificant regression which have no practical interpretive value, whether estimated by least squares or ridge regression methods. The extended inference that ridge regression is “always” better than least squares is, typically, completely unjustified.

2.6 Principal Component Regression

Principal component regression approaches the collinearity problem from the point of eliminating from consideration those dimensions of the X - space that are causing the collinearity problem. This is similar, in concept, to dropping an independent variable from the model when there is insufficient dispersion in that variable to contribute meaningful information on Y . However, in principal component regression the dimension dropped from consideration is defined by a linear combination of the variables rather than by a single independent variable.

The idea behind principal component regression, however, is to eliminate those dimensions that are causing the collinearity problem.

We reconsider the model (2.1). Let T be a $(p \times p)$ orthogonal matrix with the property

$$T'(X'X)T = \Delta = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p),$$

where λ_i , $i = 1, 2, \dots, p$, are the characteristic roots of $(X'X)$ arranged so that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Then we reparameterize the model (2.1),

$$\begin{aligned} Y &= XTT'\beta + \varepsilon \\ &= Z_*\xi + \varepsilon, \end{aligned}$$

where $Z_* = XT$ and $\xi = T'\beta$. The least square estimator of ξ , $\hat{\xi}$ is given by

$$\hat{\xi} = (Z_*'Z_*)^{-1}Z_*'Y$$

and

$$Cov(\hat{\xi}) = \Delta(\hat{\xi})$$

$$= \Delta^{-1}Z_*'\Delta(Y)Z_*\Delta^{-1}$$

$$= \sigma^2\Delta^{-1}Z_*'Z_*\Delta^{-1}$$

$$= \sigma^2\Delta^{-1}.$$

From the estimate $\hat{\xi}$, we can return to $\tilde{\beta}_*$ by the relationship

$$\tilde{\beta}_* = T\hat{\xi}.$$

Let $T = \begin{bmatrix} T_1 & T_2 & \dots & T_p \end{bmatrix}$ and $Z_{i*} = XT_i$, where $T_1, T_2, \dots, T_p$ are the columns of T . The Z_* are called the principal components of the matrix $X'X$, and the

$$\hat{\xi} = \left(\frac{1}{\lambda_i}\right)Z_{i*}'Y,$$

are the principal component regression coefficients. From

$$\tilde{\beta}_* = T\hat{\xi},$$

we have

$$\begin{aligned}\tilde{\beta}_* &= \begin{bmatrix} T_1 & T_2 & \cdots & T_p \end{bmatrix} \begin{bmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \\ \vdots \\ \hat{\xi}_p \end{bmatrix} \\ &= \sum_{i=1}^p \hat{\xi}_i T_i\end{aligned}$$

where $\hat{\xi}_i$ is the i th component of $\hat{\xi}$. Clearly, $\tilde{\beta}_*$ may be regarded as a weighted average of the T_i with the $\hat{\xi}_i$'s serving as weights.

To determine the effect of small characteristics root on the variance of $\tilde{\beta}_*$, we recall

$$\begin{aligned}\text{Cov}(\tilde{\beta}_*) &= \sigma^2 T \Delta^{-1} T' \\ &= \sigma^2 \sum_{i=1}^p \frac{T_i T_i'}{\lambda_i}\end{aligned}$$

In this form, it may be seen that a small value of λ_i , which leads to a large value of $(\frac{1}{\lambda_i})$, results in large value of the diagonal elements of $\text{Cov}(\tilde{\beta}_*)$; these, of course, are the $\text{Var}(\tilde{\beta}_*)$, the variance of the i th component of $(\tilde{\beta}_*)$. To avoid such large variance, it has been proposed to estimate

β from the first $r < p$ components - those associated with the r largest eigenvalue of $X'X$. Specifically,

let

$$T_{(r)} = \begin{bmatrix} T_1 & T_2 & \cdots & T_r \end{bmatrix},$$

$$\bar{T}_{(r)} = \begin{bmatrix} T_1 & T_2 & \cdots & T_p \end{bmatrix},$$

and

$$\hat{\xi}_{(r)} = \begin{bmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \\ \vdots \\ \hat{\xi}_r \end{bmatrix}$$

The estimate of β based on the first r principal components, $\tilde{\beta}_+$ is

$$\tilde{\beta}_+ = T_{(r)}\hat{\xi}_{(r)}$$

$$= T_{(r)}T'_{(r)}(X'X)^{-1}X'X\tilde{\beta}_*$$

$$= T_{(r)}\Delta^{-1}T'_{(r)}X'X\tilde{\beta}_*$$

The bias and the variance of $\tilde{\beta}_+$ are given by

$$\begin{aligned} Bias(\tilde{\beta}_+) &= E(\tilde{\beta}_+) - \beta \\ &= (T_{(r)}T'_{(r)} - I)\beta \end{aligned}$$

and

$$\begin{aligned} Var(\tilde{\beta}_+) &= TrCov(\tilde{\beta}_+) \\ &= \sigma^2 \sum_{i=1}^r \frac{1}{\lambda_i}, \end{aligned}$$

respectively.

It is clear that

$$Var(\tilde{\beta}_+) \leq Var(\tilde{\beta}_*)$$

Chapter 3

ERRORS OF MEASUREMENT IN THE INDEPENDENT VARIABLES

3.1 Introduction

Bias in estimation of regression coefficients may associate with

I- Measurement errors in independent variables.

II- Missing observations.

Sometimes measurement errors in independent variables are rounded up. These rounding errors also cause bias in estimation.

In the following sections we shall study the impacts of measurement errors and missing values.

Measurement Errors

In the standard regression set up, it is usually assumed that the errors of measurement occur only in the response variable, while the independent variables are measured without error. This assumption is

almost never likely to be satisfied. Therefore, it is important to examine the impacts of these errors on the regression coefficients.

The effects of measurement errors on the regression coefficients have many important practical implications. For example, in many economic policy analyses the regression coefficients are the starting point of an analysis. The regression coefficient are regarded as elasticities (or functions of elasticities) and indicate how much a change in one variable will effect the target variable. It is important in these cases to know the numerical precision of the regression coefficients.

These problems also arise in epidemiological studies in which multiple regression analysis is used. For example, in a study entitled "Does Air Pollution Shorten Lives?" Lave and Seskin (1977) estimated the mortality that may be attributed to the various pollution variables by fitting a multiple regression equation. In their study they regress the mortality rate against socioeconomic and pollution variables. These variables are usually subject to measurement errors.

Missing Values

Even in a well planned experiment sometimes it so happens that the yields on some plots are not available. They may be missing due to various reasons such as

- * natural causes like pests, flood, fire, etc.
- * man - made causes such as investigators failure to note the

yield or theft, etc.

In an industrial experiment some results can be missing because of mechanical break downs unrelated to the experimental process. In an opinion survey some individuals may be unable to express a preference for one candidate over another. Further, respondents in a household survey may refuse to report income. It is natural to treat the values that are not observed as missing, in the sense that there are true underlying values that would have been observed if technique involved had been better or equipment had been better maintained. The lack of a response is essentially an additional point in the sample space of the variable being measured, which identifies a “do not know” stratum of the population.

In this chapter we consider linear regressions in which the independent variables are fixed constants subject to rounding errors. We show that these errors will typically cause the least - squares estimators of the regression coefficients to be biased. Further we study the effects of errors of measurement on the estimates of the unknown parameters of a general linear model. It is well known that the usual least squares estimates no longer enjoy the property of unbiasedness when independent variables are subject to measurement errors. Moreover, we estimate the parameters of a simple linear regression model when some values of the independent variables are missing.

In most regression applications at least some of observations contain rounding errors. If these errors are confined to the independent variables and if the independent variables are fixed constants, then some useful results can be readily obtained. We consider the general linear regression model

$$Y = X\beta + \varepsilon, \text{ where } \varepsilon \sim N(O, \sigma^2 I),$$

$Y_{(n \times 1)}$ be vectors of observations, $\varepsilon_{(n \times 1)}$ represents random errors, $X_{(n \times p)}$ is a matrix fixed constants with rank $p < n$, and $\beta_{(p \times 1)}$ are fixed, unknown regression coefficients, O denotes the null vector of appropriate dimension and I_n stands for the $n \times n$ identity matrix. Suppose that in the recording of the j th value of the i th independent variable the actual value, x_{ij} , is rounded off according to some consistent rule and recorded as

$$x_{*ij} = x_{ij} + \varepsilon_{ij}.$$

This rounding of the independent variables can be denoted in an obvious way by

$$X_* = X + \varepsilon_*, \tag{3.1}$$

where X_* , X and ε_* are fixed matrices. We assume that the rounding errors, ε , are such that X_* (like X) has rank p .

Surely, the usual method of procedure in the face of specification errors like (3.1) is to ignore them and proceed to use the least - squares formula which defined in (2.1) on the recorded observations. Thus,

an arbitrary linear combination of the regression coefficients, say $L'\beta$, where L is a given $p \times 1$ vector, is estimated by

$$L'\tilde{\beta}_{**} = L'(X_*'X_*)^{-1}X_*'Y. \quad (3.2)$$

So, an obvious measure of the consequences in point estimation of rounding errors in fixed independent variables is given by a comparison of the properties of $L'\tilde{\beta}_{**}$ and the best linear unbiased estimator of $L'\beta$,

$$L'\tilde{\beta}_* = L'(X'X)^{-1}X'Y,$$

where $\tilde{\beta}_*$ is given in (2.2).

3.2 Errors in the Independent Variables of a General Linear Model

In most applications of regression at least some of the observations are subject to measurement errors. If these errors are confined to the independent variables and if the independent variables are fixed constants, then some useful results can easily be obtained. Suppose that instead of observing X in the general regression model we observe X_* which is defined in (3.1).

Clearly, by definition the bias of $L'\tilde{\beta}_{**}$ is

$$\begin{aligned} B(L'\tilde{\beta}_{**}) &= E(L'\tilde{\beta}_{**}) - L'\beta \\ &= L'(X_*'X_*)^{-1}X_*'X\beta - L'\beta, \end{aligned}$$

where $L'\tilde{\beta}_{**}$ is defined in (3.2).

$$B(L'\tilde{\beta}_{**}) = L'(X_*'X_*)^{-1}X_*'(X_* - \varepsilon)\beta - L'\beta,$$

where X_* is given in (3.1).

$$\begin{aligned} B(L'\tilde{\beta}_{**}) &= L'(X_*'X_*)^{-1}X_*'X_*\beta - L'(X_*'X_*)^{-1}X_*'\varepsilon\beta - L'\beta \\ &= L'\beta - L'(X_*'X_*)^{-1}X_*'\varepsilon\beta - L'\beta \\ &= -L'(X_*'X_*)^{-1}X_*'\varepsilon\beta. \end{aligned} \tag{3.3}$$

Partition ε and β in (3.3) and it is apparent that in case the i th column of ε is a null vector, i.e., in case the i th independent variable is recorded without rounding error, $B(L'\tilde{\beta}_{**})$ does not depend on the i th element of β . Thus, the least - squares estimator, $L'\tilde{\beta}_{**}$, of an arbitrary linear combination of regression coefficients is, in general, biased by an amount that depends on the size of the regression coefficients associated with just those independent variables containing rounding errors.

It is also apparent from (3.3) that a necessary and sufficient condition for the least - squares estimator, $L'\tilde{\beta}_{**}$, to be unbiased is that rounding errors, ε , satisfy

$$X'_*\varepsilon = O,$$

where O is a null matrix.

This condition typically will not be satisfied when rounding errors are present (i.e., when $\varepsilon \neq 0$).

Now the variance of $L'\tilde{\beta}_{**}$ is given by

$$\text{Var}(L'\tilde{\beta}_{**}) = \sigma^2 L'(X'_*X_*)^{-1}L.$$

The mean square error (MSE) of $L'\tilde{\beta}_{**}$ is

$$\begin{aligned} \text{MSE}(L'\tilde{\beta}_{**}) &= \text{Var}(L'\tilde{\beta}_{**}) + [B(L'\tilde{\beta}_{**})]^2 \\ &= \sigma^2 L'(X'_*X_*)^{-1}L + L'(X'_*X_*)^{-1}X'_*\varepsilon\beta\beta'\varepsilon'X_*(X'_*X_*)^{-1}L \end{aligned}$$

We define relative squared bias of $L'\tilde{\beta}_{**}$ by

$$RB = \frac{[B(L'\tilde{\beta}_{**})]^2}{\text{Var}(L'\tilde{\beta}_{**})}.$$

Therefore,

$$\begin{aligned} RB &= \frac{L'(X'_*X_*)^{-1}X'_*\varepsilon\beta\beta'\varepsilon'X_*(X'_*X_*)^{-1}L}{\sigma^2 L'(X'_*X_*)^{-1}L} \\ &= \frac{L'(X'_*X_*)^{-1}X'_*\varepsilon\beta\beta'\varepsilon'X_*(X'_*X_*)^{-1}L}{\sigma^2 L'(X'_*X_*)^{-1}X'_*X_*(X'_*X_*)^{-1}L}. \end{aligned} \tag{3.4}$$

Rao (1973, p - 62) showed that for any arbitrary vector L and a non - negative definite matrix Γ ,

$$Sup_L \left[\frac{L' \Gamma L}{L' L} \right] = \text{Largest eigenvalue of } \Gamma,$$

and

$$Inf_L \left[\frac{L' \Gamma L}{L' L} \right] = \text{Smallest eigenvalue of } \Gamma.$$

Therefore,

$$\begin{aligned} & Sup_L \frac{L' (X'_* X_*)^{-1} X'_* \varepsilon \beta \beta' \varepsilon' X_* (X'_* X_*)^{-1} L}{L' (X'_* X_*)^{-1} X'_* X_* (X'_* X_*)^{-1} L} \\ &= \text{Largest eigenvalue of } (\varepsilon \beta \beta' \varepsilon'), \end{aligned} \quad (3.5)$$

and

$$\begin{aligned} & Inf_L \frac{L' (X'_* X_*)^{-1} X'_* \varepsilon \beta \beta' \varepsilon' X_* (X'_* X_*)^{-1} L}{L' (X'_* X_*)^{-1} X'_* X_* (X'_* X_*)^{-1} L} \\ &= \text{Smallest eigenvalue of } (\varepsilon \beta \beta' \varepsilon'). \end{aligned} \quad (3.6)$$

We note that $(\varepsilon \beta \beta' \varepsilon')$ is symmetric positive semi - definite matrix of rank 1, therefore, it has exactly one non - zero eigenvalue. The remaining eigenvalues of $(\varepsilon \beta \beta' \varepsilon')$ are equal to zero. Further, since

$$Tr(\varepsilon \beta \beta' \varepsilon') = \sum \text{of the eigenvalues of } (\varepsilon \beta \beta' \varepsilon'),$$

and

$$Tr(\varepsilon \beta \beta' \varepsilon') = Tr(\beta' \varepsilon' \varepsilon \beta) = (\beta' \varepsilon' \varepsilon \beta),$$

where $\text{Tr}(\Gamma)$ denotes the trace of a square matrix Γ . Hence,

$$\text{Largest eigenvalue of } (\varepsilon\beta\beta'\varepsilon') = \beta'\varepsilon'\varepsilon\beta, \quad (3.7)$$

and

$$\text{Smallest eigenvalue of } (\varepsilon\beta\beta'\varepsilon') = 0. \quad (3.8)$$

Relations (3.4), (3.5), (3.6), (3.7), and (3.8) produce the following result:

$$0 \leq \frac{[B(L'\tilde{\beta}_{**})]^2}{\text{Var}(L'\tilde{\beta}_{**})} \leq \left(\frac{\beta'\varepsilon'\varepsilon\beta}{\sigma^2} \right). \quad (3.9)$$

Clearly, the relative bias in piont estimation of caused by errors, in the independent variables will be negligible when $\left(\frac{\beta'\varepsilon'\varepsilon\beta}{\sigma^2} \right)$ is very small.

We now consider the mean square error of $L'\tilde{\beta}_{**}$. It follows from (3.9) that,

$$\text{MSE}(L'\tilde{\beta}_{**}) = \sigma^2 L'(X_*'X_*)^{-1}L \left[1 + \frac{[B(L'\tilde{\beta}_{**})]^2}{\text{Var}(L'\tilde{\beta}_{**})} \right]. \quad (3.10)$$

Using (3.9) in (3.10) we get

$$\sigma^2 L'(X_*'X_*)^{-1}L \leq \text{MSE}(L'\tilde{\beta}_{**}) \leq \sigma^2 L'(X_*'X_*)^{-1}L \left[1 + \frac{\beta'\varepsilon'\varepsilon\beta}{\sigma^2} \right].$$

3.3 Simultaneous Estimation of Linear Functions of Parameters When Independent Variables are Subject to Measurement Errors

At times, interest is centered on certain specified linear functions of β , say the vector

$$\varphi = M\beta,$$

where M is a given $r \times p$ matrix of rank $r \leq p$.

The least squares estimator of φ is

$$\hat{\varphi} = M(X_*'X_*)^{-1}X_*'Y.$$

Clearly,

$$\begin{aligned} E(\hat{\varphi} - \varphi) &= M(X_*'X_*)^{-1}X_*'X\beta - M\beta \\ &= M(X_*'X_*)^{-1}X_*'(X_* - \varepsilon)\beta - M\beta, \end{aligned}$$

where X_* is given in (3.1).

$$\begin{aligned} E(\hat{\varphi} - \varphi) &= M(X_*'X_*)^{-1}X_*'X_*\beta - M(X_*'X_*)^{-1}X_*'\varepsilon\beta - M\beta \\ &= M\beta - M(X_*'X_*)^{-1}X_*'\varepsilon\beta - M\beta \end{aligned}$$

$$= -M(X_*'X_*)^{-1}X_*'\varepsilon\beta,$$

and

$$[E(\hat{\varphi} - \varphi)][E(\hat{\varphi} - \varphi)]' = M(X_*'X_*)^{-1}X_*'\varepsilon\beta\beta'\varepsilon'X_*(X_*'X_*)^{-1}M'.$$

The dispersion matrix of the least squares estimator, $\hat{\varphi}$, of φ is

$$D(\hat{\varphi}) = E[(\hat{\varphi} - E(\hat{\varphi}))(\hat{\varphi} - E(\hat{\varphi}))']$$

$$= \sigma^2 M(X_*'X_*)^{-1}M'.$$

Next, the mean squared error matrix of $\hat{\varphi}$ is given by

$$\begin{aligned} MSE(\hat{\varphi}) &= D(\hat{\varphi}) + [E(\hat{\varphi}) - \varphi][E(\hat{\varphi}) - \varphi]' \\ &= \sigma^2 M(X_*'X_*)^{-1}M' + M(X_*'X_*)^{-1}X_*'\varepsilon\beta\beta'\varepsilon'X_*(X_*'X_*)^{-1}M'. \end{aligned} \quad (3.11)$$

Theorem 7 For any full rank matrices X_* and M ,

$$\left| \frac{MSE(\hat{\varphi})}{\sigma^2} \right| \leq \left| M(X_*'X_*)^{-1}M' \right| \left(1 + \frac{\beta'\varepsilon'\varepsilon\beta}{\sigma^2} \right),$$

where $|E|$ denotes the determinant of a square matrix E .

Proof. We rewrite (3.11) as

$$MSE(\hat{\varphi}) = \sigma^2 Q_* Q_*' + Q_* \varepsilon \beta \beta' \varepsilon' Q_*'$$

where

$$Q_* = M(X_*'X_*)^{-1}X_*'.$$

Hence,

$$\begin{aligned} \left| \frac{MSE(\hat{\varphi})}{\sigma^2} \right| &= \left| Q_* Q_*' + \frac{Q_* \varepsilon \beta \beta' \varepsilon' Q_*'}{\sigma^2} \right| \\ &= |Q_* Q_*'| \left[1 + \frac{\beta' \varepsilon' Q_*' (Q_* Q_*')^{-1} Q_* \varepsilon \beta}{\sigma^2} \right], \end{aligned}$$

see, Rao (1973, p. 32).

Therefore,

$$\left| \frac{MSE(\hat{\varphi})}{\sigma^2} \right| \leq |Q_* Q_*'| \left[1 + \left(\frac{\beta' \varepsilon' \varepsilon \beta}{\sigma^2} \right) \text{Ch max}(G_*) \right],$$

where

$$G_* = Q_*' (Q_* Q_*')^{-1} Q_*,$$

and $\text{Ch max}(G_*)$ stands for the largest eigenvalue of G_* . Since G_* is an idempotent matrix, therefore,

$$\text{Ch max}(G_*) \leq 1.$$

Hence,

$$\begin{aligned} \left| \frac{MSE(\hat{\varphi})}{\sigma^2} \right| &\leq |Q_* Q_*'| \left[1 + \frac{\beta' \varepsilon' \varepsilon \beta}{\sigma^2} \right] \\ &= |M(X_*' X_*)^{-1} M'| \left[1 + \frac{\beta' \varepsilon' \varepsilon \beta}{\sigma^2} \right] \end{aligned}$$

This completes the proof. ■

3.4 Estimation of Parameters in a General Linear Model

When Some of the Independent Variables are Missing

A problem that is frequently occurring in statistical practices is that of dealing with regression when some of the values on the independent variables are missing. We discuss in this section the special case of regression, a dependent variable y regresses on two independent variables x_1 and x_2 , where there are missing values on x_2 only. Such a situation may arise in chemical experiments, for example, where destruction or removal of part of an experimental units prevents the investigator from making complete measurements on x_2 . The question then is how to deal properly with the incomplete sample.

We consider the linear model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3.12)$$

where the ε_i are identical independent distributed random variables with mean zero and variance σ^2 , and $(n - m)$ values on x_2 are unobserved.

The simplest approach is to apply the usual least squares analysis to the subset of complete case, one an approach equivalent to applying the method of least squares to all available observations while regarding the missing values as additional parameters to be estimated. However, many alternatives to the complete case method have been proposed.

Dear (1959) studied the substitution of the sample means for missing observations, followed by the usual application of least squares. Walsh (1959) and Buck (1960) investigated the substitution of predicted values from simple linear regression among the observed values. Beal and little (1975) applied the method of maximum likelihood to deal with the problem arising from missing data.

Properties of these procedures have been compared extensively, mainly through simulation studies by Beale and Little (1975), Donner and Rosner (1982), Haitovsky (1968), Heiberger (1977) and Timm (1970). Beal and little (1975) have found that for multinormal distributions the method of maximum likelihood is preferable to the complete case method. Donner and Rosner (1982) showed that the method of maximum likelihood is superior to other methods in estimating β_1 and β_2 of the model (3.12)

Analytic comparisons have generally not been performed, largely because of the complexity of the problem that arises when missing values are scattered over several explanatory variables. However, Donner (1982) has derived the expressions for the bias and variance associated with procedures frequently used to deal with the problem. He has found that in terms of mean squared error (MSE), the mean substitution method is more effective for estimating β_2 , the coefficient of incompletely observed variable when the correlations involving these

variables are small and the proportion of missing cases is fairly high. In estimating β_1 , the coefficient of completely observed variable, this method provides estimates which may have substantial bias unless the correlations involving these variables are small. Further, he showed that the method of linear prediction is advantageous over the complete case method for estimating β_1 while the two methods are equivalent for estimating β_2 .

3.4.1 Methods of Estimation

We are primarily concerned with estimating β_1 and β_2 under the assumptions that the $(n - m)$ values of x_2 are missing. We assume that the values (x_{1i}, x_{2i}) are fixed with zero means and unit variances over all n cases, and over the m complete case from the incomplete case in the linear model

$$Y = X\beta + \varepsilon,$$

where Y is an $(n \times 1)$ vector of observations, X is an $(n \times 2)$ input matrix, $\beta = (\beta_1, \beta_2)$ is a 2 - dimensional vector of parameters, and ε is an $(n \times 1)$ vector of random errors.

We partition Y as follows :

$$Y = \begin{bmatrix} Y_c \\ \vdots \\ Y_1 \end{bmatrix},$$

where Y_c is the vector of observations on m complete cases, Y_1 is the vector of observations for the incomplete cases.

The input matrix X and the vector of random errors , ε , are partitioned accordingly i.e.,

$$\begin{aligned} X &= \begin{bmatrix} X_c \\ \vdots \\ X_1 \end{bmatrix} \\ &= \begin{bmatrix} X_{1c} & X_{2c} \\ \vdots & \vdots \\ X_{11} & X_{21} \end{bmatrix}, \end{aligned}$$

with $X_c = (X_{1c} \quad X_{2c})$, $X_1 = (X_{11} \quad X_{21})$, and

$$\varepsilon = \begin{bmatrix} \varepsilon_c \\ \vdots \\ \varepsilon_1 \end{bmatrix}.$$

Our assumptions can be summarized as follows:

$$J'Y = J'Y_c = J'Y_1 = 0.$$

Mean of Y - values is zero.

$$J'X_{1c} = J'X_{2c}=J'X_{11}=J'X_{21}=0$$

$$m = X'_{1c}X_{1c}= X'_{2c}X_{2c},$$

$$n - m = X'_{11}X_{11}= X'_{21}X_{21},$$

$$r\ m = X'_{1c}X_{2c}$$

$$r^*\ n = X'_{1c}X_{1c}+ X'_{11}X_{21},$$

where J is a column vector of ones.

Complete Case Method :

In the complete case method the usual least squares analysis is applied to the model

$$Y_c = X_c \beta + \varepsilon_c, \text{ where } \varepsilon \sim N(0, \sigma^2 I),$$

and $X_c = (X_{1c} \quad X_{2c})$ and I is the identity matrix. For the complete case, the estimator of β is given by

$$\tilde{\beta}_{*c} = (X'_c X_c)^{-1} X'_c Y_c,$$

where

$$(X'_c X_c) = \begin{bmatrix} m & r m \\ r m & m \end{bmatrix}$$

$$(X'_c X_c)^{-1} = \frac{1}{m(1-r^2)} \begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix}.$$

The estimator $\tilde{\beta}_{*c}$ is unbiased for β and its dispersion matrix is given by

$$D(\tilde{\beta}_{*c}) = \sigma^2 (X'_c X_c)^{-1}.$$

The mean squared error (MSE) of $\tilde{\beta}_{*c}$ is given by

$$MSE(\tilde{\beta}_{*c}) = \sigma^2 \text{Tr}(X'_c X_c)^{-1}$$

$$= \frac{2\sigma^2}{\{m(1-r^2)\}}.$$

Chapter 4

REDUCTION OF BIAS BY JACKKNIFING

4.1 Introduction

The Jackknife is a general method for reducing the bias in an estimator. Quenouille(1956) and Tukey (1958) introduced the jackknife technique for reducing bias in estimation. Extensive literature dealing with this technique has been developed since then. An excellent review of the literature was presented by Miller (1974 a). In a separate article Miller (1974 b) gave detailed account of jackknifing linear model estimators. He showed that the jackknife procedures yield consistent results in large samples. Hinkley (1977) investigated the standard jackknife and a weighted jackknife in the general linear model. He found that for unbalanced data the weighted jackknife is superior to the standard jackknife.

4.2 Jackknifing

Consider the model

$$Y_1 = X_1\beta + \varepsilon_1, \quad (4.1)$$

with $E(\varepsilon_1) = 0$, $E(\varepsilon_1 \varepsilon_1') = \sigma^2 I_n$, where Y_1 and ε_1 are $n \times 1$ vectors of observations and random errors respectively, X_1 is an $n \times p$ matrix of known values, β is a $p \times 1$ vector of unknown parameters. We also assume that

$$\text{Rank}(X_1) = p < n.$$

Further, suppose that m additional observations on the response variable are available. The additional observations are of the form

$$Y = Z\beta + \mu + \varepsilon, \quad (4.2)$$

with $E(\varepsilon) = 0$, $E(\varepsilon \varepsilon') = \sigma^2 I_m$, where μ denotes a shift in mean of Y , Z is a $m \times p$ matrix of input values. Y , μ , ε are all $m \times 1$ vectors of observations, unknown constants and random errors respectively. The least squares estimator that combines (4.1) and (4.2) is given by

$$\tilde{\beta} = G^{-1}(X_1' Y_1 + Z' Y) \quad (4.3)$$

where $G = X_1' X_1 + Z' Z$. Let Y_{-i} and Z'_{-i} denote respectively the vector Y with its i th coordinate deleted and the matrix Z' with its i th column

deleted and let

$$\begin{aligned}\tilde{\beta}_{-i} &= (X_1'X_1 + Z_{-i}'Z_{-i})^{-1}(X_1'Y_1 + Z_{-i}'Y_{-i}) \\ &= (G - z_i z_i')^{-1}(X_1'Y_1 + Z'Y - z_i y_i),\end{aligned}$$

where z_i is the i th column of Z' and y_i is the i th coordinate of Y . We can rewrite $\tilde{\beta}_{-i}$ as follows:

$$\tilde{\beta}_{-i} = \tilde{\beta} - \frac{G^{-1}z_i(y_i - z_i'\tilde{\beta})}{(1 - z_i'G^{-1}z_i)}.$$

We define pseudo values as

$$w_i = m \tilde{\beta} - (m-1) \tilde{\beta}_{-i}, \quad i = 1, 2, \dots, m$$

Now, we may define a Jackknife estimator of β as

$$\begin{aligned}\bar{w} &= m^{-1} \sum_{i=1}^m w_i \\ &= \tilde{\beta} + \left(\frac{m-1}{m}\right) G^{-1} \sum_{i=1}^m z_i u_i\end{aligned}$$

where

$$u_i = \frac{y_i - z_i'\tilde{\beta}}{1 - z_i'G^{-1}z_i}, \quad i = 1, 2, \dots, m.$$

This is in line with the Jackknife method originally developed by Quenouille (1956).

Note that,

$$Var(u_i) = \frac{\left[Var(y_i) + Var\left(z_i' \tilde{\beta}\right) - 2Cov\left(y_i, z_i' \tilde{\beta}\right) \right]}{[1 - z_i' G^{-1} z_i]^2},$$

$$Var(y_i) = \sigma^2,$$

and

$$\begin{aligned} E(\tilde{\beta}) &= E\left[G^{-1}(X_1' Y_1 + Z' Y)\right] \\ &= G^{-1} E\left[X_1' (X_1 \beta + \varepsilon_1) + Z' (Z \beta + \mu + \varepsilon)\right] \\ &= G^{-1} \left[X_1' X_1 \beta + Z' Z \beta + Z' \mu\right] \\ &= G^{-1} \left[(X_1' X_1 + Z' Z) \beta + Z' \mu\right] \\ &= G^{-1} [G \beta + Z' \mu] \\ &= \beta + G^{-1} Z' \mu. \end{aligned}$$

Moreover,

$$\left[Bias(\tilde{\beta})\right]^2 = \left[E(\tilde{\beta}) - \beta\right] \left[E(\tilde{\beta}) - \beta\right]'$$

$$\begin{aligned}
&= [\beta + G^{-1}Z'\mu - \beta] [\beta + G^{-1}Z'\mu - \beta]' \\
&= [G^{-1}Z'\mu] [\mu'ZG^{-1}]. \tag{4.4}
\end{aligned}$$

Further,

$$\begin{aligned}
Var(\tilde{\beta}) &= Var [G^{-1}(X_1'Y_1 + Z'Y)] \\
&= G^{-1}X_1'\sigma^2X_1G^{-1} + G^{-1}Z'\sigma^2ZG^{-1} \\
&= \sigma^2G^{-2}(X_1'X_1 + Z'Z) \\
&= \sigma^2G^{-1}. \tag{4.5}
\end{aligned}$$

Hence,

$$\begin{aligned}
MSE(\tilde{\beta}) &= Var(\tilde{\beta}) + [Bias(\tilde{\beta})]^2 \\
&= \sigma^2G^{-1} + [G^{-1}Z'\mu] [\mu'ZG^{-1}] \\
&= G^{-1}[\sigma^2 + Z'\mu\mu'ZG^{-1}]. \tag{4.6}
\end{aligned}$$

Next,

$$Var(z_i'\tilde{\beta}) = z_i' Var(\tilde{\beta}) z_i$$

$$= z_i' \sigma^2 (G^{-1}) z_i$$

$$= \sigma^2 z_i' G^{-1} z_i.$$

Now,

$$\begin{aligned} \text{Cov} \left(y_i, z_i' \tilde{\beta} \right) &= E \left[\{ y_i - E(y_i) \} \left\{ z_i' \tilde{\beta} - E \left(z_i' \tilde{\beta} \right) \right\} \right] \\ &= E \left[\{ z_i' \beta + \mu_i + \varepsilon_i - E(z_i' \beta + \mu_i + \varepsilon_i) \} \left\{ z_i' \tilde{\beta} - z_i' E(\tilde{\beta}) \right\} \right] \\ &= E \left[\{ z_i' \beta + \mu_i + \varepsilon_i - z_i' \beta - \mu_i \} \left\{ z_i' \tilde{\beta} - z_i' (\beta + G^{-1} Z' \mu) \right\} \right] \\ &= E \left[\varepsilon_i \left\{ z_i' G^{-1} (X_1' Y_1 + Z' Y) - z_i' (\beta + G^{-1} Z' \mu) \right\} \right] \\ &= E \left[\varepsilon_i \left\{ z_i' G^{-1} \left[X_1' (X_1 \beta + \varepsilon_1) + Z' (Z \beta + \mu + \varepsilon) \right] - z_i' (\beta + G^{-1} Z' \mu) \right\} \right] \\ &= E \left[\varepsilon_i \left\{ z_i' G^{-1} \left[X_1' X_1 \beta + X_1' \varepsilon_1 + Z' Z \beta + Z' \mu + Z' \varepsilon \right] - z_i' (\beta + G^{-1} Z' \mu) \right\} \right] \end{aligned}$$

$$\begin{aligned}
&= E \left[\varepsilon_i \left\{ \begin{bmatrix} z'_i G^{-1} (X'_1 X_1 + Z' Z) \beta + z'_i G^{-1} X'_1 \varepsilon_1 \\ + z'_i G^{-1} Z' \mu + z'_i G^{-1} Z' \varepsilon - z'_i (\beta + G^{-1} Z' \mu) \end{bmatrix} \right\} \right] \\
&= E \left[\varepsilon_i \left\{ \begin{bmatrix} z'_i \beta + z'_i G^{-1} X'_1 \varepsilon_1 + z'_i G^{-1} Z' \mu + z'_i G^{-1} Z' \varepsilon \\ - z'_i \beta - z'_i G^{-1} Z' \mu \end{bmatrix} \right\} \right] \\
&= E \left[\varepsilon_i \left\{ \begin{bmatrix} z'_i G^{-1} X'_1 \varepsilon_1 + z'_i G^{-1} Z' \varepsilon \end{bmatrix} \right\} \right] \\
&= E \left[\varepsilon_i z'_i G^{-1} (X'_1 \varepsilon_1 + Z' \varepsilon) \right].
\end{aligned}$$

Recalling that y_i is the i th coordinate of Y , hence

$$\begin{aligned}
Cov \left(y_i, z'_i \tilde{\beta} \right) &= E \left[\varepsilon_i z'_i G^{-1} Z' \varepsilon \right] \\
&= E \left[z'_i G^{-1} Z' (\varepsilon \varepsilon_i) \right].
\end{aligned}$$

Next ,

$$Z' \varepsilon = \begin{bmatrix} z_1 & z_2 & \dots & z_i & \dots & z_m \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

$$\begin{aligned}
&= z_1 \varepsilon_1 + z_2 \varepsilon_2 + \cdots + z_i \varepsilon_i + \cdots + z_m \varepsilon_m \\
(Z' \varepsilon) \varepsilon_i &= z_1 \varepsilon_1 \varepsilon_i + z_2 \varepsilon_2 \varepsilon_i + \cdots + z_i \varepsilon_i^2 + \cdots + z_m \varepsilon_m \varepsilon_i.
\end{aligned}$$

Therefore,

$$E [(Z' \varepsilon) \varepsilon_i] = \sigma^2 z_i.$$

Further,

$$\begin{aligned}
Cov \left(y_i, z_i' \tilde{\beta} \right) &= z_i' G^{-1} (\sigma^2 z_i) \\
&= \sigma^2 z_i' G^{-1} z_i.
\end{aligned}$$

Hence,

$$\begin{aligned}
Var(u_i) &= \frac{\sigma^2 + \sigma^2 z_i' G^{-1} z_i - 2\sigma^2 z_i' G^{-1} z_i}{[I - z_i' G^{-1} z_i]^2} \\
&= [\sigma^2 - \sigma^2 z_i' G^{-1} z_i] [I - z_i' G^{-1} z_i]^{-2} \\
&= \sigma^2 [I - z_i' G^{-1} z_i] [I - z_i' G^{-1} z_i]^{-2} \\
&= \sigma^2 [I - z_i' G^{-1} z_i]^{-1}, \quad i=1,2,\dots,m.
\end{aligned}$$

Finally, we define pseudo values as

$$\begin{aligned}
v_i &= \tilde{\beta} - s G^{-1} z_i [I - z_i' G^{-1} z_i] u_i \\
&= \tilde{\beta} - s G^{-1} z_i [I - z_i' G^{-1} z_i] \left[\frac{y_i - z_i' \tilde{\beta}}{1 - z_i' G^{-1} z_i} \right] \\
&= \tilde{\beta} - s G^{-1} z_i \left[y_i - z_i' \tilde{\beta} \right],
\end{aligned}$$

where s is an arbitrary non - negative scalar constant .

We define a class of jackknife estimators as

$$\begin{aligned}
 \tilde{\beta}_* &= m^{-1} \sum_{i=1}^m v_i \\
 &= \tilde{\beta} - \frac{s}{m} G^{-1} \sum_{i=1}^m z_i \left[y_i - z_i' \tilde{\beta} \right] \\
 &= \tilde{\beta} - \frac{s}{m} G^{-1} \left(Z' Y - Z' Z \tilde{\beta} \right) \\
 &= \tilde{\beta} - \frac{s}{m} G^{-1} Z' \left(Y - Z \tilde{\beta} \right).
 \end{aligned}$$

In the following sections we shall study some principal properties of $\tilde{\beta}_*$.

4.3 Bias and Mean Square Error (MSE) of $\tilde{\beta}_*$

Let

$$X_* = \begin{bmatrix} X_1 \\ Z \end{bmatrix},$$

For simplicity we shall compare the bias of $X_* \tilde{\beta}_*$ with the bias of $X_* \tilde{\beta}$.

Now,

$$E \left[\tilde{\beta}_* \right] = E \left[\tilde{\beta} \right] - \frac{s}{m} G^{-1} Z' E \left[Y - Z \tilde{\beta} \right],$$

where

$$E \left[\tilde{\beta} \right] = \beta + G^{-1} Z' \mu.$$

Next,

$$\begin{aligned}
 E \left[Y - Z \tilde{\beta} \right] &= Z\beta + \mu - Z \left(\beta + G^{-1} Z' \mu \right) \\
 &= Z\beta + \mu - Z\beta - ZG^{-1} Z' \mu \\
 &= \mu - ZG^{-1} Z' \mu \\
 &= \left[I - ZG^{-1} Z' \right] \mu
 \end{aligned}$$

Finally,

$$\begin{aligned}
 E \left[\tilde{\beta}_* \right] &= \beta + G^{-1} Z' \mu - \frac{s}{m} G^{-1} Z' \left[I - ZG^{-1} Z' \right] \mu \\
 &= \beta + G^{-1} Z' \mu - \frac{s}{m} G^{-1} Z' \mu + \frac{s}{m} G^{-1} Z' ZG^{-1} Z' \mu \\
 &= \beta + G^{-1} Z' \mu - \frac{s}{m} G^{-1} Z' \mu + \frac{s}{m} G^{-1} (G - X_1' X_1) G^{-1} Z' \mu \\
 &= \beta + G^{-1} Z' \mu - \frac{s}{m} G^{-1} Z' \mu + \frac{s}{m} G^{-1} Z' \mu - \frac{s}{m} G^{-1} X_1' X_1 G^{-1} Z' \mu \\
 &= \beta + G^{-1} Z' \mu - \frac{s}{m} G^{-1} X_1' X_1 G^{-1} Z' \mu \\
 &= \beta + \left[I - \frac{s}{m} G^{-1} X_1' X_1 \right] G^{-1} Z' \mu \\
 &= \beta + RG^{-1} Z' \mu
 \end{aligned}$$

$$= \beta + RG^{-\frac{1}{2}}\Omega,$$

where,

$$R = \left(I - \frac{s}{m} G^{-1} X_1' X_1 \right)$$

and

$$\Omega = G^{-\frac{1}{2}} Z' \mu.$$

Hence,

$$\begin{aligned} E \left[X_* \tilde{\beta}_* - X_* \beta \right] &= E \left[X_* \tilde{\beta}_* \right] - X_* \beta \\ &= X_* \beta + X_* R G^{-\frac{1}{2}} \Omega - X_* \beta \\ &= X_* R G^{-\frac{1}{2}} \Omega, \end{aligned}$$

and

$$\left\{ E \left[X_* \left(\tilde{\beta}_* - \beta \right) \right] \right\} \left\{ E \left[X_* \left(\tilde{\beta}_* - \beta \right) \right] \right\}' = X_* R G^{-\frac{1}{2}} \Omega \Omega' G^{-\frac{1}{2}} R' X_*'. \quad (4.7)$$

Next,

$$\begin{aligned} E \left[X_* \left(\tilde{\beta} - \beta \right) \right] &= X_* E \left[\tilde{\beta} \right] - X_* \beta \\ &= X_* \beta + X_* G^{-1} Z' \mu - X_* \beta \\ &= X_* G^{-1} Z' \mu \\ &= X_* G^{-\frac{1}{2}} \Omega. \end{aligned}$$

Thus,

$$\left\{ E \left[X_* \left(\tilde{\beta} - \beta \right) \right] \right\} \left\{ E \left[X_* \left(\tilde{\beta} - \beta \right) \right] \right\}' = X_* G^{-\frac{1}{2}} \Omega \Omega' G^{-\frac{1}{2}} X_*'. \quad (4.8)$$

We have,

$$\begin{aligned}
\tilde{\beta}_* &= \tilde{\beta} - \frac{s}{m} G^{-1} Z' \left(Y - Z \tilde{\beta} \right) \\
&= \left(I + \frac{s}{m} G^{-1} Z' Z \right) \tilde{\beta} - \frac{s}{m} G^{-1} Z' Y \\
&= \left(I + \frac{s}{m} G^{-1} Z' Z \right) G^{-1} (X_1' Y_1 + Z' Y) - \frac{s}{m} G^{-1} Z' Y \\
&= \left(I + \frac{s}{m} G^{-1} Z' Z \right) G^{-1} X_1' Y_1 + \left(\left(I + \frac{s}{m} G^{-1} Z' Z \right) G^{-1} Z' - \frac{s}{m} G^{-1} Z' \right) Y \\
&= \left(I + \frac{s}{m} G^{-1} Z' Z \right) G^{-1} X_1' Y_1 + \left(\left(I + \frac{s}{m} G^{-1} Z' Z - \frac{s}{m} \right) G^{-1} Z' \right) Y.
\end{aligned}$$

Now,

$$\begin{aligned}
\left[Bias(\tilde{\beta}_*) \right]^2 &= \left[E(\tilde{\beta}_*) - \beta \right] \left[E(\tilde{\beta}_*) - \beta \right]' \\
&= \left[\beta + R G^{-\frac{1}{2}} \Omega - \beta \right] \left[\beta + R G^{-\frac{1}{2}} \Omega - \beta \right]' \\
&= R G^{-\frac{1}{2}} \Omega \Omega' G^{-\frac{1}{2}} R', \tag{4.9}
\end{aligned}$$

where

$$R = \left(I - \frac{s}{m} G^{-1} X_1' X_1 \right), \text{ and } \Omega = G^{-\frac{1}{2}} Z' \mu$$

Further,

$$\begin{aligned}
 Var(\tilde{\beta}_*) &= Var \left[\left(I + \frac{s}{m} G^{-1} Z' Z \right) G^{-1} X_1' Y_1 + \left(\left(I + \frac{s}{m} G^{-1} Z' Z - \frac{s}{m} \right) G^{-1} Z' \right) Y \right] \\
 &= \left(I + \frac{s}{m} G^{-1} Z' Z \right) G^{-1} X_1' \sigma^2 X_1 G^{-1} \left(I + \frac{s}{m} Z' Z G^{-1} \right) \\
 &\quad + \left(I + \frac{s}{m} G^{-1} Z' Z - \frac{s}{m} \right) G^{-1} Z' \sigma^2 Z G^{-1} \left(I + \frac{s}{m} G^{-1} Z' Z - \frac{s}{m} \right)' \\
 &= \sigma^2 G^{-1} \left[\begin{aligned} &\left(I + \frac{s}{m} G^{-1} Z' Z \right) \left(I + \frac{s}{m} Z' Z G^{-1} \right) X_1' X_1 G^{-1} \\ &+ \left(I + \frac{s}{m} G^{-1} Z' Z - \frac{s}{m} \right) Z' Z G^{-1} \left(I + \frac{s}{m} G^{-1} Z' Z - \frac{s}{m} \right)' \end{aligned} \right]. \quad (4.10)
 \end{aligned}$$

Next,

$$\begin{aligned}
 MSE(\tilde{\beta}_*) &= Var(\tilde{\beta}_*) + \left[Bias(\tilde{\beta}_*) \right]^2 \\
 &= \sigma^2 G^{-1} \left[\begin{aligned} &\left(I + \frac{s}{m} G^{-1} Z' Z \right) \left(I + \frac{s}{m} Z' Z G^{-1} \right) X_1' X_1 G^{-1} + \\ &\left(I + \frac{s}{m} G^{-1} Z' Z - \frac{s}{m} \right) Z' Z G^{-1} \left(I + \frac{s}{m} G^{-1} Z' Z - \frac{s}{m} \right)' \end{aligned} \right] \\
 &\quad + R G^{-\frac{1}{2}} \Omega \Omega' G^{-\frac{1}{2}} R'. \quad (4.11)
 \end{aligned}$$

Further, the bias, the variance - covariance matrix and the MSE of $\tilde{\beta}$ can be obtained by setting $s = 0$ in (4.9), (4.10) and (4.11) respectively.

$$\begin{aligned}
 \left[Bias(\tilde{\beta}) \right]^2 &= \left[Bias(\tilde{\beta}_*) \right]^2, \\
 Var(\tilde{\beta}) &= Var(\tilde{\beta}_*), \\
 MSE(\tilde{\beta}) &= MSE(\tilde{\beta}_*).
 \end{aligned}$$

4.4 Comparison of Bias

The following lemmas are needed to compare the biases of $X_* \tilde{\beta}_*$ and $X_* \tilde{\beta}$.

Lemma 1: The matrices $G^{-\frac{1}{2}} R' X_*' X_* R G^{-\frac{1}{2}}$ and $R' R$ have the same eigenvalues.

Proof. We observe that the eigenvalues of $G^{-\frac{1}{2}} R' X_*' X_* R G^{-\frac{1}{2}}$ are the same as the eigenvalues of $R' X_*' X_* R G^{-1}$. Next, we have

$$R = \left[I - \frac{s}{m} G^{-1} X_1' X_1 \right].$$

Hence,

$$\begin{aligned} R G^{-1} &= \left[G^{-1} - \frac{s}{m} G^{-1} X_1' X_1 G^{-1} \right] \\ &= G^{-1} \left[I - \frac{s}{m} X_1' X_1 G^{-1} \right] \\ &= G^{-1} R'. \end{aligned}$$

Therefore,

$$\begin{aligned} R' X_*' X_* R G^{-1} &= R' X_*' X_* G^{-1} R' \\ &= R' R'. \end{aligned}$$

Since,

$$X_*' X_* = G.$$

Thus, the matrices $G^{-\frac{1}{2}} R' X_*' X_* R G^{-\frac{1}{2}}$ and $R' R$ have the same eigenvalues.

■

Lemma 2: If λ is an eigenvalue of $X_1'X_1G^{-1}$ then $\left[I - \left(\frac{s}{m}\right)\lambda\right]^2$ is an eigenvalue of $G^{-\frac{1}{2}}R'X_*'X_*RG^{-\frac{1}{2}}$.

Proof. We have,

$$R' = \left[I - \frac{s}{m}X_1'X_1G^{-1}\right].$$

If λ is an eigenvalue of $X_1'X_1G^{-1}$, then $\left[I - \left(\frac{s}{m}\right)\lambda\right]$ is an eigenvalue of R' and $\left[I - \left(\frac{s}{m}\right)\lambda\right]^2$ is an eigenvalue of $R'R' = (R')^2$, (see Searle (1982), p-277).

Further, since $R'R'$ and $G^{-\frac{1}{2}}R'X_*'X_*RG^{-\frac{1}{2}}$ have the same eigenvalue, hence $\left[I - \left(\frac{s}{m}\right)\lambda\right]^2$ is an eigenvalue of $G^{-\frac{1}{2}}R'X_*'X_*RG^{-\frac{1}{2}}$.

This completes the proof. ■

Lemma 3: Let λ_1 be the largest eigenvalue of $X_1'X_1G^{-1}$ and

$$0 \leq s \leq \frac{2m}{\lambda_1}.$$

Then all eigenvalues of $G^{-\frac{1}{2}}R'X_*'X_*RG^{-\frac{1}{2}}$ lie in $[0, 1]$.

Proof. Let λ be an eigenvalue of $X_1'X_1G^{-1}$. Then it follows from lemma 2, that $\left[I - \left(\frac{s}{m}\right)\lambda\right]^2$ is an eigenvalue of $G^{-\frac{1}{2}}R'X_*'X_*RG^{-\frac{1}{2}}$.

Therefore, the eigenvalues of $G^{-\frac{1}{2}} R' X' X_* R' G^{-\frac{1}{2}}$ are non - negative. Now,

$$\left[I - \left(\frac{s}{m} \right) \lambda \right]^2 \leq 1,$$

implies

$$\frac{s}{m} \lambda \left[\frac{s}{m} \lambda - 2 \right] \leq 0.$$

A sufficient condition for

$$\frac{s}{m} \lambda \left[\frac{s}{m} \lambda - 2 \right] \leq 0,$$

is

$$\frac{s}{m} \lambda_1 - 2 \leq 0.$$

Equivalently,

$$0 \leq s \leq \frac{2m}{\lambda_1}.$$

This completes the proof. ■

Lemma 4: Let $\text{rank}(X_1) = p$. Then all eigenvalues of $(X_1' X_1) G^{-1}$ lie in $(0,1]$.

Proof. Lemma 4 follows from the fact that $G^{-\frac{1}{2}} X_1' X_1 G^{-\frac{1}{2}}$ is a positive definite matrix and

$$\begin{aligned} G^{-\frac{1}{2}} X_1' X_1 G^{-\frac{1}{2}} &= G^{-\frac{1}{2}} [G - Z' Z] G^{-\frac{1}{2}} \\ &= I - G^{-\frac{1}{2}} Z' Z G^{-\frac{1}{2}}. \end{aligned}$$

Since $G^{\frac{-1}{2}} Z' Z G^{\frac{-1}{2}}$ is a positive semi - definite matrix, all eigenvalues of this matrix are non - negative. From the relation

$$G^{\frac{-1}{2}} X_1' X_1 G^{\frac{-1}{2}} = I - G^{\frac{-1}{2}} Z' Z G^{\frac{-1}{2}},$$

it follows that all eigenvalues of $G^{\frac{-1}{2}} X_1' X_1 G^{\frac{-1}{2}}$ lie in $(0,1]$. ■

Theorem 8 Suppose that λ_1 is the largest eigenvalue of $(X_1' X_1) G^{-1}$ and

$$0 \leq s \leq \frac{2m}{\lambda_1}.$$

Then

$$\text{Max}_{\Omega} \left[\frac{\left\| X_* E \left(\tilde{\beta}_* - \beta \right) \right\|^2}{\left\| X_* E \left(\tilde{\beta} - \beta \right) \right\|^2} \right] \leq 1,$$

where $\Omega = G^{\frac{-1}{2}} Z' \mu$ and

$$\| \theta_* \| = \sqrt{\theta_*' \theta_*} = \text{Tr } \theta_*' \theta_*,$$

for any vector θ_* .

Proof. From (4.7) and (4.8) we have

$$\begin{aligned} \left\| X_* E \left(\tilde{\beta}_* - \beta \right) \right\|^2 &= \text{Tr } X_* R G^{\frac{-1}{2}} \Omega \Omega' G^{\frac{-1}{2}} R' X_*' \\ &= \text{Tr } \Omega' G^{\frac{-1}{2}} R' X_*' X_* R G^{\frac{-1}{2}} \Omega \\ &= \Omega' G^{\frac{-1}{2}} R' X_*' X_* R G^{\frac{-1}{2}} \Omega. \end{aligned}$$

and

$$\begin{aligned}
 \left\| X_* E \left(\tilde{\beta} - \beta \right) \right\|^2 &= \text{Tr } X_* G^{\frac{-1}{2}} \Omega \Omega' G^{\frac{-1}{2}} X_*' \\
 &= \Omega' G^{\frac{-1}{2}} X_*' X_* G^{\frac{-1}{2}} \Omega \\
 &= \Omega' G^{\frac{-1}{2}} G G^{\frac{-1}{2}} \Omega \\
 &= \Omega' \Omega.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \frac{\left\| X_* E \left(\tilde{\beta}_* - \beta \right) \right\|^2}{\left\| X_* E \left(\tilde{\beta} - \beta \right) \right\|^2} &= \frac{\Omega' G^{\frac{-1}{2}} R' X_*' X_* R G^{\frac{-1}{2}} \Omega}{\Omega' \Omega} \\
 &\leq \text{ch max } \left\{ G^{\frac{-1}{2}} R' X_*' X_* R G^{\frac{-1}{2}} \right\},
 \end{aligned}$$

where $\text{ch max } (Q)$ denotes the largest eigenvalue of Q .

The rest follows from lemma 3. ■

Conclusions

There are situations where some or all independent variables of a regression model cannot be measured without error. When measurements are made with errors necessary steps should be taken to improve the precision of the estimates. Unfortunately, the bias in estimation, associates with measurement errors depends on unknown parameters. Therefore, the magnitude of bias will remain unknown. This makes the estimation imprecise. The approach we propose in chapter 3, can be used to minimize the effects of measurement errors. Note from page 92 that

$$\left| \frac{\text{MSE}(\hat{\phi})}{\sigma^2} \right| \leq \left| M(X'X)^{-1}M' \right| \left[1 + \frac{\beta'\epsilon'\epsilon\beta}{\sigma^2} \right]$$

Although the expression $\left[1 + \frac{\beta'\epsilon'\epsilon\beta}{\sigma^2} \right]$ is unknown, but one can

easily minimize $\left| M(X'X)^{-1}M' \right|$ through the choice of the design matrix X . Such a design will minimize the upper bound

$$\text{of } \left| \frac{\text{MSE}(\hat{\phi})}{\sigma^2} \right|.$$

Therefore, our approach will provide protection against errors in observations.

Further, in chapter 4 we introduce the Jackknife approach. This approach is helpful in reducing the bias in estimation.

However, minimization of bias always associates with an increase in variance. In estimation minimization of bias is more important than minimization of variance. When minimization of bias is paramount important, the experimenter can employ the Jackknife Technique to address the problem.

References

Ali, M. A (1989) : “ On A Class Of Shrinkage Estimators Of The Vector Of Regression Coefficients.” *Commun . Statist .Theory Method* ,18 (12) , 4491 – 4500 .

Ali, M.A. (1991) : “ On Estimation Of Linear Functions Of Parameters Of A Linear Regression Model.” *Sankhya; The Indian Journal of statistics*, 53 , 64 – 73 .

Beale , E. M.L , and Little , R . J. A . (1975) : “ Missing Values In Multivariate Analysis.” *Journal of the Royal Statistical Society, Series, B*, 37 , 129 – 145 .

Belsley , D.A., Kuh , E and Welsch , R. E. (1980) : “ Regression Diagnostics : Identifying Influential Data And Sources Of Collinearity.” John Wiley and Sons , New York.

Buck, S. F. (1960) : “ A Method Of Estimation Of Missing Values In Multivariate Data, Suitable For Use With An Electronic Computer.” *Journal of the Royal Statistical Society, Series, B*, 22 , 302 – 307.

Casella, George. (1977) : “ Minimax Ridge Regression Estimation.” Purdue University Department Of Statistics Mimeograph Series 497, May.

Crouse , R . H ., Jin . C. and Hanumara , R . C. (1995): “ Unbiased Ridge Estimation With Prior Information And Kidge Trace .” *Commun .Statist . Theory Meth .*, 24 (9) , 2341 – 2354 .

Dear, R. E. (1959) : “ A Principal Component Missing Data Method For Multiple Regression Models.” SP - 86, Santa Monica, California, System Development – Corporation.

Donner,A.(1982): “ The Relative Effectiveness Of Procedures Commonly Used In Multiple Regression Analysis For Dealing With Missing Values.” The American Statistician 36 , 378 – 381 .

Donner , A . and Rosner , B . (1982) : “ Missing Value Problems In Multiple Linear Regression With Two Independent Variables.” Communications in statistics A, Theory and Methods 11 , 127 – 140 .

Draper, N . R . and Nostrand , R . C (1979) : “ Ridge Regression And James – Stein Estimation: Review And Comments .” Technometrics , 21 , 451-465 .

Draper, N. R. and Smith . H . (1980) : “ Applied Regression Analysis .” John Wiley and Sons , Inc.

Freund, R. J. and Minton , P.D. (1979) : “ Regression Methods A Tool For Data Analysis .” Marcel Dekker, Inc . New York and Basel.

Freund, J.E. (1992) : “ Mathematical Statistics.”(Fifth Edition). Prentice - Hall , India.

Graybill, F.A. (1961) : “ An Introduction To Linear Statistical Models .” MC Graw - Hill Book . Company , New York.

Graybill, F. A. (1976) : “ Theory And Application Of The Linear Model.” Duxbury Press.

Gray , H.L. and Schucany , W. R . (1972) : “ The Generalized Jackknife Statistic.” Volume 1, Marcel Dekker , Inc . New York.

- Gunst, R.F. and Mason, R .L. (1980) : “ Regression Analysis And Its Application A Data – Oriented Approach.” Marcel Dekker, Inc. New York and Basel.
- Haitovsky, Y.(1968) : “ Missing Data in Regression Analysis .” Journal of the Royal Statistical Society , Series B, 30,67 - 82 .
- Heiberger, R.M. (1977) : “ Regression With The Pairwise Present Covariance Matrix; A Dangerous Practice.” Technical Report No. 19, Department Of Statistics, University Of Pennsylvania.
- Hinkley, D.V. (1977) : “ Jackknifing In Unbalanced Situations.” Technometrics , 19, 285 - 292 .
- Hoerl, A.E. (1962) : “Application Of Ridge Analysis To Regression Problems .” Chemical Engineering Progress, 58 , 54 – 59 .
- Hoerl, A.E. and Kennard , R . W. (1970) : “ Ridge Regression : Biased Estimation For Nonorthogonal Problems .” Technometrics , 12 , 55 – 67 .
- Hosmane , B.S. (1988) : “ On A Generalized Stein Estimator Of Regression Coefficients.” Communications in Statistics, Theory and Methods , 17(6) , 1735 - 1740 .
- James , W. and Stein, C. (1961) : “ Estimation With Quadratic Loss ” Proc. Fourth Berkely Symp . Math . Statist. Prob . 1,361 – 379 .
- Johnson , R . A . and Wichern (1992) : “ Applied Multivariate Statistical Analysis .” (Third Edition) . Printice – Hall International , Inc.

Judge , G . G ., W. E. Griffiths, R. C. Hill and T. C. Lee, (1980) : “ The Theory And Practice Of Econometrics.” Wiley, New York.

Kendall. S.M. and Stuart. A., (1961) : “ The Advanced Theory Of Statistics .” Vol (2) (Fourth Edition). Charles Griffin and Company Limited. London and High Wycombe.

Lave, L. B. and Seskin, E. P. (1977) : “ Does Air Pollution Shorten Lives? Statistics And Public Policy,” (W. B. Fairley & F. Mosteller, eds.) Reading: MA: Addison Wesley.

Lindley, D.V. and (1962) : “ Discussion .” Journal of the Royal Statistical Society, Series, B, 24 , 285 – 287.

Little , R . J .A . and Rubin , D.B . (1987) : “ Statistical Analysis With Missing Data.” John Wiley and Sons , New York.

Manin, P.S. (1997) : “ Introduction Statistics. ” (Third Edition) . John Wiley and Sons , Inc.

Marquardt, Donald W. (1970): “ Generalized Inverses, Ridge Regression, Biased Linear Estimation, And Nonlinear Estimation.” Technometrics , 12 , 591– 612.

Mayer , L.S. and Willke , T..A (1973) : “ On Biased Estimation In Linear Models. ” Technometrics . 15 , 497 – 507

Miller , R.G. (1974a) : “ The Jackknife A Review .” Biometrika 61,1–15 .

Miller, R.G.(1974b): “An Unbiased Jackknife.” Ann, Statist.2, 880 – 891.

Neter, J , Wasserman , W. and Whitmore ,G.A. (1988) : “ Applied Statistics. ” (Third Edition). Allyn and Bacon, Inc.

Quenouille, M . H. (1956) : “ Notes On Bias In Estimation . ” *Biometrika* 43, 353-360.

Rao, C . R . (1973) : “ Linear Statistical Inference And Its Applications . ” (Second Edition) John Wiley and Sons, New York.

Rawlings, J. O. and Pantula, S.G. Dickey, D .A . (1998) : “ Applied Regression Analysis . ” (Aresearch Tool). (Second Edition) . Springer.

Searle, S . R . (1982) : “ Matrix Algebra Useful For Statistics . ” John Wiley, New York.

Seber , G.A. F. (1977) : “ Linear Regression Analysis . ” John Wiley , New York.

Stein , C.(1956) : “ Inadmissibility Of The Usual Estimator For The Mean Of A Multivariate Normal Distribution.” *Proceedings of the Third Berkeley Symposium Vol. 1*, ed. By Jerzy Neyman, PP. 197 – 206. Berkeley and Los Angeles: University of California Press.

Stein, C., (1960) : “ Multiple Regression.” *Controbutions to Probability and Statistics : Essays In Honour Of Harold Hotelling*, ed. Ingram, Olkin, PP. 424 – 443. Stanford : Stanford University Press.

Swindel , B. F. and Bower, D. R (1972) : “ Rounding Errors In The Independent Variables In A General Linear Model . ” *Technometrics* , 14, 215 - 218 .

Theil, H. (1971) : “ Principles Of Econometrics.” John Wiley, New York.

Thisted, Ronald A. (1978) : “ On Generalized Ridge Regression.” University of Chicago Department of Statistics Technical Report 57, May.

Timm, N.H. (1970) : “ The Estimation Of Variance Covariance Matrices From Incomplete Data.” Psychometrika 35, 417 - 437.

Walsh, J. F.(1959) : “ Computer Feasible General Method For Fitting And Using Regression Function When Data Is Incomplete.” SP-71, Santa Monica, California; System Development Corporation.

Weisberg, S. (1980) : “ Applied Linear Regression.” John Wiley and Sons, New York.

الخلاصة

" التقديرات المنحازة وطريقة المطواة في الانحدار "

تهدف هذه الأطروحة إلى دراسة التقديرات المنحازة وطريقة المطواة في الانحدار (الانكفاء). هناك أوضاع تكون فيها المتغيرات المستقلة في أنموذج الانحدار يصعب قياسها دون وجود أخطاء. وعندما تقاس تلك البيانات ضمن الأخطاء الموجودة، فإن درجة الدقة تكون ضرورية لتحسين المقدرات.

يتناول الفصل الأول تقديرات معاملات الانحدار المجهولة ، بالإضافة إلى خواص توزيعات التقديرات.

بينما يتطرق الفصل الثاني إلى دراسة تحيز التقديرات في الانحدار ومن ثم مناقشة التسامت (Multicollinearity) لأربع طرائق في التقديرات : جيمس شتاين (James Stein) ، التضاولية (Shrinkage) ، خط التقاطع الأعلى بين سطحين منحدرين (Ridge) وتحليل انحدار المكونات الرئيسية (Principal Component) .

يناقش الفصل الثالث تقديرات المعالم لأنموذج انحدار خطي بسيط، عندما تكون بعض قيم المتغيرات المستقلة عرضة للأخطاء، وتم إيجاد حدود للانحياز التربيعي النسبي ومتوسط تربيع الأخطاء.

عالج الفصل الرابع طريق المطواة لمتجه معاملات الانحدار المجهولة.

وتستنتج الدراسة أن العلاقة الآتية :

$$\left| \frac{\text{MSE}(\hat{\phi})}{\sigma^2} \right| \leq \left| M(X'X)^{-1}M' \right| \left[1 + \frac{\beta'\epsilon'\epsilon\beta}{\sigma^2} \right]$$

سوف تقلل الحد الأعلى لـ $\left| \frac{\text{MSE}(\hat{\phi})}{\sigma^2} \right|$